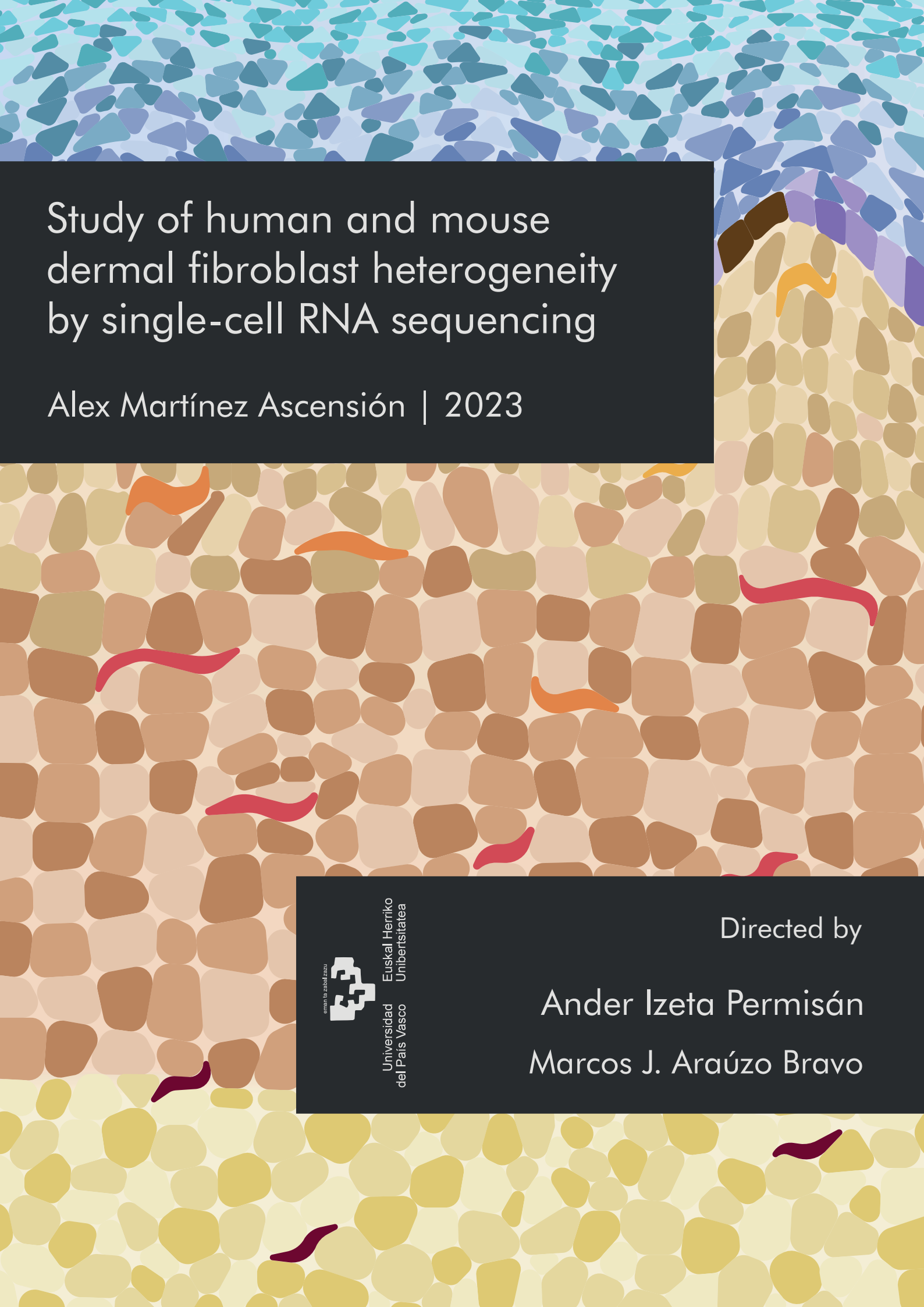
Study of human and mouse dermal fibroblast heterogeneity by single-cell RNA sequencing

Alex Martínez Ascensión | 2023



Universidad del País Vasco
Euskal Herriko Unibertsitatea

Directed by

Ander Izeta Permisán

Marcos J. Araúzo Bravo

Study of human and mouse dermal fibroblast heterogeneity by single-cell RNA sequencing

ALEX MARTÍNEZ ASCENSIÓN

July of 2023

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Study of human and mouse dermal fibroblast heterogeneity by single-cell RNA sequencing

Doctoral dissertation by

ALEX MARTÍNEZ ASCENSIÓN

Memory to achieve a PhD in Molecular Biology and Biomedicine by
The University of the Basque Country (UPV/EHU)

Doctoral Programme in Molecular Biology and Biomedicine

Directed by

ANDER IZETA PERMISÁN, Ph.D.
Tissue Engineering Department
and
MARCOS JESÚS ARAÚZO BRAVO, Ph.D.
Computational Biology and Systems Biomedicine Department

Biodonostia Health Research Institute

DONOSTIA - SAN SEBASTIÁN | July of 2023

biodonostia

osasun ikerketa institutua
instituto de investigación sanitaria

The work presented in this thesis was performed at Biodonostia Health Research Institute.

ABSTRACT

The skin is a complex structure located at the outermost part of the body. Its main function is to protect against diverse forms of trauma including thermal, chemical, and ultraviolet radiation. In addition, it prevents trans-epidermal water loss, that is, the movement of water from lower skin layers into the atmosphere; and is necessary for regulating body temperature, synthesising vitamin D and perceiving sensory stimuli from the environment.

The skin is divided into three main layers: epidermis, dermis, and hypodermis. Additionally, it includes adnexa such as hair follicles, sweat glands, and sensory terminals. The dermis is mainly composed of a rich extracellular matrix (ECM) that provides structural support and absorbs mechanical forces. This layer is further divided into the papillary dermis, located adjacent to the epidermis, cellularised and with loosely arranged collagen fibres; and the reticular dermis, a thicker layer composed mainly of arranged collagen I and III fibres, which provide structural support.

The main cell type located in the dermis is the fibroblast. Fibroblasts are traditionally known for producing the ECM of tissue stroma and were thought to serve only that function. However, further research has shown that fibroblasts are more complex than expected and are involved in the interaction with other structures, like endothelia, and a vast array of cell types, like immune cells. Despite their function heterogeneity, traditional lineage tracing studies only revealed two fibroblast types in the dermis: papillary fibroblasts and reticular fibroblasts, located in each dermis layer respectively.

Dermal fibroblast heterogeneity boomed was uncovered by the introduction of single-cell RNA sequencing (scRNAseq). In a preliminary study by Tabib et al. on healthy human dermis, fibroblast heterogeneity was divided into two main groups–$SFRP2^{+}DPP4^{+}$ and $FMO1^{+}LSP1^{+}$ fibroblasts–. The analysis revealed that these two major populations did not align in the papillary/reticular axis, and rather differed in morphology and putative function, breaking the dogma of dermal fibroblast arrangement.

Subsequent analyses confirmed previous dermal fibroblast heterogeneity results, but differences between publications hindered their replicability. Therefore, we performed a combined analysis of 4 scRNAseq datasets to merge individual observations and unravel their true diversity. We discovered that fibroblast clusters could be divided into 3 major axes–A, B, C–and 10 populations–A1-A4, B1-B2, and C1-C4–. Each axis corresponded to an individual function: namely, A were ECM-producing fibroblasts, B were immune fibroblasts and C were fibroblasts of skin adnexa.

This primary analysis was further extended into a secondary analysis, including 25 human and 9 mouse datasets, developed using a semi-supervised procedure. Human populations were extended from 10 to 15, distributed in 5 main axes–A1-A4, B1-B4, C1-C3 + C5, D1-D2 and E1–and mouse fibroblasts were separated into 17 populations, distributed in four main axes–x1-x2, y1-y5, z1-z2, w1-w5, v1–and 2 mixed axes–w/x, x/y–.

Further analysis of the markers of human fibroblasts reveals a large array of functions exerted by them. For instance, ECM-producing fibroblasts (A1-A4) could be divided into papillary fibroblasts,

reticular fibroblasts and elastic fibre-producing fibroblasts. Immune fibroblasts (B1-B4) can be differentiated based on the acute, adaptive and regulatory immune responses. Fibroblasts associated with dermal adnexa (C1-C5) could be distinguished based on cell subtypes of the hair follicle, such as dermal papilla and dermal sheath cells. Lastly, populations D1, D2 and E1 show a highly diverse marker profile, and some of them are associated with specialised structures such as peripheral nerves.

Comparison between mouse and human fibroblasts reveals the similarity of functions and a certain degree of fibroblast overlap between the two species. However, this overlap lacks finesse when looking at specific populations, probably due to the arrangement differences between human and mouse transcriptomaically-determined axes: while human axes seem to be separated on function, mouse axes separate on the dermal layer origin first, and on function second.

This *atlas* of dermal fibroblast heterogeneity is a useful tool for researchers trying to understand skin function, where fibroblast heterogeneity cannot go unnoticed; and also for researchers working on skin diseases, to use this combined dataset as a ground truth to discover differences in composition, new fibroblast types arising in disease, or simply as a tool to annotate their datasets without biases generated by experimental procedures.

RESUMEN

La piel es una estructura compleja situada en la parte más externa del cuerpo. Su función principal es proteger contra diversas formas de trauma, incluidas las radiaciones térmicas, químicas y ultravioleta. Además, previene la pérdida de agua transepidérmica, es decir, el movimiento de agua desde las capas inferiores de la piel hacia la atmósfera; y es necesaria para regular la temperatura corporal, sintetizar vitamina D y percibir los estímulos sensoriales del entorno.

La piel se divide en tres capas principales: epidermis, dermis e hipodermis. Además, incluye anexos como folículos pilosos, glándulas sudoríparas y terminales sensoriales. La epidermis está compuesta principalmente compuesta de keratinocitos, células que, conforme se diferencian, se corneifican, entrelazan entre ellos, y queratinizan, todo con el objetivo de formar una barrera impermeable contra insultos externos. La capa más extensa en volumen, y con mayor relevancia para el estudio de la tesis, es la dermis.

La dermis se compone principalmente de una matriz extracelular (MEC) rica en componentes y fibras que proporciona soporte estructural y absorbe las fuerzas mecánicas. Esta capa se divide a su vez en la dermis papilar, situada junto a la epidermis, celularizada y con fibras de colágeno dispuestas de forma laxa; y la dermis reticular, una capa más gruesa compuesta principalmente por fibras de colágeno I y III dispuestas de manera perpendicular para formar una retícula que actúa de soporte estructural.

De entre los diferentes tipos celulares que alberga la dermis, incluyendo, células epiteliales e inmunes, los fibroblástos son lás células más representativas. Los fibroblastos se conocen tradicionalmente por producir la matriz extracelular del estroma tisular y tradicionalmente se pensaba que solo cumplían esa función. Sin embargo, investigaciones posteriores han demostrado que los fibroblastos son más complejos de lo esperado y están involucrados en la interacción con otras estructuras, como el endotelio, y una amplia gama de tipos de células, como las células inmunitarias. A pesar de la heterogeneidad de su función, los estudios de rastreo de linaje tradicionales solo revelaron dos tipos de fibroblastos en la dermis: fibroblastos papilares y fibroblastos reticulares, ubicados en cada capa de la dermis respectivamente. Estos fibroblastos proceden de un linaje común que se diferencia, durante el desarrollo embrionario de ratón a día E12.5 se diferencia en los progenitores papilar y reticular, que posteriormente maduran a los dos tipos que mencionamos anteriormente.

Recientemente se ha desarrollado una metodología, denominada secuenciación de ARN en célula única (scRNAseq por su abreviación en inglés), que ha permitido analizar la heterogeneidad de diferentes tejidos a un nuevo nivel. Esta técnica combina los métodos de secuenciación masiva de ARN (RNAseq) con métodos de disgregación y aislamiento de células, de modo que el transcriptoma de cada célula es analizado independientemente. De esta manera, se pueden estudiar las diferencias transcritómicas de conjuntos de células entre si, para obtener información sobre grupos de células que tienen transcriptomas similares, y que asociamos a un tipo celular, o estadío celular independientes. Por tanto, las técnicas de scRNAseq permiten obtener, de una manera no sesgada, un panorama general de la diversidad celular individual de todo un tejido, y así poder descubrir

tipos celulares nuevos que con métodos más tradicionales basados en un conjunto limitado de marcadores no ha sido posible dilucidar. Tabib et al. (2018) fueron los primeros autores en publicar una caracterización completa de fibroblastos dérmicos humanos en piel sana utilizando scRNAseq. En el estudio determinan que la heterogeneidad de fibroblastos se resume en dos grandes grupos que expresan, por un lado, SFRP2 y DPP4, y por otro lado FMO1 y LSP1. El análisis reveló además que estas dos poblaciones principales no se alineaban en el eje papilar/reticular, y más bien diferían en morfología y función putativa (los fibroblastos SFRP2+ son alargados y están asociados a la producción de MEC, mientras que los fibroblastos FMO1+ están asociados a la respuesta inmune). Este descubrimiento rompió el dogma de la disposición de los fibroblastos dérmicos y su separación de funciones en base a la localización más apical o basal.

Diferentes análisis posteriores confirmaron los resultados previos de heterogeneidad de fibroblastos dérmicos, pero cada publicación reportaba diferencias entre las publicaciones. Puesto que estas inconsistencias dificultaba la replicabilidad para crear un modelo de fibroblastos démicos homogéneo, realizamos un análisis combinado de 4 conjuntos de datos scRNAseq para fusionar observaciones individuales y desentrañar su verdadera diversidad.

Para realizar ese primer análisis, aplicamos un método de agrupación de células (clustering) para cada conjunto de datos, y buscamos manualmente poblaciones (un cluster o grupos de clusters) que mantuvieran un perfil transcriptómico similar, es decir, que tuvieran una expresión similar de varios marcadores entre dos o más conjuntos de datos simultáneamente. Además, si varias poblaciones compartían una transcriptómica más general, y colocalizaban en los gráficos de visualización que empleamos para visualizar la heterogeneidad celular, agrupamos estas poblaciones en ejes.

En base a esta agrupación de clusters en poblaciones, y de poblaciones en ejes, descubrimos que los clusters de fibroblastos se podían dividir en 3 ejes principales (A, B, C) y 10 poblaciones (A1-A4, B1-B2 y C1-C4). Haciendo un análisis de literatura preliminar en base a los marcadores de cada eje, observamos que cada eje correspondía a una función individual: a saber, el eje A corresponde a los fibroblastos productores de MEC, el eje B corresponde a fibroblastos inmunes y el eje C corresponde a fibroblastos de anexos cutáneos, como el folículo piloso. La existencia de estas tres poblaciones fue validada mediante experimentos de inmunofluorescencia, donde se observaba que tres marcadores, uno para cada eje (SFRP2 para A, CCL19 para B y SFRP1 para C) no colocalizaban, garantizando que cada eje era independiente.

Aunque los resultados de este análisis son muy relevantes para entender la heterogeneidad de fibroblastos dérmicos, observamos diferentes carencias que subsanamos en un análisis secundario. En primer lugar, observamos que, si bien existían poblaciones con marcadores comunes a varios conjuntos de datos, la heterogeneidad inter-dataset (entre conjuntos de datos), lejos de ser despreciable, podía ser un factor que disminuyese la replicabilidad para nuevos conjuntos de datos. En segundo lugar, la metodología del análisis primario carecía de replicabilidad, al ser un proceso completamente supervisado. Además, esta metodología no es extensible a nuevos conjuntos de datos.

En base a estas dos limitaciones, decidimos extender el análisis primario a uno secundario, con dos novedades. En primer lugar, para disminuir los sesgos asociados a la heterogeneidad de conjuntos de datos individuales aumentamos significativamente el número de conjuntos de datos, incluyendo un total de 25 conjuntos de datos de dermis humana y 9 murina. Esto implica que se relaizará un

analisis independiente para cada organismo. Además, para realizar la caracterización de manera menos sesgada, diseñamos un procedimiento semisupervisado.

En este procedimiento se parte de un diccionario que tiene poblaciones con sus marcadores respectivos y diferentes conjuntos de datos divididos en clusters, y devuelve los conjuntos de datos con los clusters agrupados en poblaciones, y un nuevo diccionario con los marcadores definitivos de cada población en base a su relevancia en cada conjunto de datos individual. Este procedimiento, si bien requiere de un esfuerzo inicial y sesgado para producir el diccionario de poblaciones y marcadores, facilita la asignación de poblaciones a nuevos conjuntos de datos que se introduzcan, y actualiza la lista de marcadores correspondiente.

En base a este procedimiento, y tras varias iteraciones donde se han introducido nuevos conjuntos de datos paulatinamente y se han actualizado las poblaciones "correctas" para cada organismo las poblaciones humanas se ampliaron de 10 a 15, distribuidas en 5 ejes principales (A1-A4, B1-B4, C1-C3 + C5, D1-D2 y E1) y los fibroblastos de ratón se separaron en 17 poblaciones, distribuidas en cuatro ejes principales (x1-x2, y1-y5, z1-z2, w1-w5, v1) y 2 ejes mixtos (w/x, x/y) conformados por poblaciones que no se han podido asignar de manera concluyente a un eje en concreto.

Para compobar la eficacia y la reproducibilidad del algoritmo de asignación de poblaciones y marcadores, aplicamos un método de bootstrap, que consiste en correr el algoritmo en diferentes muestreos aleatorios de los conjuntos de datos, conteniendo un 99 % de las células originales. Después, comprobamos cómo de similares son las asignaciones de poblaciones a las células en comparación con la asignación empleando todas las células. El análisis de bootstrap revela que, en general, la asignación de poblaciones es bastante robusta, sobre todo en ratón, y que en las poblaciones donde menos robustez se aprecia es en aquellas que denominamos "poblaciones puente", es decir, poblaciones que generalmente carecen de marcadores propios, y cuyos marcadores son una mezcla de dos poblaciones separadas.

Un análisis más profundo de los marcadores de fibroblastos humanos revela una gran variedad de funciones, incluso entre poblaciones de un mismo eje. Si nos centramos en fibroblastos dérmicos humanos, concluimos que hay tres poblaciones "arquetípicas": la población A2 expresa marcadores asociados a fibras localizadas en membranas basales y, más concretamente en piel, localizados en la unión dermoepidérmica; las poblaciones A1 y A3 expresan el grueso de fibras localizadas en la MEC, así como proteínas necesarias para la maduración y degradación de estas fibras; y la población A4, derivada de A1, expresa más marcadores asociados a fibras elásticas.

En cuanto a las poblaciones del eje B, también encontramos tres poblaciones arquetípicas: la población B1, que parece estar asociada a una respuesta inmune primaria, asociada a la expresión de interleucinas agudas como IL1B o IL6, así como quemoquinas que atraen a neutrófilos y otras células inmunes; la población B2 está asociada a una respuesta inmune adaptativa, asociada a marcadores de maduración inmune, como los del sistema CD40/CD40L, así como quemoquinas de macrófagos y de linfocitos B y T, como CCL19 o CSF1; y la población B4, cuya expresión de marcadores de protección frente a estrés oxidativo y de inmunoregulación parece estar asociada a una modulación de la respuesta inmune asociada a las células inmunes y a los fibroblastos B1 y B2, así como a una reparación del tejido tras la respuesta inflamatoria.

Por último, las poblaciones del eje C están asociadas a diferentes poblaciones del folículo piloso ya descritas, como la vaina dérmica (C1) y la papila dérmica (C2); así como a posibles poblaciones que

aun no están descritas formalmente en base a marcadores, pero que pueden tener una relevancia en su homeostasis (C3, C5). Estas poblaciones, así como las poblaciones D1, D2 y E1, tienen una expresión relevante de marcadores asociados a nervio, vasos sanguíneos y metabolismo de xenobióticos, por lo que pueden tener una función mixta de soporte funcional asociada a nervios y vasos periféricos, anejos dérmicos u otras estructuras de la piel.

En lo que respecta a la caracterización de la heterogeneidad de fibroblastos dérmicos murinos, si bien no hemos realizado una caracterización funcional tan exhaustiva, basándonos en un análsis previo realizado por Joost et al. (2020) concluimos que la heterogeneidad de los ejes de ratón se caracteriza por su localización en la dermis; más concretamente, las poblaciones del eje x están localizadas en la dermis, las poblaciones del eje y están localizadas en el tejido adiposo, y las poblaciones del eje z corresponden a la fascia localizada debajo del músculo panículo carnoso.

Así pues, la comparación entre fibroblastos de ratón y humano revela una similitud de funciones y un cierto grado de superposición de fibroblastos entre las dos especies. Por ejemplo, la correspondencia x2-A2 es unívoca y asociada a la dermis papilar, mientras que A1, aunque está más asociada a x1, también muestra cierta similitud con las poblaciones x/y, y3, z2 de ratón. También parece haber cierta similitud entre la población B2 e y4, aunque no hay suficiente evidencia para asociar las poblaciones B1 y B4 a poblaciones murinas concretas. Por último, observamos una similitud entre la población C2 y las poblaciones w/x y w1/w2; la población C1 y w4; y las poblaciones D1 y D2 con las poblaciones murinas y5 y v1.

Este cierto grado de similitud es, en general, interesante, pues revela que existe cierto nexo funcional entre los fibroblastos dérmicos de ambos organismos. Sin embargo, la separación de funcional de ejes humanos y la separación por localización de ejes en ratón, sobre todo en las poblaciones asociadas a la producción de MEC y asociadas a la respuesta inmune manifiesta que hay que realizar más análisis para elucidar cual es la razón de estas diferencias entre organismos.

Los resultados obtenidos fruto del análisis de heterogeneidad de fibroblásticos dérmicos humanos han sido útiles como medio para determinar la calidad del análisis de conjuntos de datos realizados por terceros. Por ejemplo, tras un reanálisis de los datos producidos por Reynolds et al. (2021) comprobamos que, en dos grupos independientes de células, para cada uno de los grupos de células observamos que existía una población A1, otra de A2, otra de B1 y otra de B2. Es decir, para A1 había dos poblaciones transcriptómicamente diferentes con la misma etiqueta; y lo mismo para el resto de poblaciones.

Debido a la anomalía de esta caracterización, realizamos un análisis más exhaustivo de genes diferencialmente expresados entre poblaciones y observamos que aproximadamente un tercio de las células expresaban marcadores asociados a hipoxia, y otro tercio expresaba marcadores asociados a estrés celular. Además, algunos de estos efectos no pudieron ser corregidos de manera computacional, lo cual implica que las células afectadas por este efecto se encontraban en un estado transcriptómico irreversible. Estas dos condiciones de estrés celular e hipoxia se deben, principalmente, a un procesamiento exhaustivo del tejido durante la disgregación, posiblemente para asegurar la máxima captación de células. En ese proceso, las células fueron expuestas durante tiempos excesivamente prolongados (> 8h) a temperaturas elevadas (37° C), lo cual activó estos dos programas transcripcionales.

A modo de conclusión, este atlas de heterogeneidad de fibroblastos dérmicos es una herramienta útil para intentar comprender la función de la piel, donde la heterogeneidad de fibroblastos no puede pasar desapercibida; y también para el personal investigador que trabaje en enfermedades de la piel, para utilizar este conjunto de datos combinados como base para descubrir diferencias en la composición, nuevos tipos de fibroblastos que surgen en enfermedades, o simplemente como una herramienta para anotar sus conjuntos de datos sin sesgos generados por procedimientos experimentales.

This thesis is developed based on the following publications:

1. **Alex M. Ascensión**, Sandra Fuertes-Álvarez, Olga Ibañez-Solé, Ander Izeta, and Marcos J. Araúzo-Bravo. "Human dermal fibroblast subpopulations are conserved across single-cell RNA sequencing studies." *Journal of Investigative Dermatology* 141.7 (2021). doi: 10.1016/j.jid.2020.11.028
2. **Alex M. Ascensión**, Marcos J. Araúzo-Bravo, and Ander Izeta. "The need to reassess single-cell RNA sequencing datasets: The importance of biological sample processing." *F1000Research* 10 (2022). doi: 10.12688/f1000research.54864.1
3. **Alex M. Ascensión**, Marcos J. Araúzo-Bravo, and Ander Izeta. "Challenges and opportunities for the translation of single-cell RNA sequencing technologies to dermatology." *Life* 12.1 (2022). doi: 10.3390/life12010067

Although some parts of this thesis are shared across publications, roughly each major part of the thesis dissertation–i.e. Materials and Methods, Results, Discussion, etc.– will be based on each publication:

- Publication 1 describes a primary analysis of human dermal fibroblast heterogeneity. Chapters 4–Materials and Methods–, 8–Results–, and part of chapter 12–Discussion– are based on this publication. Chapters 5–materials and methods–, 9–results–, 10–results–, and 12–discussion–of this thesis are derived from a posterior analysis derived from this publication.
- Publication 2 describes the results from a reanalysis of a dataset from the secondary analysis derived from publication 1, and stresses the necessity of good quality datasets for a proper interpretation of single-cell data. Chapters 6–materials and methods–, 11–results–, and 13–discussion–this thesis are derived totally or partially from this publication.
- Publication 3 is a review of the state-of-the-art single-cell analyses focused on skin and, more importantly, disease conditions. Although we will not focus on the disease aspect in this thesis, certain elements of chapter 2–introduction–and 10–discussion–are based on this publication.

The work developed for this thesis has been presented at the following conferences:

- **Alex M. Ascensión**, Sandra Fuertez-Álvarez, Marcos J. Araúzo-Bravo, and Ander Izeta. "Dermal single-cell atlas: a novel tool to explore fibroblast heterogeneity." *ESDR 2022 Symposium* (2022).

Other works produced during the development of the thesis but which are not appearing in it are the following:

1. **Alex M. Ascensión***, Olga Ibáñez-Solé*, Iñaki Inza, Ander Izeta, and Marcos J. Araúzo-Bravo. "Triku: A feature selection method based on nearest neighbors for single-cell data." *GigaScience* 11 (2022). doi: 10.1093/gigascience/giac017
2. Olga Ibañez-Solé*, **Alex M. Ascensión***, Marcos J. Araúzo-Bravo, and Ander Izeta. "Lack of evidence for increased transcriptional noise in aged tissues." *eLife* 11 (2022). doi: 10.7554/eLife.80380
3. Javier Cabau Laporta, **Alex M. Ascensión**, Mikel Arrospide Elgarresta, Daniela Gerovska, Marcos J. Araúzo Bravo. "FOntCell: Fusion of Ontologies of Cells." *Frontiers in Cell and Developmental Biology* 11 (2021). doi: 10.3389/fcell.2021.562908

The author positions with an asterisk (*) indicate co-authorship.

The first publication describes the development of a Feature Selection method that has been extensively used throughout the analyses performed. The second and third publications refer to algorithms or analyses performed partially and jointly with other team members.

Genes and proteins throughout this thesis are depicted with the following convention: human genes/proteins are written in ALL-CAPS case–e.g., SOX2, *TGFB1*–, mouse genes/proteins are written in capitalised lower case–e.g., Sox2, *Tgfb1*; genes are written in italics–e.g., *SOX2*, *Sox2*–, and proteins are written in round type–e.g., SOX2, Sox2–.

FUNDING SOURCES

The development of this thesis was funded by *Ayudas destinadas a la financiación de la realización de una tesis doctoral mediante un contrato predoctoral*. Departamento de Educación del Gobierno Vasco. (PRE_2018_1_0008, PRE_2019_2_0233, PRE_2020_2_0081, PRE_2021_2_0111).

AGRADECIMIENTOS

Casi cinco años han pasado desde que entré en Biodonostia a dedicar mi tiempo a un proyecto, que para nada intuí que fuese a ser tan arduo, desesperante y gratificante a partes iguales. Tampoco intuí que, incluso pandemia mediante, fuese a ser una persona tan compleja y diferente a la que empezó este capítulo de mi vida. Me ha tocado explorar y redefinir el concepto de familia y relación, mi identidad y mi psicología, mis miedos y mis placeres, mi relación con la sociedad, la academia y mi papel en la ciencia.

Cinco años dan para mucho, y desde luego dan para conocer a muchas personas que me han acompañado en este camino y que, sin ellas, no sería quien soy ahora.

Como todo comienza con elles, empiezo con mis adres, que os habéis sacrificado para darme sustento y apoyo para avanzar en mi carrera. A mi madre, que me has dado la perspectiva para entenderme y mejorar como persona; y a mi padre, por hacerme ver que la paciencia es la mejor de las virtudes, y que las relaciones también pueden ir a mejor.

A Sandra, Joan y Maria José. Desde 2018 con vosotres, cada une con su camino, pero siempre con las videollamadas para contar nuestros dramas mientras jugábamos al Hanabi. Mary, me alegro mucho del tiempo que hemos compartido, en Barcelona, Donosti, Valencia, Pontevedra, … de todo lo que nos hemos contado y que nos contaremos y que, cada vez más, me hace ver lo mucho que nos parecemos.

A toda mi familia elegida, comenzando por Ari e Izei y, especialmente, Mikel. Sois mi núcleo, la razón que hace que Donosti no sea sólo donde vivo, sino mi hogar. Y sobre todo, Mikel, apareciste en mi vida en el momento que más afecto necesitaba, y gracias a ti he aprendido como poder disfrutar de una relación, en todas sus facetas. Y, por supuesto, Íñigo y John, aunque ya no estéis en Donosti, las horas y anéctodas aún se quedan.

También quiero agradecer a toda mi familia de Gehitu (Oier, Bula, Maitane, Luna, Jokin, Iñaki, Janire, Stephen…) y del grupo joven de la FELGTBI+ (Oli, Dani, Sara, Anthony, Sergio, Bea, …), y en particular a Blue, Irati y a Eva (porque unas jornadas, como una tesis, de crisis en crisis, se pueden sacar adelante). A cada une os he conocido en momentos diferentes, bien en Euskal Herria, península o islas, pero todes sois especiales para mi. Os quiero a todes.

Aunque la tesis sea un proyecto individual, sin duda el camino se hace más llevadero en un entorno de trabajo que te apoye. Por eso también quiero agradecer a mis jefes. A Marcos, por proporcioname los recursos para poder desarrollar mis investigaciones; y a Ander, por el apoyo emocional, los consejos durante la tesis y por todas las conversaciones que me han hecho ver la realidad de la academia, y que la mejor ciencia que puedo hacer es la que sigue mis valores.

A todo el labo de compus, (Olga, Javi, Koldo, Aitor, Mikel…), por las horas que hemos dedicado a hablar y sacar nuestros proyectos adelante. Y especialmente a Olga, que transformaste y cohesionaste el laboratorio; y que entre canciones italianas, viajes en Alesshprexxx, tardes de delirio en el labo,

noche en tu casa con Urki y Sahats, y horas en la *private veranda* me diste las herramientas para ser quien soy y disfrutar la vida con otros ojos.

Por supuesto, muchas gracias a todo el labo de tisu (Ainhoa, Paula, Laura, Vir, Hector, Inazio, Noelia y compañía) por el tiempo que habeís compartido conmigo; y desde luego a Sandra, por las coñitas y los dramas que han salido en estos años, y por los proyectos en los que hemos trabajado (y los que nos quedan por cerrar).

Por último, y pese a que el tiempo que llevo todavia es poco, gracias a David, a Patxi, por haberme dado la oportunidad de renovarme y seguir contribuyendo a este tortuoso mundo de la ciencia; y por supuesto a todas las currantas, que me habéis acogido como a une más (especialmente a vosotras, Andrea, Saioa, Miri y Rocío).

Y como la ciencia, si no es compartida, no es ciencia, me gustaría hacer un agradecimiento honorífico a todas aquellas personas que habéis mantenido protocolos, paquetes y, en general, contribuido al conocimiento sobre el que he desarrollado mi tesis; y también, especialmente, a Alexandra Elbakyan por ahorrarme horas de emails a PIs para pedir artículos y, sencillamente, por hacer que un conocimiento comunitario mantenido con capital público pueda seguir siéndolo. Наука и коммунизм неотделимы.

Y ya por acabar, quiero agradecer todas las personas que aun con el paso del tiempo mantenemos contacto, y que fuisteis clave en muchas etapas de mi vida; en especial a Halen e Iván, que después de tantos años siempre tenemos una oportunidad para actualizarnos. Y por último, a mis compas de la UNED, que hicisteis mis horas de mates algo más llevaderas; en especial a Hug y a Carmen.

Mila esker guztioi.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

GENERAL

AD Atopic Dermatitis

ANOVA Analysis of variance

APC Antigen-presenting cell

ATAC-seq Assay for Transposase-Accessible Chromatin using sequencing

BM Basement membrane

CAF Cancer-associated fibroblast

CCC Cell-cell communication

CDD Critical Difference Diagram

COD Curse of dimensionality

CPM Counts per million

CSPG Chondroitin sulphate proteoglycan

CTCL Cutaneous T-cell lymphoma

CTS Connective tissue sheath

CV Coefficient of variation

DAMP Damage-associated molecular patterns

DC Dendritic cell

DC Dermal condensate

DEG Differentially expressed gene

DFU Diabetic foot ulcer

DOI Digital Object Identifier

DP Dermal papilla

DRESS Drug Reaction with Eosinophilia and Systemic Symptoms

DS/CTS Dermal sheath / Connective tissue Sheath

ECM Extracellular matrix

ERCC External RNA control consortium

FACIT Fibril-associated collagens with interrupted triple helices

FACS Fluorescence-activated cell sorting

FISH Fluorescence *in situ* hybridisation

FS Feature selection

GAG Glucosaminoglycan

GAN Generative adversarial network

GF Growth factor

GLM Generalized linear model

GO Gene ontology

GRC Genome Reference Consortium

GWAS Genome wide association study

HF Hair follicle

HPV Human papillomavirus

HSPG Heparan sulphate proteoglycan

GOEA Gene ontology enrichment analysis

IBD/S Inflammatory bowel disease/syndrome

IER Immediate early response

IFE Interfollicular epidermis

IVT *in vitro* transcription

kNN k-nearest neighbours

LC Langerhans cell

lncRNA Long noncoding RNA

LPS Lipopolysaccharide

LR Ligand-receptor

MØ Macrophage

ML Machine learning

mRNA Messenger RNA

miRNA Micro RNA

NA or NaN Not a number

NB Negative binomial

NE Neighbour embedding

NGS Next generation sequencing

NK Natural killer (cell)

NMI Normalised mutual information

PAGA Partition-based graph abstraction

PBMC Peripheral blood mononuclear cells

PCA Principal component analysis

PCR Polymerase chain reaction

PCP Planar cell polarity

PPI Protein-protein interaction

QC Quality control

ROS Reactive oxygen species

RPKM Reads per kilobase per million mapped reads

(I/O)RS Inner/Outer root sheath

RT Room temperature

TI Trajectory inference

t-SNE t-distributed stochastic neighbour embedding

UMAP Uniform manifold approximation and projection

UMI Unique molecular identifier

sc Single-cell

scRNAseq Single-cell RNA sequencing

SSc Systemic sclerosis

SVD Singular Value Decomposition

VE Vascular endothelium

(d/s)WAT (dermal/subcutaneous) white adipose tissue

GENES AND PROTEINS

α-SMA/ACTA2 alpha smooth muscle actin/Actin alpha 2

APC Adenomatosis polyposis coli

AREG Amphiregulin

ASPN Asporin

BMP Bone morphogenic protein

CAV Caveolin

CAMKII Calmodulin-dependent kinase II

CCN Cellular Communication Network

CKIα Casein kinase Iα

DAAM DVL-associated activator of morphogenesis 1

DCN Decorin

DKK Dickkopf

DMKN Dermokinin

DPT Dermatopontin

DSH/DVL Disheveled

EGFR Epidermal growth factor receptor

EN1 Engrailed 1

ERK Extracellular signal-regulated kinase

FAP Fibroblast activation protein

Fz/Fzd Frizzled

FSP/S100A4 Fibroblast-specific protein

GPCR G-protein-coupled receptor

GSK3 Glycogen synthase kinase 3

HA Hyaluronic acid

IFI27 Interferon α-inducible protein 27

IGF Insulin-like growth factor 1

JNK c-Jun amino-terminal kinase

LRRC15 Leucine-rich repeat containing 15

LRP Lipoprotein-related protein

MAPK mitogen-activated protein kinase

MMP Matrix metalloproteinase

MRTF-A Myocardin-related transcription factor A

mTORC1 Mammalian target of rapamycin complex 1

MUSK Muscle skeletal receptor Tyr kinase

NFAT Nuclear factor of activated T cells

PDGF Platelet-derived growth factor

PDGFRA Platelet-derived growth factor receptor-alpha

PKC Protein kinase C

PLC Phospholipase C

PP2A Protein phosphatase 2A

PTK7 Protein Tyr kinase 7

ROCK Rho-RhoA/Rho-associated coiled-coil protein kinase

ROR Receptor Tyr kinase-like orphan receptor

RSPO R-spondin

RYK Receptor Tyr kinase

SRF Serum response factor

SFRP Secreted frizzled-related protein

TAZ PDZ-binding motif

TCF T cell factor

TGF-β Transforming growth factor-β

TGFR TGF-β receptor

TEAD TEA domain family member

TIMP Tissue inhibitor of MMP

TLR Toll-like receptor

VCAN Versican

VIM Vimentin

YAP Yes-associated protein

ZNRF3 Zinc and RING finger 3

Part I

INTRODUCTION

1

SINGLE-CELL TECHNOLOGIES AND ANALYSIS

The concept of *cell* is one of the most influential developments in modern biology. Since the improvements made in microscopy by Robert Hooke and Anton van Leeuwenhoek, who first described the cell as the structural unit of life back in the 17h century, became the vertebrating concept for many biologists for centuries (Gest, 2004). Cell theory was developed initially around 1840, with Theodor Schwann and Matthias Jakob Schleiden postulating two of the three common tenets currently used: (1) all living organisms are composed of one or more cells, and (2) the cell is the most basic unit of life (Schwann, 1839; Sharp, 1984). 15 years later, Rudolf Virchow added the third tenet–although Robert Remak suggested it some years before–, famously known as *omnis cellula e cellula*: (3) all cells arise from pre-existing cells (Mazzarello, 1999). Modern cell theory includes additional elements such as the existence of nucleic material, cell metabolism or organism activity derived from its cells, but the three classical tenets still hold more than 150 years afterwards (Wolfe, 1985).

Cells and tissues have traditionally been studied under the histological lens; that is, they were described based on morphology. After the decades of 1950 and 1960, with several biochemistry developments such as the discovery of DNA structure by Franklin, Watson and Crick (Watson et al., 1953), the location of the genetic information within it alongside its transcription and translation (Crick, 1970; Jou et al., 1972; Roeder et al., 1969), cells were assigned their biological function based on their gene expression patterns. Cells are different not only because of the morphology but because they have different gene expression patterns; tissues are composed of different cell types and their secretions which, if they change, will induce changes in the tissue.

Therefore, one of the main aspects of modern cell and tissue biology is finding ways to study cells and tissues via their gene expression. Some common forms to analyse the expression of a few genes are based on the direct measurement of mRNA via the Polymerase Chain Reaction (PCR), Fluorescence *in situ* hybridisation (FISH); or the measurement of protein quantitatively or qualitatively with western blots, immunohistochemistry or immunofluorescence. However, these methods can only analyse a limited number of genes at a time and are not scalable to large numbers of genes or even the whole genome.

The first method to analyse a large number of genes was the microarray, developed by Schena et al., 1995, where 45 probes were designed to detect transcripts from *A. thaliana*. Later developments by Agilent, Affymetrix or Illumina have expanded the number of probes to detect thousands of transcripts

across different species, which made them affordable and widely used by researchers, especially between 2005 and 2016.

More recently, next-generation sequencing (NGS) methods such as RNA sequencing (RNA-seq) have emerged as an alternative approach for gene expression profiling (Emrich et al., 2006; Lister et al., 2008). In RNA-seq, mRNA is extracted, cDNA from that mRNA is synthesised and amplified, and adaptors are attached for sequencing. This process, unlike microarrays, allows for full sequencing of the whole transcriptome, while the former only profiles predefined transcripts/genes through hybridisation. Similar to microarrays, companies like Illumina, Pacific Biosciences or Oxford Nanopore have *democratised* the use of RNA-seq for a wide range of researchers.

The main limitation of *standard* RNA-sequencing protocols is that they were designed for bulk RNA extraction from tissues. Although this would allow the characterisation and comparison of tissues under different conditions, the information at the cellular level is lost since it is *averaged* across the different cell types in the sample (Grün et al., 2015). To minimise the differences between cell types, some bulk RNA-seq studies used Fluorescence-activated cell sorting (FACS) to select specific cell types or tissues based on a select set of markers, which is achieved more easily on immune cell types (Carithers et al., 2015). However, the full, unbiased characterisation of the heterogeneity within tissue was far from achievable.

Single-cell methods are a solution to the limitation of bulk tissue analysis. By tissue disaggregation and cell isolation, single-cell methods analyse the biological properties –i.e. transcriptome, proteome, or epigenome–that may be described within each cell. Although many current single-cell methods combine NGS approaches to analyse whole -omic properties of the cells, technically, a method is considered "single-cell" if cells are individually analysed, regardless of the analysis method-e.g. some publications consider FACS within the family of single-cell methods–.

1.0.1 *History of single-cell methods*

The first event recorded of genetic profile characterisation using single-cell methods was by Eberwine et al., 1992, who measured the expression of a handful of individual genes from single cells for the first time in 1992, using a sophisticated approach based on *in vivo* reverse transcription (RT) followed by amplification through *in vitro* transcription (IVT).

Almost 20 years afterwards, the first protocol for single-cell sequencing was published by Tang et al., 2009, who adapted NGS and single-cell technologies to make them compatible with high-throughput DNA sequencing, thus allowing completely unbiased transcriptome-wide investigation of the mRNA in a single cell for the first time (Figure 1). They subsequently used this method to trace the derivation of mouse embryonic stem cells from the inner cell mass with single-cell resolution (Tang et al., 2010a).

The use of embryonic stem cells facilitated the use of the Tang et al., 2009 protocol, considering these cells are relatively easily separable compared to cells from adult tissues. Furthermore, cells were previously classified and sorted. A shift in the field came when Guo et al., 2010 demonstrated that distinct cell types could be identified without pre-sorting. The authors used RT-qPCR of 48 genes in parallel on more than 500 embryonic cells.

Islam et al., 2011 published the first single-cell RNA-seq protocol for multiplexing cells from a single 96-well plate using a unique template-switching oligo (TSO) in each well via their STRT-seq method. This method exploits the template-switching property of the reverse transcriptase to tag the 5' ends of poly-Adenylated mRNA molecules. Following PCR amplification, the tagged ends are pulled down and sequenced.

Ramsköld et al., 2012 published SMART-seq, the first protocol that allowed the full-transcript characterisation of gene expression in single-cells. This method later evolved into SMART-seq2, which is currently used in many single-cell studies (Picelli et al., 2013). In the same year, Kivioja et al., 2012 reported using unique molecular identifiers (UMI) to tag independent mRNAs, although the first use in a single-cell context did not occur until 2014 (Islam et al., 2014).

Brennecke et al., 2013 released SMARTer (C1) in 2013, the first microfluidic C1 system used for passive cell capture in a scRNAseq context. Fluidigm later commercialised this method. In this protocol, cells are loaded onto the chip and are passively captured in up to 96 isolated chambers in about half an hour.

So far, all experiments have been performed in cell cultures, embryonic cells or tumour cells. The first use of a scRNAseq method applied for a solid tissue was registered by Jaitin et al., 2014, who studied the composition of 4,468 mouse spleen cells using their own method, MARS-seq. They could perform an unbiased characterisation of different states of dendritic cell differentiation, B cells and macrophages, among other hematopoietic cells.

Lastly, in 2015 two different protocols to isolate cells in droplets and carry out barcoded cDNA preparation within each droplet were published: inDrop (Klein et al., 2015) and Drop-seq (Macosko et al., 2015). Droplet-based isolation methods have become one of the main strategies to massively isolate hundreds to millions of cells in a single study. One of these methods, Chromium, developed by 10X genomics based on an adaptation the Dropseq method, was developed in 2017 and is the current gold standard in scRNAseq methods (Zheng et al., 2017).

1.0.2 *Fast scaling of single-cell analysis*

In the following years, many different incremental improvements have contributed to increases in scale, such as decreases in required reagent volumes and consumable costs facilitated by the democratisation of microfluidic technologies, random capture methods, *in situ* barcoding, and reduction in sequencing costs (Figure 1) (Illumina, 2021; Kolodziejczyk et al., 2015; Picelli et al., 2013).

As a result of these improvements, there has been a boom in the number of datasets produced and the number of cells per dataset. Looking at the database of single-cell datasets compiled by Svensson et al. (Svensson et al., 2020), Figure 2 shows the scaling of the number of cells per dataset with time. There is an inflexion point in the number of publications in 2018–the density of points visibly increases–, possibly due to the commercialisation of different library preparation kits by companies like 10X Genomics.

One of the consequences of this boom in dataset size is the publication of a range of single-cell atlases, that is, large datasets consisting of the sequencing of one or several organs from different organisms by a coordinated set of laboratories, generally using a common, standardised protocol for sample processing and dataset analysis. Some of the examples of these atlases are (1) the Human

Cell Atlas initiative, which recently claimed to have profiled more than 50 million cells from over 30 human organs (Lindeboom et al., 2021), (2) the Tabula Sapiens Consortium, with 500,000 cells from 24 human organs (T. Sapiens Cons. et al., 2022), (3) or the Mouse Cell Atlas, with more than 1 million cells from more than 10 mouse tissues (Wang et al., 2022).

There are also several large datasets of more than 1 million cells, like the one focused on mammalian organogenesis from Cao et al., 2019–with more than 2 million cells–, the large analysis of 1.5 million immune cells related to COVID-19 by Ren et al., 2021; or the analysis of the differentiation of 1 million dopaminergic neurons by Jerber et al., 2021.

Figure 2 shows a clear, steady expansion of the number of cells per dataset, with the regression trend indicating an expected 2.4x increase in the number of cells per year, similar to the famous Moore's Law (Moore, 1998). Similar to the original predictions by Moore, whose law is predicted to reach a fundamental limit around 2030 (Kumar, 2015), it is possible that in a near future, the number of cells per dataset will also reach a fundamental limit based on the ability to do proper and meaningful dataset analysis–this will be discussed in Chapter 13 of the Discussion–.

The large amounts of data and research goals associated with single-cell analysis require efficient computational methods and sound statistical analysis. This aligns with the definition of data science, leading us to conclude that we have entered an era of "Single-Cell Data Science" (SCDS). SCDS amplifies the challenges already present in traditional data science techniques (Lähnemann et al., 2020).

1.1 METHODS IN SINGLE-CELL SEQUENCING

In the nearly 10 years that the single-cell sequencing realm has been developing and expanding, many methods have arisen. In this section, we will focus on explaining how the most common scRNAseq methods work and will also mention other relevant types of methods, like the spatial or proteomic single-cell methods. A full comprehensive description and analysis of single-cell methods is available at Ding et al., 2020; Mereu et al., 2020; Stuart et al., 2019a; Ziegenhain et al., 2017.

1.1.1 *Single-cell RNA sequencing*

Single-cell RNA sequencing (scRNAseq) is the most frequently used method within the family of single-cell methods. In scRNAseq, mRNA from individual cells is extracted, amplified and sequenced. Figure 3 shows a summary of the steps of some common scRNAseq methods.

The first step within the scRNAseq processing, similar to other methods, is a proper experimental design. Single-cell methods include a set of variables that can confound the experimental procedure, such as the biological or technical replicates, sequencing runs, flow cells and lanes, as in bulk RNA-seq (Hicks et al., 2015). Therefore, it is important to randomise over as many of the previously stated factors as possible so that putatively interesting features observed in the data are not due to artefacts imposed during sample preparation and/or data collection. Nonetheless, conducting a fully randomised experiment in many cases is not realistic. Limited samples, a fixed number of single-cell isolation platforms and sequencers, time constraints, and budgets often impede the theoretically ideal experiment from being realised in practice (Bacher et al., 2016).

1.1.1.1 *Cell isolation*

The first step after sample extraction and tissue disaggregation is cell isolation. One optional previous step performed in bulk experiments is sample purification using FACS. With unrestricted sorting gates, random samples of cells can be purified, and dead cells and debris can be removed (Grün et al., 2015). Additionally, with a suitable FACS instrument that performs *index sorting*–i.e. data about the fluorescence levels of proteins and cell size for each well can be stored–cell subtypes can be selected based on morphology or on the expression of a set of markers (Hayashi et al., 2010). To maximise capture rates of FACS-based methods, diluted cell suspensions should be sorted at low speeds–e.g., 100 cells/s–(Lafzi et al., 2018).

Before the appearance of high-throughput cell isolation methods, cells were separated by micropipetting, that is, the selection of individual cells and transfer into contained compartments, or by laser capture microdissection (LCM), where a laser attaches individual cells from a tissue to a thin film that can then be removed (Frumkin et al., 2008; Keays et al., 2005). Both of these methods assure the selection of the desired cells, but the recovery[1] is extremely low (< 100) and time-consuming.

High-throughput methods solve these problems by automating cell isolation by several mechanisms. Droplet methods, also known as bead-based methods, are commonly used. Here, two flows of liquid—one containing reagents–including lysis buffer and reverse transcriptase–and beads with poly(T) RT primers, and the other containing cells in the buffer are merged into a combined flow. This flow is separated into droplets by adding oil at set intervals. By calibrating the relative rate of the two flows and controlling the creation of droplets, it can be ensured that, in most cases, only single cells will be isolated in droplets (Svensson et al., 2018).

Droplet methods can increase the recovery by up to millions of cells. However, because of the random nature of this process, it requires a large number of cells and thus is not suited to samples with limited availability of cells. Additionally, although droplets are of nanoliter volumes, relatively large amounts of reactants are required to ensure that the co-occurrence of multiple cells in the same droplet–a doublet– is minimised, which requires a large proportion of empty droplets (Grün et al., 2015).

Although for some of these methods, such as CEL-seq or Drop-seq, the set-up for neither of these methods are commercially available, and the user is required to build a microfluidic device based on the information provided by the authors (Grün et al., 2015); other methods such as Chromium are commercially available (Zheng et al., 2017).

A similar set of cell isolation methods are microfluidic methods. There, cells are captured using integrated fluidic circuits, which currently enable an analysis of up to 96 cells per chip. These captured cells can subsequently be inspected under a microscope. The main limitations of microfluidic methods are that (1) they require a specific setup, and the final number of cells captured is relatively low compared to droplet methods; (2) cells have to be of a relatively homogeneous size since the capture sites are tuned to specific ranges, and (3) the capture efficiency can be low for sticky or nonspherical cells (Kolodziejczyk et al., 2015). For these reasons, microfluidic methods have been overshadowed by droplet methods.

1 Technically *recovery* refers to the final amount of cells whereas *throughput* is the number of cells isolated per unit time. Since *throughput* and *recovery* are related, *throughput* is usually used for the meaning of *recovery* as well (Tomlinson et al., 2013).

The last family of cell isolation methods is plate-based methods. In plate-based methods, cells can be sorted directly into 96- or 384-well plaques for subsequent single-cell sequencing. Most of these methods rely on flow cytometry, which might be useful since the parameter information can be allocated to each well (Kolodziejczyk et al., 2015). However, the main drawbacks of using plate-based methods are that (1) reactions cannot still be downscaled to nanolitre volumes, which can entail a higher reagent cost per cell (Jaitin et al., 2014)–up to 8-12$ per cell (Lafzi et al., 2018)–; and (2) the cell throughput is still relatively low compared to droplet-based methods, where the number of cells ranges in the few thousands at most.

The main advantage of plate-based methods and why some are used so far is that the library preparation methods allow for full transcriptomic capture, whereas droplet-based methods currently have a 3' enrichment bias.

CELL BARCODING AND POOL BARCODING Cell barcodes (CB) are small unique oligo fragments attached to all the mRNA molecules of a cell so that the gene expression information of that cell is unique. This step, applicable to plate-based and droplet-based methods, is necessary for the later after-cell-lysis and pooling of all the droplets into a single volume (Grün et al., 2015). Cell barcoding assignment depends on the method. For instance, in 10x chromium, CB information is appended to individual beads that collocate with a cell in an individual droplet.

CB offers a great advantage due to the possibility of sequencing large numbers of cells. However, CB imposes some limitations when creating compatible sets of barcodes. For instance, each pair of barcodes should differ in at least two positions to avoid CB collapse, that is, when two independent CBs are so similar that algorithms merge them into one when accounting for CB sequencing errors (Grün et al., 2015). Additionally, due to Poisson statistics of cell capture, to ensure that mostly single cells are isolated during random cell-isolation methods means that there will always be large inefficiencies in terms of cell isolation, and the pool of barcodes will always have to be substantially larger than the number of cells captured to avoid barcode duplication (Svensson et al., 2018).

Traditional cell barcoding is done *at once*; that is, full CBs are first synthesised, and then they are attached to the mRNAs from individual cells. Recently, a new family of methods called *pool-barcoding* was developed: sci-RNA-seq (Cao et al., 2017) and SPLIT-seq (Rosenberg et al., 2018). In these methods, single cells are pooled into mini-pools of 10–100 cells by FACS, distributed over multiwell plates with unique barcodes in each well. First-strand synthesis labels all cells in the well with a first barcode. Cells are then pooled and again randomly split into mini-pools in plate wells by FACS, and a second well-specific barcode is added. This procedure can be repeated multiple times before the final pooling of the cells to amplify the material and create a sequencing library. This results in an arbitrarily low probability that any two cells will co-locate in the same sequence of wells, so RNA from each cell is uniquely labelled, circumventing the necessity to create an inherently large set of CBs from traditional barcoding.

SPIKE-IN ADDITION RNA spike-ins are manually manufactured RNA transcripts of known sequence and quantity used to calibrate measurements in RNA-seq studies. These spike-ins can also be used in scRNAseq to normalise differences in expression across different cells by adding the same amounts of external spike-in RNAs to each sample (Lin et al., 2020). Different types of spike-ins can be added to the sample, although the most commonly used ones are the ERCC spike-ins, which consist of 92 different spike-in sequences at a wide range of concentrations (Lee et al., 2016).

Although spike-ins are widely used in certain protocols–e.g. SMART-seq2–there are certain drawbacks to their use: (1) spike-ins are typically added in at very high relative concentrations and, consequently, they take up a relatively large proportion of reads (Bacher et al., 2016); (2) their length ranges in 500-2000 nucleotides, which is shorter than the average mRNA (Stegle et al., 2015), which together with (3) their secondary structure and shorter poly(A) tail compared to endogenous mRNA, the efficiency of amplification may be lower as well (Svensson et al., 2017); and (4) external spike-ins can vary significantly even between technical replicates, which means that, apart from being an additional variable to control within the experimental design, technical-replicate-derived batch-effects may still occur (Lin et al., 2020). For these reasons, spike-in use has declined for the last years (Kharchenko, 2021).

1.1.1.2 *Reverse transcription and 2nd strand synthesis*

The amount of RNA present in a single cell is limited and ranges from 1–50 pg depending on cell type (Boon et al., 2011), with an estimated mean value of around 10 pg of DNA (Grün et al., 2015). Due to the small amounts of initial material, mRNA amplification is necessary to quantify it after sequencing.

Before amplification, because mRNA molecules are single-stranded, a second-strand synthesis is performed. To do it, a poly-T oligo with complementary information–e.g. CB or UMIs– is added in the mixture–i.e. well plates, droplets, etc.–, which binds to the poly-A 3' tail of the mRNA. There are two methods to extend this secondary strand: poly-A tailing or template switching (Kolodziejczyk et al., 2015).

In poly-A tailing, after the poly-T oligo is mixed, the first strand is synthesised with a poly-A end tail so that the poly-T oligo used in the first strand, or another poly-T oligo added at this second step, binds to that poly-A tail during the second strand synthesis. In template switching, after the first poly-T addition, a template switching oligo (TSO) is added. The reverse transcriptase *switches* template strands from cellular RNA to the TSO and continues replication to the 5' end of the TSO (Ramsköld et al., 2012). The key advantage of using TSOs is that full-length transcript coverage can be obtained, reducing 3' coverage biases originating from incomplete reverse transcription.

After reverse transcription, minute amounts of cDNA have to be amplified. This step is essential for scRNAseq, as it increases the number of cDNA copies of a gene, making it more likely to be detected by sequencing (Jiang et al., 2022). Either PCR or IVT can achieve cDNA amplification. PCR uses the effects of the Taq DNA polymerase to duplicate each de-hybridised molecule of cDNA, whereas IVT uses a T7 polymerase to transcribe the same cDNA molecules continuously.

Both methods have their advantages and disadvantages. The PCR amplification process is non-linear, which means that the ratio between the copy numbers of two differentially expressed genes is artificially distorted. This can result in the ratio becoming even larger or smaller than it originally was before PCR (Figure 4) (Jiang et al., 2022; Mereu et al., 2020). To avoid that, a balance between a lower/higher number of cycles and fewer/more copies has to be considered. On the other hand, IVT linear amplification reduces these biases since the same molecule is used in each amplification round and not molecules from previous rounds, but it requires an additional round of reverse transcription of the amplified RNA, which results in additional 3' coverage biases. Despite its disadvantages, PCR amplification is the most commonly used.

UNIQUE MOLECULAR IDENTIFIERS The main drawback of the amplification is the impossibility of differentiating two reads of different mRNAs from two reads of the same mRNA molecule derived after amplification. Unique molecular identifiers (UMI) *barcode* each individual mRNA molecule within a cell during reverse transcription, and therefore reduce this amplification bias (Grün et al., 2015; Kolodziejczyk et al., 2015). UMIs generally consist of 6 to 20 bp oligos that are usually attached at the 3' end or 5' end of the mRNA (Kivioja et al., 2012). After sequencing, the amplification bias can be eliminated by counting each label only once instead of the reads derived from all amplicons (Grün et al., 2015).

Although UMIs offer a great advantage in reducing noise resulting from amplification bias, protocols that implement UMIs sequence only the 5' or 3' end of each transcript, making them unsuitable for studies of isoforms or allele-specific expression (Bacher et al., 2016). Svensson also observed a UMI-saturation phenomenon whereby even in long (> 10bp) UMIs, the amplification bias could no the fully resolved, and absolute mRNA quantification was limited (Svensson et al., 2017).

Similar to BCs, UMIs can also suffer from different mutations or even *shifting* due to PCR stutter of low-entropy templates. This effect can induce an overestimation of transcript abundances due to shifted or mutated UMIs being considered differentiated entities. Nonetheless, computational corrections during sample preprocessing should alleviate these biases (Sena et al., 2018).

1.1.1.3 *Library construction*

Once full cDNAs have been constructed, they are sequenced. Due to the length of the amplified cDNA, it needs an additional processing step where shorter sequencer-ready fragments are built to guarantee a robust amplification due to the limitations of sequencer read length (Natarajan et al., 2019). The vast array of library construction methods is out of this thesis's scope and reviewed more in-depth in (Ziegenhain et al., 2017).

Once libraries are constructed, they are sequenced. Generally, the sequencing depth for libraries constructed for transcriptomics studies should be based on the specific research question and the desired resolution of cell type diversity. A shallow depth, such as 1000 to 10,000 reads per cell, may be sufficient for resolving major cell types, while a higher depth may be necessary for resolving cell subtypes or states (Heimberg et al., 2016). However, beyond a certain threshold, usually around 500,000 or 1 million reads, there is little change in the number of reliably expressed genes (Wu et al., 2013). Other authors state that even ~50,000 reads per cell are sufficient to distinguish relatively complex cell types (Pollen et al., 2014).

1.1.2 *Single-cell proteomic methods*

The characterisation of mRNA molecules has been traditionally easier than the characterisation of proteins. Therefore RNA and DNA multiplexed techniques became the gold standard of analysis of cell state, compared to protein multiplexed techniques, which are not as widely available. This bias in the transcriptome analysis leads to the tacit assumption of proteome and transcriptome being at similar levels when making inferences about the function of the transcriptomic data.

However, wide evidence supports that transcriptomic and proteomic profiles are ever slightly correlated. For instance, Schwanhäusser et al., 2011 determined that the correlation between mRNA and

protein abundances in mammalian cells was 0.41; and that about 40% of the variance of protein levels between different proteins could be explained by mRNA levels. A follow-up study re-analysing the same dataset with a different statistical model concluded that about 56%–84% of the protein variance could be explained by mRNA variance (Li et al., 2014).

This effect occurs for several reasons: (1) changes in mRNA concentration due to transitions between steady-states, (2) the signal delay between mRNA transcription and translation into protein, (3) translation on demand–mRNA levels are kept at a basal rate, but a signal activates the translation dedicating translation machinery from other mRNAs (differential ribosome usage)–, or (4) housekeeping genes, where protein levels are kept the same but mRNA levels change across cells because of transcriptional bursting (Liu et al., 2016).

As a result, single-cell proteomic methods are good candidates for obtaining more accurate descriptions of the activity profiles of the different cells. Current single-cell proteomic methods can be divided into three main types.

Mass spectrometry (MS)-based single-cell methods combine isolation of cells with MS. In both methods, cells are isolated in nanowells, where different reagents are mixed to perform then the MS/MS of each nanowell (Cong et al., 2021; Ctortecka et al., 2021). The main advantage of these methods is their high-resolution proteomic profile (>1000 peptides), but they suffer from low cell throughput values (~10–100 cells) and lack of robust data integration pipelines (Vanderaa et al., 2021).

The second family of methods is based on cell surface protein analysis, in which antibodies against membrane proteins are coupled with DNA tags, sequenced as the rest of the elements within the droplet (Mimitou et al., 2019; Peterson et al., 2017; Stoeckius et al., 2017). Methods from Stoeckius et al., 2017 and Peterson et al., 2017 target common immune markers, and they observe significant overexpression of certain markers in protein form compared to RNA–CD56, CD11c, CD14 by Stoeckius et al., 2017, and CD56, CD19, CD33, Cd11b and CD155 by Peterson et al., 2017–. While those methods have a higher cell throughput (~ 10,000) and are interpreted as "a wider version of FACS", the array of antibodies to be used–about 100, biased towards immune populations–is far from a gold standard for general tissue analysis.

Notably, most of these proteomic methods are technically multi-omic since they capture protein and mRNA information from each cell. Although the rise of multi-omic methods is promising, they are out of the scope of this dissertation. Stuart et al., 2019a thoroughly explains the different methods and the integration workflows.

1.1.3 *Single-cell spatial methods*

Cell location plays a fundamental role within tissue structure due to the juxtacrine and paracrine signalling. There are many dedicated methods to study the histological structure of a tissue, as well as to show the location of DNA, RNA or protein markers within it. However, most of these methods are usually suited for analysing a handful of markers.

To solve this issue, single-cell spatial techniques merge histology sampling with RNA-seq or proteomic signatures from individual cells or small areas within the tissue sample. These techniques connect the expression signatures of hundreds to thousands of genes areas specific to within the tissue, thereby providing pertinent spatial information.

Similar to scRNAseq methods, spatial transcriptomics workflows and bioinformatic frameworks to analise the data have emerged fairly recently and are an area of rapid evolution. Several reviews extensively inform about the history, methods, and pros and cons (Longo et al., 2021; Moses et al., 2022).

Currently, two main types of spatial methods exist: high-plex RNA imaging and spatial barcoding.

High-plex RNA imaging has evolved from traditional *in situ* hybridization and sequencing methods and has extended the number of targets to be observed at once. Most of these techniques–seqFISH+ (Eng et al., 2019), MERFISH (Moffitt et al., 2018) or STARmap (Wang et al., 2018b)–use fluorescent probes designed for an array of genes, which bind specifically to the mRNAs, and the probe position is identified using a sequencer. The differences between methods generally rely on the probe composition or the method of transforming sequencing information into a barcode associated with a gene. For instance, MERFISH uses a combination of fluorescence on-and-off signals to a binary code, which is then assigned as a gene, whereas seqFISH+ uses a colour-coded assignment where each gene has an associated colour sequence.

High-plex RNA methods allow for subcellular resolution and have greater depths per transcript compared to spatial barcoding methods (Longo et al., 2021). Although the theoretical number of barcode combinations would allow for a quasi-complete transcriptome characterisation–seqFISH allows for $4^8 = 65,536$ combinations and MERFISH could allow >10,000 genes–in reality, current methods still require a pre-selection of gene targets (Moses et al., 2022). For instance, STARmap offers a 160-gene panel and captures ~250 reads per cell; and MERFISH offers a 135-gene panel and captures ~100 reads per cell. Additionally, these methods suffer from more read-out noise and require specialised equipment.

On the other hand, spatial barcoding methods such as Slide-seq2 (Stickels et al., 2020), Visium (Ståhl et al., 2016) or HDST (Vickovic et al., 2019) use custom slides with immobilised probes, set up to capture spatial information. When tissue sections are deposited on the slides, the mRNA molecules of the tissue are hybridised to the probes on the slide, amplified, and sequenced. These probes are located in specific locations, called spots, and include a barcode that can be demultiplexed to provide the spatial location.

These methods show a greater field of view and greater coverage in the number of genes that can be detected. However, they show 2 main disadvantages. First, the spot diameter–2 μm for HDST, 10μm for Slide-seq and 55μm for Visium–tends to include several cells, and thus, transcriptomic data do not present single-cell resolution. In complex tissues with intertwined, diverse cell types, this resolution is not enough–although novel iterations such as Visium HD claim near subcellular resolution–. The second drawback relies on the lack of resolution. Although these methods promise an unbiased transcriptomic representation, there is a strong bias towards housekeeping genes, and only a handful of marker reads are captured (Schäbitz et al., 2022).

1.2 PROCESSING OF SCRNASEQ DATASETS

Once single-cell RNA sequencing (scRNA-seq) data has been acquired; the next step is to process it for analysis. While some aspects and tools used for scRNA-seq data processing are similar to those

used for bulk RNA-seq data, the unique structure of scRNA-seq data requires specific methods and tools to address the specific challenges it poses (Adil et al., 2021).

A crucial decision before data processing is the choice of programming language, which will impact the selection of analysis tools. Many tools are designed in a specific programming language (Luecken et al., 2019). The most popular languages for scRNA-seq analysis are R and Python. However, cross-environment support is increasing (Scholz et al., 2018).

In recent years there has been an explosion of different tools for scRNA-seq analysis. According to the *scrna-tools* website (Zappia et al., 2018), as of the writing of this text, there are more than 1400 tools available; 48% developed in R, 36% in Python, and 16% in other languages. Additionally, there are platforms such as Seurat (Butler et al., 2018), Scater (McCarthy et al., 2017), or Scanpy (Wolf et al., 2018) that provide integrated environments for developing pipelines and contain extensive analysis toolboxes. Most of these platforms, and their underlying methods, try to address scalability to the higher numbers of cells being sequenced and the higher resolutions per cell (Lähnemann et al., 2020).

In this section, we will expand upon the steps involved in a single-cell analysis pipeline, presented in their recommended order of execution–some of these are shown in Figure 5. However, it is worth noting that the order may vary in some cases. To clarify, we will denote the essential steps for the analysis with an asterisk (*). The remaining steps are recommended based on the data quality or other constraints, or they may be downstream steps following clustering.

1.2.0.1 *Technical aspects of single-cell data: noise and zeros*

Before delving into the processing of single-cell data after sequencing, it is important to acknowledge common issues about noise and high proportions of zeros in scRNA-seq data and other single-cell technologies, which are well-known limitations of single-cell methods. scRNAseq data may appear similar to those obtained from bulk expression experiments, but the limited amount of starting material (<10 pg) and higher resolution in scRNA-seq can result in distinct characteristics, such as a large number of zeros (both biological and technical), greater variability, and complex patterns of gene expression (Bacher et al., 2016; Saliba et al., 2014). While the high levels of noise in scRNA-seq data can be offset by the larger number of cells typically analysed in these studies, compared to bulk RNA-seq studies, it remains a significant challenge (Stegle et al., 2015).

Even deeply sequenced datasets might have 50% zeros, whereas shallowly sequenced datasets can have 99% zeros. By contrast, in typical bulk RNA sequencing datasets, <20% of data entries are zeros (Andrews, 2020; Haque et al., 2017; Jiang et al., 2022). The levels of noisiness and zero presence can greatly affect downstream analysis because of a gene having up to hundreds of counts in one cell and none in a similar cell. It is observed that even cells of the same type can show substantial gene expression variability leading to technical variations–subtypes–that are artefacts derived from the noisiness (Eldar et al., 2010; Munsky et al., 2012). This effect is less pronounced in genes with a high read count, which shows similar expression profiles across similar cells (Heimberg et al., 2016).

Recently, Jiang et al., 2022 reviewed the sources of zeroes in scRNA-seq data by analysing the steps of mRNA transcription and library preparation, as outlined in Figure 6. Zeroes in scRNA-seq data can be divided into biological and non-biological. Biological zeroes contain valuable information about cell states, while non-biological zeros include technical zeros, which occur during the preparation

of biological samples for sequencing, and sampling zeros, which arise due to limited sequencing depths. These non-biological zeroes can lead to false negative results because they are difficult to distinguish from biological zeroes without additional knowledge.

Of note, transcriptomic noise also comes from the same sources as zeros; thus, we could use the terms "biological noise" and "technical noise". However, to avoid confusion, we will only explain the outcome of zeros as the reference of the section. Additionally, the single-cell community also uses the term *dropout*, a data-driven concept, to refer to zeros. However, the use of this term has been inconsistent and could be referred to all zeros (Gong et al., 2018; Pierson et al., 2015), non-biological zeros (Li et al., 2018b; Ran et al., 2020), or even non-biological zeros and low expression measurements (Zand et al., 2020; Zhu et al., 2018). Thus, we will avoid the use of this term throughout the dissertation.

Biological zeros in scRNAseq data reflect the true absence of a gene's transcripts or messenger RNAs (mRNAs) in a cell. Biological zeros occur for two reasons: First, the gene is unexpressed in that cell, cell state or cell type, because it is not part of its transcriptomic program (gene 1 in Figure 6). Second, the gene is temporarily not expressed due to the transcription bursting process (gene 2), a well-known phenomenon in gene regulation (Raj et al., 2006; Raser et al., 2004; Suter et al., 2011). mRNA processing is transcribed intermittently due to the stochastic effects of transcription factor binding, chromatic conformation, and other factors and, as such, a gene switches between active and inactive states, and its transcription only occurs during the active state (Paszek, 2007).

Non-biological zeros reflect the loss of information about truly expressed genes due to the inefficiencies of the technologies employed from sample collection to sequencing. There are two types of non-biological zeros (Silverman et al., 2020): technical and sampling zeros.

Technical zeros arise from library-preparation steps before cDNA amplification (gene 3). One cause of technical zeros is the imperfect mRNA capture efficiency in the reverse transcription (RT) step from mRNA to cDNA. The efficiency varies widely across protocols and may be as low as 20% (Islam et al., 2014; Kolodziejczyk et al., 2015). For example, if an mRNA transcript has an intricate secondary structure or is bound to proteins, it might not be reversely transcribed to cDNA efficiently (Kharchenko et al., 2014; Svensson et al., 2017), and thus would not be amplified nor sequenced.

The other type of non-biological zeros, sampling zeros, occurs due to a constraint on the total number of reads sequenced, i.e., the sequencing depth determined by the experimental budget and sequencing machine. During sequencing, cDNAs are randomly captured and sequenced into reads. Hence, this random sampling makes a gene with fewer cDNAs (gene 4) inefficiently amplified (gene 5) more likely to be undetected. As a result, the genes with small cDNA proportions in the sequencing library are likely to be missed by sequencing and thus result in sampling zeros, as showcased in Figure 4.

If all the steps go "correctly", a sufficiently expressed gene is retro-transcribed, amplified and sequenced, then it will show a non-zero count (gene 6). However, the biases associated with each step may alter the final number of retrieved reads and increase the noise of the non-zero counts. Some of these biases may be lessened, for instance, by using UMIs to minimise amplification biases in non-zero gene expression measurements (Islam et al., 2014).

Svensson, 2020 shows that non-zero-inflated distributions can describe the variation in droplet scRNA-seq data using droplet-based ERCC spike-in data. Jiang et al., 2022 replicated this study and

observed that, for samples from Drop-seq and 10x Genomics, read distribution could be fitted with non-zero-inflated distributions (Poisson and NB) correctly, whereas, for samples from Smart-seq2 data, half of the genes' read distribution could be better fitted with zero-inflated distributions (ZIP and ZINB). These results confirm that the main source of the increase in the proportion of zeros is purely technical.

Interestingly, Qiu, 2020 presents a new approach for analysing transcriptomic variability in single-cell RNA sequencing data. The authors propose using only binarised counts, where all non-zero counts are truncated to one. This method eliminates the need for normalising individual cells' sequencing depths and suggests that zeros in the data are biologically meaningful. The binarisation process removes expression differences between highly and lowly expressed genes, making it less ideal for certain analyses, such as clustering. It also highlights co-expression patterns that can be used for marker gene selection and gene network construction.

1.2.1 *FASTQ processing* *

The output of sample processing and library construction for scRNAseq is a FASTQ file that contains raw reads from the sequencer–or a derived format that can be ultimately transformed into a FASTQ–. This step aims to create a matrix of cells[2] and corresponding gene expression counts. Each cell of the matrix indicates the number of UMIs mapped for a specific gene expressed in that particular cell. This step is crucial and depends entirely on the sample quality (Adil et al., 2021).

The common pipeline for FASTQ processing involves several steps: (1) Quality control (QC) on FASTQ files to remove low-quality reads and trim adapters. (2) Mapping the reads to a genome using an aligner or a pseudoaligner. (3) Demultiplexing CBs and/or UMIs–if they are used–and then (4) counting the UMIs assigned to each CB to construct the matrix by setting specific thresholds to differentiate empty droplets from filled ones, in case they were used (Luecken et al., 2019).

The alignment in step (2) can be done with aligners such as STAR (Dobin et al., 2012), or with pseudoaligners such as *alevin* (Srivastava et al., 2019), or *kallisto* (Bray et al., 2016). The difference between an aligner and a pseudoaligner is that the latter utilises a k-mer-based search to identify genomic locations that are likely to be transcribed and then aligns the RNA-seq reads to those locations, allowing for the detection of novel or unannotated transcripts. For a proper alignment, a minimum remaining read length (> 35 bp for mouse or human) after trimming should be required to avoid false-positive hits (Grün et al., 2015). The advantage of pseudoaligners is that they are faster and usually are constructed using the whole genome, or the extended transcriptome, thus allowing for the capture of mRNA isoforms.

The objective of demultiplexing in step (3) is to *collapse* reads with similar CBs to the same cell and similar UMIs to each of the original mRNAs the UMI came from. Due to amplification or sequencing errors, CBs and UMIs that are *similar* are *collapsed* into one, to avoid false negatives–i.e. reads that belong to a specific cell/mRNA but which, due to the errors, would be assigned to another cell or as another mRNA–(Parekh et al., 2018).

2 Technically barcodes because there might be instances of more than one cell with the same barcode due to an artefact during processing.

Although these steps can be done individually with different packages–e.g. bustools (Melsted et al., 2019) or zUMIs (Parekh et al., 2018) for step (3) and DropletUtils (Griffiths et al., 2018) for step (4)–, other pipelines perform the pre-processing from beginning to end–e.g. CellRanger for 10X chromium data, kallisto (Bray et al., 2016) for droplet-based data of specialised pipelines from nfcore for droplet-based and plate-based data–.

1.2.2 *Sample QC **

In addition to FASTQ quality control, filtering out low-quality cells and genes in scRNAseq protocols is essential. The preparation of scRNAseq libraries may capture stressed, broken, or dead cells, which we refer to as "low-quality" cells. These cells can lead to misinterpretation of the data and should be removed from the analysis (Figure 5a) (Ilicic et al., 2016).

QC on cells is commonly performed by identifying and removing cells with higher mitochondrial genes, genes with higher non-exonic reads, high proportions of unmapped or multi-mapped reads, and a low number of counts and genes assigned per barcode (Andrews, 2020). The distributions of these QC covariates are analyzed to detect outliers, which are then removed by setting threshold values. These outliers may correspond to dying cells, cells whose membranes are broken, or doublets (Luecken et al., 2019). These variables are considered during cell QC because it is hypothesized that when a cell's membrane is broken, cytoplasmic RNA is lost, while RNAs enclosed in the mitochondria and nucleus–pre-mRNA with exonic reads–are retained (Ilicic et al., 2016). Removing these deceptive cells improves the quality of PCA and a higher number of significant differentially expressed genes (Ilicic et al., 2016).

There are several aspects to consider of QC on cells. First, there is no gold-standard threshold for each of the variables, and thus cut-offs are manually selected or arbitrarily defined in terms of absolute deviations of the distributions of quality metrics (Andrews, 2020). Especially in samples with highly heterogeneous cell types, these cutoffs should be chosen carefully (Amezquita et al., 2019).

Also, "sufficient data quality" cannot be determined *a priori*, so it is judged based on a downstream analysis performance–e.g., cluster annotation–. Thus, it may be necessary to revisit quality control decisions multiple times when analysing the data. Often it is beneficial to start with a permissive QC threshold and investigate the effects of these thresholds before going back to perform more stringent QC. If datasets are known *a priori* to be of bad quality, stringent QC may be necessary to avoid false positives (Luecken et al., 2019).

Additionally, if there are uncertainties about the data quality, some aligners provide the fraction of reads that map back to the genome. If this value is low, it might indicate that RNA has degraded (Stegle et al., 2015) and therefore, any analysis will yield bad quality results.

Lastly, considering any of these three QC covariates in isolation can lead to the misinterpretation of cellular signals. For example, cells with a comparatively high fraction of mitochondrial counts may be involved in respiratory processes. Likewise, other QC covariates also have biological interpretations. Cells with low counts and/or genes may correspond to quiescent cell populations, and cells with high counts may be larger in size (Luecken et al., 2019).

Regarding QC on genes, raw count matrices typically contain over 20,000 genes. It is important to filter out genes that are not expressed in a significant number of cells, as they do not provide meaningful information about cellular heterogeneity. Like cell QC, this step is performed arbitrarily, and the threshold values should be reevaluated based on downstream analysis results and the overall quality of the cells (Luecken et al., 2019).

1.2.3 *Dataset denoising and imputation*

As commented on section 1.2.0.1 single-cell datasets show higher levels of zeros and noise in the data. Especially regarding zeros, and considering that many downstream methods do not alter the zeros, it is tempting to assume that they represent missing values and to fill in an estimate derived mathematically from the detected transcripts. In principle, removing zeros could reduce the noise and make it easier to identify the underlying structure of the data (Andrews, 2020); however, we can never be sure which of the observed zeros represents "missing data" and which accurately represents a true absence of gene expression in the cell (Lähnemann et al., 2020).

Dataset denoising, also named imputation–as used in traditional ML methods to fill gaps in the data–is applied in single-cell data in response to the previously mentioned sparsity. Denoising estimates each gene's true expression level in each cell (Figure 7).

A vast array of imputation methods are grouped into three main groups (Hou et al., 2020; Lähnemann et al., 2020). (1) model-based methods that use probabilistic models to model the sparsity in the data directly and may or may not distinguish between biological and technical zeroes–SAVER (Huang et al., 2018), scImpute (Li et al., 2018b)–; (2) data-smoothing methods that adjust all values by smoothing or diffusing the gene expression values in cells with similar expression profiles–MAGIC (Dijk et al., 2018), kNN-smoothing (Wagner et al., 2017)–; (3) data-reconstruction methods that use techniques such as (a) low-rank matrix-based methods and (b) deep-learning methods to identify a latent space representation of the cells and then reconstructs the observed expression matrix from the low-rank or estimated latent spaces, resulting in a non-sparse matrix–DCA (Eraslan et al., 2018), scVI (Lopez et al., 2018). The latter methods generally strongly rely on the choice of hyperparameters, which can deter users from using them (Lähnemann et al., 2020).

Imputation methods have been shown to improve the estimation of gene–gene correlations (Luecken et al., 2019) and, depending on the amount of sparsity, could potentially improve the result of dimension reduction, visualisation and clustering applications (Lähnemann et al., 2020). However, these methods suffer from the circularity that arises when imputation solely relies on information internal to the imputed dataset, which frequently results in many false-positive signals being introduced by distorting that underlying structure (Andrews et al., 2018; Lähnemann et al., 2020). As a result, spurious structural patterns in low dimensional representations of imputed data arise, greatly affecting downstream processing (Hou et al., 2020). For instance, distortions in expression distributions may cause imputed data to violate assumptions of some statistical tests and a larger number of false-positive DEGs appear (Andrews et al., 2018).

There is currently no consensus on how denoised data should be used based on its pros and cons. Due to the modification of the data, it is still recommended not to use it for steps such as DEG, but it can be used to improve the structure of plots of dimensionality reduction methods (Luecken et al., 2019).

1.2.4 *Ambient gene expression removal*

Ambient gene expression refers to counts that do not originate from a barcoded cell but from other lysed cells whose mRNA contaminated the cell suspension before library construction (Luecken et al., 2019). As a result, many cell types may express genes from the lysed cells, resulting in a false positive signal (Angelidis et al., 2019).

To remove ambient expression profiles in scRNAseq data, cells that are unlikely to represent intact individual cells should first be excluded (Andrews, 2020). Tools like *EmptyDrops* first estimate the background levels of RNA present in empty wells or droplets and then identify cell barcodes that significantly deviate from the background, which indicates the presence of a cell (Lun et al., 2019).

A secondary step is to directly correct the count data of contaminated cells with tools such as *SoupX*, which models the ambient RNA expression and removes it from the contaminated cells (Young et al., 2020).

1.2.5 *Doublet removal*

Doublets are a common artefact in single-cell RNA sequencing experiments caused by cell sorting or capture errors. Doublets are created by capturing two cells instead of one and can be particularly prevalent in droplet-based protocols (Amezquita et al., 2019). For instance, in certain methods, inferred rates ranged from 2% (2500 cells) to 8% (10,000 cells) (Lafzi et al., 2018). Doublets can lead to misinterpretation of results, as they may be mistaken for intermediate populations or transitory states that do not exist.

To remove doublets, the best way is to perform cell isolation at higher dilutions, resulting in increased reagent costs. Computationally, traditionally–and often nowadays–cells with higher gene counts were removed. However, it has been observed that this method is not correct due to the assumption that doublets will express a more diverse array of genes than singlets; however, the number of counts per cell is compositional, limited by the capacity of the sequencer (Andrews, 2020).

Tools such as *scrublet* (Wolock et al., 2019a) and *DoubletFinder* (McGinnis et al., 2019) simulate possible doublets from the dataset itself and then calculate the similarity of real droplet barcodes to the simulated doublets and define a threshold to distinguish the inferred doublets from the assumed singlets (Andrews, 2020; Luecken et al., 2019).

1.2.6 *Data integration **

Batch or data integration responds to the necessity of joining datasets across batches that arise from various biological and technical sources–e.g. the time of the experiment, the person carrying out the experiment or differences in reagents–(Andrews, 2020). If not properly accounted for, batch effects can be mistaken for true biological signals, but through careful experimental design, they can be avoided altogether (Baran-Gale et al., 2017). Data can be integrated by keeping genomic features or cells as anchor (Argelaguet et al., 2021).

Keeping genomic features as an anchor, also known as *horizontal integration*, occurs in experimental designs where the same data modality is profiled from independent groups of cells–e.g. two samples

of the same scRNAseq experiment–. This horizontal integration is also known as *batch-effect correction*, although this word has different uses depending on the authors. In some cases, batch-effect correction refers to the correction of samples from the same experiment or with a high degree of similarity (Luecken et al., 2019), whereas for others, it is considered for all scenarios where horizontal integration applies (Argelaguet et al., 2021), or there is no clear distinction (Andrews, 2020). In this work, we will use either *batch-effect correction* or *horizontal integration*.

To perform horizontal integration, some authors claim that in datasets with high similarity–e.g. samples from the same experiment–it can be performed with linear correction methods derived from bulk methods such as *ComBat* (Johnson et al., 2006), which outperforms methods specifically designed for single-cell such as MNN (Büttner et al., 2018; Luecken et al., 2019); while other authors state the opposite, mainly because these methods implicitly assume identical (or at least known) cell type composition across batches (Andrews, 2020; Haghverdi et al., 2018).

Most horizontal integration methods developed for single-cell data rely on nonlinear–or locally linear–strategies. These methods can be classified into three types depending on the output types: a corrected feature matrix–MNN/fastMNN (Haghverdi et al., 2018) and Seurat CCA (Stuart et al., 2019b)–, a batch-corrected embedding–Harmony (Korsunsky et al., 2019) and Scanorama (Hie et al., 2019)–, or an integrated cell-cell similarity graph–e.g. *bbknn* (Polanski et al., 2020) and *Conos* (Barkas et al., 2019). An extensive comparison of horizontal integration methods is available at (Luecken et al., 2021). Depending on the output of the integration method, the downstream steps order will differ.

Batch-effect correction methods pose several challenges, especially non-linear methods. First, a classical problem of nonlinear integration methods is overcorrection; that is, the algorithm forcibly merges nonmatching subpopulations of cells. An optimal method should be able to detect this and prevent the merging of datasets when no common biological variation exists. Second, high-dimensional observations–e.g. gene expression counts–can be severely distorted due to the batch alignment, and other downstream gene-based analyses, such as gene marker detection or differential expression analysis, can be affected. When extensive biological variability exists across batches–e.g. when samples are profiled across a developmental time course–disentangling batch effects from the underlying biological signal of interest is more challenging, and the previous two undesired effects become more apparent.

To assess the efficacy of integration, *de visu* observations on low-dimensionality representations are usually performed. With them, samples that are not integrated or cell types in which integration is more complex are discerned. This assessment can also be done quantitatively using different tools such as kBET, which performs a statistical test of whether the label composition of a kNN of a cell is similar to the expected (global) label composition (Büttner et al., 2018).

Other types of data integration are *vertical integration*, where cells are kept as anchors, and it is applied in experimental designs with multiple data modalities-e.g. integrating scRNAseq data with single-cell methylome data from the same samples–; or *diagonal integration*, if cells and features are different–e.g. integrating scRNAseq data with single-cell ATAC seq data and spatial data from different samples–. None of these scenarios is covered in the work of this thesis and are extensively described, alongside integration strategies, by Argelaguet et al., 2021 and (Longo et al., 2021).

1.2.7 *Normalisation **

Each count in a count matrix represents the successful capture, reverse transcription and sequencing of a molecule of cellular mRNA, as shown in Section 1.2.0.1. Count depths for identical cells can differ due to the variability inherent in each of these steps (Luecken et al., 2019; Stegle et al., 2015). Thus, when gene expression is compared between cells based on count data, any difference may have arisen solely due to sampling effects. Additionally, reads obtained from a sequencing experiment will vary between cells, due to cell cycle stage or other biological factors, even within the same cell type, and these differences may need to be corrected as well (Andrews et al., 2019). Therefore, a normalisation procedure to account for these effects is necessary.

Most basic normalisation methods in scRNAseq data are inherited from traditional methods of bulk RNA-seq (Luecken et al., 2019). These approaches share the same motivation—to bring cell-specific measures onto a common scale by standardising a quantity of interest–e.g., total read counts per sample–across cells while assuming, for example, that most genes are not differentially expressed. One of these methods is counts per million (CPM), which consists of dividing the number of counts of a gene in a cell by the number of counts in the cell and then multiplying by 1,000,000 (Mortazavi et al., 2008). Variations of this method scale the size factors with different factors of 10 or by the median count depth per cell in the dataset.

Other less-used traditional methods exist, like the downsampling protocol, which is the process of random sampling reads or counts from the data to leave all cells with a pre-specified number of counts or fewer. While downsampling throws away data, it also increases technical zeros and noise, which CPM and other global scaling normalisation methods do not. Other methods, like RPKM (Reads per kilobase per million mapped reads) or account for the length of the gene, although this method is only useful for full-transcript scRNAseq methods (Kowalczyk et al., 2015; Soneson et al., 2018).

All these previously mentioned methods are called global scaling/normalisation methods because they apply a common transformation by assuming that all cells in the dataset initially contained an equal number of mRNA molecules and count depth differences arise only due to sampling. There is some debate on using these methods, based on the preference between whether all genes should be weighted equally for downstream analysis or whether the magnitude of gene expression is an informative proxy for the importance of the gene (Luecken et al., 2019; Street et al., 2018).

Some methods are designed for scRNAseq data, like *scran* (L. Lun et al., 2016), which achieves a robust outcome using pools of cells to estimate size factors. Cells are normalised in each pool, and the size factors used in the normalisation are used to build a system of linear equations to define individual cell factors. Alternatively, spike-in RNAs from the External RNA Control Consortium or housekeeping genes can be used to estimate size factors. These, as well as the previous methods, are linear normalisation methods because they perform linear transformations on the data (Luecken et al., 2019).

Although useful, it has been observed that non-linear normalisation methods have been shown to outperform global scaling and linear methods, especially in situations with strong batch effects (Cole et al., 2019). A new set of normalisation methods was recently created based on fitting counts to a model. The methods by Mayer et al., 2018 and Hafemeister et al., 2019 fit an NB model to count data, using technical covariates such as the read depth and the number of counts per gene to fit

the model parameters, and the residuals of the model fit to serve as a normalised quantification of gene expression.

An additional step after normalisation is applying a variance-stabilisation transformation. The most common transformation is log1p, also known as log(1+x), adding a pseudo count to the normalised counts and applying the logarithm of that value. Adding a pseudo count is necessary for zero counts so that their transformed value is 0, not $-\infty$. This transformation has three important effects. Firstly, distances between log-transformed expression values represent log-fold changes, which are the canonical way to measure changes in expression. Secondly, log transformation mitigates (but does not remove) the mean-variance relationship in single-cell data (Brennecke et al., 2013)[3]. Finally, log transformation reduces the skewness of the data to approximate the assumption of many downstream analysis tools that the data are normally distributed (Luecken et al., 2019).

This usefulness is highlighted theoretically and practically. Theoretically, (1) when performing PCA on the gene expression matrix to find a reduced-dimensional representation that captures the variance, it is desirable that all genes contribute equally, which is the case of log transformation; and (2) it converts multiplicative relative changes to additive differences. In the context of PCA, this allows for interpreting the projection axes in terms of a relative, rather than an absolute, abundance of genes (Booeshaghi et al., 2021). Practically, log transformation is useful in downstream applications for differential expression testing (Finak et al., 2015) or batch correction (Büttner et al., 2018; Johnson et al., 2006), which commonly use the log transformation for these purposes.

The use of log-transformation, however, poses two main problems (Booeshaghi et al., 2021): (1) differences in log1p transformation are apparent for lowly-expressed genes, which can introduce spurious differential expression effects into the data (Lun, 2018); and (2) highly-expressed genes, despite being thoroughly-expressed, still have higher variance compared to genes with lower numbers of counts, so it is not a completely variance-stabilising function. To minimise this gap, a lower factor–instead of 1,000,000–can be used during normalisation and a larger pseudo count during log-transformation (Lun et al., 2019).

Although less popular, other variance-stabilising functions can be used such as the square root transformation ($\sqrt{x}$) or their related transformations such as the Anscombe transformation $2\sqrt{x+3/8}$ (Anscombe, 1948) and the Freeman-Tukey transformation $\sqrt{x}+\sqrt{x+1}$ (Freeman et al., 1950). Normalising UMI counts by sequencing depths followed by one of the square-root transformations has been advocated for UMI data processing (Wagner, 2020a,b). However, neither of these transformations is sufficiently variance stabilising (Lause et al., 2021).

1.2.8 *FS*

Feature selection (FS) is the process of choosing a subset of features–or variables–from the model– or data– to reduce the dimensionality of the data and improve the downstream efficacy of a machine learning (ML) pipeline or any other method (Saeys et al., 2007). The use of FS within ML is advised for a series of reasons: (1) FS allows for a simplification of the model, which reduces training times; (2) avoids overfitting and improves model performance; (3) allows for an easier interpretation of the model by the researcher, (4) may be required to fit within the input space of the training algorithm;

3 The aim behind the choice of a variance-stabilising transformation is to transform the values such that the variability of the transformed values is not related to their mean value (Bartlett, 1947; Yeo, 2000)

(5) may reduce noise in the model, and therefore improve the accuracy of the training algorithm and (6) reduces the curse of dimensionality (COD).

Alongside FS, feature extraction (FE) is also typically used in ML. FE differs from FS in that while FS selects a subset of original features from the model, FE creates a new subset of features by modifying and combining the original features. One example of FE is PCA, which is described in Section 1.2.9 of the Introduction. In this case, PCA is a FE method because a new set of variables–components–is chosen by finding the best linear combination of the original features that maximises the variance (Jolliffe et al., 2016).

1.2.8.1 *The curse of dimensionality (COD)*

One of the reasons to use FS and FE methods is the curse of dimensionality (CODÇ). The rationale behind the COD is that, when looking at the distribution and number of equally-spaced points, the higher the dimension, the exponentially higher the number of points needed to fill that space evenly. For instance, if we consider a line (1D) with 10 points, the number of points in a plane (2D) will be 100, and 1000 in a cube (3D). Therefore, to conserve the same density of points, we will need 10 times the amount of points from the previous dimension for each increment in the dimension. If we do not keep up with this increment, the density of points becomes much sparser across the space, and the dataset becomes "less informative".

Another studied phenomenon that relies on COD and justifies the necessity of FS and FE is the *peaking phenomenon* (Hughes, 1968). This phenomenom states that in a dataset with a fixed number of training samples, a trained algorithm's predictive power, or classification accuracy, first increases with the number of features but then decreases, indicating a number of features at which this algorithm performs best. This effect goes against the primary intuition that, with more information, the algorithm should perform better. An example of how COD affects DR and clustering is observed in Figure 8.

1.2.8.2 *The need for FS in scRNAseq datasets*

Gene expression datasets are highly dimensional, as the expression of tens of thousands of genes is measured in any given experiment. A direct consequence is that these datasets suffer from the COD, resulting in a sparsity of the data (Bellman, 2015). Additionally, this sparsity is exacerbated by the low capture efficiency of mRNA in single-cell experiments, owing to the tiny amounts of genetic material to be amplified, even though there are considerable recovery differences across methods (Bzdok et al., 2018).

The phenomenon of COD has been recently studied in scRNAseq data (Imoto et al., 2022). It was observed, in several datasets, that the use of higher dimensions showed a loss of closeness during hierarchical clustering–samples were mixed–, PC contribution rates decreased drastically, 2D PCA plots showed less separation of samples, and UMAP plots failed to discern many samples. These results were partially explained due to the noise within scRNAseq datasets, which, at higher dimensions, exacerbate the consequences of the COD. Thus, only a limited amount of genes showed a relevant expression across many cells, and many were only expressed in a few.

As a result of the COD, methods that rely on distances, such as the kNN graph construction, may yield contradictory results because the distances between similar and dissimilar points collapse. One way

to minimise the effects of COD is to use a FS or a FE method–or a combination of both–to reduce the dimensionality of the data.

Early methods for FS in scRNA-seq data were based on the idea that genes whose expression shows a greater dispersion across the dataset are the ones that best capture the biological structure of the dataset (Brennecke et al., 2013; Osorio et al., 2019). Conversely, genes that are evenly expressed across cells are unlikely to define cell types or cell functions in a heterogeneous dataset and should not be considered relevant. The most straightforward way of selecting genes that are not evenly expressed is to look at a measure of the dispersion of each gene's counts and select those genes that have a dispersion over a threshold.

Newer FS methods have arisen after different corrections, like the one originally described by Satija et al. implemented in Seurat (Satija et al., 2015), later adapted to *scanpy* (Wolf et al., 2018). This method calculates the average expression and dispersion (variance/mean) across all cells; and the genes are placed into 20 bins based on their average expression. For each bin, a z-normalised measure is calculated to correct for deviations in the dispersion, and the top genes with the highest measure across all bins are selected.

Another dispersion-based FS method is the one implemented in *scry* (Townes et al., 2019). In their study, the authors used a GLM-PCA model instead of the standard PCA to avoid the implicit assumptions of PCA and selected the genes based on the binomial approximation to the multinomial deviance.

1.2.8.3 *Metrics based on the proportion of zeros to discriminate valid features*

Early studies observed that the read distribution of most of the single-cell studies could be fitted to a negative binomial (NB) (Vieth et al., 2017). More specifically, read counts produced a zero-inflated bimodal distribution, whereas UMI counts produced a NB distribution (Chen et al., 2018b; Kharchenko, 2021). These results were later replicated by Svensson, stating that the proportion of zeros in droplet-based scRNAseq data, originally assumed to be technical zeros, was tightly related to the mean expression of genes, following a NB curve (Svensson, 2020).

From this observation, Svensson, 2020 concludes that genes with an expected lower percentage of zeros tend to have an even expression across the entire set of cells. Conversely, genes with a higher-than-expected percentage of zeros might possess biological relevance because they are expressed in fewer cells than expected, and these cells might be associated with a specific cell type or state.

This finding opened the path for new FS methods that would rely on genes that showed a greater-than-expected proportion of zeros, according to their mean expression. These methods are based on a null distribution of some dataset property, and genes whose behaviour differs from the expected are selected.

Two methods following this premise were developed by Andrews et al., 2018: *nbumi* and *m3drop*. Both methods work by fitting different zero-count probability distributions. After parameter fitting, they used a t-test with the expected values based on the fitting equation to calculate the p-value associated with a gene with a mean expression having a higher proportion of zeros than expected.

1.2.8.4 *triku*

The FS method used in most of the proceedings of this thesis is *triku*, which was developed during the thesis period (Ascensión et al., 2022). In summary, *triku* favours features that are *locally-expressed*, that is, that a set of transcriptomically-related cells expresses it. In contrast, sparsely expressed features on transcriptomically-unrelated cells, or widely expressed in most of the cells, are not favoured for selection.

To perform the selection, for each gene *triku* calculates two distributions: (1) the kNN count distribution, that is, for each cell, its gene expression and the expression values of the kNN are summed, and the distribution refers to the kNN counts across cells; and (2) the null distribution, which would resemble the distribution of reads using a randomised kNN graph.

If a gene is locally expressed, its kNN count distribution will be highly skewed because most of the counts are condensed in a set of correlated cells, whereas, in the rest of the scenarios, this skewness is not that visible. Then, the (1) and (2) distributions are compared using the Wasserstein distance, which increases with the divergence between the two distributions. This distance is later corrected for average expression values, and the features with higher distances are selected. A scheme of the algorithm can be found in Figure 9.

1.2.9 *Summarisation dimensionality reduction (PCA) **

After FS, dimensionality reduction (DR) is often performed as the next step. This step involves generating a matrix with the same number of cells but with a reduced number of features, typically between 2 and 50, which capture the underlying structure in the data. Embedding the expression matrix into a low-dimensional space is useful for several reasons and is typically done to achieve one of two objectives: visualisation or summarisation (Luecken et al., 2019). In the context of single-cell data, summarisation methods are often used as an intermediary step before using visualisation methods. These methods aim to avoid major distortions in the data but are ineffective at visualising the data in 2D or 3D space, due to the large diversity of cell types. On the other hand, visualisation methods are useful for capturing the heterogeneity of the data in a 2D or 3D space, but at the expense of high distortions, which may strongly bias further steps such as differentially expressed gene (DEG) calculation or trajectory inference (Luecken et al., 2019; Xiang et al., 2021).

One of the most used summarisation methods is principal component analysis (PCA). PCA aims to find a limited number of linear combinations[4] of transcripts that capture as much variance in the dataset as possible. In that context, capturing variance stands for emphasising variation and similarity and bringing out strong patterns in a dataset.

Thus, the new linear combinations, called principal components (PCs), describe a low-dimensional space where the main transcriptional variation between cells occurs. The number of PCs determines the reliability of the information preserved; that is, a higher number of PCs will capture more information about the data; but will require more memory to save and time to compute and may

4 A linear combination is a mathematical operation that involves taking a set of numbers (coefficients) and multiplying them by a set of variables (genes) and then summing the results: $\mathrm{LC} = a_1 \cdot gene_1 + a_2 \cdot gene_2 + \cdots + a_n \cdot gene_n$. Linear combinations are useful because they can represent a wide range of mathematical objects concisely and computationally efficiently. They can express a new vector as a linear combination of basis vectors. They are also used in many areas of mathematics, science and engineering and can model a wide range of real-world phenomena.

capture unwanted elements like noise derived from "non-relevant" genes. Therefore, the number of top PCs is usually selected by plotting the fraction of variance explained by each component and then identifying the point where the curve makes a sharp bend, often referred to as the 'knee', and keeping only those components above the knee. This process can be done manually or automatically (Kharchenko, 2021; Luecken et al., 2019). Additionally, a number between 25 and 50 components is usually selected as a rule of thumb.

It has been observed that the first few principal components (PCs) often encode undesired effects such as batch effects, size effects or cell cycle due to poor upstream correction. While these components can be removed to eliminate these effects, it is not recommended as they may also contain relevant biological variability. A common practice is to use this step to identify these patterns and make the necessary upstream corrections (Barron et al., 2016).

PCA, similar to other methods, can be useful, although it has some limitations that must be considered. These limitations can be broken down into two categories: technical issues related to the computation of PCA and data-related issues stemming from the mathematical limitations of PCA concerning the structure of scRNA-seq data.

On the technical aspect, although PCA computation is deterministic–i.e. the computation follows a set of well-defined rules and is guaranteed to produce a specific and predictable outcome–its calculation is only feasible with small matrices–the product of the number of cells and genes is smaller than 10^7–, due to time and memory constraints (Tsuyuzaki et al., 2020). For larger matrices, if the data matrix can be loaded onto memory as a sparse matrix [5], some implementations for these algorithms are available (in-memory and sparse matrix), but they may require very long calculation times. Other algorithms, such as SVD update, repeatedly perform SVD–a *parallel* of PCA–using subsets of the data sampled from the data matrix and incrementally updating the result. In both cases, these implementations are heuristic and might overlook some important differential gene expression patterns. In the case of large-scale scRNA-seq studies aiming to find novel cell types, this property may cause a loss of clustering accuracy and not be acceptable (Tsuyuzaki et al., 2020).

Since PCA preserves Euclidean distances between cells, one of the biggest assumptions is that data follows a Gaussian distribution structure. Although some of the transformations performed as a first step to PCA like *centring the data*–i.e. subtracting by the mean and divided by the variance–try to turn the data to fit into a Gaussian-like distribution, in reality, single-cell data fits better into an NB distribution (Section 1.2.7) and is non-continuous (Andrews, 2020). As a result, it may exhibit artefacts when applied to data with gradients or non-continuous data (such as counts); one such artefact, called the "arch" or "horseshoe" effect, occurs when PCA is applied to scRNAseq data without log-transformation (Hsu et al., 2023).

Another problem related to PCA is that the variance of an individual transcript depends heavily on its expression magnitude–even after variance-stabilising transformations–, and the top PCs will focus on the detailed fluctuations of highly expressed transcripts at the expense of broader transcriptional patterns (Kharchenko, 2021).

5 A sparse matrix is a matrix with mostly zeros while a dense matrix contains mostly non-zero elements. Dense matrices are stored in 2D arrays, whereas sparse matrices store the non-zero elements in three arrays: the 2D coordinates of the dense matrix and the value.

Still, PCA remains the most widely used method largely due to its simplicity, speed, and computational efficiency. In a comparison of 18 dimension reduction methods, PCA ranked highly when accuracy and performance in the downstream analysis were considered with computational scalability (Sun et al., 2019).

Despite that, several other approaches have been made to create alternatives to PCA. One option is to include more complex nonlinear relationships by autoencoder neural networks, which provide a convenient computational approach for learning complex nonlinear multidimensional functions. An autoencoder is a neural network designed to learn an effective low-dimensional representation by finding functions that map data to and from low-dimensional space in a way that yields optimal reconstruction of the original data (Amodio et al., 2017; Eraslan et al., 2019). Because these functions can be nonlinear, the resulting reduced dimensions can more effectively capture the underlying structure of the populations than can linear approaches. Nonlinear mapping, however, makes it more difficult to interpret the latent states (Kharchenko, 2021).

Butler et al., 2018 showed batch effect removal could be formalised as canonical correlation analysis (CCA), which is mathematically very similar to PCA. CCA is a way of inferring information from cross-covariance matrices by finding linear combinations of the two matrices which have a maximum correlation with each other (Tsuyuzaki et al., 2020). This, in the context of batch-effect correction, is a way to find combinations of variables that merge the datasets by minimising these batch effects while effectively reducing the dimension, like in PCA.

1.2.10 *Manifold graph representation **

After applying a summarising dimensionality reduction technique, the next steps involve further reducing the dataset to a 2D or 3D space easily visualisable by researchers. Then, clustering is performed to find distinct populations within the data. However, due to constraints on the data input type for some of these algorithms, as well as other reasons, the next step is to construct a $\mathbf{k}$-nearest neighbours ($\mathbf{k}$NN) graph (Luecken et al., 2019).

The $\mathbf{k}$NN graph is often constructed in single-cell data to help identify patterns and relationships in the data. The $\mathbf{k}$NN graph is a type of similarity graph, where each cell is represented as a node and edges are drawn between cells that are similar to each other based on their gene expression profiles (Kharchenko, 2021). To compute the similarity between cells, different metrics can be used. Although the Euclidean and cosine distances are commonly used, a recent benchmark suggests using correlation distances–e.g. Spearman correlation–and two proportionality measures (ϕ_s and ρ_p) defined by (Quinn et al., 2017) to compute the similarity matrix. Once the similarity matrix is computed, for each cell, its most $\mathbf{k}$ similar cells are selected, and the graph is constructed (Kharchenko, 2021).

In practice, computing the full distance matrix is intractable because the matrix would contain ${n_{cells}}^2$ entries which, for large datasets, is not memory-efficient. Additionally, using a heuristic method avoids the calculation of distances of cells that are *already known* to be distant and thus not within the $\mathbf{k}$NN of certain cells. For instance, McInnes et al., 2018 developed an approximate nearest neighbour search algorithm based on the Nearest-Neighbour-Descent algorithm by Dong et al., 2011.

One of the main advantages of constructing a kNN graph is that it allows researchers to uncover patterns and relationships in the data that may not be immediately apparent through other methods. The kNN graph provides a visual representation of how the cells are related to each other based on their gene expression profiles, which can help researchers identify groups of cells that are similar to each other and understand how they differ from other groups (Kharchenko, 2021). In practice, its ability to capture the structure of the data can be readily appreciated when looking at embeddings of visualisation DR (Maaten et al., 2008a; McInnes et al., 2018).

1.2.11 *Visualisation dimensionality reduction **

It was previously mentioned that DR methods could be distinguished as summarisation or visualisation. The aim of visualisation DR algorithms is to *reduce* the manifold, or structure, of the data into a 2D or 3D space, which is more interpretable for the researchers. Most of these methods are non-linear, which implies that the transformation they apply to the data is less mathematically "interpretable". The ideal aim of these methods is to avoid overcrowding of data in certain spots, which is a common phenomenon of linear methods in datasets with a large number of entries, and instead, find a projection in which cells similar to each other are positioned near while preserving the global structure of the dataset (Hie et al., 2019). However, preserving the embedding without minimal distortion is almost impossible when the number of entries is large, and the reduced dimension is low (Johnson et al., 1984). Therefore, all the DR visualisation methods will have significant distortions in their projections, which will be more apparent depending on what variable they favour: the local resemblance of similar cells or global structure.

If the DR method operates specifically on a kNN graph, it is called neighbour embedding (NE) (Böhm et al., 2022). NE algorithms optimise the layout using attractive forces between all pairs of points connected by a kNN graph edge, thus placing them closer to the low-dimensional embedding. In addition, every point feels a repulsive force to every other point, which prevents trivial solutions, such as positioning all points on top of each other.

One of the first NE methods widely used in single-cell methods is t-distributed stochastic neighbour embedding (t-SNE) (Maaten et al., 2008b), which learns a low-dimensional embedding in which the distribution of pairwise distances among cells forms a reasonably good information approximation of the distribution of pairwise distances in the original, high-dimensional space. In general terms, t-SNE compares two probabilities of, simplifying, cells being similar: the probabilities obtained by transforming the distances between cells in high dimensionality and the probabilities by transforming distances of projected cells in low dimensionality. This projection, called *initialisation*, can be random or done with any other DR technique–generally a linear method–. Then, iteratively, it tries to match the probabilities in the projection with the probabilities at the higher dimension by moving the position of the projected cells by a differential amount. The direction of the movement is a combination of attractive forces of the positions of the kNN and the repulsive forces of the rest of the cells. The aim is to minimise the divergence between the high-dimensional probabilities and the low-dimensional probabilities (Kobak et al., 2019). A second commonly used NE method is UMAP (McInnes et al., 2018). UMAP and t-SNE core workflow is the same, with changes in how the probabilities based on the distances are calculated, how the divergences between high-dimensional and low-dimensional probabilities are calculated, how the change of position of cells in low dimension-

ality is executed and how their positions are set during initialisation. The consequences of these changes will be discussed later.

Generally speaking, both UMAP and t-SNE work by preserving the local structure, that is, cell-to-cell similarities, at the cost of losing global structure, that is, general interactions between large groups of cells, or the more overall disposition of the data as a manifold. Other methods, such as *DiffMap* (Coifman et al., 2005), *ForceAtlas2* (Jacomy et al., 2014) of PHATE (Moon et al., 2019), try to preserve more the global structure of the dataset. For instance, PHATE works by calculating the distance between cells and transforming them into probabilities, as well as *diffusion* probabilities between long-range relationships. Lastly, the informational distance that measures the dissimilarity between diffused probabilities is used to indicate a general scale of the data. These distances are embedded using multidimensional scaling (MDS). *ForceAtlas2* uses the kNN graph and, by using a function of attractive and repulsive forces between the nodes, finds a projection of the graph into the lower dimension representation.

Since the creation of UMAP, there have been multiple comparatives stating that UMAP works better than t-SNE at preserving global information and with faster computation times (Becht et al., 2018), which made UMAP replace t-SNE as the *de facto* DR method. However, recent analyses by Kobak et al. (Kobak et al., 2019, 2021) proved that both methods function similarly. Regarding computation times, although it is true that classical integrations like Barnes-Hut t-SNE (Maaten, 2014) suffered from high computational times, newer implementations like FIt-SNE (Linderman et al., 2019) or Multicore-tSNE (Ulyanov, 2016) now work up to 4 times faster than UMAP on certain datasets (Kobak et al., 2019).

The second concern is that t-SNE seems to capture less global structure than UMAP and that t-SNE initialisation was random and, therefore, less reproducible than UMAP. While this may not be a problem in some situations, scRNA-seq datasets often exhibit biologically meaningful hierarchical structure, and thus the larger the data set, the more severe this problem becomes (Kobak et al., 2019). Considering that both algorithms work on the same kNN graph, there is no reason to believe that global structure preservation should be worse (Kobak et al., 2019). Kobak et al., 2021 argue that this phenomenon is entirely explained by t-SNE's random initialisation versus UMAP's initialisation using a linear DR method. Assigning a PCA-initialisation to t-SNE and a random initialisation to UMAP completely reverts the outcome, and their performances are similar when both use the same initialisation strategy (Kobak et al., 2019). This effect is represented with a toy dataset in Figure 10. Additionally, Kobak et al., 2019 recommend a set of easy parameter changes to provide more compact cluster visualisations, as UMAP does.

Recently, Böhm et al., 2022 provided a mathematical unification scheme of NE algorithms. The scheme is based on a parameter, *exageration* (ρ), which explains the trade-off between representing continuous manifold structures–stronger attraction, higher ρ–and discrete cluster structures–weaker attraction, lower ρ–. For instance, Laplacian Eigenmaps were generated using $\rho \to \infty$, *ForceAtlas2* with $\rho = 30$, UMAP with $\rho = 4$ and t-SNE with $\rho = 1$ (Figure 11). The authors conclude that no algorithm is "best" overall and it depends on the dataset and the context of use (Johnson et al., 2022).

1.2.12 *Clustering and community detection methods **

One of the most popular algorithms used in single-cell analyses and derived from ML and data exploratory analysis is clustering (Luecken et al., 2019). The goal of clustering is to partition a set of

data points into clusters such that the points within a cluster are more similar to each other than to points in other clusters. The ability to define cell types through clustering based on transcriptome similarity has emerged as one of the most powerful applications of scRNAseq. Defining cell types based on the transcriptome is attractive because it provides a data-driven, coherent and unbiased approach that can be applied to any sample (Kiselev et al., 2019).

The most common clustering methods are based on cell-cell distance matrices, like the *classical* Euclidean distance and cosine similarity (Haghverdi et al., 2018), or more refined metrics (Quinn et al., 2017). One of the simplest methods is k-means clustering, in which the data is partitioned into k centres (centroids) specified by the user, and each cell is assigned to the closest centroid. Then, iteratively, the position of the centroids is optimised to, ideally, go to the centres of the clusters (MacQueen, 2019). Certain implementations of k-means, such as Lloyd's algorithm, have the advantage of scaling with the number of points, which makes it a very fast algorithm. However, all the implementations of this algorithm are heuristic, which means that it usually falls in local minima, and thus yield suboptimal solutions (Kiselev et al., 2019). Additionally, k-means tends to produce equally-sized partitions, which for many single-cell datasets with cell types in a range of proportions, is not ideal. These drawbacks can be overcome by repeated application of k-means using different initial conditions. The SC3 algorithm exploits this process and finds a consensus, averaged solution between all clustering solutions (Kiselev et al., 2017).

Another widely used generic clustering algorithm frequently adapted for scRNA-seq is hierarchical clustering, which sequentially combines individual cells into larger clusters (agglomerative) or divides clusters into smaller groups (divisive). The algorithm's output is a data partition, as well as a dendrogram that explains the different partition steps of the clustering solution and the relationship between the partitions. An important shortcoming of hierarchical clustering is that both time and memory requirements scale at least quadratically with the number of data points, which means that it is prohibitively expensive to use hierarchical clustering for large datasets (Kiselev et al., 2019).

A third algorithm family has recently entered the field of single-cell analysis: community detection methods. These methods work on the kNN graph of the data and aim to find communities. A community is defined as a set of nodes (cells) with many edges (connections) between them and fewer edges between the nodes outside that community. In complex networks, like the kNN graph of biological samples, cells tend to group into communities based on similarity, and a community can be inferred as a group of cells that are similar to each other. Thus, community detection methods aim to find these communities in an unsupervised manner. Technically, community detection methods are different from clustering methods because the former use the kNN graph to perform the partition, whereas the latter use the cells or a distance matrix. However, it is common to abuse the term and call communities detected by these methods *clusters* because often, the representations to visualise the communities, like visualisation DR methods, do not show the kNN graph, but the cells instead.

To perform community detection, two elements are needed: (1) a measurement of *strengh* of the division of a graph into its communities and (2) an algorithm that optimises the measurement, also called *quality function*. The most common *quality functions* are modularity and its improved version, Constant Potts Model. These measurement tries to explain the difference between the actual number of edges in a community and the expected number of edges using a random graph with no communities as the reference. Therefore, optimising the measurement will yield a partition of the graph into its communities (Traag et al., 2019). Since that optimisation is an NP-hard problem,

community detection algorithms optimise modularity and the other measurements using a heuristic method.

One of these algorithms is the *louvain* algorithm (Blondel et al., 2008). This algorithm optimises in two phases: (1) local moving of nodes; and (2) network aggregation. In the local moving phase, individual nodes are moved to the community yielding the largest increase in the quality function. The first iteration starts with as many communities as nodes. In the aggregation phase, a secondary aggregate network is created based on the partition obtained in the local moving phase. Each community in this partition becomes a node in the aggregate network. These steps are iterated until convergence.

The main problem with *louvain* is that a node may be moved to a different community while acting as a bridge between different components of its old community (Traag et al., 2019). Removing that node disconnects the old community. Therefore, it is possible to have groups of cells that belong to the same community but are not connected, which goes against the definition of the community. To solve that problem, a more recent algorithm called *leiden* applies a similar but more refined procedure which avoids reassigning already-assigned cells to other clusters. Additionally, despite being more complex than *louvain*, some optimisations make *leiden* have slightly better computation times than *louvain* (Traag et al., 2019).

One of the reasons community detection algorithms have gained traction and are already the gold standard in the single-cell analysis is because they can be effectively run on graphs containing millions of cells (Andrews, 2020; Kharchenko, 2021). Additionally, they are often faster than clustering algorithms as only neighbouring cell pairs have to be considered to belong to the same cluster and not all cells. This approach thus greatly reduces the search space for possible clusters (Luecken et al., 2019).

Regardless of the clustering or community detection algorithm, there are some common pitfalls that depend on the properties of the genes selected–to calculate the distance matrix or the kNN graph–, which are described in Figure 12: underclustering, overclustering, and cluster splitting (Hie et al., 2019).

- Cells are underclustered if cells of different types are assigned to the same cluster, masking variation in the data (Figure 12c). The omission of key genes that distinguish cell types leads to underclustering if these cells are similar to one another.
- Cells are overclustered if multiple clusters represent the same cell type (Figure 12e). Including genes whose variation across cells is not informative to the biologically meaningful clustering could lead to overclustering.
- Cluster splitting occurs when some cells of the same type are scattered between different clusters dominated by other cell types (Figure 12g), a common pitfall in discovering rare cell subsets.

Most of these pitfalls have their solutions. Underclustering generally is solved by DR visualisation of the clusters by using algorithms such as Silhouette that give a measure of the goodness of clustering (Rousseeuw, 1987). In that case, usually isolating the populations and performing a subclustering is the best solution. Overclustering can be partially overcome by a correct FS that removes non-informative genes. Cluster splitting can usually be solved by using a combination of the under- and

overclustering solutions, as well as using algorithms designed for rare cell detection, like *RaceID* (Grün, 2019).

Another common problem with data partitioning is that all algorithms use a parameter related to the number of partitions. For instance, k-means clustering uses k as the number of clusters, and *louvain* and *leiden* use the *resolution* parameter, which increases the number of clusters as it increases. The choice of parameter often has a large effect on the outcome, and therefore there are computational methods available to help guide the choice (Kiselev et al., 2018). Many of these methods are based on the idea of calculating a cluster quality score and identifying an 'elbow,' that is, the point where the score plateaus. These scores favour a fairly coarse resolution, with clearly separated clusters rather than closely related or overlapping cell types. However, these methods tend to work adequately with simple datasets and *break* in datasets that have subtypes, trajectories or large disparities in cluster sizes, which ultimately requires an informed judgement from the researcher and entails a risk of bias (Kiselev et al., 2018). This phenomenon can be observed in Figure 13a, where the partitions shown in subfigures b and c show that a more coarse partition is useful for devising general characteristics of the graph, and a secondary partition helps retrieve detailed elements.

The general recommendation for the analysis of cellular populations in single-cell data is to use a range of partitions in a divide-and-conquer manner to have a structure that recapitulates both tissue development and tissue organisation in major cell types; while accounting for continuous cell states in addition to discrete cell types, also named as minor cell types (Kharchenko, 2021; Lähnemann et al., 2020). In some cases, secondary cell partitioning can be done using all cells at once or focusing on specific populations and dividing them individually. This secondary approach might be preferred because when isolating each population, only the most relevant genes are selected for DR and clustering (Figure 13d-g). Doing the secondary clustering with all populations at once may lead to an underclustering of the subpopulations (Hie et al., 2019).

1.2.13 *Differentially expressed gene analysis*

One important goal, after separating cells into distinct groups using computational methods or cell-surface markers, is to identify the genes that most effectively distinguish between the different groups. A common approach to achieving this is identifying genes that show different levels of expression between pairs of groups, also known as differentially expressed genes (DEGs) (Stegle et al., 2015). DEG analysis is typical of bulk methods, where DEGs between groups of samples from different conditions were calculated as leads of the process or condition by running a statistical test and obtaining the p-values. In scRNAseq analysis, DEG analysis adopts a new dimension by not only being able to calculate DEGs between samples but between groups of cells. In fact, in the most basic scenario of analysing one sample, typical DEG analysis is used to find markers of different cell types or states.

While DEG analysis is commonly performed in scRNAseq pipelines and may seem straightforward, many factors can impact the interpretation of the results. Therefore, it is important to consider these aspects when conducting DEG analysis. One of the implicit assumptions in the null hypothesis of DEG tests is that genes have the same distribution of expression values between the two groups. Therefore, two genes with similar gene expression but which have different unimodal or multimodal distributions are likely to falsely reject the null hypothesis (Adil et al., 2021; Luecken et al., 2019). If

intrinsic variation between biological replicates exist, the most frequently used methods can identify differentially expressed genes even in the absence of biological differences (Squair et al., 2021)

The central assumption of DEG testing is to find DEGs between two groups of cells that are not known to be different from each other beforehand, and if they were the same, the test would not report any DEGs. However, if DEG testing is performed after clustering, the populations being compared are already assumed to be different, and thus, significant DEGs are expected to be found (Andrews, 2020; Lähnemann et al., 2020; Luecken et al., 2019). In fact, this assumption is flawed as it has been demonstrated that unsupervised clustering followed by DEG analysis can result in artificially low p-values (Zhang et al., 2019b). Therefore, it is important to be cautious when interpreting results obtained this way. Vandenbon et al., 2020 recently developed a clustering-independent method for finding differentially expressed genes in single-cell transcriptome data, which uses Kullback–Leibler divergence to find genes that are expressed in subsets of cells that are non-randomly positioned in a multidimensional space.

The sparsity and heterogeneity of single-cell data have encouraged the development of specialised statistical methods to identify differentially expressed mRNAs (Squair et al., 2021). However, recent comparisons have concluded that the methods developed for bulk RNA sequencing, as well as the non-parametric Wilcoxon test, perform well compared to purpose-built methods (Soneson et al., 2018). This is particularly true when larger cell populations are being compared, as the advantages of sophisticated parametric models fade and standard parametric tests can provide sufficient statistical power while making fewer assumptions (Kharchenko, 2021). In cases where single-cell methods were used, there was a bias to label highly-expressed genes as differentially-expressed, and the opposite to lowly-expressed genes (Squair et al., 2021).

1.2.14 *Trajectory inference and RNA velocity*

The dynamics of transcriptional states are of primary interest in many biological contexts, such as organism development and response to stimuli (e.g. immune response or cancer expansion) (Kharchenko, 2021; Tritschler et al., 2019). Current scRNA-seq protocols capture a static snapshot of cellular states at a particular time, but some features of the underlying dynamic process can be deduced using trajectory inference (TI) methods (Saelens et al., 2019). These methods order cells along a trajectory based on similarities in their expression patterns, creating a continuous trajectory where each cell can be assigned a specific position (Andrews, 2020).

TI methods reconstruct paths through cellular space by finding paths that minimise transcriptional changes between neighbouring cells. It is assumed that the process of interest dominates this transcriptomic similarity and that the more similar two cells are, the closer their stage of progression along this process. First, a similarity measure quantifies the similarity between the gene expression profiles of two cells. Then, by ordering the cells according to this similarity measure, a hypothetical trajectory development can be inferred (Tritschler et al., 2019). The ordering of cells along these paths is described by a pseudotime variable, which is often interpreted as a proxy for developmental time (Luecken et al., 2019). The resulting trajectories are often linear, bifurcating or tree-shaped, but more recent methods also identify more complex trajectory topologies, such as cyclic or disconnected graphs (Luecken et al., 2019; Saelens et al., 2019). Depending on the measurement defined to

describe intercellular similarity, and the types of trajectories that are allowed, different results may arise.

Early TI methods typically fixed the topology algorithmically, for example, linear (Bendall et al., 2014; Shin et al., 2015) or bifurcating trajectories (Haghverdi et al., 2016; Setty et al., 2016), or through parameters provided by the user (Matsumoto et al., 2016; Trapnell et al., 2014). These methods mainly focus on correctly ordering the cells along the fixed topology. More recent methods also infer the topology (Qiu et al., 2017; Street et al., 2018; Wolf et al., 2019), which increases the difficulty of the problem at hand but allows the unbiased identification of both the ordering inside a branch and the topology connecting these branches (Saelens et al., 2019). An alternative to classical visualisation on the cell level is partition-based graph abstraction (PAGA) (Wolf et al., 2019). This tool has been shown to adequately approximate the topology of the data while coarse-graining the visualisation using clusters–traditional TI methods do not rely on cluster information to construct the trajectories–. In combination with any of the above visualisation methods, PAGA produces coarse-grained visualisations, which can simplify the interpretation of single-cell data, especially with large numbers of cells (Luecken et al., 2019).

Due to the active research on TI, more than 40 TI widely used methods have been created recently (Saelens et al., 2019). In a recent comparison of these methods, it was concluded that no individual method performs optimally for all trajectories (Luecken et al., 2019). For instance, *Monocle* and PAGA scored better on the topology scores, whereas other methods, such as *Slingshot*, were better at ordering the cells and placing them into the correct branches. Using one method resulted in getting a top model about 27% of the time, and up to 74% with the addition of six other methods (Saelens et al., 2019). As a result, as a method's performance is heavily dependent on the trajectory type being studied, the choice of method should currently be primarily driven by the anticipated trajectory topology in the data. When inferring cellular trajectories, it is important to use multiple methods to check the robustness of the inferred trajectory. This is important even if the topology of the trajectory is known beforehand because the underlying biology is often more complex than anticipated by the user. Using multiple methods, including those that make fewer assumptions about the trajectory topology, can provide a more comprehensive understanding of the biological processes at play. (Saelens et al., 2019).

TI methods are extremely complex and are based on several assumptions. Therefore, they should be used considering their limitations as well. The main assumption of TI methods is that they assume that the data contains a trajectory, even where there might not be one. For instance, Kharchenko, 2021 shows computationally optimal spanning trees for independent PBMC populations, yet interpreting it as a dynamic process would be incorrect. It is, therefore, important to determine the identity of the subpopulations first and restrict the trajectory modelling to the appropriate part of the dataset. This false positive occurs because, technically, inferred trajectories only denote transcriptional similarity, which can be mistaken for a biological trajectory (Luecken et al., 2019). Some TI methods include an uncertainty evaluation in their model (Griffiths et al., 2018). Additionally, experimental validation should be performed on inferred trajectories (Tritschler et al., 2019).

Another assumption of TI methods is that all the cells captured in the pseudotemporal ordering are being captured, which requires sufficient sampling from the underlying process. Therefore, less frequent specific cell types should be enriched to ensure a reliable inference, but with enough caution not to distort the developmental landscape (Tritschler et al., 2019). Additionally, TI methods

also assume ergodicity[6] (Kharchenko, 2021). However, this might not be the case due to temporary stages not being captured in the dataset, and therefore the inferred trajectory may lack steps of the biological process.

Another set of methods, based on the same principles as TI methods, was recently developed. These methods are constructed around the concept of RNA velocity, which relates the balance of spliced and unspliced forms of the genes in a cell to its future state (La Manno et al., 2018). During a dynamic process, an increase in the transcription rate results in a rapid increase in unspliced mRNA, followed by a subsequent increase in spliced mRNA until a new steady state is reached. Conversely, a drop in the transcription rate first leads to a rapid drop in unspliced mRNA, followed by a reduction in spliced mRNAs. These quantities can be combined in a simple model of mRNA kinetics to estimate the first-time derivative of the transcriptional state, which can then be used to approximate what the state of the cell would have been shortly before or after the time of the measurement. Therefore, a positive RNA velocity indicates that a gene is upregulated, which occurs for cells that show a higher abundance of unspliced mRNA for that gene than expected in a steady state. Conversely, negative velocity indicates that a gene is downregulated. The combination of velocities across genes can then be used to estimate the future state of an individual cell.

The advantage of RNA velocity with respect to TI methods is that RNA velocity is based on the fundament of spliced and unspliced mRNA, and therefore, despite its assumptions and the noisiness of the dataset, inferences on the velocity have a *biological* basis. However, the major computational limitation of RNA velocity, shared with TI methods, is its reliance on a DR method for visualization, which may lead to a loss of interpretability. Tritschler et al., 2019 makes an important claim regarding TI methods that can be extended RNA velocity, other aspects of single-cell analysis, or even in bioinformatics as a whole: if we put too much prior knowledge–and assumptions–into a model, we will not unbiasedly generate novel information but will recapitulate known information, reducing the model to a visualisation tool.

1.2.15 *Cell-cell communication analysis*

Cell-cell communication (CCC) is a fundamental aspect of biology in which cells communicate with each other to maintain the proper functioning of an organism. There are four main types of CCC: (1) autocrine, in which a cell sends signals to itself, (2) juxtacrine, in which a cell communicates with a physically adjacent cell, (3) paracrine, in which a cell communicates with nearby cells, and (4) endocrine, in which a cell secretes molecules that travel through extracellular fluids to other parts of the organism. Cell-cell communication is also referred to as ligand-receptor (LR) communication, as the communication typically occurs through the binding of ligands to receptors (Armingol et al., 2020). Analysis of CCC in single-cell shifts from focusing on which cells are present to further focusing on what relationships between cells are present (Almet et al., 2021).

Cell-cell communication is not limited to single-cell studies and has been extensively studied using various techniques that focus on the protein-protein interactions (PPIs) that mediate communication. These techniques include yeast two-hybrid screening, co-immunoprecipitation, proximity labelling

6 Ergodicity expresses the idea that a point of a moving system, either a dynamical system or a stochastic process, will eventually visit all parts of the space that the system moves in, in a uniform and random sense. This implies that the average behaviour of the system can be deduced from the trajectory of a "typical" point.

proteomics, fluorescence resonance energy transfer imaging, and X-ray crystallography (Rao et al., 2014).

CCC in single-cell is studied with specific methods, which require databases of existing PPI. Due to the technical and biological noise, the ground truth of interactions is masked, and CCC is often defined probabilistically (Almet et al., 2021).

The most common methods in single-cell CCC analysis work with clusters of cells and compute a communication score for each ligand–receptor pair. To compute that score, generally, a null distribution is computed–e.g. off LR expression or z-score–through cluster label permutation, and a non-parametric test assesses the differences with the null model. The most commonly used CCC methods in single-cell are *CellChat* (Jin et al., 2021) and *CellPhoneDB* (Efremova et al., 2020).

While the basic concept of inferring cell-cell communication (CCC) is relatively straightforward, the method can be overly simplistic and may overestimate communication activity. This is due to several factors (Almet et al., 2021).

First, CCC occurs at the protein level, not the gene level. This means that a gene may be expressed, but its corresponding protein may not be present, leading to an overestimation of CCC. To address this, single-cell proteomic methods may be useful, but currently, they are not widely used.

Second, CCC is spatially constrained. Many signalling pathways are activated through ligands diffusing from sender cells to nearby receiver cells. As a result, the number of cells that a sender cell can communicate with is limited by the finite spatial range of the ligand. This means that CCC determined using ligand or receptor gene expression data alone may not reflect the actual communication activity that occurs at the protein level and within a specific spatial context. To address this factor, CCC analysis can be used in spatially-resolved methods (Almet et al., 2021).

1.2.16 *Gene ontology enrichment analysis*

Gene Ontology enrichment analysis (GOEA) is a method used to identify statistically significant overrepresentation of specific biological terms in a set of genes (Ashburner et al., 2000). The purpose of the analysis is to determine whether the genes in a given set are functionally related and to identify the biological processes or molecular functions that are most associated with those genes. GOEA in the context of scRNAseq does not vary from the rest of the contexts. In general, GO is composed of defined terms and the structured relationships between them (GO); and the associations between gene products and the terms (GO annotations) (Yon Rhee et al., 2008).

The p-value is calculated by comparing the observed number of genes in the set that are annotated with a specific term to the expected number of genes that would be annotated with that term by chance. The calculation of the p-value in a GOEA typically involves two steps. First, the expected number of genes annotated with a specific term by chance is calculated based on the total number of genes in the set and the proportion of genes in the entire genome that are annotated with that term.

Next, the observed and expected numbers of genes are compared using a statistical test such as the χ^2 test or Fisher's exact test. The p-value is the probability that the observed difference between the number of genes in the set that are annotated with a specific term and the expected number of genes is due to chance. A small p-value indicates that the overrepresentation of the term in the

set of genes is statistically significant and unlikely to have occurred by chance. The common use of GOEA is to find GO terms with statistically significant p-values from the list of GO terms available at the ontology. To obtain the terms, a p-value is calculated for each term, and then p-values are corrected using multiple hypothesis testing methods like the FDR.

GOEA can be performed with several tools, such as *goatools* (Klopfenstein et al., 2018), or *Enrichr* (Kuleshov et al., 2016)–a web served that performs GOEA from a vast array of GO sets–. In Python, the module *gseapy* can automate the performance of GOEA by using *Enrichr* API (Fang et al., 2021).

Despite the relevance and wide use of GOEA in bioinformatic analysis, it has some limitations. The most relevant one is that a very high number–over 95%–of GO annotations are computationally derived and have not been manually curated. Therefore, there is a high false positive rate of genes being falsely associated with a higher number of GO terms.

1.3 TABLES AND FIGURES

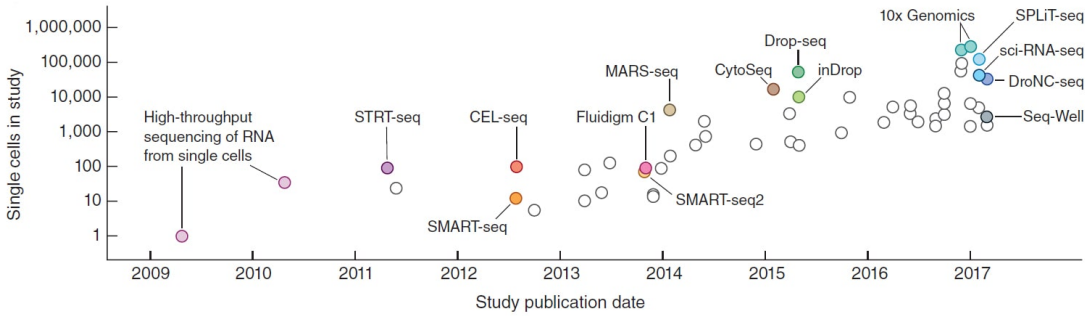


Figure 1: **Representative single-cell publications**. Each publication shows the number of cells detected. The coloured publications represent novel methods that are relevant to the single-cell field. Adapted from Svensson et al., 2018, Figure 1B.

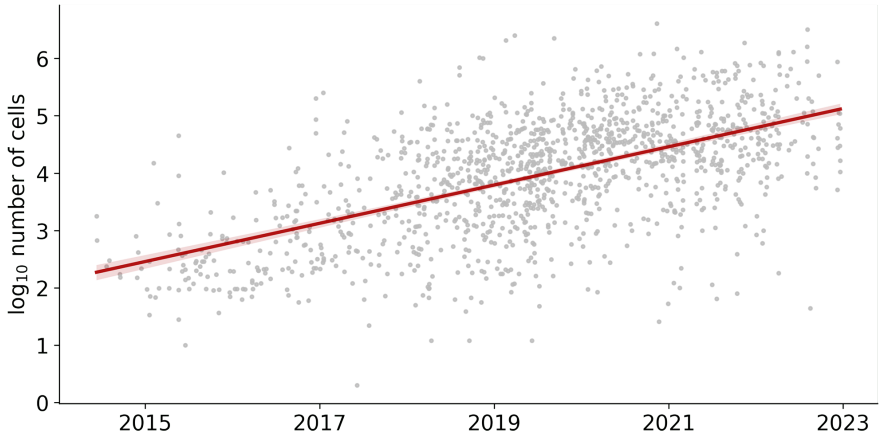


Figure 2: **Trend of number of cells per publication**. Each dot in the scatter plot represents a publication from Svensson et al., 2020 database. Datasets from 06/2014 backwards were removed to facilitate the interpretation of the figure. The red line represents the linear regression fitted to the scatter plot.

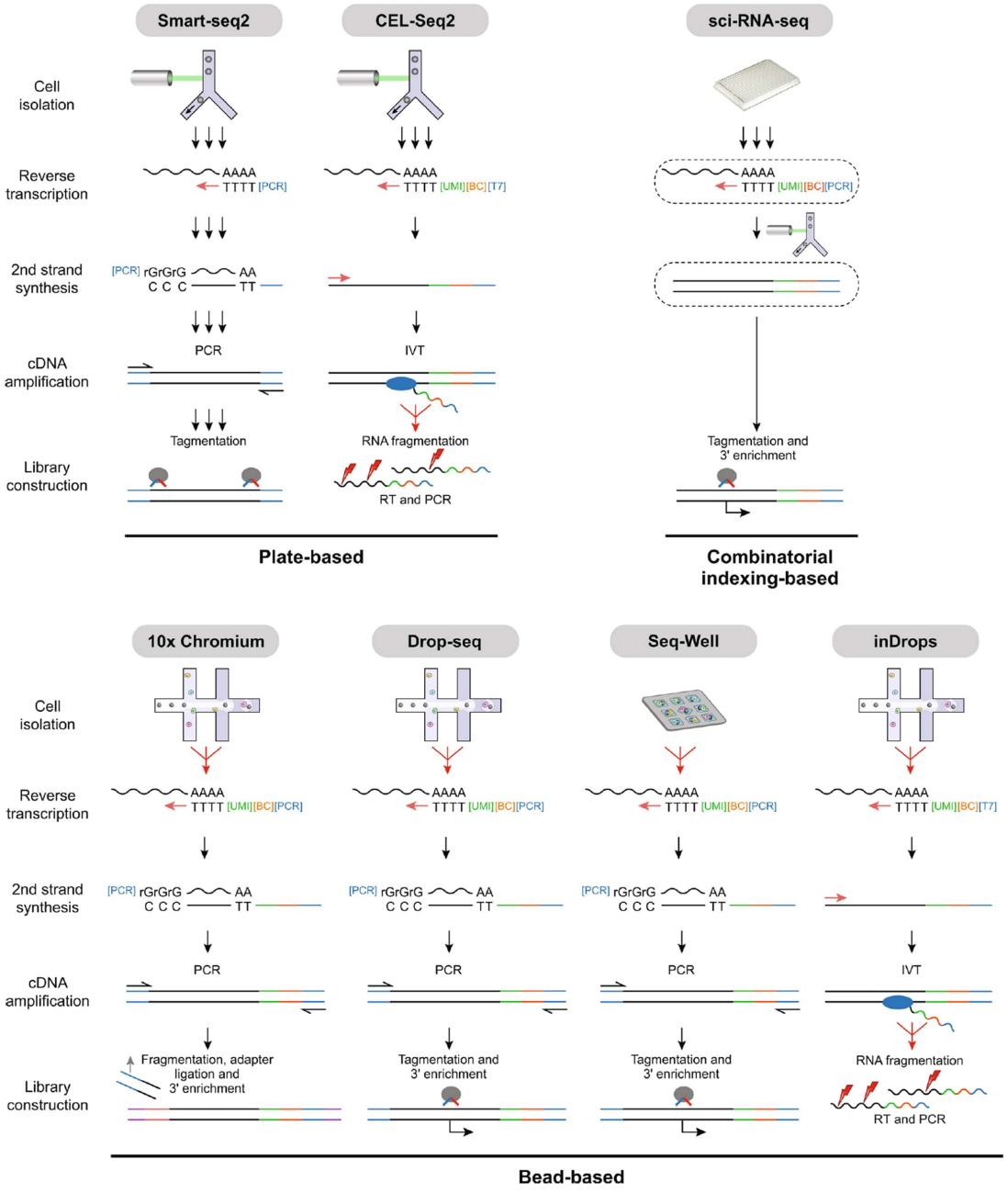


Figure 3: **Scheme of scRNAseq methods**. Each method has a graphical representation of the main steps in the scRNAseq sample processing pipeline. Adapted from Ding et al., 2020 (Extended Data Fig. 1).

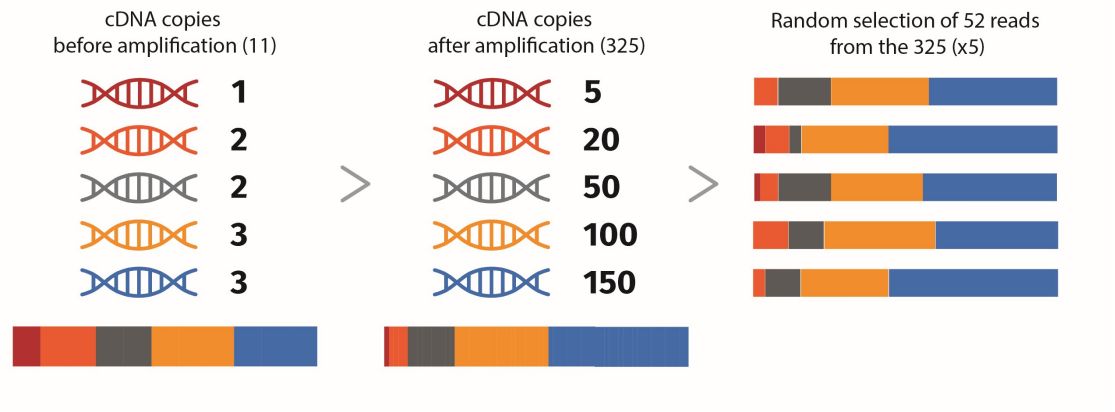


Figure 4: **Scheme of PCR amplification bias**. Five genes have their cDNAs amplified by PCR. After the non-linear amplification, their relative proportions change. If the sequencing depth is limited to 52 reads, the first gene has sampling zeros in three out of five hypothetical sequencing experiments. Adapted from Jiang et al., 2022 (Figure 3)

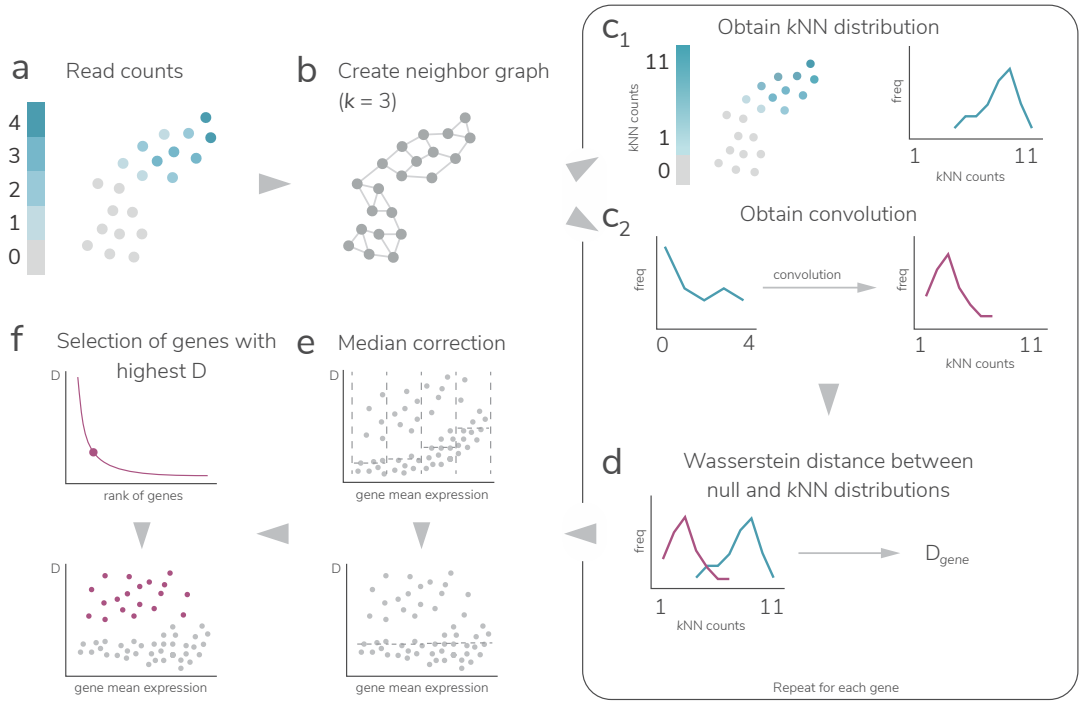


Figure 9: **Graphical abstract of *triku* workflow.** a) DR representation of the gene expression from the count matrix from a dataset, where each dot represents a cell. b) kNN graph representation with 3 neighbours. For each cell, the k transcriptomically most similar cells are selected (3 in this example). c1) Considering the graph in b) for each cell with positive expression, the expression of its k neighbours is summed to yield the kNN distribution in blue. c2) With the distribution of reads (blue line), the null distribution is estimated by sampling k random cells. d) The null and kNN distributions of each gene are compared using the Wasserstein distance. e) For each gene, its distance is plotted against the log mean expression, and divided into w windows (4 in this example). For each window, the median of the distances is calculated and subtracted from the distances in that window. f) All corrected distances are ranked and the cutoff point is selected.

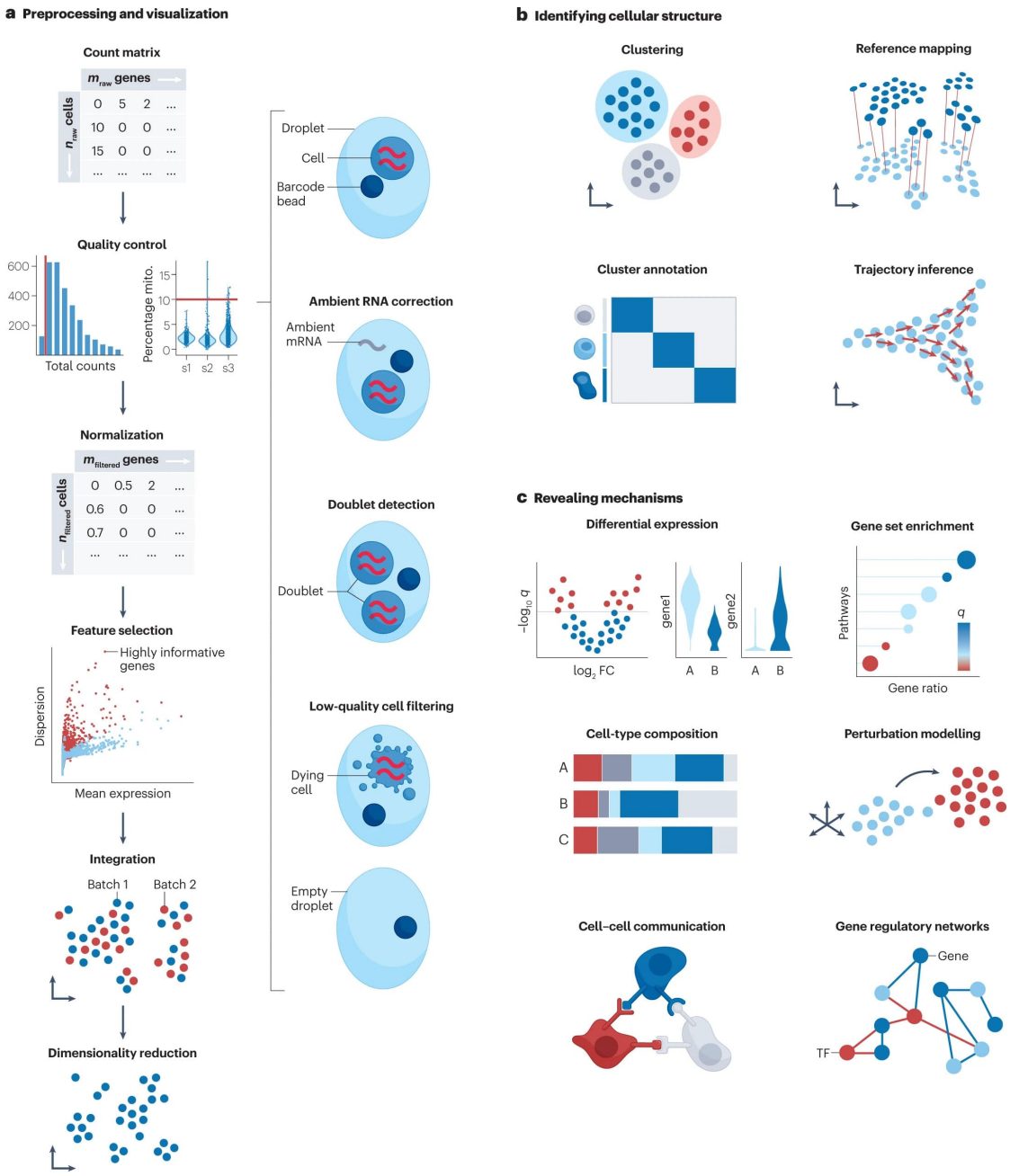


Figure 5: **Scheme of scRNAseq data processing.** (a) Basic data processing steps, including count matrix generation, cell and gene QC, ambient RNA correction, doublet detection, normalisation, feature selection, and dimensionality reduction. (b) Methods of direct population annotation based on discrete clusters and continuous identities. (c) Complementary analyses performed in several single-cell scenarios. Adapted from Heumos et al., 2023 (Figure 2).

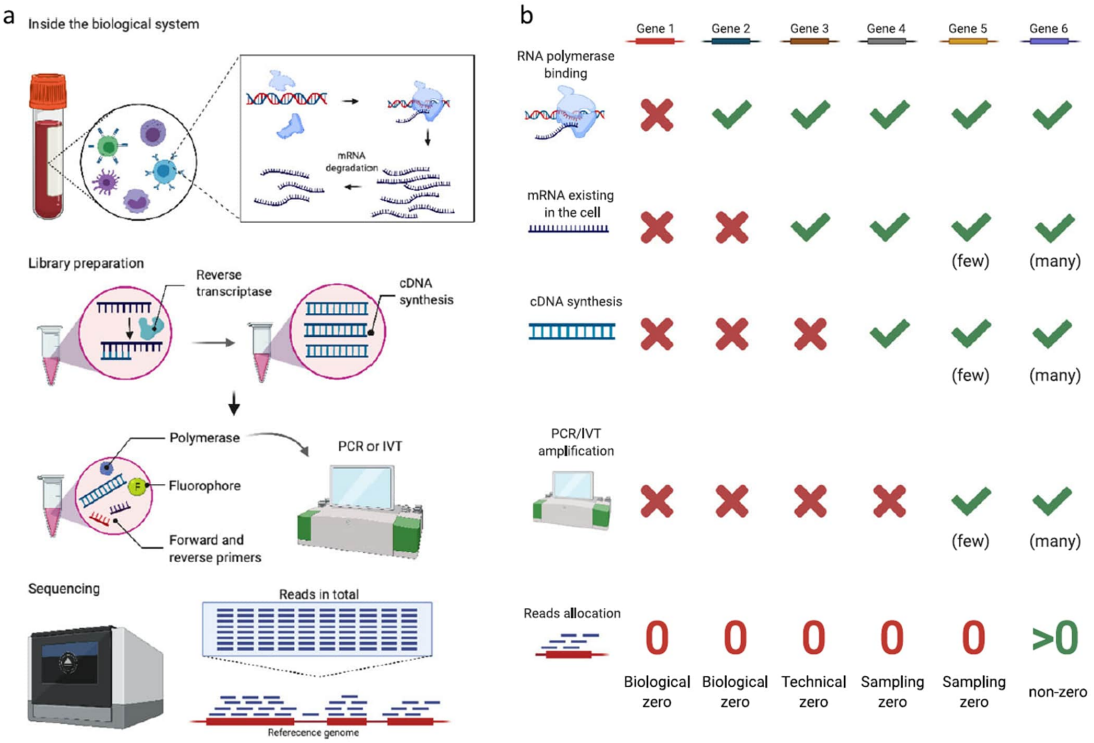


Figure 6: **Sources of zeros in scRNA-seq data.** (a) An overview of a scRNA-seq experiment. Biological factors determining true gene expression levels include transcription and mRNA degradation (top panel). Technical procedures that affect gene expression measurements include cDNA synthesis, PCR or IVT amplification, and sequencing depth (bottom three panels). Finally, every gene's expression measurement in each cell is defined as the number of reads or UMIs mapped to that gene in that cell. (b) How the biological factors and technical procedures in (a) lead to biological, technical, and sampling zeros in scRNA-seq data. Red crosses indicate occurrences of zeros, while green checkmarks indicate otherwise. Biological zeros arise from two scenarios: no transcription (gene 1) or no mRNA due to faster mRNA degradation than transcription (gene 2). If a gene has mRNAs in a cell, but its mRNAs are not captured by cDNA synthesis; the gene's zero expression measurement is called a technical zero (gene 3). If a gene has cDNAs in the sequencing library, but its cDNAs are too few to be captured by sequencing, the gene's zero expression measurement is called a sampling zero. Sampling zeros occur for two reasons: a gene's cDNAs have few copies because they are not amplified by PCR or IVT (gene 4), or a gene's mRNA copy number is too small so that its cDNAs still have few copies after amplification (gene 5). If the factors and procedures above do not result in few cDNAs of a gene in the sequencing library, the gene would have a non-zero measurement (gene 6). Adapted from Jiang et al., 2022 (Figure 1)

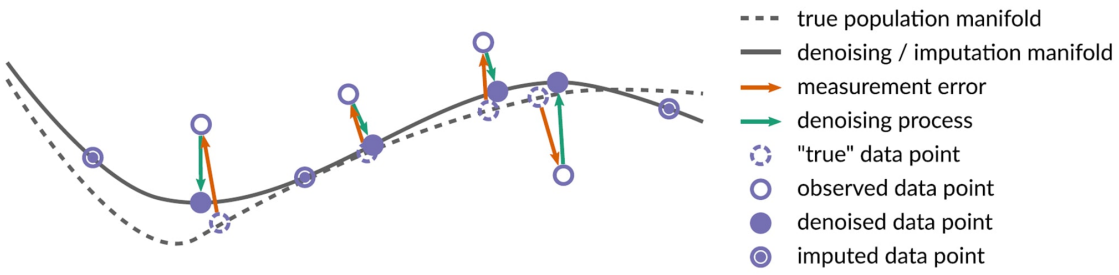


Figure 7: **Scheme of data imputation (zeros and non-zeros).** Measurement error requires denoising methods or approaches that quantify uncertainty and propagate it down analysis pipelines. Where methods cannot deal with abundant missing values, imputation approaches may be useful. While the true population manifold that generated data is never known (dashed line), one can usually obtain some estimation of it that can be used for both denoising and imputation (continuous line). Adapted from Lähnemann et al., 2020 (Figure 2).

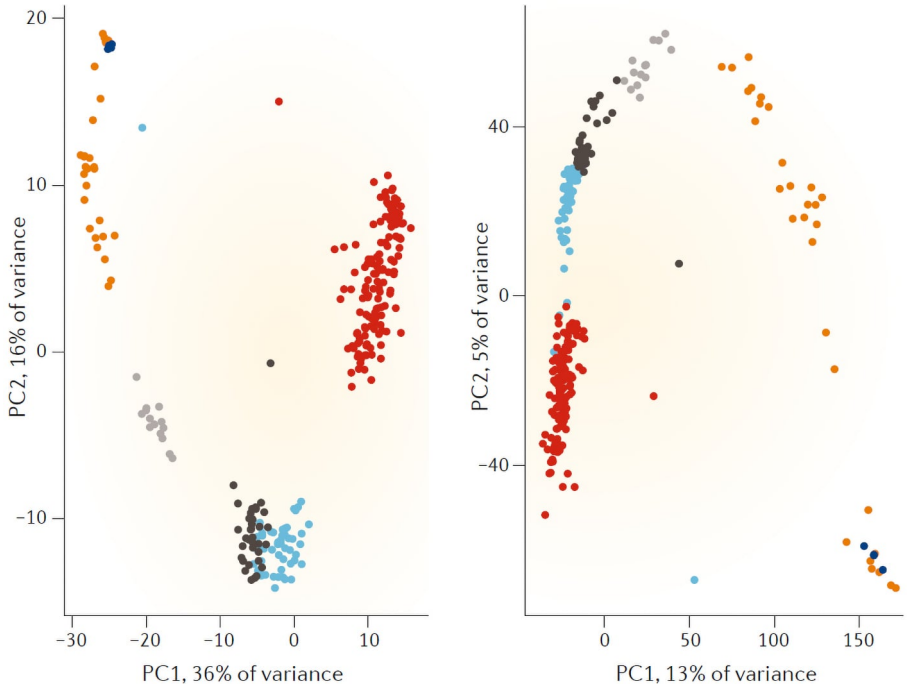


Figure 8: **Illustration of the curse of dimensionality.** PCA plots of the Deng data set using 500 (left) and 20,000 (right) of the most variable genes. Ideally, the six separate populations of cells should be distinguishable. When using a large number of features, clusters are less distinct, as indicated by the shorter distances between clusters (for example, the 4-cell stage is not as isolated). Consequently, unsupervised clustering becomes more challenging. Adapted from Kiselev et al., 2019, Figure 2.

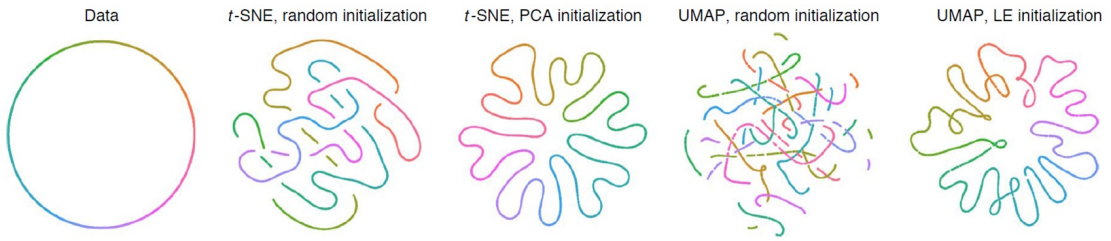


Figure 10: **t-SNE and UMAP with random and non-random initialisation.** Embeddings of n = 7,000 points sampled from a circle with a small amount of Gaussian noise (σ = r/1,000, where r is the circle's radius). Random and PCA initialisation were used for t-SNE and random and LE initialisation for UMAP. Adapted from Kobak et al., 2021, Figure 1.

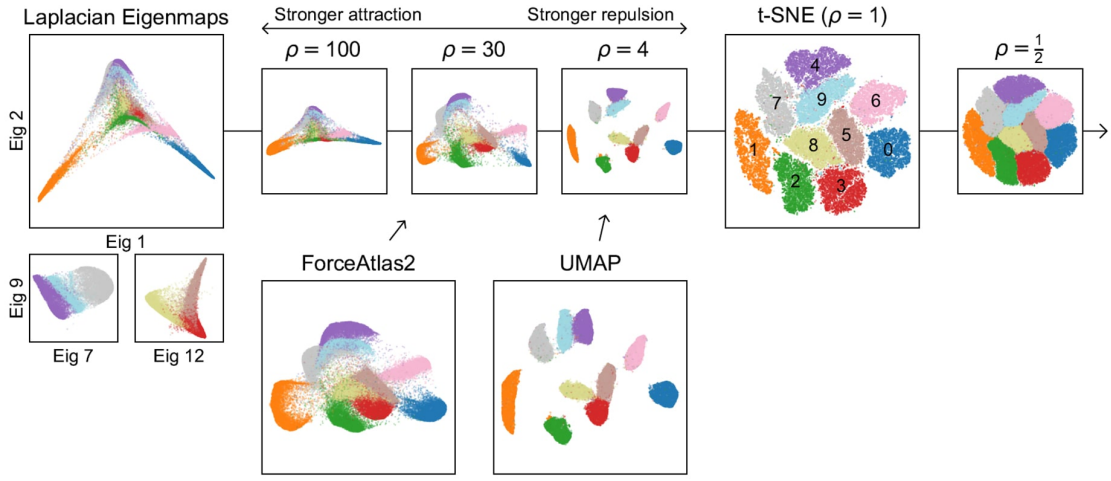


Figure 11: **Attraction-repulsion spectrum for the MNIST data.** Different embeddings of the MNIST data set of hand-written digits (n = 70 000); colours denote digits as shown in the t-SNE panel. Multiplying all attractive forces by an exaggeration factor ρ yields a spectrum of embeddings. Values below 1 yield inflated clusters. Values above 1 yield more compact clusters. Higher values make multiple clusters merge, with $\rho \to \infty$ approximately corresponding to Laplacian eigenmaps. UMAP is similar to $\rho \sim 4$. ForceAtlas2 is similar to $\rho \sim 30$. Adapted from Böhm et al., 2022, Figure 1.

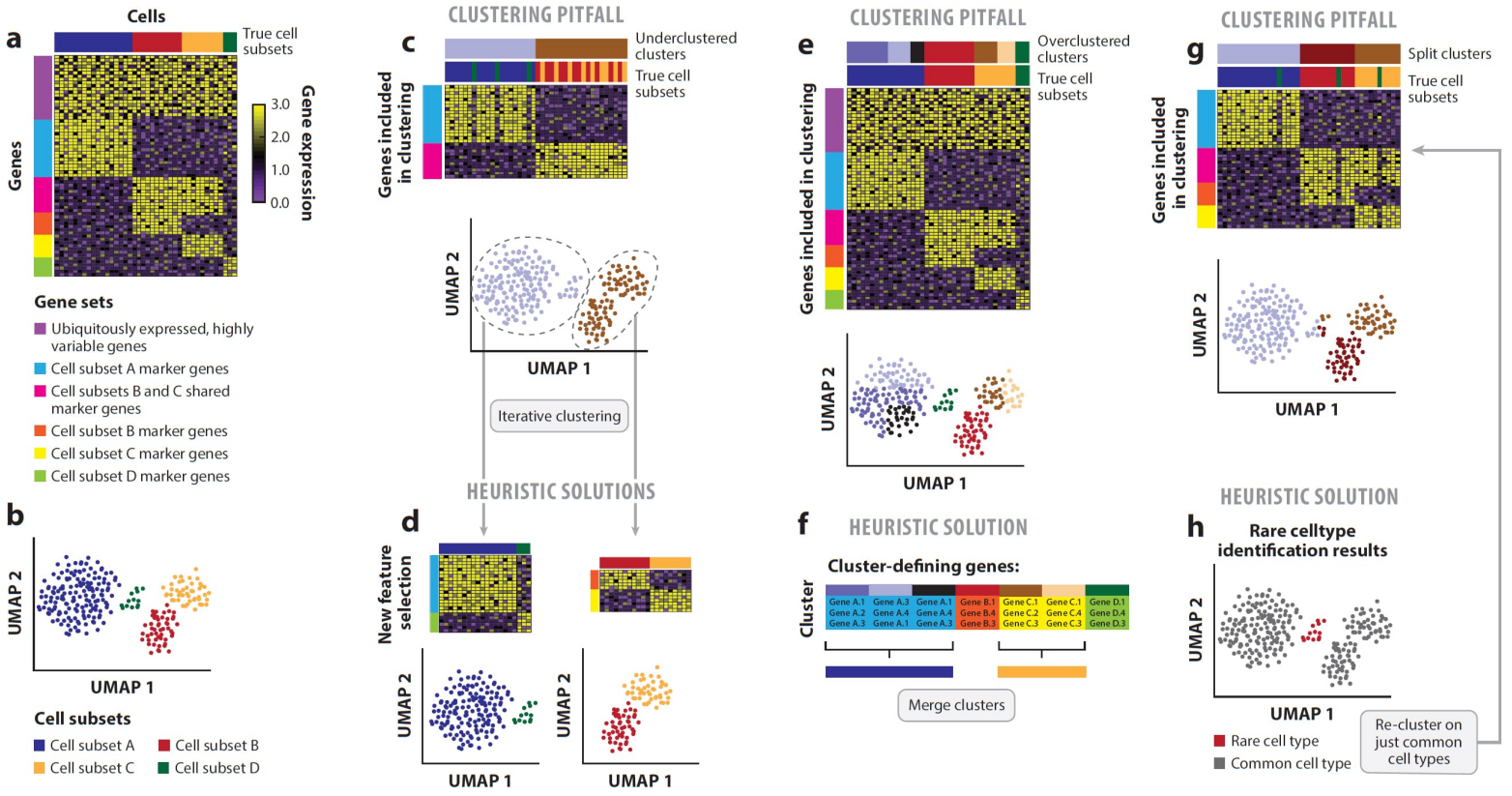


Figure 12: **Illustration of clustering pitfalls.** (a) Heatmap of gene expression by cell, including marker genes for each cell type and uninformative genes. (b) Cells are coloured by their true cell subset labels in dimensionality-reduced space. (c) Underclustering can occur when marker genes that distinguish between similar cell subsets (e.g., B and C) or are unique to one cell type (e.g., D) are not included as features. (d) Iterative clustering can split clusters from panel c when variable genes are reselected for the cells in these clusters. (e) Overclustering can occur when uninformative but variable genes are included, resulting in clusters with no distinguishing marker genes. (f) Overclustered cells in panel d can be merged based on shared marker genes. (g) Rare cell types are split between clusters when their marker genes are excluded from clustering. (h) Rare cell type discovery methods identify rare cells before clustering. Adapted from Hie et al., 2020, Figure 3.

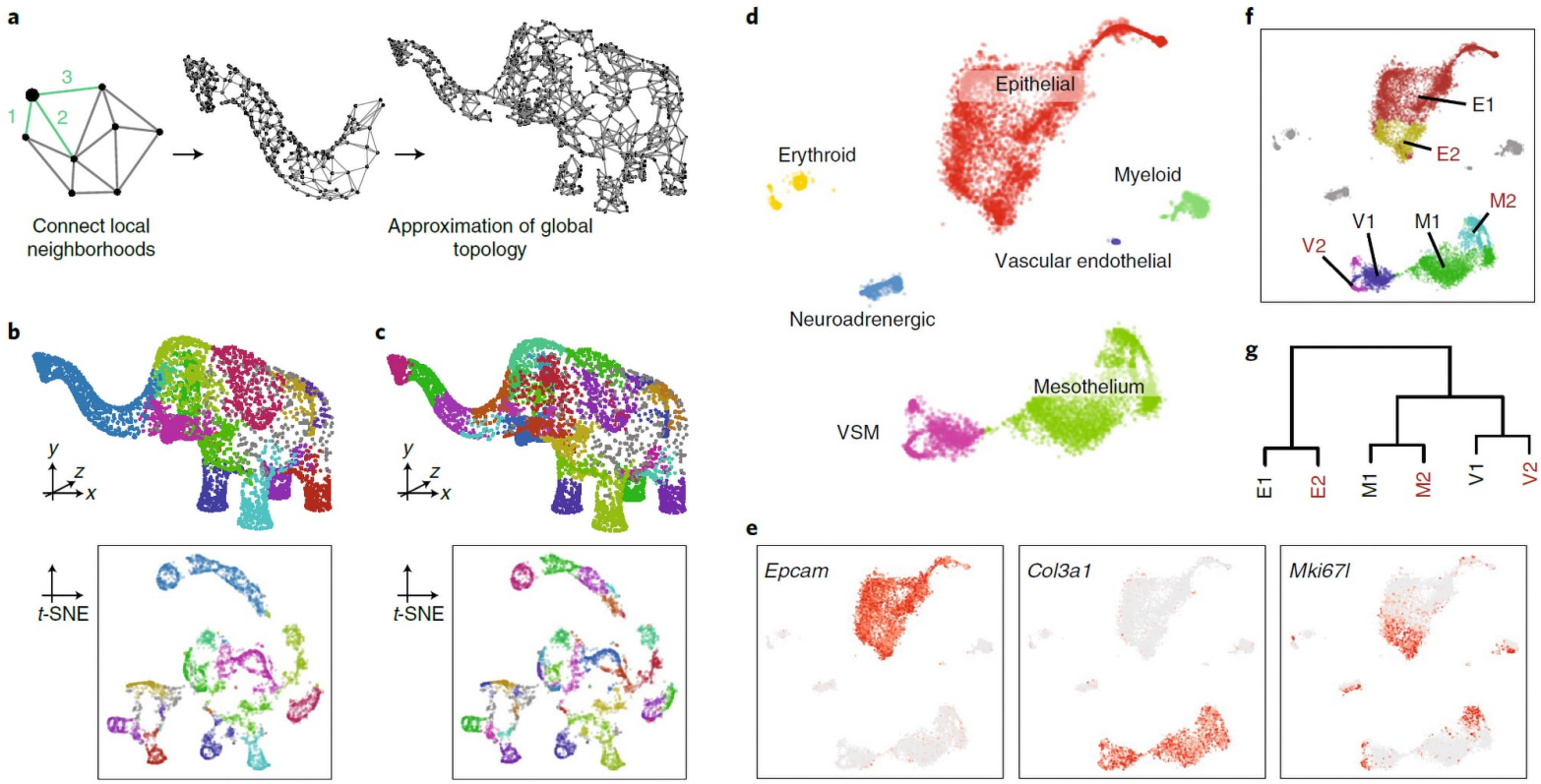


Figure 13: **Approximating and partitioning complex manifolds.** (a) Complex, curved surfaces can be well approximated by neighbourhood graphs. A simple graph connects each point with its k closest neighbours (kNN graph). As more points and regions are measured, the complex structure of the object can be revealed. (b) The elephant graph (in a) is clustered using the *leiden* clustering algorithm (resolution r = 0.5). The resulting clusters are shown as colours on the data's 3D model (top) and t-SNE embedding (bottom). (c) Clustering resolution is arbitrary. Similar to b, the plots show clustering with increased resolution (r = 3). The clusters are smaller but capture equally valid anatomical elements. (d) Overview of the E12.5 fetal pancreas dataset shown on a t-SNE embedding. (e) Expression of genes illustrating epithelial (*Epcam*), mesenchymal (*Col3a1*) and cell cycle (*Mki67*) signatures. (f) Coarse clustering of the mesenchymal and epithelial populations separating cycling cells within each group. (g) Hierarchical clustering of the coarse clusters (f) on the basis of the cosine similarity of the whole-transcriptome average profiles is driven by the cell-type differences. Neither t-SNE embedding nor hierarchical clustering reveal that the cycling clusters (E2, V2, M2) are distinguished from their corresponding populations by a very similar cell cycle signature. Adapted from Kharchenko, 2021, Figure 4.

2

SKIN BIOLOGY

2.1 SKIN MORPHOLOGY AND FUNCTION

The skin is an important organ in most animals, in the form of a soft and flexible layer, accounting for up to 5% of total body weight (Sontheimer, 2014). It has been shaped by the adaptation of animals to various environments, resulting in a highly variable morphological organ with specialised structures such as hair, scales, nails, hooves, and spurs (Montagna, 1967).

Skin serves a vast array of functions, mainly protection against diverse forms of trauma, including thermal, chemical, and ultraviolet radiation. In addition, it prevents trans-epidermal water loss (TEWL), that is, the movement of water from lower skin layers into the atmosphere. Other relevant associated functions are regulating body temperature, synthesis of vitamin D and perceiving sensory stimuli from the environment through its network of nerve endings (Farage et al., 2010; Yannas, 2014). These functions demonstrate how the skin is a vital component of the organism, contributing to overall health and well-being. The importance of the skin as a barrier is illustrated by the mortality associated with large surface area burns, where increased transepidermal water loss culminates in dehydration, renal failure and shock (Abdo et al., 2020).

2.1.1 *Epidermis*

The epidermis is the outermost layer of the skin. In humans, its thickness averages 0.4mm ranging from 0.03 mm on the eyelids to 1.5 mm on the palms and soles. The epidermis is composed of four major cell types: keratinocytes, melanocytes, Langerhans cells, and Merkel cells, with keratinocytes making up more than 90% of cells. The epidermal structure is layered into four main strata: *stratum basale*–basal layer–, *stratum spinosum*–spinous layer–, *stratum granulosum*–granular layer–and *stratum corneum*–cornified layer–(Barbieri et al., 2014). In palms and soles, a thin layer between the granular and cornified layers is present: the *stratum lucidum*–clear layer– (McGrath et al., 2004). The lower portion of the dermis shows downward invaginations, called *rete ridges*, and upward interdigitating structures, *dermal papillae*, which tridimensionally resemble the bottom of an egg carton. These structures increase the surface area and increase oxygen, nutrient and waste exchange rate, as well as strengthen the connection between the epidermis and dermis (Barbieri et al., 2014).

Keratinocytes are ectodermal-derived and can be distinguished from melanocytes and Langerhans cells in the epidermis by their larger size, intercellular bridges, and ample cytoplasm (Abdo et al., 2020; Barbieri et al., 2014). Keratinocytes proliferate to heal wounds, transport water and urea through aquaporins, receive melanin from melanocytes, control water permeability and participate in innate and adaptive immunity through antimicrobial peptide secretion and the presence of Langerhans cells, respectively (Horst et al., 2018). Additionally, they release neurotransmitters, neuropeptides, hormones and cytokines which transmit and process sensory information in the CNS, regulate skin permeability barrier function and can affect the physiology of multiple systems (Baroni et al., 2012; Denda et al., 2007).

2.1.1.1 *Epidermal layers*

STRATUM BASALE The stratum basale is the bottom layer of the epidermis, next to the dermis, composed of cuboidal keratinocytes, and has the most mitotic activity. Keratinocytes from the basal layer move upward through the spinous and granular layers, undergoing changes as they mature and reach the stratum corneum over 30-50 days (Barbieri et al., 2014). Only 15% of basal layer cells are constantly involved in renewing the epidermis; the rest are resting and activate during skin rejuvenation or injury (Baroni et al., 2012). At the tips of epidermal rete, slow-cycling keratinocytes with stem cell characteristics can rapidly proliferate in response to injury (Barbieri et al., 2014).

Keratinocytes from stratum basale are arranged in a single layer along the basement membrane and contain large oval nuclei with abundant organelles and a basophilic cytoplasm due to the active synthesis of protein used by cells in upper layers. These keratinocytes often contain melanin pigment acquired from nearby melanocytes (Barbieri et al., 2014).

Keratins are structural proteins produced by keratinocytes and form a web-like pattern of intermediate filaments that originate from a perinuclear ring, extend through the cytoplasm, and connect at junctional desmosomes and hemidesmosomes. During differentiation, keratins interact with filaggrin to form highly ordered arrays (Baroni et al., 2012). Basal keratinocytes mainly produce keratins 5 and 14, while suprabasal keratinocytes mainly produce keratins 1 and 10. Mutations in keratins can cause blistering diseases. Hemidesmosomes and desmosomes secure keratinocytes to the basement membrane and to each other. Tonofilaments made of keratins provide internal structure and can comprise up to 80% of the total protein in differentiated keratinocytes (Barbieri et al., 2014).

Another function of keratinocytes in basal and the spinous layers is the synthesis of vitamin D. The precursor 7-dehidrocholesterol produced in the keratinocytes is transformed by sun UV light into pre-vitamin D3, which is further converted into vitamin D3. This molecule goes through 2 hidroxilations, the first one in the liver and the second one in the kidneys until reaching its final form– 1,25-dihidroxyvitamin D– (Abdo et al., 2020). Vitamin D is a fat-soluble prohormone steroid primarily acknowledged for its endocrine role in calcium homeostasis and plays a lesser role in the regulation of skin differentiation, immune function, hair follicle cycling, photo-protection, upregulation of innate immune defences and wound healing (Mostafa et al., 2015).

STRATUM SPINOSUM The stratum spinosum layer is typically 5 to 10 cell layers thick and is located above the stratum basale. The cells in this layer adhere to each other by desmosomes, which give the cells a spiny appearance due to the shrinking of the microfilaments. The cells in this layer express keratins 1 and 10 and have large, pale-staining nuclei, as they are active in synthesizing cytokeratin,

which forms tonofibrils and desmosomes for strong connections between cells. Desmosomes play a significant role in maintaining the structural integrity of the epidermis, as evidenced by blistering disorders due to mutations. This layer also contains Langerhans cells (Barbieri et al., 2014).

STRATUM GRANULOSUM The stratum granulosum thickness ranges from 1 to 10 cells. Keratinocytes in the stratum granulosum are flatter and more irregular in shape than those in the stratum spinosum, and they have deeply basophilic keratohyalin granules. Keratinocytes also release lamellar granules and keratinosomes, which create a permeability barrier to water and facilitate cell adhesion in the stratum corneum (Barbieri et al., 2014).

STRATUM CORNEUM The outermost layer of the epidermis, the stratum corneum, provides a permeability barrier that prevents TEWL and protects against external agents, including bacteria, fungi, and chemicals (Abdo et al., 2020; Baroni et al., 2012). This layer is composed of corneocytes, keratinocytes that, in their differentiation process, lose their nuclei and other organelles and become completely keratinised–up to 80%–(Abdo et al., 2020; Barbieri et al., 2014). The desmosomes that had held the keratinocytes together begin to disappear or become nonfunctional in this layer, and the filaggrin released from the granules of the granular layer aggregates the keratin into bundles that promote their characteristic flattened shape (Baroni et al., 2012). The skin in this layer has an acidic pH that may have an important function in preventing infection (Wilhelm et al., 1990).

In the upper stratum corneum, enzymes such as steroid sulfatase, break down the lamellar granules' components from the granulocytes' apex that facilitate cell adhesion in the stratum corneum, resulting in the desquamation of the uppermost cells. The function of corneocytes and the contents of the lamellar granules acts as a "brick and mortar" wall, where the corneocytes are the bricks, and the content of the lamellar bodies acts as the mortar (Abdo et al., 2020). The main contents of the lamella bodies are lipids such as ceramides, cholesterol and free fatty acids, most of them produced from the hydrolysis of sphingolipids. Ceramides and fatty acids help fill and compact the space between the corneocytes, acting as a barrier, and cholesterol may also help add a degree of fluidity and flexibility to the membrane system (Baroni et al., 2012).

2.1.1.2 *Minor epidermal cell types*

MELANOCYTES Melanocytes play a major role in protecting the skin from ultraviolet radiation through the production of melanosomes that are transferred to keratinocytes to protect their nuclei (Barbieri et al., 2014). These cells are located mainly at the basal and suprabasal layers of the epidermis, although the melanocyte stem cell reservoir lies in the bulge of hair follicles (HFs), from where they migrate to the epidermis and mature (Abdo et al., 2020).

After sun exposure, it is thought that p53 activation in keratinocytes induces the production of a melanocyte-stimulating hormone resulting in increased melanin production by melanocytes. A complex tyrosine metabolism pathway produces melanin at the melanosomes, which later is transferred to the cell periphery along microtubules. Melanin is further transferred from melanocytes to basal keratinocytes, with one melanocyte supplying 30-40 keratinocytes (Costin et al., 2007).

LANGERHANS CELLS Langerhans cells are a type of dendritic antigen-presenting cell (APC) found in the epidermis. They have clear cytoplasm and can be difficult to differentiate from other cell

types such as T cells, macrophages, and melanocytes, on hematoxylin- eosin-stained preparations. Langerhans cells are anchored to neighbouring keratinocytes by E-cadherin, and they phagocytose antigens in the epidermis and then migrate to draining lymph nodes by downregulating this cadherin (Barbieri et al., 2014).

Langerhans cells present antigens to T cells using MHC class II molecules and CD1a, which can present microbial lipids to T cells. The Langerhans cell marker CD1a also promotes effector functions and the secretion of cytokines in the dermal layers. After T cells have been activated in the lymph nodes, they migrate back to the epidermis as memory T cells and upregulate inflammatory pathways. Langerhans cells also express an antigenic peptide with an affinity towards MHC-I molecules, which promotes the function of cytotoxic T cells ($CD8^+$) (Abdo et al., 2020).

MERKEL CELLS Merkel cells are specialised epidermal cells found in the skin and hair follicles (Abdo et al., 2020). They are rare and located in the basal layer of the epidermis and around the bulge region of hair follicles. They are believed to arise from epidermal ectoderm rather than neural crest. Merkel cells have an established function as slowly adapting mechanoreceptors that detect light touch, as well as spatial features such as curvature and edge.

Recent research suggests that Merkel cells may have endocrine and paracrine functions. This idea is supported by the fact that Merkel cells contain neuroendocrine markers such as chromogranin A and a variety of neurosecretory substances such as neuropeptides. Furthermore, some Merkel cells are not associated with a nerve, supporting the idea of additional functions beyond mechanoreception. Despite this, much about the role of Merkel cells remains to be discovered (Barbieri et al., 2014).

2.1.2 *Dermo - epidermal juntion*

The skin must maintain its mechanical stability to prevent the shearing of the epidermis from the dermis. To achieve this, neighbouring keratinocytes are bound to each other by desmosomes and a meshwork of tonofilaments, and the epidermis is further stabilised by its interlocking projections dermal rete and papillae that increase the surface area for binding interactions between these layers (Abdo et al., 2020).

The dermo-epidermal junction (DEJ), which represents the interface between the epidermis and dermis, consists of an aggregation of proteins and structures known as the basement membrane (BM). One of the classical functions of the BM is mechanical stability of skin, but it also allows for diffusion of molecules such as growth factors and cytokines, including keratinocyte and platelet-derived growth factors (PDGFs), that regulate both keratinocyte and dermal fibroblast functions through regulation of activation and release. In addition, BM plays an important role during the remodelling process after injury and damage to BM by cancer leads to cell activation in the stroma (Saikia et al., 2018).

The basement membrane can be divided into the superior *lamina lucida* and the inferior *lamina densa* (Barbieri et al., 2014). Hemidesmosomes at the basal keratinocytes bind through several molecules–e.g., integrin $\alpha 6\beta 4$, bullous pemphigoid antigens or collagen XVII–to the laminin 332 at the lamina densa, primarily composed of type IV collagen arranged in a chicken-wire-like lattice (Barbieri et al., 2014; Saikia et al., 2018). Some of these components, like integrin $\alpha 6\beta 4$, have been found to modulate basal cell cytoskeleton organisation, proliferation, differentiation, and apoptosis.

Then, anchoring fibrils, which contain type VII collagen, link the type IV collagen of the lamina densa to the underlying anchoring plaques in the dermis, which consist primarily of aggregates of collagen VII (Barbieri et al., 2014; Goldsmith et al., 2012). Mutations in several components of the BM, like laminin 332 or collagen VII, result in different types of epidermolysis bullosa (Goldsmith et al., 2012). A detailed schematic of the BM can be observed in Figure 14.

2.1.3 *Dermis*

Below the BM lays the dermis, which is 15-40 times thicker than the epidermis. The main aim of the dermis is to reinforce the skin and provide structural support by absorbing mechanical forces. Additionally, through the dermis pass vascular structures that nourish the rest of the skin, neural structures that communicate signals from within the dermis and epidermis to the nervous system, and lymphatic vessels that facilitate immune communication. The dermis is mainly composed of an extracellular matrix produced by fibroblasts generally producing collagen and elastic fibres, but it also contains a wide range of cell and molecule types (Abdo et al., 2020; Barbieri et al., 2014).

2.1.3.1 *Cell composition of the dermis*

FIBROBLASTS Fibroblasts are the primary cell type in the dermis and are mesenchymal cells producing extracellular structural proteins, glycosaminoglycans, collagen and elastin fibres that provide structural support (Abdo et al., 2020). Dermal fibroblast origins, functions and other aspects are extensively discussed in Section 2.2.1.

DENDRITIC CELLS AND MACROPHAGES Macrophages and dermal dendritic cells (DCs) are bone marrow-derived cells that play important roles in innate and adaptive immunity. Under basal conditions, macrophages and dermal dendritic cells are most commonly found around the perivascular space. Although macrophages can act as antigen-presenting cells (APCs), their main function is to engulf proteins and cellular structures to preserve ECM integrity. Additionally, they can secrete several cytokines to generate adaptive immune responses. On the other hand, dermal dendritic cells capture, process, and present antigens to T cells, similar to Langerhans cells in the epidermis. After encountering antigens in the dermis, dermal dendritic cells migrate to the lymph nodes, presenting antigens to T cells (Barbieri et al., 2014).

MAST CELLS Mast cells are cells that are found in the bone marrow and are important in the initiation of both immune and inflammatory responses. They are most commonly located around blood vessels, especially postcapillary venules, and contain a variety of granules, including histamine, heparin, cytokines, chymase, and tryptase. Mast cells serve pro-inflammatory and anti-inflammatory roles, and one of their main functions is to enhance the recruitment of neutrophils during the innate immune response. Additionally, mast cells play a role in Th2 adaptive immunity and are involved in immediate hypersensitivity reactions through the mediation of IgE and degranulation (Barbieri et al., 2014).

2.1.3.2 *Dermal microvascular unit*

The cutaneous tissues are characterised by the presence of a complex network of blood vessels, lymphatic vessels and nerves known as neurovascular plexi or, specifically within the skin, the dermal microvascular unit (Figure 15) (Nolano et al., 2012). Generally, all three components can be found intertwined in each plexus, although they can also surround other structures, like HFs or glands. These plexi provide essential functions such as circulation, nutrition, and cutaneous sensation to the skin (Barbieri et al., 2014).

VASCULAR PLEXUS The dermal vasculature of the skin comprises three major structures, each with distinct locations and functions. The superficial vascular plexus is the first structure located at the boundary between the papillary and reticular dermis. It contains arterioles and venules with capillary loops extending into each dermal papilla to provide nutrition and waste removal to the epidermis.

The second structure, the deep vascular plexus, defines the border between the reticular dermis and the subcutaneous fat, and contains larger-calibre vessels. The third structure, the subcutaneous plexus, is located in the hypodermis and contains larger arteries and veins. All three structures are connected vertically with ascending arterioles and descending venules (Barbieri et al., 2014).

In addition to its role in nutrient communication, dermal vasculature has a crucial role in temperature regulation. The vasculature contains specialised structures called glomus bodies, a form of vascular anastomosis. These bodies play a key role in temperature regulation by controlling blood flow to the area. When the anastomoses are open, large increases in blood flow occur, allowing heat dissipation. On the other hand, vasoconstriction and closure of the anastomoses cause an abrupt cessation of heat loss (Abdo et al., 2020).

NERVE PLEXUS The cutaneous nerve plexus belongs to the peripheral nervous system and includes afferent fibres and autonomic fibres of the sympathetic nervous system responsible for executing involuntary responses. The afferent fibres are responsible for perceiving external stimuli and transmitting them to the central nervous system for interpretation. Like the vascular plexus, nerve fibres form vast meshes within each plexus and play an important role in various physiological processes, including thermoregulation, blood vessel tone, and the regulation of glandular secretions (Barbieri et al., 2014).

The terminals of the nerve fibres can take on three forms. They can form mechano-receptors that are highly specialised for the perception of mechanical stimuli, they can penetrate the keratinocytes and end up in the epidermis or they can innervate structures such as glands, blood vessels, and HFs (Iribar et al., 2018; Weddell, 1945).

The innervation of HFs is especially relevant, since it is necessary for mechanosensory purposes. In HF, nerves surround each hair into a structure called *lanceolate complex*, in which nerve endings protrude from the *ring* in the hair upwards, arranged parallel to the hair follicle long axis. It has been observed that epidermal stem cells in that area secrete a protein, EGFL6, necessary for proper patterning, touch responses, and αv integrin-enrichment of lanceolate complexes (Cheng et al., 2018a).

LYMPHATIC PLEXUS The dermal lymphatics play a critical role in the drainage of immune cells, fluids, and macromolecules, including lipids and proteins. These lymphatic structures parallel major vascular structures and can be found just below the superficial vascular plexus (Barbieri et al., 2014). The main immune populations trafficking through the dermal lymphatic vasculature are Langerhans cells and DCs. In addition, the lymphatic vasculature plays a major role in reverse cholesterol transport by recirculating high-density lipoprotein-bound cholesterol from the interstitial space (Petrova et al., 2017).

It has been observed that poorly functioning dermal lymphatics can result in several adverse effects, including impaired immunity and recurrent infections, as well as fibroplasia secondary to chronic lymphedema (Barbieri et al., 2014).

2.1.3.3 *Extracellular matrix*

The extracellular matrix (ECM) is the large embedding produced and re-structured by dermal fibroblasts. The most abundant components of ECM are collagen I and III (Lynch et al., 2018), although there are other collagens, as well as other matrix molecules produced by these fibroblasts, including fibronectin, elastin, periostin, and tenascin-C. Other key ingredients of the ECMs are bridging proteins and matrix-modifying enzymes–there are more than 300 proteins–which, together, are referred to as the matrisome (LeBleu et al., 2020; Lendahl et al., 2022).

COLLAGEN Collagens are the main components of ECM, which provide the skin with tensile strength and resistance to plastic deformation (Cole et al., 2019; Gosline et al., 2002). They account for around 30% of protein mass in humans, and more than 70% of dermis dry weight (Bauer et al., 1979). In the adult dermis, the majority of collagens are fibrillary collagens, including type I (80%), type III (15%), and type V collagen (5%), and comprise up to 70–80% of the skin dry weight (Haydont et al., 2019). During infancy, a higher proportion of collagen III changes to a predominant collagen I deposition (Haydont et al., 2019). The collagen turnover is infrequent, with an estimated half-life of 15 years in human skin (Verzijl et al., 2000).

Mature collagen molecules comprise three intertwined alpha chains, each characterised by the distinctive amino acid triplet repeat Gly–X–Y. The amino acid variability at the X and Y positions and the frequency and length of non-collagenous domains are responsible for determining the specific type of collagen (Cole et al., 2019). The production of collagen starts with the synthesis of pre-procollagen α chains in the rough endoplasmic reticulum, some of which residues then undergo C-vitamin-dependent hydroxylation. The α chains form a triple helical structure called procollagen, secreted from the cell and cleaved by metalloproteases to form tropocollagen. The tropocollagen molecules self-assemble and are cross-linked to form collagen fibres, providing tensile strength (Barbieri et al., 2014).

The formation and assembly of collagen fibres are carried out through nucleation and regulation. The “nucleating” collagens, mainly type V and XI, are located in the centre of the fibres and initiate the incorporation of other types of collagen. The architecture of the dermal ECM can be further modulated by the family of ‘fibril-associated collagens with interrupted triple helices’ (FACIT), including type XII, XIV, and XVI collagens. These proteins help bridge collagen I and III fibrils with proteoglycans like decorin and perlecan, which regulate fibrillogenesis (Haydont et al., 2019; Reed et al., 2002).

Post-translational modifications to type I collagen include the previously mentioned intracellular conversion of lysine to hydroxylysine and the extracellular conversion of lysine and hydroxylysine into aldehydes through lysyl oxidase (LOX). This LOX-directed conversion of amino acids into aldehydes enables non-reducible crosslinking that stabilises the dermal collagen and makes it resistant to proteolytic cleavage (Cole et al., 2019).

OTHER FIBERS Aside from collagen, the ECM also contains other components, like glycosaminoglycans (GAGs), proteoglycans, the elastic fibre network, and water (Haydont et al., 2019).

GAGs are a group of long, negatively charged polysaccharides. They play important roles in various biological processes, including lubrication of joints, tissue hydration and elasticity maintenance, and cell growth and division regulation. In the dermis, GAGs like hyaluronic acid, which have a high water-binding affinity, help keep hydration levels and confer stiffness and resistance to deformation (Juhlin, 1997). Proteoglycans are macromolecules composed of a protein nucleus, to which GAGs–e.g. chondroitin sulfate, dermatan sulfate–adhere and branch. Proteoglycans have great molecular variability–e.g. lumican (LUM - 38 kDa) or decorin (DCN - 36 kDa) to aggrecan (ACAN - 220 kDa) or versican (VCAN - 260-370 kDa). This variability confers them a wide range of functions, mainly the regulation of size and assembly of collagen–by perlecan, lumican, decorin, versican and tenascin X (Kadler et al., 2008)–or interaction with factors and cytokines, facilitating their retention within the ECM (Sorrell et al., 2009).

Elastic fibres, akin to their name, confer elasticity to the skin and other structures like vascular vessels by allowing deformation forces and returning to their original state (Naylor et al., 2011). Elastic fibres form an organised and highly structured network that runs from the DEJ to the deep dermis. For instance, oxytalan fibres penetrate the DEJ perpendicularly and are composed of fibrillin-rich microfibrils; and elaunin fibres, composed of fibrillin and elastin in equal proportions; also run perpendicular to the epidermis and are slightly thicker than oxytalan fibres. In the lower layers of the dermis, the fibres are essentially composed of elastin, thicker, and oriented parallel to the epidermis (Haydont et al., 2019).

There are other large molecules, like fibronectins or laminins. Fibronectins are a set of glycoproteins that connect the collagens in the ECM with the integrins at the cell surface to facilitate the movement of the cells through the ECM. Laminins are located at the basal membrane and connect components like integrins to collagens or form dense nets that absorb tensional forces within the dermis (Lewin et al., 2006).

MMPS AND TIMPS Another key part of the ECM are matrix metalloproteinases (MMPs) and their respective tissue inhibitors (TIMPs). The proper balance of MMPs and TIMPs is crucial for the normal deposition and degradation of the ECM. MMPs are calcium-dependent zinc-containing endopeptidases and degrade a large array of molecules like different collagens, aggrecan, laminin, elastin or fibronectin (Van Linthout et al., 2014).

For instance, in immune settings, MMP-1 initiates cleavage of type I and type III collagen fibrils at a specific location within their central triple helix, and, once cleaved, the fibrils are further degraded by MMP-3 and MMP-9 (Fisher et al. 1996; Brennan et al. 2003). MMP-9 is the most relevant pro-fibrotic MMP, and its inhibition or deletion has been shown to reduce fibrosis in models of dilated cardiomyopathy and myocardial infarction (Van Linthout et al., 2014).

MMPs in the skin are relevant to maintain homeostasis, and dysregulation can lead to different diseases. CCN1 has been shown to up-regulate MMP-1 via interaction with $\alpha V\beta 3$ integrin and thereby contributes to aberrant dermal collagen homeostasis in human skin (Qin et al., 2013). TGF-β/Smad signalling pathway promotes the synthesis of collagen, fibronectin, DCN, and VCAN, and inhibits the synthesis of ECM-degrading metalloproteinases (MMPs), functioning as the major regulatory pathway for ECM production (Quan et al., 2015).

2.1.3.4 *Papillary and reticular dermis*

The dermis can be divided into two main parts, the papillary and reticular dermis, easily distinguishable in histological samples. The papillary dermis is a thin–300-400 μm–zone that approximates the epidermis and is composed of loosely arranged collagen. The thicker reticular dermis is located beneath the papillary dermis and connects the subcutaneous fat and tissue (Barbieri et al., 2014). The two layers differ regarding the secretion of several extracellular matrix components, growth factors, cytokines, MMPs and TIMPs, as well as their fibroblasts–discussed in Section 2.2.1– (Stunova et al., 2018).

The papillary dermis comprises primarily type III collagen fibres arranged in thin and poorly organised bundles with a high density of active fibroblasts (Barbieri et al., 2014; Stunova et al., 2018). Within the nuclear collagens, type V collagen is the predominant (Nauroy et al., 2017); regarding FACIT collagens, types XII and XVI are more abundant. Moreover, the elastic fibre disposition, mainly of elaunin, is perpendicular to the epidermis (Haydont et al., 2019).

On the other hand, the reticular dermis is composed primarily of large and well-organised type I collagen bundles, with a lower density of fibroblasts and other cell types (Barbieri et al., 2014). Moreover, type IX nucleating and type XIV FACIT collagens are more predominant in this layer (Nauroy et al., 2017). Regarding elastic fibres, they are thicker than in the papillary dermis, and run parallel to the epidermis (Haydont et al., 2019). It has been shown that a higher level of ECM deposition correlates with a higher promotion of vascularization by cells near adipose tissue, which might correlate with papillary dermis cells being most involved in human skin wound healing, compared to cells located at the reticular dermis (Stunova et al., 2018).

2.1.4 *Adnexa*

2.1.4.1 *Hair follicle and pilosebaceous unit*

Hair is a characteristic feature of mammals. It involves physical protection, camouflage, thermal insulation, perception of stimuli, and sweat dispersal. In humans, hair has a mainly social and aesthetic role, covering almost the entire body, except the palms of the hands and feet, as well as discrete genital areas. Two types of hair are found on the body: terminal hair that is thick, long and pigmented; and fine, short and poorly pigmented hair, called vellus hair (Marks et al., 2019). In mice, more than four types of HFs differ in length, thickness, and in the presence or absence of kinks in the hair shaft (Schlake, 2007).

Anatomically, the HF is an invagination of the epidermis with a channel through which the hair exits the skin. The distal region (closest to the epidermis), named *infundibulum*, spans the epidermal opening to the insertion of the sebaceous gland. Below this is the isthmus, where the arrector pili

muscle (APM) joins. In its lower zone are the epithelial and melanocytic stem cells that support the regeneration of the follicle and the pigmentary unit; this region is known as the bulge. Further down the bulge, the deepest part of the HF contains the hair bulb, which contains the dermal papilla cells (DP). DP comprises packed fibroblasts that exert inductive signals to determine cycle length, size, and hair type (Barbieri et al., 2014; Rahmani et al., 2014).

If the HF is cut transversely, there are three main layers can be seen–from inwards to outwards–: (1) the inner root sheath (IRS), a thin layer composed of several sublayers that helps shape and mould the developing hair; (2) the outer root sheath (ORS), a thick layer that joins the epidermis at the upper portion of the HF, and (3) the connective tissue sheath (CTS), made up of loose collagen fibres and elongated cells that surround the entire follicle in transversal and longitudinal dispositions towards the near and far ends of the CTS, as well as intertwined blood vessels and nerves (Martino et al., 2020; Müller-Röver et al., 2001). The cells that conform the CTS are called dermal sheath (DS) cells (Martino et al., 2020). CTS/DS is physically separated from the ORS by a cell-free glassy membrane (Müller-Röver et al., 2001; Schneider et al., 2009). Although not considered a layer *per se*, the medulla or cortex is the hair root, and is surrounded by the IRS.

The HF is made up of cells derived from the ectoderm (keratinocytes), the mesoderm (fibroblasts of the PD and CTS/DS) and the neural crest (melanocytes), which are in constant communication through epithelial-mesenchymal interactions and allows maintaining basic processes such as the hair growth cycle, or morphogenesis during the embryonic period (Sennett et al., 2015). The HF, together with the sebaceous gland (SG) and the APM, form the pilosebaceous unit.

The SG is an exocrine gland that opens into the HF and secretes sebum, an oily compound that lubricates the skin and hair. Sebaceous secretions also play an important thermoregulatory role in conjunction with apocrine glands. In hot conditions, the secretions emulsify the sweat produced by the eccrine glands, producing a sheet of sweat that is not readily lost in drops of sweat (Zouboulis, 2004). Additionally, sebaceous lipids help maintain the integrity of the skin barrier, and its slightly acidic pH acts as a barrier to microbes (Zlotogorski, 1987).

The APM attaches to the bulge region of the hair follicle, which also contains the follicular stem cells, and comprises smooth muscle innervated by the autonomic nervous system. The function of the APM is to cause the hair to raise in response to stresses such as cold or fear (Barbieri et al., 2014).

HAIR CYCLE In all mammals, HFs follow a cycle of hair regeneration consisting of three main phases: anagen, catagen and telogen (Martino et al., 2020). The anagen phase, or growth phase, is the longest phase where the hair matrix is actively produced by the cells from the DP. In catagen, or regression phase, the hair matrix is liberated and the dermal canal where the hair grows is shrunk. During the catagen phase the DS contracts and DP retracts to the new base of the regressed hair. Lastly, during telogen, or resting phase, the DS surrounds the DP only, and not the rest of the HF; and both DS and DP remain *dormant* until a new signal enters the HF in the anagen phase. During the early anagen, the DP cells migrate towards the lower end of the dermis, and the DS surrounds the new hair canal. This process is schematised in Figure 16.

2.1.4.2 *Sweat glands*

Apart from the SGs adjoint to the HFs in the pilosebaceous unit, sweat glands are the other major type of gland in the hairy skin (Barbieri et al., 2014).

Sweat glands can be divided into two types: eccrine and apocrine glands. Eccrine glands function in thermoregulation and are widely distributed over the entire skin surface. The sweat these glands produce is mainly water, salt, and metabolic waste products. These cells are richly innervated by cholinergic fibres, which are thought to regulate sweat production in response to thermal, emotional, and gustatory stimuli. There are three million eccrine glands present in the skin that are capable of producing 1.8 l of sweat per hour (Barbieri et al., 2014). In addition, cytokines and proteolytic enzymes can be found in eccrine secretions, suggesting that eccrine glands may also play a role in skin immunity.

Apocrine glands have an unclear function and are composed of three segments: the intraepidermal duct, the intradermal duct, and the secretory portion. These cells are innervated by adrenergic and cholinergic fibres and are inactive until puberty. Apocrine secretions are a complex mixture of lipids and proteins, like sialomucin. Epidermal bacteria act on these molecules to produce odiferous molecules with biological effects. In lower mammals, apocrine secretions act as pheromone signals; however, their function is poorly understood in humans and may represent evolutionary vestiges (Barbieri et al., 2014).

2.1.4.3 *Sensory terminals*

There are two types of sensory receptors in the dermis: bare nerve endings, for nociception and thermal sensation; and encapsulated mechanoreceptors, such as Meissner's corpuscles, which detect moving touch and Pacinian corpuscles, which perceive vibration and brief touch (Abdo et al., 2020).

Regarding bare nerve endings, nociceptors are activated by noxious stimuli such as heat, pressure, or chemicals, and they play a crucial role in transmitting pain signals to the brain. They help protect the body by detecting harmful stimuli and triggering a withdrawal reflex, which helps to prevent further injury. Thermoreceptors, on the other hand, are specialised nerve endings that respond to changes in temperature. They help regulate body temperature by detecting changes in skin temperature and transmitting this information to the brain, which then adjusts the level of sweating or shivering in response. Both are type C fibres, which are unmyelinated and have the lowest response (Glatte et al., 2019).

On the other hand, Meissner and Paccini fibres are of type Aβ, that is, thick, highly myelinated, and rapid conducting with mechanoreceptive properties (Glatte et al., 2019). Meissner's corpuscles are found most commonly in the dermal papillae of the fingertips, composed of laminar cells functioning similarly to Schwann cells. Meissner's corpuscles function as rapidly adapting mechanoreceptors important for light touch sensation. Pacinian corpuscles are commonly found in the dermis of weight-bearing skin and subcutaneous fat, and they are rapidly adapting mechanoreceptors important for pressure and vibration sensation.

Additionally, the skin can detect a pattern of mechanical stimuli on a smaller scale than would be expected merely by the location of the nerve terminals (Abdo et al., 2020). Keratinocytes contain

sensors that can recognise and respond to various physical and chemical environmental factors. For example, TRPV1 is strongly expressed near the skin's surface, consistent with its role in detecting temperature changes. TRPV3 and 4 are activated by mechanical stress and changes in osmotic pressure (Becker et al., 2005; Caterina et al., 1997; Vallbo et al., 1999). Subepidermal Schwann cell networks are likely to be involved in the skin sensory role (Iribar et al., 2018).

2.1.5 *Hypodermis*

Below the dermis is the subcutaneous fat, which comprises lobules of adipocytes separated by fibrous septa, which have arteries, veins, and nerves that supply the subcutaneous fat and dermis running through them (Barbieri et al., 2014). This layer provides mechanical support, insulation, and thermoregulation and serves as a store for energy (Abdo et al., 2020).

In mice and humans, this layer is divided into two distinct layers: the dermal white adipose tissue (dWAT) and the subcutaneous white adipose tissue (sWAT). In mice, both types of fatty tissue are separated by the *panniculus carnosus* (PC) muscle. Humans lack PC, although it is evolutionarily conserved in discrete locations (Naldaiz-Gastesi et al., 2018). In humans, dWAT and sWAT are contiguous, making their histological identification difficult; however, various clinical observations assert that the *fascia superficialis* demarcates both adipose compartments (Zwick et al., 2018).

The adipocytes that make up dWAT or sWAT differ in their metabolism, morphology, and embryonic origin (Driskell et al., 2014). In response to tissue remodelling cycles coupled to the hair cycle, dWAT has a high cell turnover rate, while other fatty deposits such as sWAT or vWAT (visceral WAT) have extremely limited cell turnover (Rivera-Gonzalez et al., 2016).

It has been seen that dWAT is involved in processes as diverse as hair cycle regulation, ECM remodelling, healing, hormonal segregation and innate immune defence against infections (Guerrero-Juarez et al., 2018; Zwick et al., 2018). Generally, dermal adipocytes might be considered "ROS-sensing adipocytes", able to orchestrate a dedifferentiation and/or an immune response depending on the finest ROS tuning within those cells (Segalla et al., 2021).

In humans, the association of dWAT adipocytes with pilosebaceous units is particularly intriguing. In these structures, dermal adipocytes form clusters of cells exhibiting a *cone* geometry–observable under electron microscopy–extending to the first half of the hair follicle, a structure that might suggest a sort of "sensing" offshoots. In that sense, dWAT is probably a fine sensor of epidermal stress towards sWAT or a sensitive projection of sWAT towards epidermis (Segalla et al., 2021).

Within the WAT, a secondary classification can be performed. Within the lobules, two types of adipocytes can be found. Most adipocytes in the subcutaneous fat are white adipocytes: spherical cells with a large central droplet and a peripherally located nucleus. White adipocytes function in the storage of energy. Brown adipocytes, on the other hand, more commonly found in visceral adipose tissue than subcutaneous fat, are polygonal cells with several lipid droplets and a round nucleus. These adipocytes express uncoupling protein 1 (UCP1), which uncouples oxidative phosphorylation from ATP generation. This process is responsible for brown adipocyte's role in thermoregulation (Barbieri et al., 2014).

2.1.6 *Differences between mouse and human skin morphology*

Both humans and mice are mammals, which gives them a high degree of similarity in many aspects, including skin. Using mice as a research model has greatly contributed to understanding normal and pathological cutaneous repair, but there are challenges in bridging the gap between preclinical and clinical studies because skin morphology and repair in mice do not perfectly mirror that of humans. To properly interpret the results, it is important to understand the particularities of each species (Zomer et al., 2018) (Figure 17).

Mice have small bodies, shorter life expectancy, and differences in physiology compared to humans, making it difficult to effectively reproduce the pathogenesis of certain human diseases such as diabetes (Cibelli et al., 2013). In light of these limitations, other models, such as pigs, which are physiologically closer to humans, are being used, but they pose greater handling difficulties than mice. Despite these limitations, most studies on skin wound healing are still performed in mice (Zomer et al., 2018).

Although human and mouse skin have the same layers of cells in the dermis and epidermis, they greatly differ in thickness and number. Human skin is relatively thick, firm, and adheres to underlying tissues, while mouse skin is thinner and looser. Each major layer in the human epidermis contains more layers–5-10 each–than mice epidermis–2-3 each– (Abdullahi et al., 2014), which increases its barrier function and increases percutaneous absorption (Bronaugh et al., 1982). The rete-papillae structure between the human dermis and epidermis is lacking or less apparent in mouse (Zomer et al., 2018).

Mouse skin is covered by hair follicles, while human skin has sparse and uneven hair follicle distribution, and most of the epidermis is classified as interfollicular (Gerber et al., 2014). This affects healing since areas with high hair density heal faster than those with few follicles (Martinot et al., 1994). Hair follicles in mice are regenerated after skin healing, but this process is limited to the fetal stage in humans (Gurtner et al., 2008).

In addition, *panniculus carnosus* is present in mouse subcutaneous tissue but not in humans, which influences skin biomechanics (Gerber et al., 2014). That thin layer of muscular tissue gives contraction potential to the skin since large wounds heal mainly by contraction and union of edges. In fact, up to 90% of excisional wounds in mice close by contraction. The human dermis, in contrast, is firmly attached to the subcutaneous tissues, and contraction is highly variable and much less pronounced than in mice (Fang et al., 2008).

Eccrine sweat glands, responsible for body temperature control and playing a role in innate immune response, are present in humans but not in mice (Gerber et al., 2014). The percentages of peripheral blood leukocytes also differ between humans and mice, which affects innate and adaptive responses (Doeing et al., 2003). Moreover, Langerhans and $CD8^+$ T cells populate the human epidermis, whereas the murine epidermis contains a specific population of $\gamma\delta$ dendritic epidermal T cells, which are fundamental for skin homeostasis and tissue repair, and could explain hair follicle regeneration in mice but not in humans (Smithey et al., 2014).

2.2 FIBROBLAST BIOLOGY

Fibroblasts are cells responsible for forming and maintaining connective tissues rich in ECM that supports various vital organ functions. These functions include protecting the skin from blunt and sharp injuries and facilitating elasticity and recoil in the lung during breathing (Plikus et al., 2021).

Fibroblasts were first described as a distinct cell type in 1859 by German pathologist Rudolf Virchow (Virchow, 1859), who called them *Spindelzellen des Bindegewebes*—"spindle-shaped cells of the connective tissue", and the term "fibroblast" was first proposed by Ernst Ziegler and Alexander Maximov to describe cells that produce new connective tissue upon healing (Maximow, 1905; Ziegler, 1895). In 1879, Matthias Duvall distinguished fibroblasts from epithelial cells within the mesenchyme of chick embryos (Duval, 1879). In 1896, this observation was replicated by Santiago Ramón y Cajal, who observed *células fusiformes* or *fibro-células* as essential producers of granulation tissue in healing skin wounds and scars (Ramón y Cajal, 1896).

It is relevant to understand the etymology of fibroblast further down, which explains why we coin the suffix *-blast* to a cell type that might be quiescent. Ziegler and Maximov used the term *-blast* as an indicator of active protein synthesis in these cells. In these studies, the quiescent fibroblasts of uninjured organs were sometimes referred to as fibrocytes, which displayed a small cytoplasm, few ribosomes and condensed chromatin, suggesting low levels of protein synthesis (Porter, 1964). It was assumed that fibrocytes give rise to fibroblasts upon wounding, and conversely, fibroblasts differentiate into fibrocytes during healing (Porter, 1964). The original meaning of fibrocyte was progressively lost, and fibroblast was adopted as the common name, irrespective of its state of activity. Nowadays, fibrocytes, bone marrow-derived circulating mesenchymal stem cells, have conferred a new meaning to the old *fibrocyte* term (Chong et al., 2019).

Fibroblasts have been very relevant cells in biology and are widely used. In the early days of cell culturing, fibroblasts were sturdy and easy to propagate on artificial surfaces such as glass and plastic. Thus, many cell lines used in research are derived from fibroblastic lineages. Additionally, the Hayflick limit, discovered in 1965 by Leonard Hayflick, who observed that some cultured cells stop dividing and senesce after a finite number of population doublings in vitro, was found on fibroblasts (Hayflick, 1965). This discovery disproved the belief that cultured cells could be passaged indefinitely.

FIBROBLAST IDENTITY AND HETEROGENEITY Fibroblasts are one of the few cell types found across the body's major organs. Thus, one of fibroblast biology's hallmarks is characterising fibroblast heterogeneity and discovering common aspects shared across all fibroblasts. However, the research on this matter has been hampered by ambiguous and diversified cell classification about, for example, lineage, morphology, location or growth characteristics (Lendahl et al., 2022).

Classically, one of the common physical characteristics of fibroblasts is that, in adult tissues, they appear as spindle- or stellate-shaped with an oval nucleus and a distinct endoplasmic reticulum (Lendahl et al., 2022); but this phenotype might also be shared with other cell types with a mesenchymal origin, and it is not precise enough for characterisation. One common way to characterise cell types is via their marker expression. Although not uniquely expressed in fibroblast, certain proteins, such as vimentin (VIM) and fibroblast-specific protein 1 (FSP1/S100A4), have served as useful markers to identify fibroblasts by immunohistochemical techniques (Dulbecco et al., 1983; Strutz et al., 1995).

Following the first cDNA microarray-based study of fibroblast heterogeneity, transcriptional analysis of fibroblast was conducted using bulk mRNA isolates obtained from multiple cells (Heruth et al., 2012). From these and other studies, certain molecular markers were proposed to distinguish fibroblasts from other cell types: platelet-derived growth factor receptor-alpha (PDGFRA), fibroblast activation protein alpha (FAP), and CD90 (Thy1), as well as the "classical" VIM and FSP1 (Lendahl et al., 2022). However, none of the markers listed above is specific for fibroblasts, as their expression is also observed in other cell types, such as perivascular, epithelial, and endothelial, and in substantially reduced amounts in some immune cells. For instance, Xie et al., 2018 observed that PDGFRA is a clear marker of lipofibroblasts in the lungs. Other markers have also been proposed, like COL1A1/COL1A2 (Philippeos et al., 2018), CAV1 (caveolin 1), DDR2 (discoidin domain receptor 2), platelet-derived growth factor receptor-beta (PDGFRB), podoplanin (PDPN) and alpha-smooth muscle actin (α-SMA/ACTA2) (LeBleu et al., 2020). Similarly, most of these markers are shared with perivascular cells, endothelial cells, CAFs, or other cell types in different organs.

FIBROBLAST INTRA-ORGAN HETEROGENEITY Despite the partially shared markers across tissues and organs, fibroblast heterogeneity spans species, organs and developmental stage boundaries. It has long been recognised that different fibroblast populations reside simultaneously in the same organ (LeBleu et al., 2020; Plikus et al., 2021). Several connective tissues harbour highly specialised cell types: adipocytes in fat, chondrocytes in cartilage, and osteocytes/blasts in bone. This translates into fibroblasts with both *in vivo* and *in vitro* data suggesting that these cell types are lineage-related and share a common mesenchymal origin and evolution (Lendahl et al., 2022). For instance, comparing fibroblasts from mouse heart, skeletal muscle, intestine, and bladder revealed that less than 20% of fibroblast-enriched genes overlapped between these four organs (Muhl et al., 2020). In fact, it has been observed that fibroblasts are more transcriptomically similar to their local tissue structural cells than to fibroblasts from other locations (Krausgruber et al., 2020).

The anatomy of the skin reveals that fibroblasts from different areas of the body have distinct transcriptional programs. The fibroblasts in various parts of the adult body come from various embryological sources, such as the neural crest, lateral plate mesoderm, and dermatomyotome (Fontaine et al., 1977; Houzelstein et al., 2000). This is reflected in the differences in the expression patterns of Hox genes, which determine the body plan along major axes such as craniocaudal, dorsal-ventral, and proximal-distal, including in humans (Chang et al., 2002; Rinn et al., 2006). Thus, different "HOX codes" of skin fibroblasts result in regional variations in matrisome and signalling factors, serving as "information" for nearby cells. For instance, HOXA13, which regulates distal identity, is expressed specifically by finger and foreskin fibroblasts in humans (Rinn et al., 2006).

An extensive review has recently been published by Elmentaite et al., 2022 indicating cell heterogeneity, including fibroblast heterogeneity, across organs based on single-cell atlases (Figure 18). Some examples of this heterogeneity are:

- Distinct intestinal fibroblast types exist in the villus tip (WNT5A$^+$), crypt base (WNT2B$^+$) and lamina propia (ADAMEC1$^+$) of the human gut. These cells function to collaboratively support stem cell and individual enterocyte differentiation along the crypt–villus axis (Elmentaite et al., 2020; Kinchen et al., 2018).

- In the endometrium, the transcriptional programme of specialised fibroblasts changes with the menstrual cycle phase. A recently identified population of $C7^{+}$ fibroblasts in the basal layer is present in both proliferative and secretory phases (Garcia-Alonso et al., 2021).
- In the bone marrow, $LEPR^{+}$ CXCL12-abundant reticular (CAR) cells have been observed to be primed to become adipocytes or osteoclasts (Dolgalev et al., 2021; Tikhonova et al., 2019; Wolock et al., 2019b).

The characterisation of heterogeneity across tissues is necessary to create a working atlas for the research community, and single-cell technologies have advanced significantly in their creation (Lendahl et al., 2022). However, certain caveats have to be addressed. Firstly, there is a loss of certain cell types during single-cell sample preparation; or cells firmly embedded within the ECM, such as pericytes and mesangial cells, are hard to extract as intact individual cells, leading to under-representation or contamination by transcriptomes from other cell types. Secondly, the cell type terminology in atlases should be as accurate as possible to avoid misclassification into other or new cell types. For instance, when fibroblasts were collected and sequenced as part of broader atlas projects, they were occasionally called by other names, such as stromal cells, mesenchymal stem cells, myofibroblasts and unknown mesenchymal cells, illustrating the ambiguities regarding cell nomenclature (Green et al., 2018; Tabula Muris Consortium, 2018).

FIBROBLAST ROLES Beyond gene markers, functional properties should be the ultimate discriminator of shared versus unique themes in fibroblast biology (Plikus et al., 2021). A major shared function of fibroblasts, as we described in Section 2.1.3.1, is ECM synthesis to create connective tissue by depositing fibre- and sheet-forming collagens, proteoglycans, elastin, fibronectin, microfibrillar proteins, and laminins, which collectively comprise the *matrisome*; as well as its remodellation via crosslinking, proteolysis and other processes. Although this aspect is well known in the skin, it is equally relevant for the rest of the organs. Therefore, another perspective of comprehending fibroblast heterogeneity is understanding its common and different functions in different tissues and organs. Figure 19 describes some of these functions.

Apart from the *canonical* ECM creation and remodelling function, numerous studies now suggest fibroblasts are heterogeneous in their origins, molecular markers, and functions, particularly during the pathological remodelling of organ tissue; and therefore have a wide range of functions extended beyond ECM, like that of immune and angiogenic coordinators, dynamically remodelling a changing microenvironment (McGee et al., 2013).

Fibroblasts secrete numerous cytokines, adipokines, and growth factors whose properties, including diffusion dynamics, are modulated by the ECM and converge to create signalling niches and positional cues for diverse other cells, including, but not limited to, tissue-resident stem cells and immune cells (McGee et al., 2013). In this light, fibroblasts' role in encoding positional information for other cells is particularly important for proper embryonic development, as revealed in classic tissue recombination studies (Dhouailly et al., 1975). One biomechanical means to generate proper ECM niches is by *tug-and-pulling* on the ECM, resulting in tissue-level mechanical forces and matrix polarisation (Huang et al., 2012b).

Quiescent fibroblasts have also been shown to function as progenitors that can be induced to divide rapidly to produce many more ECM-secreting fibroblasts and additional distinct mesenchymal lineages, such as adipocytes in response to injury and hair cycling in the skin, or as bone-forming

osteoblasts or lipid-filled adipocytes during embryonic development, adult homeostasis, and injury, repair, and remodelling (Junker et al., 2009; Rivera-Gonzalez et al., 2016).

2.2.1 *Dermal fibroblasts: origins, types and functions*

In vitro studies in humans (Janson et al., 2012; Sorrell et al., 2004) and *in vivo* studies in mouse skin (Abbasi et al., 2020; Driskell et al., 2013; Jiang et al., 2018a,b; Rinkevich et al., 2015) have shown the existence of different subpopulations of fibroblasts in the dermis, associated to the papillary and reticular dermis, as well as specialised populations in the HF and APM. Most of the analyses regarding the origin of fibroblasts have been performed using mouse lineage tracing experiments. Thus the development of human fibroblasts should be inferred with caution.

Lineage tracing experiments (Driskell et al., 2013) revealed that dermal fibroblast populations, including papillary fibroblasts, reticular fibroblasts, and dermal papillae, are derived from a multipotent mesenchymal progenitor (Pdgfrα^+ Dlk1$^+$ Lrig1$^+$) that is present at embryonic day 12.5 (E12.5), with lineage commitment occurring by E16.5 (Figure 20). In that stage, the fibroblast progenitor is differentiated into a papillary dermal progenitor (Pdgfrα^+ Dlk1$^-$ Blimp1$^+$) and a reticular/hypodermal progenitor (Pdgfrα^+ Dlk1$^+$ Blimp1$^-$). At postnatal day 2, the dermal progenitor is differentiated into the papillary dermal fibroblast (Pdgfrα^+ Dlk1$^-$ Sca1$^-$ Cd26$^+$), which further differentiates into the APM fibroblast (Pdgfrα^+ Cd26$^-$ Itga8$^+$). The reticular/hypodermal fibroblast differentiates into the reticular fibroblast (Pdgfrα^+ Dlk1$^+$ Sca1$^-$), and the adipocyte precursor (Pdgfrα^+ Dlk1$^{+/-}$ Sca1$^+$ Cd24$^+$), which further differentiates into adipocyte (Pdgfrα^+ Dlk1$^-$ Perilipin$^+$). Supporting studies confirmed that Lrig1$^+$ papillary fibroblasts give rise to pericytes in the upper dermis, whereas pericytes in the lower dermis are primarily derived from reticular Dlk1$^+$ fibroblasts (Ganier et al., 2022).

Using similar approaches, distinct mouse embryonic fibroblast lineages were also described within the dorsal dermis (Rinkevich et al., 2015). The Engrailed-1 (En1)-positive (En1$^+$) lineage primarily contributes to connective tissue deposition and organisation during embryonic development, cutaneous wounding, radiation fibrosis, and cancer stroma formation. The En1$^+$ fibroblast lineage highly expresses *Cd26*. This lineage is found to be responsible for the characteristic fibrotic potential of fibroblasts. In fact, ablation of *Cd26*$^+$ fibroblasts diminishes connective tissue deposition in wounds in adult murine skin. In contrast, En1$^-$ fibroblasts enhance dermal development and regeneration.

Of note, these markers are not useful for identifying fibroblasts in adult skin. For instance, the *Cd26* marker that defines the papillary fibroblast lineage is found to be also expressed in reticular fibroblasts (Rinkevich et al., 2015).

CHARACTERISTICS OF PAPILLARY AND RETICULAR FIBROBLASTS Normal dermal fibroblasts consist of at least two distinct cell subpopulations – papillary fibroblasts and reticular fibroblasts. They occupy unique dermis niches and differ in physiology and functional characteristics summarised in Table 1.

Papillary fibroblasts are small, spindle-shaped cells that are more rapidly proliferating and populate the superficial dermis (Wang et al., 2004; Wang et al., 2008). They express high levels of collagen type XVI and DCN but have low expression of VCAN (Sorrell et al., 2004; Sriram et al., 2015). Papillary fibroblasts play a crucial role in immune responses and are required for new hair follicle formation following wounding (Janson et al., 2012).

On the other hand, reticular fibroblasts are large, stellate cells surrounded by a thicker extracellular matrix and occupy the deeper dermis (Wang et al., 2004; Wang et al., 2008). They have low expression of collagen type XVI and DCN but high secretion of VCAN (Sorrell et al., 2004; Sriram et al., 2015). The main function of reticular fibroblasts during wound healing is cytoskeletal organisation and cell motility expressing fibroblast activation markers such as α-SMA (Janson et al., 2012). They are primarily involved in producing ECM during homeostasis and healing and respond to paracrine signals from the epidermis. Despite their relative distance from the epidermis, reticular fibroblasts respond to paracrine signals: the production of epidermal TGF-β2 (Transforming growth factor β2) influences their proliferation, ECM deposition and differentiation (Lichtenberger et al., 2016).

DS AND DP FIBROBLASTS Apart from papillary and reticular fibroblasts, fibroblasts that regulate epithelial activities such as stem cell quiescence, proliferation, and differentiation are also found in skin appendages (Plikus et al., 2021). This includes DP fibroblasts located at the base of hair follicles and dermal sheath (DS) fibroblasts surrounding the hair follicle. These fibroblasts develop from embryonic neural crest (Sriram et al., 2015), more specifically from papillary precursors through an intermediary dermal condensate ($Tbx18^{+}$) progenitor (Mok et al., 2019; Sennett et al., 2015) and express specific marker genes upon full formation (Driskell et al., 2011; Rezza et al., 2016). Several markers have been described for human and mouse DP and DS fibroblasts, including CD133–in human and mouse–and nonspecific alkaline phosphatase–in human–(Anan et al., 2003; Ito et al., 2007; Johnson et al., 1945). Marker expression changes according to the stage of the hair cycle and, in the mouse, it also depends on the hair follicle type (Driskell et al., 2009, 2012; Rendl et al., 2005).

DP fibroblasts play a central role in hair development and the coordination of the hair cycle (Driskell et al., 2011; Sennett et al., 2012). Specifically, they regulate bulb size, hair length and thickness, and the duration of the hair growth phase (anagen) (Sriram et al., 2015). Additionally, they have several distinctive features (Plikus et al., 2021). First, they are highly aggregative, including *in vitro*, where they form mound-like colonies before reaching confluence. Second, they associate with and function as the signalling niche for epithelial hair stem cells. Third, their gene expression changes prominently and periodically in synchrony with the hair growth cycle. Interestingly, there is evidence that, in certain circumstances, DP fibroblasts can differentiate to other cell types, including nerve and cartilage (Toma et al., 2001).

DS fibroblasts surround the entire hair follicle, forming the CTS. Their function has been associated with interfollicular skin regeneration, hair follicle regeneration, and hair cycle-associated angiogenesis (Higgins et al., 2017; Yoshida et al., 2019). Gene expression by DS fibroblasts is distinct and prominently includes contractile proteins paramount for hair growth termination, as smooth muscle-like contraction by the DS helps the shrinking hair follicle remodel properly (Heitman et al., 2020). Moreover, DS fibroblasts interface directly with skin resident immune cells and produce immune-suppressive factors such as the TGF ligands b1 and b2 and immunomodulatory molecules such as programmed death-ligand 1 (PDL1) and CD200, which may contribute to the immune privilege of hair follicles, a property that guards them against autoimmune reactions (Paus et al., 2003).

2.2.2 *Main cell signalling pathways related to dermal fibroblast functioning*

2.2.2.1 *Wnt*

Many signalling pathways have been implicated in skin fibrosis and fibroblast homeostasis, and the Wnt pathway is one of them. This complex pathway has been studied for over 30 years (Nusse et al., 1982). Wnt signalling activates many intracellular transduction pathways to control cell fate. The main two pathways activated by Wnt are the canonical Wnt pathway (cWnt) or Wnt/β-catenin-dependent pathway and the noncanonical (ncWnt) or β-catenin independent pathway (Figure 21). Two of the most common ncWnt pathways are planar cell polarity (PCP) and the Wnt/Ca^{2+} pathways, reviewed in (Griffin et al., 2021; Niehrs, 2012).

Wnts are the secreted glycoproteins that trigger the specific cascades upon binding to their respective receptors. Wnt receptors, or frizzled (Fz), are a family of transmembrane receptors resembling GPCRs. These receptors act jointly with different co-receptors, such as lipoprotein-related proteins (LRP), specifically LRP5/6. There is no clear Wnt-Fz relationship or relationship between Wnt ligands and activated pathways. For instance, it is more likely that Wnt1, Wnt3a and Wnt8 are linked to canonical Wnt signalling, whereas Wnt5a and Wnt11 are predominantly involved in non-canonical Wnt signalling (Kikuchi et al., 2011).

CANONICAL WNT PATHWAY This pathway is based on the activation and translocation of β-catenin into the nucleus, targeting an array of genes. Without Wnt activation, β-catenin is degraded by the β-catenin destruction complex, composed of Axin, APC, PP2A, GSK3, and CK1α. When Wnt binds to its receptor, Axin and DVL are activated, which inhibits GSK3 and allows the translocation of β-catenin to the nucleus. In the nucleus, β-catenin binds to T-cell factor/lymphoid enhancer factor (TCF/LEF) to regulate the expression of Wnt target genes (e.g., *Wnt2b*, *Wnt3*, *Wnt1*).

It has been observed that DVL can form oligomers, which may cluster Wnt-FZR complexes and promote endocytosis of downstream signalling. This endocytosis process gathers more recruiting complexes than at the membrane without endocytosis. As a result, larger amounts of β-catenin are free in the cytoplasm to be translocated.

NON-CANONICAL WNT PATHWAYS The ncWnt signalling comprises numerous pathways, the most studied being the PCP pathway and the Wnt/Ca^{2+} pathway.

In the PCP pathway, Wnt activates DVL, which binds to and activates DAAM or RAC1. DAAM activation leads to the activation of ROCK, which induces cytoskeleton rearrangement. This pathway is prominently involved in regulating cell polarity and morphogenic processes. It also uses ROR and PTK7 as co-receptors instead of LRP5/6. On the other hand, the Wnt/Ca^{2+} pathway also leads to DVL activation, which binds to and activates PLC signalling, increasing intracellular Ca^{2+} concentration. One interesting outcome of this pathway is that CAMKII is activated, which inhibits β-catenin activity in the nucleus. This pathway is more involved in cancer, inflammation and neurodegeneration.

REGULATION OF WNT SIGNALLING We have observed that different Wnt pathways compete with each other, like Wnt/Ca^{2+} pathway *antagonising* with Wnt/β-catenin pathway. It is also widely observed that Wnt/PCP pathway and Wnt/β-catenin pathways are also virtually exclusive, and inhibiting one usually upregulates the other. This antagonistic crosstalk occurs, in part, at the lig-

and–receptor level. For example, WNT5A, which preferentially activates PCP signalling, competes for and inhibits the binding of WNT3A to Frizzled 2, thereby suppressing the β-catenin-dependent pathway (Sato et al., 2009).

Wnt signalling is also controlled by several upstream endogenous inhibitors, such as secreted frizzled-related proteins (SFRPs), Dickkopf (DKK) proteins, and the Wnt inhibitory protein (WIF). SFRPs and WIF bind to Wnt proteins, thereby preventing their interaction with their receptor. On the other hand, DKK proteins are thought to bind to LRP co-receptors, thereby disturbing the formation of the Fz receptor complex.

There is a secondary activatory regulation of Wnt signalling by a family of proteins called R-spondins (RSPOs). RSPOs are necessary for β-catenin signalling to occur and promote PCP signalling (Kazanskaya et al., 2004; Kim et al., 2008; Ohkawara et al., 2011). Under no RSPO presence, the Zinc and RING finger 3 surface protein (ZNRF3) ubiquitylates FZR, targeting it for degradation and further internalising other receptors. RSPOs bind to their receptors, such as LGRs, which capture ZNRF3 to make a complex. With no free ZNRF3, FZRs accumulate on the plasma membrane again.

EFFECTS OF WNT SIGNALLING IN FIBROBLASTS Wnt pathway has been shown to regulate the fate specification of fibroblast progenitors into various lineages, notably in the skin. One of the main targets of this pathway is HF formation. It was observed that Wnt signalling is activated in the papillary fibroblast progenitors of the upper skin layer and then becomes progressively restricted to the so-called dermal condensate cells of developing hair follicles (DasGupta et al., 1999; Zhang et al., 2019c). In adults, overexpression of β-catenin in basal epidermal leads to *de novo* hair follicle morphogenesis in adult skin.

In the context of skin wounding, transient Wnt activity in myofibroblasts promotes regenerative healing with new hair follicles (Lim et al., 2018), whereas its chronic activity drives fibrotic response and failure to regenerate hairs (Gay et al., 2020). Rognoni et al., 2016 observes contrary results, showing that Wnt/β-catenin expression in the adult dermis inhibits *de novo* hair follicle formation in healing wounds, and postnatal ablation of β-catenin in dermal fibroblasts enhances new hair formation.

It has also been observed that Wnt signalling has an inhibitory effect on adipocyte differentiation of mesenchymal cells *in vitro* (Plikus et al., 2021). *In vivo* studies show that Wnt/β-catenin signalling in the epidermis impairs adipogenesis (Donati et al., 2014) and, in dermal fibroblasts, it leads to fibrosis and impairs adipogenesis too (Mastrogiannaki et al., 2016).

2.2.2.2 *TGF-β/BMP*

TGF-β is a growth factor that plays a crucial role in fibrotic processes and belongs to a family of related growth factors. It is a protein secreted and usually bound to ECM components but can be activated by cleavage from proteases such as plasmin, or MMP2 and MMP9. The active form of TGF-β is a homodimer, while the latency-associated peptide (LAP) and the latent TGF-β binding protein (LTBP) form the inactive form by sequestering the homodimer. After cleavage, TGF-β binds to TGF-β receptor 2 (TGFR2) and activates it, which activates receptor signalling by a combination of both isomers–TGFR1 and TGFR2– (Figure 22a) (Meng et al., 2016).

The activated TGF-β receptor 1 phosphorylates Smad2 and Smad3, which form a complex with Smad4 and move to the nucleus. Smad3 in the complex directly binds to gene promoters, triggering the

transcription of profibrotic molecules such as α-SMA, connective tissue growth factor (CTGF/CCN2), collagen I, and TIMP, leading to myofibroblast activation and matrix deposition. Smad3 also induces the transcription of profibrotic microRNA (miRNA) and long noncoding RNA (lncRNA), indirectly inhibiting antifibrotic miRNAs. Furthermore, Smad3 increases the transcription of profibrotic molecules by impacting epigenetic modifications of DNA and histones (Meng et al., 2016).

Smad2 and Smad4 do not have DNA binding domains, but rather serve as regulators of Smad3-driven gene transcription. Different genes are influenced by the transcription of Smad3-containing complexes, as they are bound by non-Smad co-activators such as histone acetyltransferase p300, AP-1, Sp-1, and CREB binding protein, or co-suppressors such as Ski proto-oncogene, SnoN, and Smad nuclear-interacting protein-1. Other members of the Smad family act as negative regulators of Smad2/3/4, such as Smad7; inhibiting fibrosis (Meng et al., 2016).

A secondary pathway involves the activity of the bone morphogenic protein 7 (BMP7), which upon binding to different receptors–BMPRII, and Alk2/3/6–phosphorylates Smad 1 and Smad5 proteins, which bind Smad4 and the complex migrates to the nucleus (Villapol et al., 2013). This pathway can be inhibited by Smad7 too, as well as Smad6 (Villapol et al., 2013). Interestingly, Smad1-5 may indirectly inhibit the "classical" pathway to avoid redundant TGF-β signalling (Meng et al., 2016), in a similar fashion to the Wnt pathway.

Many signalling pathways regulate TGF-β/Smad signalling and vice versa, resulting in a complex set of interactions known as pathway crosstalk (Figure 22b), often referred as the *non-canonical* TGF-β pathways (Meng et al., 2016):

- Mitogen-activated protein kinases (MAPKs): MAPKs interact with TGF-β/Smad signalling at several levels, and these interactions are the most clearly established example of pathway crosstalk with TGF-β/Smad signalling. TGF-β1 can directly induce ERK activation in a Smad-independent manner. In addition, TGFR1 can induce the activation of p38 and JNK via a pathway involving TNF receptor-associated factor 6 and TGF-β activated kinase-1/M3K7.
- Wnt/β-catenin: Crosstalk between the Wnt/β-catenin and TGF-β/Smad pathways is poorly understood. On one hand, Wnt3a may induce the production of TGF-β and collagen I in fibroblasts. On the other hand, TGF-β may activate Wnt signalling, for example, by inhibiting GSK3, or by suppressing the production of DKK1 in a p38-dependent manner.
- BMP-7: BMP-7, a member of the TGF-β superfamily, is a well-characterised antifibrotic factor that can counteract the effects of TGF-β. BMP-7 binds to several receptors–activin A receptor type I and BMP receptors type 1A and 1B–to activate Smad1 and Smad5 and thereby inhibit Smad3-dependent gene transcription as well as Smad-independent pathways.

TGF-β is a canonical marker of myofibroblast activation and is widely observed in fibrotic processes. This aspect is discussed further in Section 2.2.4.

2.2.3 *Fibroblast interaction with other cells*

Fibroblasts play a critical role in regulating tissue homeostasis and wound healing by interacting with various other cell types in the body, including immune cells, epithelial cells, and smooth muscle cells, to maintain tissue integrity and respond to injury or disease. In this section, we will study the major

interactions between fibroblasts and other cell types, some of which have been extensively reviewed by Van Linthout et al., 2014 and Stunova et al., 2018.

2.2.3.1 *Immune system*

Fibroblasts in fibrotic or malignant lesions play an important role in modulating immune reactions by exerting their regulatory functions through their inflammatory secretome. This secretome includes chemokines, interleukins, and growth factors that influence the activation of adjacent immune microenvironments (LeBleu et al., 2020; Van Linthout et al., 2014).

Fibroblasts control the recruitment of immune cells and regulate their behaviour, retention, and survival in damaged tissue. This crosstalk between fibroblasts and leucocytes sometimes depends on the interaction between the surface antigen CD40 on fibroblasts and its ligand, CD40L, on immune cells (Yellin et al., 1995). These interactions are similar to those between lymphocytes and antigen-presenting cells–CD40-CD40L between APCs and T cells, for instance–, indicating fibroblasts' crucial role in the crosstalk (Van Linthout et al., 2014).

As per some general observations, fibroblasts from fibrotic lesions express higher levels of the MCP-1 receptor, CCR2, compared to those from non-fibrotic lesions (Hogaboam et al., 1998). MCP-1 itself stimulates collagen expression and endogenous up-regulation of TGF-β expression in fibroblasts, leading to autocrine or juxtacrine stimulation of collagen gene expression (Gharaee-Kermani et al., 1996). An example of this regulation is in periostin-expressing cardiac fibroblasts, which, when genetically modified to lose IL-17 signalling, reduce the production of the pro-fibrotic growth factor GM-CSF, effectively limiting cardiac inflammation and tissue death (Chen et al., 2018a).

T CELLS T cells are indispensable, proved by the finding that T-cell-deficient mice develop in a "permanent" fibrosis state (McLarren et al., 2011). Different studies have shown that fibroblasts interact with a range of T cells, including Th2 and Th17.

Generally, the adhesion of T-lymphocytes to fibroblasts appears to be mediated by IL-1R, which is associated with the cell membrane on both T-lymphocytes and fibroblasts (Qwarnström et al., 1991). Consequently, dermal fibroblasts produce increased IL-6 and IL-8, and T lymphocytes are activated to secrete soluble IL-1β and IL-6 (Spörri et al., 1996). A general intermediary observed to mediate T-cell interaction is TNFα. TNF-α-activated dermal fibroblasts stimulate T-lymphocyte proliferation, and that growth-promoting activity requires cell–cell interactions between IL-15 on the wound fibroblast surface and the IL-15 receptor (IL-15RA) on T-lymphocytes (Bodnár et al., 2008). Physical interactions between activated T-lymphocytes and dermal fibroblasts reduce the production of type I and III collagen by fibroblasts (Rezzonico et al., 1998).

Regarding Th2 cells, characterised by the secretion of the cytokines IL-4, IL-5, and IL-13, with wound healing and fibrosis, IL-13 has been observed to be a mediator of fibrosis, and in combination with IL-4 is capable of inducing the phenotypic transition of human fibroblasts to myofibroblasts. In addition, IFNα and IFNγ have been shown to inhibit fibroblast collagen production by inhibiting IL-4- and IL-13-promoted differentiation into myofibroblasts (Spörri et al., 1996).

Lastly, IL17A-producing Th17 cells have been suggested to inhibit Type I collagen production in dermal fibroblasts of healthy and systemic sclerosis patients (Hatamochi et al., 1984). Additionally, they induce the expression of a variety of pro-inflammatory mediators, such as IL-1, IL-6, TNF-a, CXCL8,

granulocyte colony-stimulating factor, and granulocyte–macrophage colony-stimulating factor by endothelial and epithelial as well as fibroblasts, which ultimately results in the recruitment and activation of neutrophils (Saalbach et al., 2000).

APCS Fibroblasts may also play a role in immune modulation by acting as *amateur* antigen-presenting cells, possibly devoid of costimulatory signals (Sprent, 1995). Additionally, APCs such as macrophages interact directly within dermal fibroblasts and other immune types to elicit an adaptive response. Macrophages stimulated by TLR ligands and IFNγ undergo classical M1 activation, while those stimulated by IL-4 and IL-13 become M2 macrophages. M2 macrophages are involved in wound healing, tissue remodelling, fibrosis, and inflammatory responses (Martinez et al., 2009; Martinez, 2008).

It has been shown that M1 macrophages induce the release of fibroblast pro-inflammatory cytokines and chemokines (CCL2, IL-6 and CCL7), MMPs (MMP-1, MMP-2, MMP-3 and MMP-14) and TIMP-1 (Ploeger et al., 2013). On the other hand, M2 macrophages secrete PDGF-CC, which accelerates the differentiation of wound fibroblasts into myofibroblasts (Glim et al., 2013), or arginase 1, capable of controlling L-proline production essential for the collagen synthesis of activated myofibroblasts. Additionally, the stimulation of dermal fibroblasts with M2 macrophage paracrine factors leads to higher fibroblast proliferation than stimulation via M1 macrophage soluble factors (Hesse et al., 2001).

Looking into the macrophage activation by fibroblasts, macrophage colony-stimulating factor (M-CSF), secreted by keratinocytes and fibroblasts, is thought to support the growth of dermal dendritic and Langerhans cells (Takashima et al., 1995).

2.2.3.2 *Keratinocytes*

Keratinocytes are dynamic cells within the skin and are actively bound to the dermis via the DEJ. Therefore, cell-cell communication between keratinocytes–mainly from the basal layer–and fibroblasts is active and bidirectional.

Fibroblasts influence keratinocytes via paracrine and juxtacrine signalling. For instance, fibroblast-derived soluble factors, including KGF, hepatocyte growth factor (HGF), granulocyte-macrophage colony-stimulating factor (GM-CSF), IL-6, IL-19 and FGF-10, diffuse into the epidermis and influence both keratinocyte growth, differentiation and synthesis of the basement membrane (Gron et al., 2002; Marchese et al., 2001; Smola et al., 1998; Sorrell et al., 2004; Sun et al., 2013). Other factors, like the fibroblast-secreted heparin-binding EGF-like growth factor (HB-EGF), upon binding to its receptor, stimulate keratinocyte migration and proliferation in direct contact. Lastly, it has already been mentioned that the origin of dermal fibroblasts is important. For instance, keratin 9, responsible for the thickening of the epidermis, is not expressed by trunk-derived keratinocytes; however, the co-culturing of trunk-derived keratinocytes with fibroblasts derived from palms and soles have been found to induce the expression of keratin 9 (Yamaguchi et al., 2005).

Reversely, there is a large amount of evidence supporting keratinocytes influencing fibroblast biology. For instance, IL-1α, secreted by keratinocytes, is thought to act as an inducer of the production of IL-6 and IL-8 in dermal fibroblasts (Boxman et al., 1996). Additionally, keratinocyte-derived stratifin (also known as 14-3-3σ protein) promotes the degradation of ECMs by dermal fibroblasts through the secretion of MMPs – primarily MMP-1 and MMP-3 (Ghaffari et al., 2006). In a related study, the activities of both MMP-2 and MMP-9, produced by fibroblasts and keratinocytes, are significantly

enhanced in co-culture systems. This effect replicates with the fibroblast-produced expression of TIMP-1 and TIMP-2 (Stunova et al., 2018).

2.2.3.3 *Endothelial cells*

Endothelial cells are necessary for a healthy dermis due to their function in nutrient transport. Therefore, it is expected that vascular endothelial cells, as well as perivascular cells, will be highly reliant on ECM and fibroblast-derived signalling. Angiogenesis–the development of new blood vessels from existing ones–is highly dependent on fibroblast-derived factors. In fact, endothelial cell adhesion to the ECM through integrins is essential for proliferation, which requires the ECM-remodelling capacity of fibroblasts. It is thought that the ECM acts as a scaffold for forming capillaries and a "trap" that retains and concentrates growth factors in the micro-environment surrounding cells. MMPs modify the ECM via degradation, which supports endothelial cell migration, vessel sprouting and the cleavage of ECM proteins, which leads to the formation of anti-angiogenic fragments (Berthod et al., 2006); for example, MMP-9 cleaves collagen type IV, one of the components that make up the endothelial basement membrane, into the anti-angiogenic peptide tumstatin. Therefore fibroblast-produced TIMP1, which inhibits MMP9, indirectly leads to a pro-angiogenic environment (Belotti et al., 2011; Liu et al., 2008).

Other classical factors produced by fibroblasts, like TGF-β, lead to the expression of VEGF under non-physiological conditions, leading to vessel formation (Berthod et al., 2006). VEGF stimulates endothelial cell proliferation, migration and NO release by activating the VEGF receptors Flt1 (VEGF-R1) and KDR (VEGF-R2/Flk-1). Other factors secreted by fibroblasts that are tightly related to vessel formation are HGF, FGF-2, angiopoietin-1 (Ang-1), angiopoietin-2 (Ang-2) and IL-8 (Kroll et al., 2000; Martin et al., 2001; Orlandini et al., 2001; Staton et al., 2010). Ang-1 acts via the PI3K/Akt signalling pathway and while Ang-1 acts as an agonist for Tie-2, Ang-2 constitutes a Tie-2 antagonist (Kim et al., 2000). Following binding, Ang-1 induces endothelial migration and survival and tube formation in vitro and, moreover, promotes blood vessel structural integrity in vivo (Saito et al., 2003; Staton et al., 2010).

2.2.4 *Myofibroblast*

In adult organs, fibroblasts are relatively quiescent unless tissue repair mechanisms or dynamic structural changes are initiated. Gabbiani et al., 1971 reported that fibroblasts respond to tissue injury by assuming a contractile phenotype found in granulation and fibrotic tissue. Later on, Majno et al., 1971 coined the term *myofibroblast* to these fibroblasts "…that, under certain conditions, are capable of modulating toward a cell type that is structurally and functionally close to smooth muscle; for these cells, the name 'myofibroblast' may be appropriate". Much research has been done since, and myofibroblasts are currently well established as the cell type associated with fibrosis in several tissues and organs, as well as participating in the wound healing process without leading to fibrosis with an appropriate environment. Extensive reviews on myofibroblasts can be found at Schuster et al., 2022; Shaw et al., 2020; Tai et al., 2021.

Despite carrying "fibroblast" in their name, myofibroblasts can be generated from different cell sources, including different fibroblast types (Chang et al., 2002; Correa-Gallegos et al., 2021; Jiang et al., 2018a,b; Rognoni et al., 2018), adipocytes (Festa et al., 2011; Marangoni et al., 2015; Shook et al.,

2018), pericytes, MSC (Jiang et al., 2020; Soliman et al., 2021; Yokota et al., 2021); and other more *exotic* sources such as vascular endothelial cells (Zhao et al., 2021) and circulating $CD45^+$ cells–e.g. fibrocytes, macrophages–(Chong et al., 2019; Sinha et al., 2018). This section will focus on the fibroblast-to-myofibroblast transition, although other transitions share common characteristics (Schuster et al., 2022).

Fibroblast-to-myofibroblast differentiation typically occurs in injury environments, such as wound healing in the skin (Figure 23). In this process, TGF-β, Wnt, PDGF, TNF-α, IL-1, or IL-6–i.e. canonical inflammatory markers–are present and signal quiescent fibroblasts to transform intro myofibroblasts, which are specialised to produce collagen faster than quiescent fibroblasts (Tsukui et al., 2020). Myofibroblast precursors with fibroblast characteristics typically express collagen I, VIM, PDGFRα (Iwayama et al., 2015; Li et al., 2018a), β1 integrin (CD29) Hic1 (Abbasi et al., 2020), Thy-1–cluster of differentiation, CD90–(Jiang et al., 2018a; Worthen et al., 2020), and FSP1 (Li et al., 2020b).

After this signalling, fibroblasts undergo an intermediary proto-myofibroblast stage in which they synthesise and deposit ECM components such as collagen I and III, which replace the provisional matrix, to reduce the extent of the injury and create the granulation tissue (Darby et al., 2007). At this stage, proto-myofibroblasts are still subjected to TGF-β1 transformation and lack α-SMA expression (Gabbiani et al., 1971; Hinz et al., 2007). Upon further activation, proto-myofibroblast transform into α-SMA^+ myofibroblasts, which have a higher contractile force and adhere to the ECM to rearrange it (Schuster et al., 2022). When the damage is healed, myofibroblasts disentangle their focal adhesions from the ECM and revert to quiescence (Thannickal, 2013). Sometimes, they also prune by undergoing apoptosis or senescence (Demaria et al., 2014; Wilkinson et al., 2019).

Depending on tissues and conditions, activated myofibroblasts also express high levels of certain cell membrane proteins, PDGFRβ) (Henderson et al., 2013), α11β1 integrin–binds fibrillar collagen–(Zeltz et al., 2020), αv integrins (Conroy et al., 2016), and cadherin-11 (Lodyga et al., 2019; To et al., 2019). LRRC15 is expressed by CAFs activated into a contractile phenotype ("myCAF") (Purcell et al., 2018).

It has also been observed that myofibroblasts can revert to the original cell type under the proper conditions (Hinz et al., 2012), or display broader lineage plasticity and convert into other specialised mesenchymal lineages upon injury resolution. For example, myofibroblasts in large murine skin wounds terminate their contractile behaviour and reprogram into new lipid-filled adipocytes in response to BMP ligands secreted by hair follicles (Plikus et al., 2017).

Although the transition to the myofibroblast state is essential for tissue repair after injury, its aberrant and sustained switch critically drives fibrosis and contributes to cancer progression (Plikus et al., 2021). The excessively stiff and compositionally abnormal ECM disrupts the microarchitecture and results in the loss of other tissue-resident cells, causing organ dysfunction. Moreover, myofibroblasts are usually absent in normal interstitial spaces but increase in number in fibrotic or cancerous lesions; thus, they are often thought of as pathology-associated fibroblasts (LeBleu et al., 2020). Fibrosis is estimated to contribute to almost 50% of all deaths in the developed world (Friedman et al., 2013), and despite the existence of drugs that can delay its progression, to date, there is no truly effective treatment (Dempsey et al., 2019).

TGF-β IN MYOFIBROBLASTS The TGF-β pathway dominantly controls activation of the myofibroblast state, originating from the latent TGF-β in the LTBP1 form bound to ECM components. This acti-

vation by TGF-β leads to the transcription of CCN2 and ACTA2; the former binds to ECM components such as HSPG, and the latter constitutes the α-SMA fibres that strengthen the cell. A summary of TGF-β pathway in myofibroblasts is available in Figure 23.

Different elements of the TGF-β pathway contribute to the myofibroblast activation or differentiation in different tissues and are a clear indicator of fibrosis after prolonged exposure.

For instance, it has been reviewed that human fibrotic tissue displays an elevated expression of TGF-β ligands in the lungs, skin, and skeletal muscle (Lodyga et al., 2020). Another TGF-b superfamily member, Activin A, which, like TGF-β ligands, activates SMAD2/3 (Pangas et al., 2000), is upregulated in human scars and other fibrotic diseases. Other members, like BMP, appear during skeletal muscle regeneration, where BMP signalling converts fibroblasts to myofibroblasts (Lima et al., 2020), and during wound healing in the skin, where it reprograms myofibroblasts to adipocytes (Plikus et al., 2017).

2.3 TABLES AND FIGURES

Table 1: **Characteristics of papillary and reticular fibroblasts.** More characteristics are available at (Stunova et al., 2018)

	Papillary fb	**Reticular fb**
Size	small	large
Shape	spindle	stelated
Proliferaion rate	high	low
Focal adhesions	rare	abundant
Angiogenesis	promote	no effect
Collagen I	low	high
Collagen VI	high	low
Versican	low	high
Decorin	high	low
α-SMA	low	high
TGF-β	low	high
PDPN	high	-
TGM2	-	high

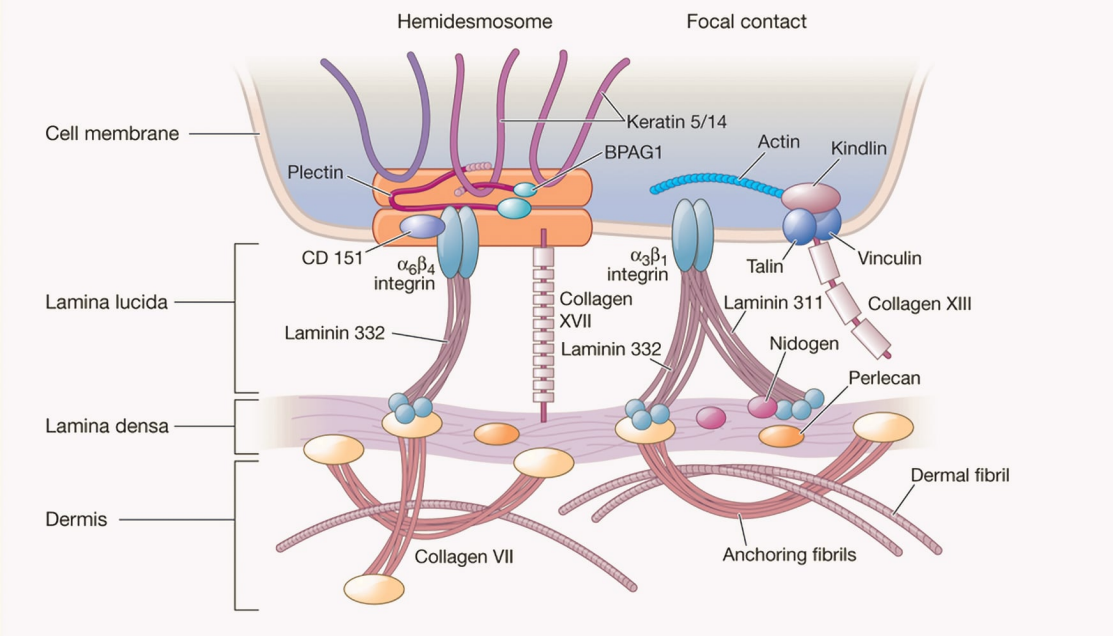


Figure 14: **Schematic diagram of the dermal basement membrane.** The main components located at the DEJ cell membrane, dermis, and lamina lucida and densa are depicted. Adapted from (Goldsmith et al., 2012), Figure 53-9.

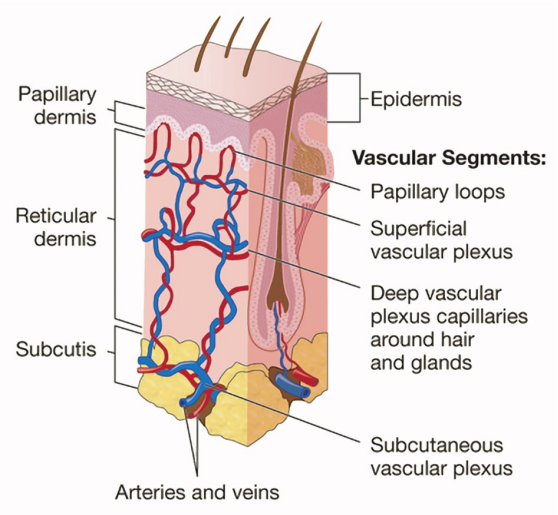


Figure 15: **Schematic diagram of the dermal microvascular unit.** Adapted from Goldsmith et al., 2012, Figure 6-1.

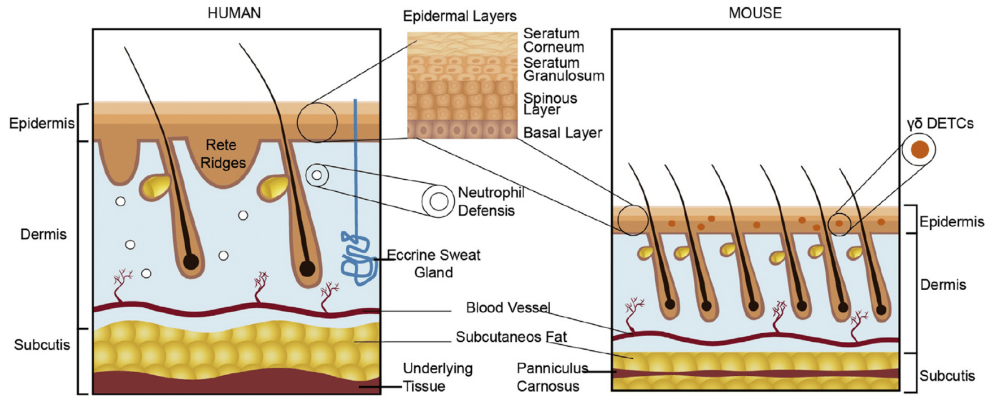


Figure 17: **Differences between mouse and human skin morphology.** Human skin is thicker, contains more epidermal layers than murine skin, and adheres to the underlying tissues. Rete ridges, eccrine sweat glands and neutrophil defensins are present in humans but absent from mice skin. On the other hand, murine skin is richer in hair follicles, presents $\gamma\delta$ dendritic epidermal T cells (DETCs) and the *panniculus carnosus*, a muscle layer with important contraction potential. Adapted from Zomer et al., 2018, Figure 2.

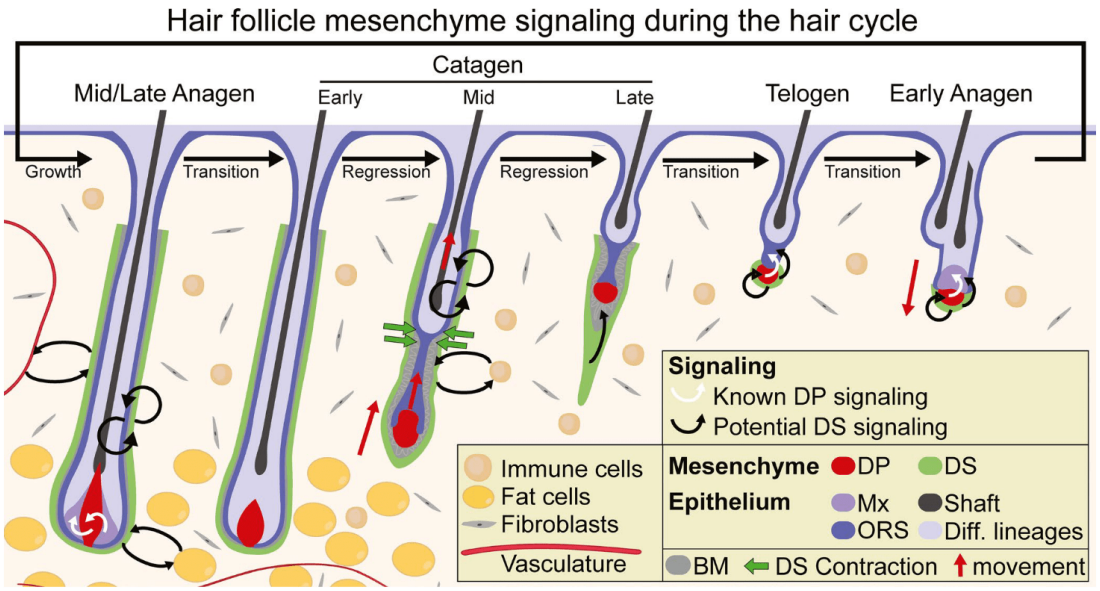


Figure 16: **Schematic diagram of the HF cycle.** During the growth phase (anagen), the DP signals to epithelial progenitors to instruct their proliferation and upward differentiation. During the regression phase (catagen), dermal sheath contraction at the interface of the club hair and epithelial strand (the "bottleneck") powers follicle regression and upward movement of the DP. DS contraction facilitates the relocation of the DP to its stem cell-adjacent position by the resting phase (telogen). During follicle regeneration at the onset of the next growth phase, the DP provides activating signals to the SCs to trigger their proliferation. Adapted from Martino et al., 2020, Figure 3.

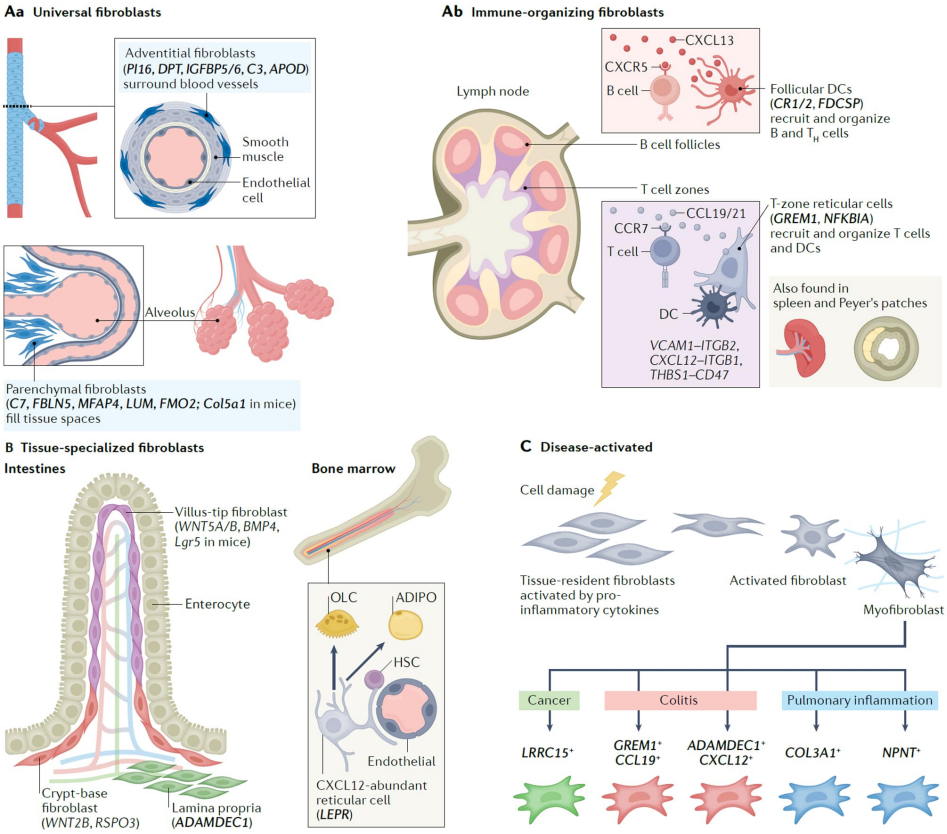


Figure 18: **Universal and specialised fibroblasts found across human organs.** (A) The first class of fibroblasts is ubiquitous or pan-tissue fibroblasts, including adventitial fibroblasts, that form a niche for blood vessels; parenchymal fibroblasts that fill tissue spaces such as those found in the lung (part Aa); and immune- organising fibroblasts found across lymphoid organs (part Ab). (B) Tissue- specialised fibroblasts occupy specialised niches, for example, compartmentalised intestinal crypt–villus fibroblasts and CXCL12-abundant reticular cells in the bone marrow. (C) Disease-activated fibroblasts are found in chronically inflamed or cancer-invaded tissues. These activated fibroblast states are transcriptionally distinct from fibroblasts found in healthy tissues. Markers described using single-cell approaches are highlighted in bold. ADIPO, adipocyte; DC, dendritic cell; HSC, haematopoietic stem cell; OLC, osteoclast; T_H cell, T helper cell. Adapted from Elmentaite et al., 2022, Figure 3.

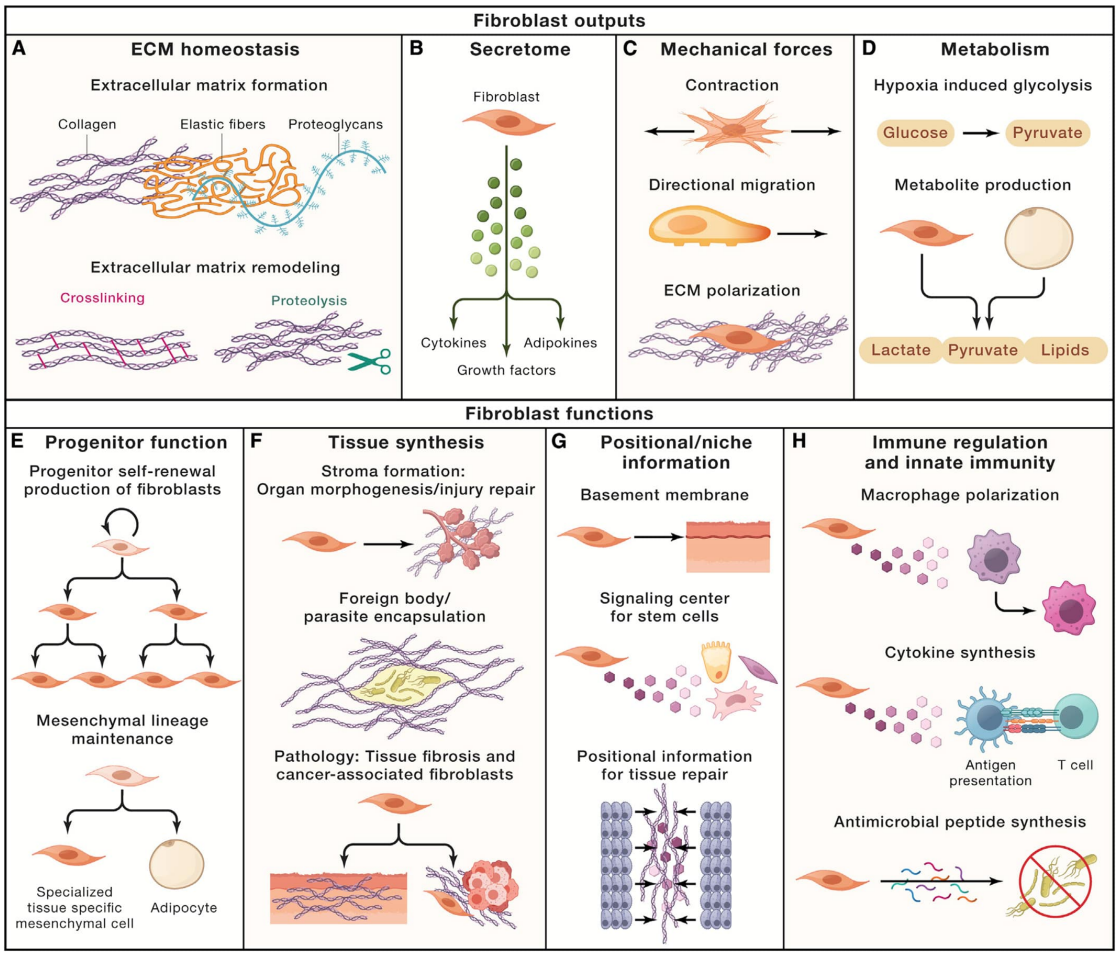


Figure 19: **Summary of fibroblast outputs and functions.** (A–D) Key outputs of fibroblasts and their mesenchymal lineages include extracellular matrix (ECM) secretion and remodelling (A), secretion of signalling factors for surrounding cells (B), mechanical force generation (C), and regulation of tissue metabolism and metabolite secretion (D). (E–H) Fibroblasts also function as progenitor cells for mesenchymal lineages (E), as "makers" of new tissue during organ morphogenesis and tissue repair and upon various pathological conditions (F), as sources of positional information across distinct anatomical regions of the same organ and as key signal contributors toward stem cell niches (G), as well as target cells and reciprocal modulators of diverse innate and adaptive immune functions (H). Adapted from Plikus et al., 2021, Figure 1.

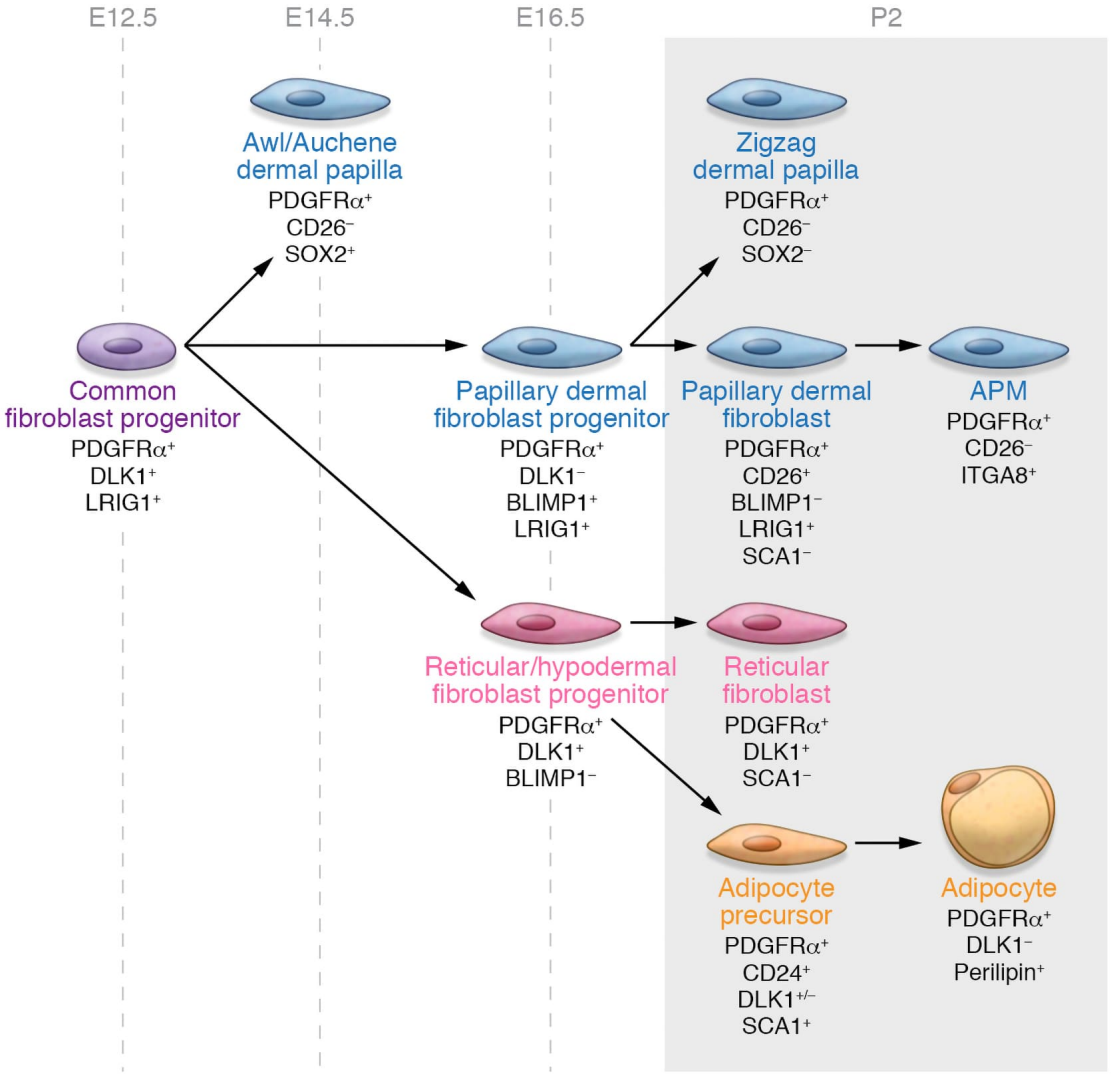


Figure 20: **Mouse dermal fibroblast lineages.** Dermal fibroblasts derive from common fibroblast progenitor cells and differentiate into specific lineages by postnatal day 2 (P2). These subtypes display distinct functions. Adapted from Lynch et al., 2018, Figure 3.

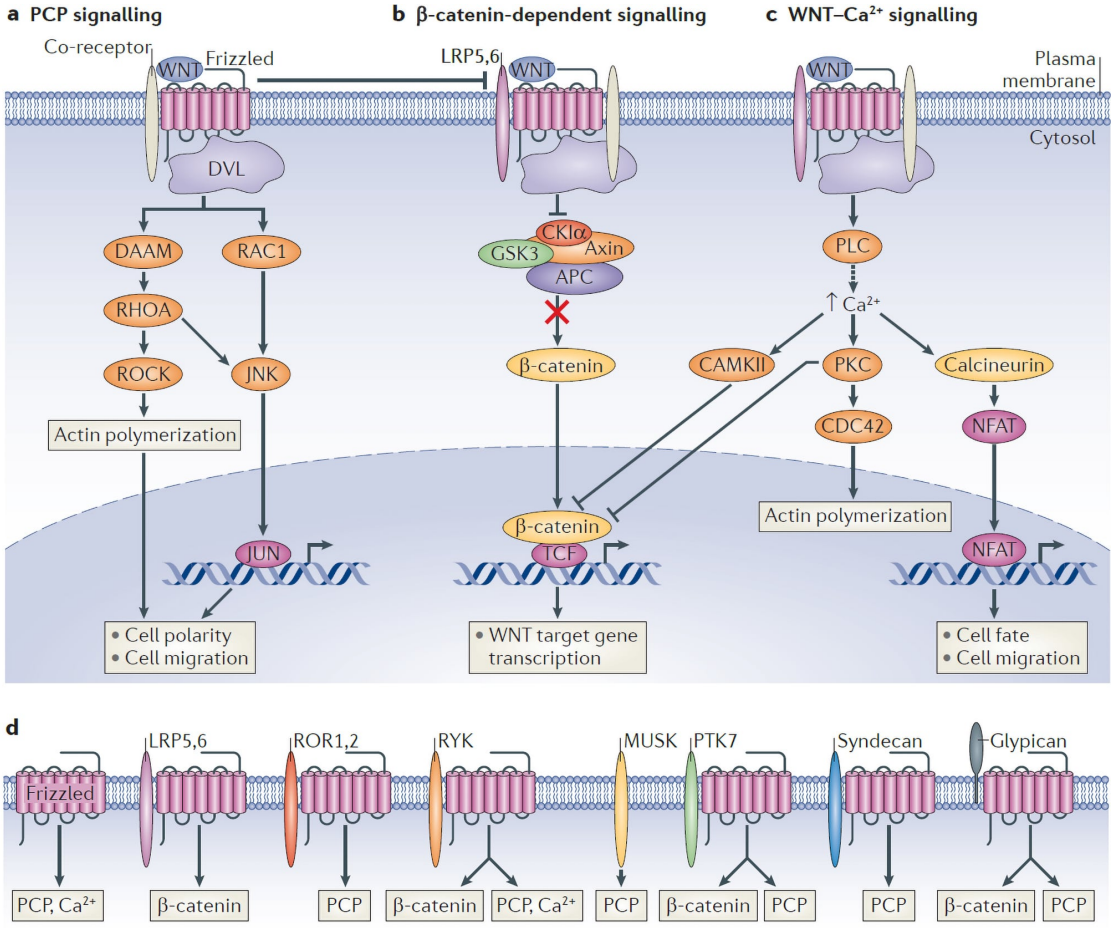


Figure 21: **Wnt signalling pathways.** Simplified schemes showing the main Wnt pathways directed by specific Wnt, Frizzled and Wnt co-receptor interactions. (a) Planar cell polarity (PCP) signalling triggers activation of the small GTPases RHOA and RAC1, which in turn activate RHO kinase (ROCK) and JUN-N-terminal kinase (JNK), respectively, leading to actin polymerization and microtubule stabilization. This pathway prominently regulates cell polarity, motility and morphogenetic movements. (b) Under steady-state conditions, glycogen synthase kinase 3 (GSK3) phosphorylates β-catenin, which triggers its degradation. However, in the presence of Wnt ligand, the destruction complex (comprising GSK3, casein kinase Iα (CKIα), Axin and adenomatosis polyposis coli (APC)) is recruited to the Wnt–receptor complex and inactivated. This allows β-catenin to accumulate and translocate to the nucleus, where it activates the transcription of target genes under the control of T cell factor (TCF), among others. (c) The Wnt–Ca2+ pathway activates Ca2+- and calmodulin-dependent kinase II (CAMKII), protein kinase C (PKC) and calcineurin. Calcineurin activates the nuclear factor of activated T cells (NFAT), which regulates the transcription of genes controlling cell fate and cell migration. PCP and Ca2+ pathways antagonise β-catenin signalling at various levels. (d) Major pathways used by Wnt receptors and co-receptors. Only the three best-characterized Wnt pathways are shown: the β-catenin-dependent pathway (β-catenin); the PCP pathway; and the Wnt–Ca2+ pathway (Ca2+). Adapted from Niehrs, 2012, Figure 3.

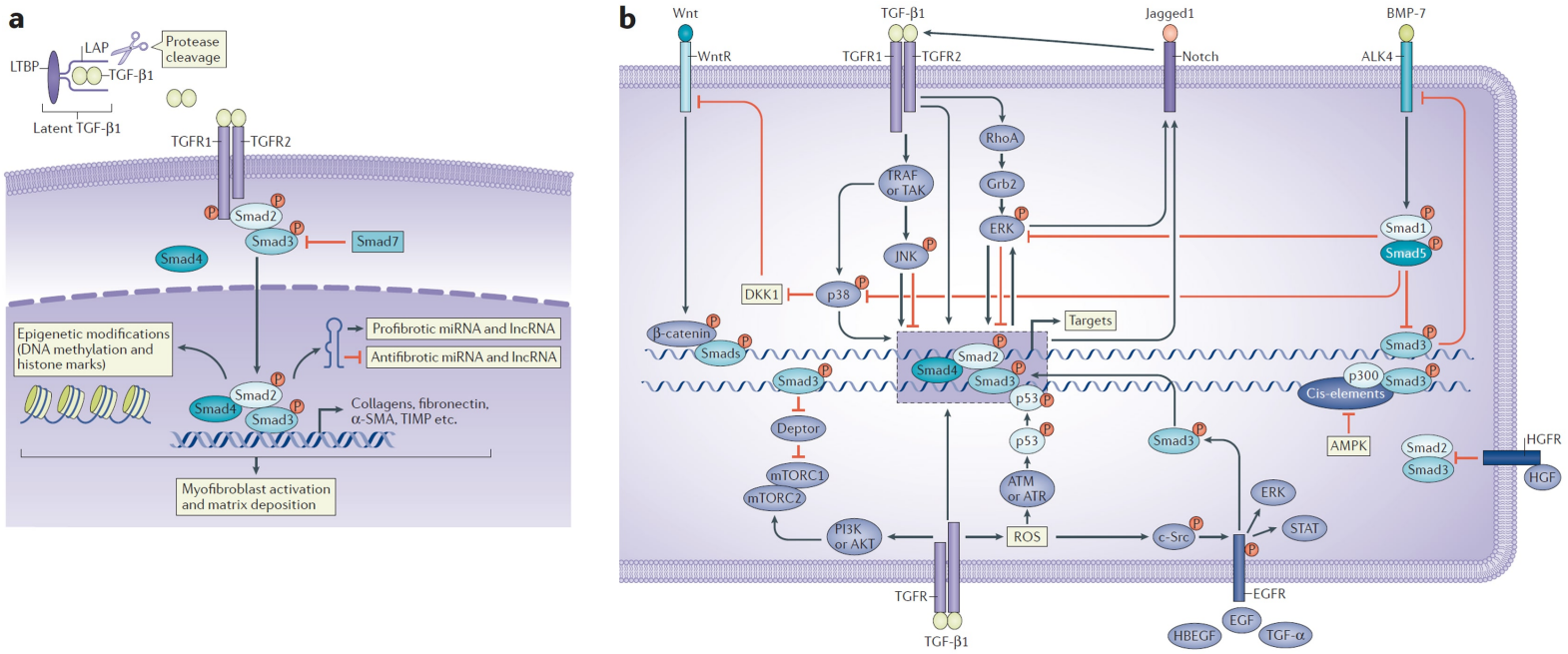


Figure 22: **TGF-β signalling pathways.** (a) Overview of canonical TGF-β/Smad signalling in tissue fibrosis. Once TGF-β1 is released from the LTBP complex, it binds to TGFR2, which recruits and activates TGFR1. TGFR1 phosphorylates Smad2 and Smad3, which are complex with Smad4 and translocate to the nucleus. Smad3 can increase the transcription of profibrotic molecules. Smad7 is a negative regulator of Smad2/3 and inhibits fibrosis. (b) Crosstalk between TGF-β/Smad and other pathways in tissue fibrosis. In addition to Smad2/3/4 complex pathway, TGF-β1 interacts with other signalling pathways. It can activate the MAPKs, p38, JNK and ERK in a Smad-independent manner. MAPKs can modulate Smad3 transcriptional activity. In addition, other signalling pathways can activate MAPKs and thereby modify Smad phosphorylation. Wnt-induced β-catenin stabilisation can enable β-catenin to complex with Smad proteins and enhance gene transcription of profibrotic molecules. TGF-β can also transactivate the EGFR via a ROS-dependent mechanism. BMP7 activates Smad1 and Smad5, negative regulators of Smad3-based gene transcription.Adapted from Meng et al., 2016, Figures 1 and 2.

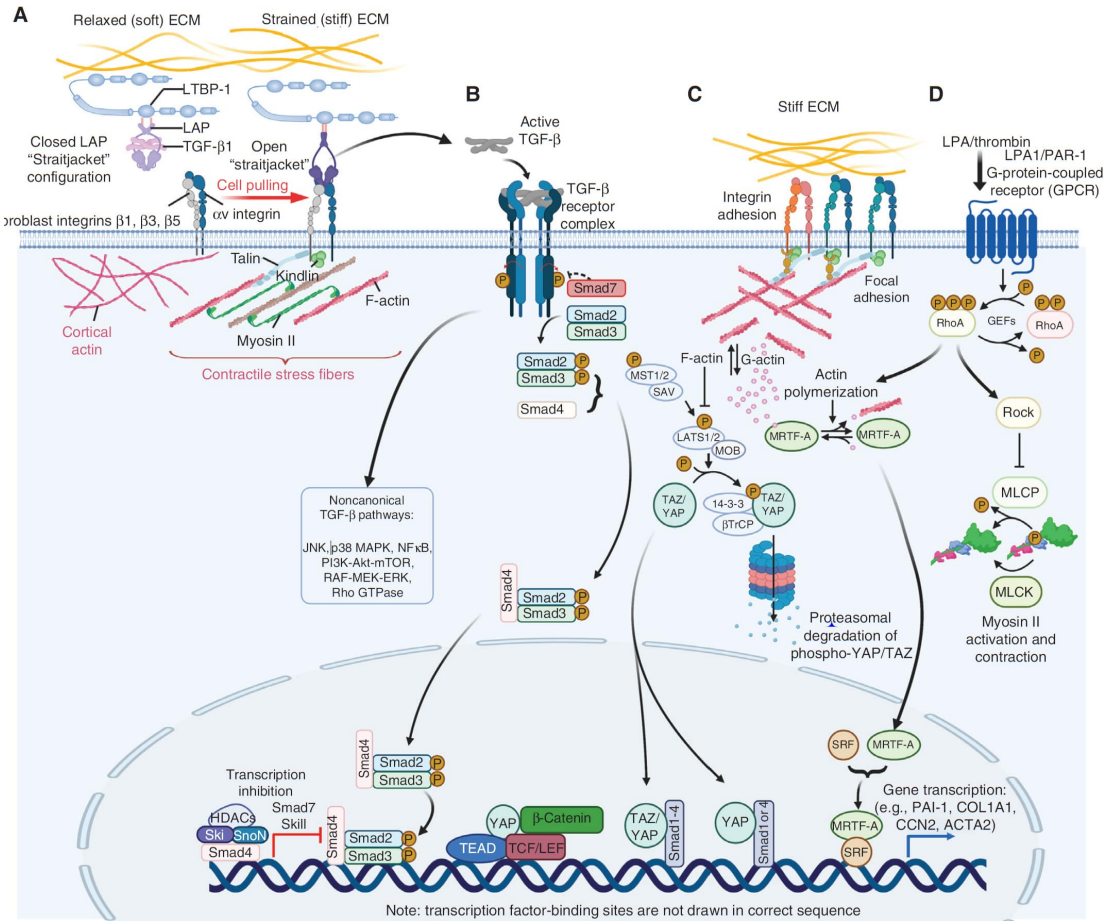


Figure 23: **Mechanical regulation of myofibroblast activation.** (A) Association of TGF-β1 with its latency-associated peptide (LAP) creates a small latent complex (SLC) that is stored in the extracellular matrix (ECM) by binding to the latent TGF-β1-binding protein 1 (LTBP1). A contractile cytoskeleton and mechanical resistance in the ECM shifts integrin αvβ1 into the active conformation. ECM releases active TGF-β1. (B) Active TGF-β1 binds to the TGF-β receptor complex to promote canonical (Smad) or non-canonical (JNK, ERKs, MAPK, or ROCK). Myofibroblast signalling pathways converge at multiple intersection points. (C) RhoA-mediated ROCK activation contributes to actin polymerisation following integrin-mediated mechanosignalling, or (D) GPCR ligand binding. Actin polymerization liberates myocardin-related transcription factor A (MRTF-A) from G-actin to translocate into the nucleus. MRTF-A and SRF drive the transcription of CCN2 and ACTA2. ROCK further regulates gene expression through the transcriptional coactivators YAP and TAZ. Translocation of YAP and TAZ and association with transcription factors such TEAD results in the transcription of genes that promote cell proliferation and fibrogenesis, like CCN2 and miR-21. Adapted from Schuster et al., 2022, Figure 3.

3

THE SKIN BIOLOGY FROM A SINGLE-CELL PERSPECTIVE

In this section we are going to describe, based on the information explained in chapters 1 and 2 about scRNAseq and skin structure and composition respectively, the new discoveries about epidermis and dermis cell composition using scRNAseq. To simplify the understanding, each paper will be discussed individually.

3.1 STATE-OF-THE-ART OF EPIDERMIS AND HF COMPOSITION FROM SINGLE-CELL RNA-SEQ DATASETS

3.1.1 *Human*

CHENG ET AL., 2018C The first study of the epidermis in human samples was performed by Cheng et al., 2018c, who researched the heterogeneity of epidermal cell types on the human scalp, trunk, and foreskin tissue, as well as psoriatic skin. In a primary analysis where all 3 tissues were merged, the *traditional* epidermal strata could be recovered: basal keratinocytes expressed *COL17A1*, *KRT5* and *KRT14*; spinous layer keratinocytes expressed *KRT1*, *KRT10*, *DSG1* and *DSP*; and granular keratinocytes expressed *SPINK5*, *FLG* and *LOR*. This latter layer showed a very reduced cell count due to the loss of these cells during processing.

A secondary analysis of only basal keratinocytes revealed differences across tissues and samples. Basal keratinocytes could be separated intro 3 main categories: *basal1* enriched for *CXCL14* and *DMKN*, *basal2* enriched for *CCL2* and *ILR2*, and *basal3*, enriched for amphiregulin (*AREG*) and *EGFR*. *Basal1* was more or less consistent across tissues and was more prominent in the trunk and scalp. On the other hand, *basal2* was enriched in the foreskin and psoriasis, possibly indicating an immune modulation of these keratinocytes, and *basal3* was enriched in the foreskin, with *AREG* and *EGFR* indicating a higher proliferative rate of neonatal skin (Gilchrest, 1983). For instance, AREG has been described to promote keratinocyte dedifferentiation and proliferation by binding to EGFR and triggering the MAPK/ERK pathway (Seykora, 2010).

Apart from the classical layer view of the epidermis, Cheng et al., 2018c also identified 4 other populations: *WNTI*, which expressed a range of Wnt antagonists–*SFRP1*, *FRZB*, *DKK3*, *WIF1*–; *follicular*, which expressed transcripts known to be expressed in HF and sebaceous glands–*S100A2*, *APOE*, *KRT17*, *MGST1*, *APOC1*–; *channel*, which was upregulated for ion channel and cell-cell communication

transcripts–*GJB2*, *GJB6*, *ATP1B3*, *ATP1B1*, *FXYD3*–and mitochondrial channels–*VDAC2*, *SLC25A5*–; and *mitotic*, which overexpressed different DNA synthesis and cell cycle markers–*PCNA*, *KI67*–.

Considering that *WNTI* and *follicular* clusters were more prominent in the scalp sample, a secondary reanalysis with that sample was performed. *WNTI* transcripts were replicated in the reanalysis, and these cells may represent outer bulge cells, which have been shown to secrete Wnt inhibitors (Lim et al., 2016). For instance, *SFRP1* was shown to be expressed in cuboidal cells of ORS at the base of the HF, and is potentially active in regressing telogenic HFs (Geyfman et al., 2014). *MGST1* and *APOE* transcripts of the *follicular* cluster were also identified in a scalp subpopulation named *sebaceous*. Finally, an additional cluster was identified, *UHF diff*, which expressed *CST6*, *KRT17* and *KRT79*, and might correspond to differentiated upper HF cells. Lastly, within the spinous layer, an additional subcluster was identified expressing *S100A7/8/9* and *IFI27*. These genes have been expressed in clinically normal scalp (Ruano et al., 2016; Suomela et al., 2004), and the authors suggest that may be primers of typical inflammation, pruritus and scaling in the area.

Finnegan et al., 2019 A later reanalysis of the neonatal foreskin data by the same group recapitulated the epidermal differentiation into 8 discrete clusters based on an unsupervised clustering method (Finnegan et al., 2019). This method fitted each cluster with different continuous stages based on the basal-to-granular differentiation model. Stages 1 to 3 were related to basal keratinocytes, and stages 5 to 7 to differentiated keratinocytes. Stage 8 related to the *channel* population discovered by Cheng et al., 2018c, and did not fit within the linear stage differentiation trajectory.

Interestingly, stage 4 corresponded to the previously characterised *mitotic* population by Cheng et al., 2018c, which had a high expression of basal markers *KRT5* and *KRT14*, and an intermediate expression of early differentiation markers *KRT1* and *KRT10*, indicating that the population is rapidly cycling from a basal state into a differentiated state. TF analysis revealed that there were sets of TFs which were either expressed in basal stages or differentiated stages, and a subset of TFs specific to the mitotic stage had the highest expression and relevance compared to the rest of TFs–e.g. *EZH2*, *DNMT1*, *UHRF1*–indicating that this mitotic population might be an inflexion point within the differentiation trajectory. Depletion of some of these transcripts, such as *ETV4* and *ZBED2* induced a significant increase in differentiation marker *KRT10*.

The authors performed an additional analysis to group different transcription factors into modules based on similar functions. This analysis revealed, for instance, a module consisting of calcium-binding and cell adhesion genes (*CDH3*, *FAT1*, *DSG3*); a module enriched in mitotic stage-associated keratins (*KRT6A*, *KRT6B*); and a module that contained genes involved in barrier function (*DEGS2*, *CERS3*), cell adhesion (*DSC1*, *PERP*), tight junctions (*CLDN1*, *CLDN8*), and desquamation (*KLK8*, *KLK11*). Based on this analysis, they observed that basal keratinocytes showed enrichment of antioxidant proteins located at organelles like mitochondria, whereas differentiated keratinocytes expressed an antioxidant core program located at cytoplasmic vesicles. Authors argue that increasing ROS levels in mitochondria promote basal keratinocyte differentiation, and thus these antioxidant proteins protect against endogenous ROS, whereas the location of antioxidant proteins in vesicles may be important for epidermal barrier function.

Wang et al., 2020 A recent study by Wang et al., 2020 used human neonatal epidermis samples to understand basal stem cell transition states. Their analysis observed 4 basal populations (BAS-I to BAS-IV), two spinous populations (SPN-I and SPN-II), and a granular population in addition

to melanocytes and Langerhans cells. The authors developed an estimate of cell type transition likelihood during the process of reconstructing the differentiation trajectory of keratinocytes. From that model, they observed that BAS-III and BAS-IV were the populations with higher likelihood of differentiating into spinous populations. Based on that, the authors concluded that BAS-III and BAS-IV were putative basal stem cells that differentiate into spinous keratinocytes.

Furthermore, the researchers discovered that the presence of *PTTG1* in BAS-I and HELLS, as well as *UHRF1* in BAS-II, is essential for maintaining the epidermal homeostasis. This finding suggests that these genes play a crucial role in epidermal homeostasis. Moreover, the study indicates that the spatial distribution of putative stem cell populations within the epidermis is more significant than previously believed. The authors emphasize the importance of the specific localization of these cell populations in relation to the epidermal rete ridges for maintaining the overall health and function of the epidermis. BAS-III population was defined by the expression of *ASS1*, *COL17A1*, *POSTN*, and located at the top and side of the rete ridges, whereas BAS-IV was defined by the expression of *GJB2*, *KRT6A* and *KRT16*, and located at the bottom of the rete ridges. This observation may agree with the previous ones from Cheng et al., 2018c and Finnegan et al., 2019, who observed that the *mitotic* population was located between basal and spinous keratinocytes.

However, the main difference is that while Cheng et al., 2018c and Finnegan et al., 2019 observe one mitotic population, Wang et al., 2020 observes two. They argue that the need for multiple stem cell populations may be protective against harmful insults by requiring a redundancy or complementary mechanism to ensure proper cell proliferation and differentiation in homeostasis.

Takahashi et al., 2020 In Cheng et al., 2018c HF populations derived from scalp epidermis are present in the sample. However, the enrichment for IFE during sample processing limited the heterogeneity of HF populations recovered. Takahashi et al., 2020 harvested hair grafts from patients undergoing hair transplantation to analyse the heterogeneity of populations within the HF. The unsupervised clustering revealed a range of populations. The classical IFE populations were appearing–2 basal, 3 spinous and 1 granular layer–, as well as HF-associated populations: *infundibulum–RCAN1, CXCL14–*, *isthmus–EPCAM, CYR61–*, *bulge–CXCL14–*, *lower bulge–ANGPTL7, COMP, CHI3L1–*, *matrix/cortex/medula–KRT85, KRT35–*, *IRS cuticle–KRT73, KRT28–*, *IRS Huxley and Henle's layers–TCHH, FABP9–*, *ORS companion layer–KRT75–*, *ORS basal–KRT16, KRT6A–*, *ORS suprabasal–KRT16–*, *sebaceous/apocrine cells–DCD–*, as well as immune, endothelial and melanocyte cells.

When ordering the HF types in pseudo time, they observe that the cortex/medulla/matrix population branches off, terminating in ORS-derived companion layer cells and a further differentiated subset of cells, likely representing hair shaft medulla and cortex components. These data are consistent with a prior report in which early matrix progenitors give rise to the companion layer, and later matrix progenitors give rise to the IRS and lower hair shaft components (Mesler et al., 2017).

3.1.2 *Mouse*

Joost et al., 2016 Mouse epidermis and HF heterogeneity have been long studied, and the first paper to thoroughly analyse epidermal populations was by Joost et al., 2016. In their paper the authors analyse 1,422 epidermal cells, revealing 25 distinct populations, most related to HF. In a primary analysis, cells are divided into 13 populations: IFE cells–consisting on $Krt14^{hi}Mt2^{hi}$ IFE basal cells (IFE

B), *Krt10*$^{+}$*Ptgs*$^{+}$ IFE differentiated cells (IFE DI and DII), and *Lor*$^{+}$*Flg2*$^{+}$ IFE keratinised cells (IFE KI and KII)–, upper hair follicle cells –*Krt79*$^{+}$*Krt17*$^{+}$ divided into three types depending on the expression level (uHF I/II/III)–, sebaceous gland –*Mgst1*^{hi}*Scd1*^{hi} (SG)–, outer bulge–*Postn*^{hi}*Cd34*^{hi} (OB)–, inner bulge–*Krt6a*^{hi}*Krt75*^{hi} (IB)–, T cells–*Cd3*^{hi}*Thy1*^{hi} (TC)–, and Langerhans cells–*Cd207*^{hi}*Ctss*^{hi} (LH)–.

Delving into subtypes, each major cell type–OB, IB, uHF and IFE B–could be further divided, leading to 25 subtypes.

The upper HF cells were categorised into four (uHF IV–VII), one ambiguous (uHF III), and two new (uHF I and uHF II) populations, distinguishable by their location and gene expression. The new populations were located around the SG opening and could be distinguished by *Rbp1* expression as well as high levels of *Defb6* and *Cst6*. uHF I cells, expressing *Klk10*, were located in two suprabasal rings around the SG opening, while uHF II cells, expressing Krt14 (but not Krt5), were linked to the SG duct. The other uHF subpopulations (uHF IV–VII) showed a typical uHF signature with gene expression associated with the basal–*Krt5*, *Krt14*–, suprabasal–*Krt10*, *Ptgs1*–, and the keratinised layer of the IFE–*Flg2*, *Lor*–.

The outer bulge, characterised by high expression of *Cd34*, *Krt15*, and *Lgr5*, is the most well-investigated HF compartment. Most outer bulge cells belong to either OB I or OB II, distinguished by their gene expression and location in the proximal–*Cd34*^{hi}*Postn*^{hi}*Lgr5*^{hi}*Krt24*^{hi}–or central–*Cd34*^{hi}*Postn*^{hi}*Lgr5*^{dim} *Krt24*^{dim}–part of the outer bulge, respectively. OB III, IV, and V are additional populations at the distal end of the bulge area and lower isthmus. OB III is unique in its gene signature, including *Aspn*, *Nrep*, and *Robo2*, and shows strong expression of *Gli1* and *Lgr6*, indicating inclusion of cells from both *Gli1*$^{+}$ and *Lgr6*$^{+}$ populations. OB IV is marked by an overlapping outer bulge and upper HF signature, while OB V is a population of suprabasal cells expressing both an outer bulge signature and differentiation markers.

The majority of inner bulge cells belong to IB I, which exclusively expresses the typical inner bulge signature with high levels of *Krt6a*, *Krt75*, *Timp3*, and *Fgf18*. IB II is a population of cells expressing inner and outer bulge markers and is located in the outer bulge. IB III co-expresses an inner bulge and a differentiation signature, including *Krt10* and *Ptgs1*, and is located at the distal end of the inner bulge compartment.

Regarding FE basal cells, the authors discovered a subpopulation that expresses a unique combination of signatures, including low levels of upper HF markers like *Krt79* and *Postn*, pan-HF markers such as *Sostdc1*, *Aqp3*, and *Fst*, in addition to the IFE basal signature. This signature marks basal cells of the infundibulum, which connects the HF to the IFE and was previously not transcriptionally identified. Thus, the population is named INFU B. Two distinct basal IFE populations (IFE BI and II) were also discovered, both expressing high levels of *Krt14* and *Krt5*, with IFE BI additionally expressing high levels of *Avpi1*, *Krt16*, *Thbs1*, and the transcription factor *Bhlhe40*.

Based on this demarcation of cell types, the authors conclude that *classical* markers of specific *homogeneous* populations such as *Gli1* and *Lgr5* in the bulge, *Lgr6* in the isthmus and *Lrig1* in the infundibulum are not that specific, but are expressed in more than one population.

Interestingly, the authors could classify most of the cell types alongside axes: a spatially-defined axis–distal bulge cells to proximal IFE– and a differentiation axis–basal to differentiated–. The first axis mainly separated OB and IB from uHF and SG, and the IFE group (basal, differentiated and keratinised); whereas the second axis joined OB, IB and uHF into a basal state together with IFE B, and separated it from IFE D and IFE K populations. Similarly to the observations performed in human

keratinocytes, the authors postulate that epidermal basal cells reach a no-return differentiation state that commits to the keratinisation of the epidermis, while cells not reaching that state provide long-term renewal.

Lastly, the researchers noted that traditional markers used to identify stem cell progenitors (SCPs) such as *Cd34*, *Lgr5*, *Lgr6*, or *Lrig1* were not adequate for accurately distinguishing basal cell populations. This is because a significant proportion of suprabasal cells, up to 33%, expressed these markers, while up to 27% of SCPs did not exhibit any of these markers. Therefore, the study highlights the limitations of relying solely on these classical markers to identify and characterize basal cell populations accurately.

JOOST ET AL., 2020 The previous analysis by Joost et al., 2016 was performed in telogenic HFs. However, many HFs are often found in other phases, mainly in the anagen or growing phase. Consequently, in the study published by Joost et al., 2020, the authors extend the previous study to include samples of skin in the anagen (5-week-old mice) as well as telogen (9-week-old mice), extending the number of populations studied. They also characterise other major cell types such as dermal, vascular, immune, and neural crest-derived cells, making it one of the most complete studies of mouse skin.

The major cluster of cells in anagen were HF-derived cells. Upon clustering, they have two small populations and one major branched population. The two minor populations corresponded to suprabasal and basal outer layer–*Barx2*^{hi}*Il1ra1*^{lo} (sbOL) and *Barx2*^{lo}*Il1ra1*^{hi} (bOL)–. The sbOL was divided into three continuous subtypes representing mainly different companion layers. The bOL was divided into a larger cluster of ORS and a minor cluster of lower proximal cup cells (bottom of the bulge, near DP cells).

The third, major *Msx2*$^{+}$ cluster corresponded to the cells forming the IRS, the cuticle and the medulla of the HF. Each was a distinctive branch, parting from a core of cells forming the germinative layer (GL). These cells were divided into 4 clusters (GL1-GL4) expressing *Mt2* and *Dcn*, from which the different branches arise based on RNA velocity and trajectory inference analysis. For instance, IRS comes mainly from the *Id3*^{hi} GL1, whereas the medulla comes mainly from the *Lef1*^{hi} GL4; although all GLs contribute to all three branches. In the IRS branch, which could be divided into IRS1-IRS6, differentiation from IRS1 cells was demarcated into a IRS3/4 branch expressing *Krt71* and *Tchh*, or the IRS5/6 branch expressing a more cuticle identity marked by *Krt72* and *Krt73*.

SHIN ET AL., 2020 In this paper, Shin et al., 2020 characterise the loss of functionality of mesenchymal progenitors to repopulate DP during ageing, leading to mammal hair loss. For part of their experiments, the authors perform scRNAseq of HFs from mice at the early anagen stage, using FACS to enrich for HF mesenchymal cells (DP and DS). After computationally selecting specific cell types, they obtained a total of 1,084 *Sostdc1*$^{+}$, *Rspo3*$^{+}$, *Fgf7*$^{+}$ DPs and 702 *Cd200*$^{+}$, *Stmn2*$^{+}$, *Ednrb*$^{+}$ DS/CTS cells. There is a third population of HF mesenchymal stem cells, termed progenitor cluster (PRG), enriched for cell cycle genes such as *Cdk1* and *Top2a*.

DP cells could be divided into four major clusters–*Notum*$^{+}$ DP1, *Cyr61*$^{+}$ DP2, *Sostdc1*$^{+}$ DP3, and *Frzb*$^{+}$ DP4; and CTS into two clusters–*Mgp*$^{+}$ CTS1, *Ednrb*$^{+}$ CTS2–. Trajectory inference analysis and RNA velocity showed that DP cells show a DP fate–led by DP2, then DP4, and finally DP1 and DP3–and a CTS fate–led partially by CTS1–. Additionally, based on the RNAscope of some of the markers, the

authors state that CTS1 colocalises partially with PRG progenitor cells in the lower bulge, whereas CTS2 resides in the middle-upper part of the HF. Regarding DP cells, DP2/4 are located in the lower part of the DP and are termed as *supplementary* DP, whereas DP1/3 are located in the upper portion of the DP and are termed as the *definitive* DP, which produces the medulla, cortex and IRS cells among others. These findings were congruent with the ones described by Yang et al., 2017.

MA ET AL., 2020 Most single-cell papers use scRNAseq methods during their analyses. The paper by Ma et al., 2020 uses a novel method, SHARE-seq, that combines RNA-seq and ATAC-seq data from the same cell. ATAC-seq reveals domains of regulatory chromatin (DORCs), regions where chromatin is open and thus accessible for transcription. Chromatin potential has been observed to have a longer reach in early stages, while RNA velocity has a farther reach in late pseudotimes, predicting the future of individual cells on a timescale of hours (La Manno et al., 2018). Chromatin potential is established before gene expression, and it is expected to predict future cell states on a longer timescale than RNA velocity, particularly during differentiation. Thus, using ATAC-seq is usually useful to determine early transcriptional dynamics.

The authors use SHARE-seq with brain, lung and skin tissue in the study. In skin, they focus on the development of hair matrix. The authors found that DORCs become accessible before the associated gene's expression onset, indicating lineage priming. One of the target genes is *Wnt3*, a marker detectable at the late stages of hair shaft differentiation. SHARE-seq shows that *Wnt3* DORC becomes accessible before *Wnt3* gene expression, with sequential activation of peaks in the *Wnt3* DORC, followed by nascent and mature RNA expression. Additionally, the pseudo-temporal ordering of RNA expression and TF motif activity showed that *Lef1* is the lineage-priming TF, followed by *Hoxc13* and *Wnt3* expression.

3.2 STATE-OF-THE-ART OF DERMIS COMPOSITION FROM SINGLE-CELL RNA-SEQ DATASETS

3.2.1 *Human*

PHILIPPEOS ET AL., 2018 The first analysis of the human dermis using single-cell methods was performed by Philippeos et al., 2018. In their analysis, they sequenced 184 cells from the abdominal skin of a donor. To guarantee that fibroblasts were representative, cells were filtered to be $CD90^+$, although some $CD90^-$ cells were also included. In the analysis, they reported 4 major groups: 1) A $COL6A5^+$, $COL23A1^+$, $HSPB3^+$ group that corresponded to papillary fibroblasts; 2) a $CD26^+$, $MFAP5^+$, $PRG4^+$ population that, for being $CD39^+$ but $COL6A5^-$, could be identified as reticular fibroblasts; 3) a $CD74^+$, $HLA\text{-}DR4^+$, $CLDN5^+$ that lacked expression of *LUM* and *DCN*, and could be macrophages or dendritic cells; and a 4) $RGS5^+$ group that could be identified as perivascular cells. Therefore, the analysis yielded only 2 effective fibroblast populations: papillary and reticular fibroblasts. Although the analysis was relevant, it lacked enough cells to constitute a representative sample of dermal fibroblasts.

TABIB ET AL., 2018 The first single-cell RNA sequencing (scRNAseq) study showing a previously unexpected degree of heterogeneity of human dermal fibroblasts was published by Tabib et al., 2018. The dataset amounted to 8522 across six samples, 2742 of which were fibroblasts. Unsupervised clustering revealed several clusters that could be divided into two major populations: (1) $SFRP2^+$

DPP4$^+$ cells and (2) *FMO1*$^+$ *LSP1*$^+$ cells. Immunofluorescence of *SFRP2*$^+$/*DPP4*$^+$ cells revealed small elongated cells with elongated nuclei and overexpressed ECM- and fibrosis-related genes, suggesting their classical role as a cell involved in ECM deposition and/or orientation. On the other hand, *FMO1*$^+$/*LSP1*$^+$ cells were relatively large and had lower collagen expression. They also strongly expressed CXCL12, possibly indicating a role in the retention of CXCR4+ immune cells.

The two major populations could be further subdivided. *SFRP2*$^+$ fibroblasts could be divided into the *WIF*$^+$ *COMP*$^+$ *NKD2*$^+$ population and the *PCOLCE2*$^+$ *CD55*$^+$ *FSTL3*$^+$ population. The *WIF*$^+$ cells expressed higher levels of collagen I. Additionally, they found additional smaller populations that were neither *SFRP2*$^+$ nor *FMO1*$^+$: a *CRABP1*$^+$ that could be DP cells; a closely related *COL11A1*$^+$ *DPEP1*$^+$ *RBP4*$^+$ cluster that could be related to pluripotency; a *SFRP4*$^+$ *PRG4*$^+$ population that clustered near the *SFRP2*$^+$ population; and an independent *SFRP4*$^+$ *ANGPTL7*$^+$ *C2orf40*$^+$ population.

Interestingly, the two major fibroblast subtypes were found with no recognisable similarity with papillary and reticular fibroblasts, and *classical* markers like *DPP4*, *LRIG1* or *DLK1* were not expressed, of were confined to a set of populations. The authors postulate that papillary and reticular matrices are established during development and do not require specialised fibroblast types for maintenance.

Vorstandlechner et al., 2020 While looking for the fibroblast subset responsible for fibrosis, Vorstandlechner et al., 2020 characterised three large fibroblast clusters, one of which could be subdivided into four smaller groups, yielding a total of 6 different subtypes (FB1 to FB6). These fibroblasts were characterised by the expression of *LUM*, *FBLN1*, *PCOLCE*, *PLPP3*, *SDC2* and *MXRA8* genes. FB1 (*FAP*, *SFRP2*, *MFAP5*, *LOX*, *HAS*, *FAP*, *MMP14*) was associated with ECM assembly and wound healing; FB2 (*FMO1*, *APOE*) with immunological process; FB3 (*SFRP2*, *WIF1*) was associated with cartilage development and leptin signalling; FB4 (*B4GALT1*) was characterised by a response to growth factors; FB5 (*CXCL1*, *APOE*) with immunological processes; and FB6 (*WIF1*, *CXCL1*, *SFRP2*, *COL14A1*, *FN1*, *MMP14*) was associated with IFNγ response and p38 and NKκB signalling, as well as with matrix assembly–due to the expression of *FN1*, *HAS1*,. *MMP14* and *COL14A1*–.

Additionally, they observe that *traditional* markers of papillary and reticular fibroblasts such as *NTN1*, *PDPN* or *MGP* were detected in all fibroblasts clusters. Their dataset differed substantially from that of Tabib et al., 2018 because it showed virtually no expression of FMO1 in any cell and SFRP2 expression in almost every fibroblast subset. RNAscope assessed SFRP2 expression in the papillary and reticular dermis. They attributed the variations with Tabib et al., 2018 dataset to technical reasons and donor site-specific differences.

Solé-Boldo et al., 2020 In a study on dermal ageing, Solé-Boldo et al., 2020 determined four main young fibroblast subpopulations that could be discriminated into five after secondary analysis: (1) secretory reticular–WISP2, *SLPI*, *CTHRC1*, *MFAP5*, *TSPAN8*–, (2) secretory papillary–*APCDD1*, *ID1*, *WIF1*, *COL18A1*, *PTDGS*–, (3) mesenchymal–*ASPN*, *POSTN*, *GPC3*, *TNN*, *SFRP1*–, and (4) proinflammatory–*CCL19*, *APOE*, *CXCL2*, *CXCL3*, *EFEMP1*–, which in turn was composed of two subclusters–one with predominant expression of *CXCL1* and the other with predominant expression of *CCL19*–. Looking at collagen expression patterns, the authors observed cartilage-associated *COL11A1* and *COL24A1* more expressed in the mesenchymal cluster, and papillary dermis and DEJ collagens *COL13A1*, *COL18A1* and *COL23A1* in the secretory papillary population.

To confirm some of these patterns at the biological level, the authors used RNAscope to detect specific transcripts to establish the location of each population–secretory-reticular *CTHRC1*, secretory papillary *APCDD1*, pro-inflammatory *CCL19* and *APOE*, and mesenchymal *ASPN*–. As expected, secretory papillary and secretory reticular were predominantly, but not exclusively, expressed in their respective locations; immune was located throughout all the dermis and in association with the vasculature, and mesenchymal was predominantly located in the reticular dermis in association with hair follicles.

When compared with Tabib et al., 2018 dataset, and similar to Vorstandlechner et al., 2020, they found that the expression of FMO1 was very low in all dermal fibroblasts, whereas SFRP2 was expressed by both secretory subpopulations and a subgroup of the proinflammatory subpopulation; similarly, as discovered by Vorstandlechner et al., 2020. The authors found no overlap between their dataset and the dataset from Philippeos et al., 2018. They attributed these discrepancies to variations in the experimental approaches used in the two studies.

He et al., 2020 Lastly, He et al., 2020 performed a joint analysis of skin from healthy- and AD-derived samples. In their analysis, they separated fibroblasts into 4 main populations: (1) $FBN1^{+}MFAP5^{+}$ (2) $APOE^{+}ABCA6^{+}$, (3) $COL11A1^{+}LAMC3^{+}$ –also expressing *POSTN* and *EDNRA*–, and (4) $COL6A5^{+}COL18A1^{+}$. This last population, which also expressed cytokines *CCL19* and *CCL2*, was predominant in AD samples, indicating a putative role in the pathogenesis of this disease.

They noted parallels between the primary fibroblast clusters of healthy skin and those found in earlier studies, albeit with varying gene expression patterns. They ascribed the plasticity of fibroblasts or their transcriptional responses to atopic dermatitis. For example, they discovered low FMO1 levels in fibroblasts, as reported in Vorstandlechner et al., 2020 and Solé-Boldo et al., 2020. In discussing the $COL6A5^{+}$ cluster identified by Philippeos et al., 2018, they emphasized its absence of proinflammatory characteristics. Interestingly, *COL6A5* and *COL18A1* are also markers of fibroblast population from other datasets, such as the previously reported papillary population from Philippeos et al., 2018, or the secretory papillary from Solé-Boldo et al., 2020.

3.2.2 *Mouse*

Salzer et al., 2018 One of the first papers to perform a scRNAseq of fibroblast cells is by Salzer et al., 2018. Their study analyses changes in skin fibroblasts that lead them to acquire an adipogenic profile. This profile is partially reverted by caloric restriction. Their study observes that pairwise distances of old clusters are less separated than those from young clusters, indicating less transcriptomic heterogeneity of old cells. Additionally, they observed a more marked *Sca1* and *Pparg* expression in old samples, as well as other genes related to a pro-adipogenic signature. Using different tracing methods, they observed that the old fibroblast lineage was related more to the lower reticular dermis than the papillary dermis.

Although the results from this study are relevant, the low number of cells hindered a complete characterisation of the fibroblast heterogeneity, mainly consisting of a major cluster per sample and condition.

GUERRERO-JUAREZ ET AL., 2019 The first mouse study in which the fibroblast heterogeneity is studied *per sé* is in Guerrero-Juarez et al., 2019. The authors study the fibroblast populations arising after wounding (12 days post-wound). Single-cell of microdissected tissue reveals 5 main fibroblast types: FIB-1 to FIB-5. A separate analysis of these cells yielded 12 subclusters, sC1 to sC12. All fibroblasts shared a common TF signature, comprised of 20 TFs such as *Cebpb*, *Hif1a*, *Klf4/6/9*, *Nfat5*, *Nr4a1/2*, *Prrx1* or *Runx1*. Some of these factors–*Runx1*, *Tcf4*, and *Zeb2*–were reported to be implicated in myofibroblast differentiation.

Fibroblasts in clusters sC1/2/4-8/10/12 showed a marked *Pdgfra* expression, and sC1/4/5/7/8 also showed *Crabp1* expression. Some of these clusters were small, like *Cyp26b1*$^{+}$ *sC7* or *Gos2*$^{+}$ cluster. Regarding TGF-β signalling, two major populations were described: one comprised approximately 24% of wound fibroblasts on sC3/9/C11. This population expresses low levels of TGF-β receptors, *Tgfbr2*, *Tgfbr3*, and PDGF receptor *Pdgfra* but high levels of *Pdgfrb*. They also stain with RGS5 and were located throughout the wound dermis. The second population represents about 76% of wound fibroblasts and is more heterogeneous, consisting of nine subclusters expressing intermediate to high levels of *Tgfbr2*, *Tgfbr3*, and high levels of *Pdgfra* but not *Pdgfrb*.

The main drawback of this study is the lack of a control population to compare. Therefore, it is not possible to ascertain the fibroblast populations on healthy, uninjured dermis based on this dataset.

JOOST ET AL., 2020 The first full characterisation in the uninjured dermis is performed by Joost et al., 2020, who discovered 8 main clusters: FIB1-FIB4, DS1, DS2, aDP and tDP. aDP and tDP are DP cells present in anagen and telogen phases, respectively, with aDP expressing *Corin*, *Nrg2* and *Cntn1*, and tDP expressing *Crabp1*, *Notum* and *Pappa2*. DS1 cells are associated with telogen and anagen cells–marked by *Abi3bp*, *Ramp1* and *Mylk*– and DS2 cells are mainly associated with anagen –marked by *Acta2*, *Tagln* and *Grem2*.

The FIB1-FIB4 populations demarcate different layers of the skin. FIB1 and FIB2 are thoroughly located in the epidermis, FIB1–*Sparc*^{hi}*Dcn*$^{+}$–being more prominent in anagen and FIB2–*Sparc*^{lo}*Dcn*$^{+}$–in telogen; FIB3 was associated with cells in the hypodermis–*Plac8*^{lo}*Cxcl12*^{hi}*Gpx3*$^{+}$–, and FIB4 to cells in the adventitia, below the *panniculus carnosus*–*Plac8*^{hi}*Mfap5*^{hi}*Gpx3*$^{+}$–. Although FIB1 and FIB2 are independent populations based on the hair follicle phase, a classification algorithm in the validation dataset failed to replicate the differences and combines them into FIB1/FIB2. (Figure S5H in the original publication).

Based on image analysis, both FIB1/FIB2 cells were stable, with around 40% cells in both hair cycle stages. FIB1 cells have increased expression of genes linked to collagen production such as *Sparc*, as well as core TFs *Creb3*, *Creb3l3* and *Mxd4*. On the other hand, FIB2 cells exhibited an ECM signature dominated by proteoglycans such as *Dcn*, *Lum* and *Mfap4*.

JACOB ET AL., 2022 Skin development has been an active research field, with many relevant discoveries, like the fibroblast fate determination within papillary and reticular dermis described by Driskell et al., 2013 and commented on in Section 2.2.1. Driskell et al., 2013 observed that the divergent fate of papillary and reticular fibroblasts is established at E16.5. Although Rinkevich et al., 2015 have proposed the existence of potential fate-specification before E16.5, the biology of fibroblast development is still far from being completed.

In this recent study, Jacob et al., 2022 used scRNAseq on E12.5, E13.5 and E14.5 embryos to study cell development of the dermis. One of their main focuses is dermal fibroblasts, where they observed that, even at E12.5, a lineage commitment is already established. At E12.5, there are two major fibroblast subsets: *FIB Origin* and *FIB Deep*. The *FIB Origin* cells constitute a Wnt-pathway-activated *Lef1*$^{+}$*Axin2*$^{+}$ fibroblast subset that maps to the upper dermis; and is divided into 6 subgroups–*FIB Origin 1/6*–with different signatures and lineage fates. *FIB Origin 1/2* express *Lef1*, *Tcf7*, *Dkk1* and differentiates into upper dermal fibroblasts, whereas *FIB Origin 5/6* expresses *Epha4*, *Creb5* and *Thbs2* and differentiates into muscle and inter-muscle fibroblasts in the lower layers of skin. *FIB Origin 3/4* are a *mixture* of the 1/2 and 5/6 subsets. On the other hand, *FIB Deep* population is divided into three subpopulations that map mainly to the lower dermis (which was no muscle structure yet). This population does not seem to have a clear differentiation trajectory at E13.5 and E14.5.

FIB Origin 1/2 differentiates into two populations: *FIB Upper* and *FIB Lower*. *FIB Upper* shows an active Wnt pathway based on the expression of *Lef1* and *Axin2*, and, based on other marker expressions, it is mainly localised to the upper dermis. Some of these cells exit cell cycle and start expressing *Sox2*, indicating the early settling of the dermal condensate (DC), from which HFs stem. Additionally, they observe that *Dkk2* is downregulated, which is, in fact, present in non-hairy skin while absent in hair-bearing skin, thus supporting *FIB Upper* differentiation into *FIB DC*. *FIB Origin 1/2* also differentiates into *FIB Lower*, a population that lacks unique marker expression but which, based on the location of surrounding markers, or its lack thereof, is putatively located in the upper dermis, below *FIB Upper* and above the *panniculus carnosus*.

FIB Origin 5/6 differentiates into two populations: *FIB Muscle* and *FIB Inter*. *FIB Muscle*, is a group of cells expressing *Nppc*, *Rgcc* and *Gfra1* and is exclusively located within the developing muscle layers. clustering of this population reveals two subgroups, characterised by the expression of *Aspn* and *Wnt4*–*FIB Muscle 1*–or *Ebf2* and *Igfbp3*–*FIB Muscle 2*–. The former is believed to maintain satellite cell quiescence, while the latter is expected to support myofibroblast differentiation. On the other hand, *FIB Inter* expresses *Mfap5* and *Gata6*, and is divided into 3 subgroups, two of them expressing fascia markers *Nov*, *Dpp4*, *Plac8*, or *Col14a1*, and an additional one expressing adipogenic TFs *Pparg* and *Cebpa*, and expected to be a differentiated state of the two fascia-associated populations. These *FIB Inter* are believed to be fascia and pre-adipocyte-related fibroblasts and are located beneath the *panniculus carnosus*.

Part II

THESIS HYPOTHESES AND OBJECTIVES

THESIS HYPOTHESIS AND OBJECTIVES

Fibroblasts in the dermis have been traditionally separated into "papillary fibroblasts" and "reticular fibroblasts" based on their location in the papillary and reticular dermis. Single-cell RNA sequencing expanded this characterisation into disconnected populations across few-specific publications. Based on this fact, we hypothesised that dermal fibroblast heterogeneity was not dataset-specific, but rather general and consistent across studies. To study this hypothesis we followed these objectives:

1. Perform a characterisation of human and mouse dermal fibroblast populations, separately. For each organism, single-cell datasets from several sources will be jointly studied to find similarities and differences in fibroblast populations.
2. Analyse the extent to which this characterisation shows dataset-specific fibroblasts.
3. Study the functions of individual populations based on their most representative markers.
4. Analyse the similarities between mouse and human fibroblasts to understand common fibroblast function patterns.
5. Study how inter-dataset heterogeneity may affect results provided by individual datasets, and how this heterogeneity may impact the results reported in the scientific literature.

Part III

MATERIALS AND METHODS

4

PRIMARY OF FIBROBLAST HETEROGENEITY IN HUMAN AND MOUSE SKIN USING SCRNASEQ

4.1 HUMAN DATASET DOWNLOADING AND FASTQ PROCESSING

The count table from Tabib et al. (2018) dataset was downloaded in comma-separated values format from `https://dom.pitt.edu/wp-content/uploads/2018/10/Skin_6Control_rawUMI.zip`. Metadata table including authors' computationally assigned clusters was downloaded from `https://dom.pitt.edu/wp-content/uploads/2018/10/Skin_6Control_Metadata.zip`. The count table from Solé-Boldo et al. (2020) dataset was downloaded in mtx format from the Gene Expression Omnibus database (dataset GSE130973). The count table from the Vorstandlechner et al. (2020) dataset (V) was kindly provided by the authors after personal request. The count table from He et al. (2020) dataset (H) was elaborated from the original fastq read files from healthy donors. Sequence Read Archive accession numbers were the following: SRR11396171XX, where XX is 62, 64, 66, 67, 68, 70, 71, and 75. Each fastq file contained Chromium 10X, version 2, Cell Barcode + Unique Molecular Identifier in the first 26 bases and the read on the rest. From each fastq, one fastq with the Cell Barcode + Unique Molecular Identifier and a second fastq with the read were produced according to 10X file standards. All samples were jointly processed (read pseudomapping and counting) using *loompy* 3.0.6, which depends on *kallisto* 0.46.2, using default parameters.

4.2 HUMAN DATASET PREPROCESSING

All datasets were preprocessed using *scanpy* 1.4.6 (with *pandas* 1.0.5, *numpy* 1.18.5, and *matplotlib* 3.2.0 as dependencies) (Wolf et al., 2018). From each raw dataset, the preprocessing was done in two steps. The first step produced a general representation of the dataset, from which fibroblasts would be selected. The second step produced a representation of the fibroblast populations. In the first step preprocessing, all genes without expression were removed. Datasets were log-transformed with a pseudocount (log1p) and normalized using *scanpy*'s `normalize_total` method. Feature selection was done with *triku*, version 1.1.0 (Ascensión et al., 2022), using default parameters. Next, PCA was applied, and then kNN graph construction, UMAP dimensionality reduction (McInnes et al., 2018), and Leiden community detection (Traag et al., 2019) with default values were performed. Fibroblast communities were selected using *LUM*, *PDGFRA*, *COL1A1*, *DCN*, and *FBLN* genes. In the Vorstandlechner et al., 2020 dataset, UMAP and *leiden* were substituted by the t-SNE Maaten et al., 2008b coordinates

and cell labels provided in the metadata. In the Tabib et al. dataset, *leiden* communities were also substituted by the labels in the metadata. Once the fibroblast populations were isolated from each dataset, the second preprocessing step was done as follows. Genes not expressed in that subset were filtered, and feature selection with *triku* was reapplied. PCA was calculated with 30 components, and the kNN graph was constructed with k as half the square root of the number of fibroblasts, using cosine distance. UMAP was run with `min_dist` values of 0.2 for Tabib, 0.3 for Solé-Boldo, 0.3 for Vorstandlechner, and 0.05 for He. Then, *leiden* community detection was run with resolution values (higher values yield more communities) of 2 for Tabib, 1.8 for Solé-Boldo, 2 for Vorstandlechner, and 1.3 for He.

4.3 MOUSE DATASET DOWNLOADING AND PROCESSING

Refer to 5.1 for dataset download and fastq file preprocessing, and to 5.3 for dataset processing. These are used during the secondary analysis, but the methodology is the same compared to the primary analysis phase.

Out of all the available datasets used in the secondary analysis, the ones selected for the primary analysis were Abbasi et al., 2020, Buechler et al., 2021, Haensel et al., 2020, Phan et al., 2020, Shook et al., 2020.

In these datasets, due to being in a preliminary analysis phase, no metadating was applied.

4.4 MAPPING OF CLUSTERS AND AXES ACROSS DATASETS

Once communities were obtained, (DEGs) were calculated using *scanpy*'s `rank_genes_groups` method with the Wilcoxon method. If two or more communities did not express different DEGs, they were joined. Applying this method, we reduced the Tabib communities from 21 to 14 (T0–T13), the Solé-Boldo communities from 17 to 11 (S0–S10), the Vorstandlechner communities from 19 to 12 (V0–V11), and the He communities from 18 to 9 (H0–H8). To further reduce these communities to clusters A1–A4, B1 and B2, and C1–C4 as well as to axes D, E, and F, we recalculated the DEGs between the new communities and manually curated for genes that would map across two or more datasets using T communities as reference.

A gene was considered a good candidate for a cluster marker if it was mainly expressed in that cluster, and other genes with the same expression pattern in one dataset followed the same expression pattern in the rest of the datasets. In some cases, one gene expressed in two or more communities from one dataset would map to one community in another dataset. If this pattern was repeated with the rest of the markers, those communities were candidates to be merged into one cluster, setting a common denominator for all datasets. For example, *CCL2*, *ITM2A*, and *PLAG2A* are markers of T2 and T5 communities, which map to the S3; V3, V5, and V10; and H3 communities. Therefore, all these communities were merged as B1. However, T2 and T5 do express different DEGs (T2 overexpresses *NFKBIA* and *GSN*, whereas T5 overexpresses *CXCL12* and *ITM2A*). The same process was applied to the construction of axes.

If two or more clusters share a set of markers and those markers are shared across datasets, these clusters were merged to the same axis. Again, those markers were manually curated to avoid selecting DEGs that were more sparsely expressed or did not share a common expression across datasets. Axes

D, E, and F could technically be considered clusters. However, we decided that the axis nomenclature fits them better.

4.5 DATASET INTEGRATION

To integrate datasets from different experiments, we applied the `scanpy.AnnData.concatenate()` function on Tabib et al., 2018, Solé-Boldo et al., 2020, and Vorstandlechner et al., 2020 datasets to concatenate them into a single dataset. He et al., 2020 dataset was not integrated owing to poor cell viability inducing possible artefacts during integration. Batch effect correction on the concatenated dataset was applied using *bbknn* (Polanski et al., 2020) with standard parameters. Downstream analysis was performed by normalising, applying PCA, feature selection, and neighbour calculation as described earlier. *leiden* clustering was reapplied to obtain new clusters from which the proportion of each dataset and individual dataset cluster was calculated.

4.6 ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Abdominal tissue remnants used in histology and immunofluorescence were obtained from healthy donors undergoing abdominoplasty, following all legal and ethical regulations and after protocol approval by the Research Ethics Committee of Policlínica Gipuzkoa. Donors were informed of the procedure and voluntarily signed written informed consent.

4.7 HISTOLOGY AND IMMUNOFLUORESCENCE

Abdominal tissue was embedded in optimal cutting temperature medium (Tissue-Tek, Sakura Finetek, Barcelona, Spain, 4583), frozen by immersion in isopentane previously cooled in liquid nitrogen, and stored at −80 °C until usage. For immunofluorescence, cryostat sections (10 μm) from optimal cutting temperature blocks were dried and fixed in Histofix solution (PanReac AppliChem, Barcelona, Spain, 256462.0967) for 10 minutes at RT and then permeabilized and blocked with 10 % normal donkey serum (Merck, Kenilworth, NJ, D9663) and 0.3% Triton X-100 (Merck, T8787) in PBS for 1 hour at RT. Primary antibodies were incubated for 1 hour 30 minutes at RT (diluted in blocking solution).

Slides were rinsed twice with PBS, and secondary antibodies were incubated for 1 hour at RT (diluted in blocking solution). The slides were rinsed twice with PBS, and the nuclei were stained with 1 μg/ml DAPI solution (Thermo Fisher Scientific, Waltham, MA, D3571) for 2 minutes at RT. Slides were rinsed again with PBS and mounted with Fluoromount G mounting media (Electron Microscopy Sciences, Hatfield, PA, 17984-25). Sections were analysed using a Zeiss LSM900 confocal laser scanning microscope (×20 objective, Zeiss, Oberkochen, Germany). Antibodies used are listed in Supplementary Table 2.

4.8 MAPPING OF CLUSTERS AND AXES ACROSS MOUSE DATASETS

Instead of the fully manual cluster labelling and mapping done in human datasets, the mechanism proposed for mouse datasets is semi-supervised. Once all mouse datasets' clusters are labelled, populations are matched using the population matching algorithm from section 5.4. In this part,

Table 2: **Antibodies used in the primary study immunofluorescence staining.**

Antibody	Host Species	Target Species	Clonality	Dilution	Brand	Cat.No.
CXCL12	Mouse	Human	Monoclonal (clone 79018)	1:100	Research and Diagnostic Systems	MAB350
SFRP1	Rabbit	Human	Polyclonal	1:100	Abcam	ab4193
SFRP2	Rat	Human	Monoclonal (clone 8H1)	1:100	Merck-Millipore	MABS330
Donkey anti-mouse IgG AF488	Donkey	Mouse	Polyclonal	1:500	Thermo Fisher Scientific	A21202
Donkey anti-rabbit IgG AF647	Donkey	Rabbit	Polyclonal	1:500	Thermo Fisher Scientific	A32795
Goat anti-rat IgG AF555	Goat	Rat	Polyclonal	1:500	Thermo Fisher Scientific	A21434

instead of only matching the populations of each dataset within its own datasets, the algorithm is applied cross-datasets. For instance, markers of cluster 1A in dataset A are mapped in datasets B and C, markers of cluster 1B are mapped in datasets A and C, and markers of cluster 1C are mapped in dataset A and B.

With this step we seek to find cluster correspondences between datasets, and generate a cluster graph that explains the relationships of clusters between and within datasets. To generate this graph, between-dataset clusters are mapped manually to seek correspondences. There are several correspondence cases that can be considered to construct this graph:

- One-to-one correspondence (Figure 24A): all (or most) of the cells from one cluster in dataset A correspond to the cells of a cluster from dataset B, when mapped into dataset A. We consider these two clusters to be a match.
- One-to-several correspondence (Figure 24B): more than one cluster from dataset A correspond to one cluster of dataset B. This is possible because cell types from dataset A are more diverse, due to being more cells, or having a more diverse transcriptomic profile. This information is considered in the graph, indicating that the clusters from dataset A belong to dataset B.
- Partial overlap (Figure 24C). Some cells of one cluster from dataset A overlap with dataset B, but not fully. This overlap is considered a match, but the partial non-overlapping clusters are included in the graph.

An additional step to construct the relationship graph is to consider the relations between clusters within the same dataset. To do that, the PAGA algorithm is run within each dataset and, combined with the structure of the UMAP plot, the relationships of clusters within datasets used in the graph are completed. Once the graph is constructed, a second round of processing is done, where excess clusters are manually pruned and a consensus set of labels is obtained. The markers of each new merged label are gathered from the sets of markers of similar clusters, trying to find the ones that are associated with the most of the clusters.

4.9 TABLES AND FIGURES

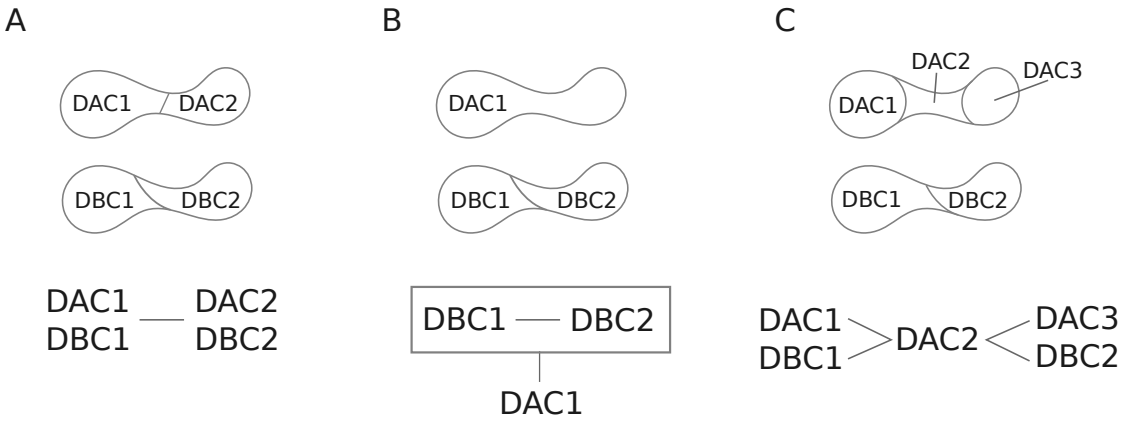


Figure 24: **Scheme of overlap types in cluster correspondence graph construction.** (A) In one-to-one correspondence, cluster 1 from dataset A (DAC1) maps to cluster 1 from dataset B (DBC1). (B) In one-to-several correspondence, when mapping dataset A clusters to dataset B clusters, DAC1 maps to DBC1 and DBC2. Therefore DBC1 and DBC2 are joined in a inferior level to DAC1. (C) In partial overlap, we see that DAC2 maps partially to DBC1 and DBC2, so this relationship is also mapped, as if DAC2 was a bridge, either between DAC1 and DAC3, and DBC1 and DBC2.

5

SECONDARY ANALYSIS OF FIBROBLAST HETEROGENEITY IN HUMAN AND MOUSE SKIN USING SCRNASEQ

5.1 DATASET DOWNLOADING AND FASTQ PROCESSING

In this section, we are going to describe the download and preprocessing of raw samples. Additionally, for many datasets that are already preprocessed at the time of download, we will explain the preprocessing steps that were applied. Unless otherwise stated, the traditional preprocessing performed for already preprocessed files was running 10X CellRanger software on raw fastq files, with hg38 or mm10 genomes as the reference.

5.1.1 *Human datasets*

The human datasets used in this analysis are the following:

- Ahlers et al., 2022: fastq files were downloaded from SRA archive (accession number PRJNA754272) and were processed using kallisto, embedded in loompy pipeline.
- Billi et al., 2022: files were downloaded in 10X mtx form (unprocessed filtered file) from GEO database (accession numbers GSE186476).
- Boothby et al., 2021: files were downloaded in h5 form (unprocessed filtered file) from GEO database (accession number GSE183031).
- Burja et al., 2022: fastq files were downloaded from SRA archive (accession number ERR9121019) and were processed using kallisto, embedded in loompy pipeline.
- Deng et al. (2021) (Deng et al., 2021): files were downloaded in 10X mtx form (unprocessed filtered file) from GEO database (accession numbers GSM4994382, GSM4994383 and GSM4994384 for scar and GSM4994379, GSM4994380 and GSM4994381 for scar).
- Gao et al., 2021: files were downloaded as a unprocessed filtered loom from GEO database (accession number GSE162183).
- Gaydosik et al., 2019: files were downloaded in csv form (unprocessed filtered file) from GEO database (accession numbers GSM3679038, GSM3679039, GSM3679040, GSM3679041).

- Gur et al., 2022:files were downloaded in txt form (unprocessed filtered file) from GEO database (general accession number GSE195452).
- He et al., 2020: fastq files were downloaded from SRA archive (accession numbers SRR11396162, SRR11396164, SRR11396166, SRR11396167, SRR11396168, SRR11396170, SRR11396171, SRR11396175) and were processed using kallisto, embedded in loompy pipeline.
- Hughes et al., 2020: files were downloaded in csv form (unprocessed filtered file) from GEO database (accession number GSE150672). Metadata files indicating sample and disease were downloaded from this website (https://tinyurl.com/4e345da2).
- Kim et al., 2020a:fastq files were downloaded from SRA archive (accession numbers SRR9307706, SRR9307707, SRR9307708, SRR9307709, SRR93077010, SRR9307711) and were processed using kallisto, embedded in loompy pipeline.
- Liu et al., 2021a: fastq file processing produced a faulty count matrix and, therefore, the count matrix was extracted from the Seurat object provided under personal request to the authors, which can be downloaded from the following link (https://tinyurl.com/3bvzajkn).
- Mariottoni et al., 2021: files were downloaded in tsv form (unprocessed filtered file) from GEO database (accession number GSM5352395).
- Mirizio et al., 2020: fastq files were downloaded from SRA archive (accession numbers SRR12955136 to SRR12955159), were trimmed using `seqtk` (arguments `-e 124 -e 52`) and were processed using kallisto, embedded in loompy pipeline.
- Reynolds et al., 2021: dataset download and processing is described in Materials and Methods section 6.1.
- Rindler et al. (2021) (Rindler et al., 2021): files were downloaded in tsv form (unprocessed filtered file) from GEO database (accession numbers GSM5534590 to GSM5534593).
- Solé-Boldo et al., 2020: fastq files were downloaded from SRA archive (accession numbers SRR9036396 and SRR9036397) and were processed using kallisto, embedded in loompy pipeline.
- Tabib et al., 2018: data file in csv form was downloaded from Pittsburgh University website (link) as well as the metadata file (link).
- Tabib et al., 2021: files were downloaded in h5 form (unprocessed filtered file) from GEO database (accession numbers GSM4115868, GSM4115870, GSM4115872, GSM4115874, GSM4115875, GSM4115876, GSM4115878, GSM4115880, GSM4115885, GSM4115886).
- Jones et al., 2022: files were downloaded in fastq form from Amazon Web Services platform (under previous access request). Fastqs from samples *TSP10_Skin*, *TSP14_Abdome*, *TSP14_Chest* were downloaded and processed using kallisto, embedded in loompy pipeline.
- Theocharidis et al., 2020: 10X mtx raw filtered files were obtained by personal request, and downloaded from their Dropbox service. Sample IDs from controls were *H1_080717*, *H2_091117*, *H3_091117*, *H4_100317*.
- Theocharidis et al., 2022: files were downloaded in csv form (unprocessed filtered file) from GEO database (accession numbers GSM5050521, GSM5050534, GSM5050538, GSM5050540, GSM5050542, GSM5050548, GSM5050552, GSM5050553, GSM5050555, GSM5050556, GSM5050560, GSM5050564,

GSM5050567, GSM5050568, GSM5050574, GSM5050522, GSM5050524, GSM5050525, GSM5050526, GSM5050528, GSM5050529, GSM5050562, GSM5050565, GSM5050570, GSM5050572, GSM5050523, GSM5050527, GSM5050531, GSM5050532, GSM5050536, GSM5050537, GSM5050539, GSM5050541, GSM5050547, GSM5050566, GSM5050569, GSM5050573, GSM5050530, GSM5050533, GSM5050535, GSM5050557, GSM5050558, GSM5050559, GSM5050563).

- Vorstandlechner et al., 2020: processed loom file was obtained under personal request.
- Vorstandlechner et al., 2021: files were downloaded in tsv form (unprocessed filtered file) from GEO database (accession numbers GSM4729097 to GSM4729102).
- Xu et al. (2021) (Xu et al., 2021b): files were downloaded in 10X mtx form (unprocessed filtered file) from OMIX database (accession number: OMIX691).

5.1.2 *Mouse datasets*

The mouse datasets used in this analysis are the following:

- Abbasi et al., 2020: files were downloaded in 10X mtx form (unprocessed filtered file) from GEO database (accession number: GSM2910020).
- Boothby et al., 2021: files were downloaded in 10X mtx form (unprocessed filtered file) from GEO database (accession numbers: GSM5549901 and GSM5549902).
- Buechler et al., 2021: fastq files were downloaded from ArrayExpress database (accession number E-MTAB-10315) and were processed using kallisto, embedded in loompy pipeline.
- Haensel et al., 2020: files were downloaded in 10X mtx form (unprocessed filtered file) from GEO database (accession numbers: GSM4230076 and GSM4230077).
- Joost et al. (2020) (Joost et al., 2020): files were downloaded in 10X mtx form (unprocessed filtered file) from GEO database (accession numbers: GSM4186888 to GSM4186893).
- Phan et al., 2020:files were downloaded in loom format (unprocessed filtered file) from GEO database (accession numbers: GSM4647788 to GSM4647790).
- Shin et al., 2020: files were downloaded in mtx form (unprocessed filtered file) from GEO database (accession numbers: GSM3177991, GSM3177992, GSM4155928, GSM4155929).
- Shook et al., 2020: fastq files were downloaded from SRA database (accession numbers SRR10480641 and SRR10480643 to SRR10480646) and were processed using kallisto, embedded in loompy pipeline.
- Vorstandlechner et al., 2021: unprocessed filtered mtx files were provided under personal request.

5.2 DATASET METADATA

Due to the large volume of samples, we decided to apply a unified metadata scheme to all the datasets. This unified schema will help detect sample biases based on an array of factors: age, eth-

nicity, sex, library preparation method, etc. The accessibility of this table with metadata is available at Materials and Methods Chapter 7.

The metadata table includes the following columns: *Author*, *Year*, *DOI*, *Accession (General)*–if there is more than one sample in the dataset–, *Accession (Sample)*, *Accession (SRR)*–if data is downloaded in fastq form–, *Data origin*–e.g. FASTQ, count matrix in h5 format, mtx format, csv format or other formats–, *Aligner*, *Genome*, *Donor identifier*–if more than one sample comes from the same donor–, *Sample identifier*–usually obtained from the source of the dataset–, Internal sample identifier–may differ from the Sample Identifier if it is too large or misleading–, *Downloaded*–yes or no, if it is considered or not for the analysis–, *Analysed*, *Exclusion reason*, *Library preparation*, *Sequencer*, *Organism*, *Age*, *Age (mean)*–sometimes Age might be a range if not completely specified in the paper–, *Age format (y/m)*, *Gender*, *Race*, *Ethnicity*, *Sample location*, *Condition*, *Condition (other)*–additional information such as treatment, comorbidities, etc.–. A detailed description of each variable is available at Section 7.3 from Chapter 7.

When each dataset is downloaded and transformed into a compatible AnnData h5ad format, this information is added to the AnnData, so it remains throughout the processing pipeline. This allows having a unified standard for dataset processing and analysing larger numbers of datasets in the same pipeline.

5.3 DATASET PREPROCESSING

All datasets, regardless of organism, are processed using a common scheme:

1. Dataset loading, and common gene changing: depending on the mapping database, some relevant genes in the analysis with two names are merged into one. For example, CCN5/WISP2 is set to WISP2. Most of this items were transformed due to personal observations.
2. Basic QC metric calculation: `sc.pp.calculate_metrics` is used to create basic metrics such as the number of expressed genes per cell, the number of counts per cell, etc. Additionally, the percentage of mitochondrial counts is calculated.
3. Plotting QC metrics and thresholding: `pct_counts_mt` VS `total_counts` and `n_genes_by_counts` VS `total_counts` scatter plots are used to select the threshold of number of genes and percentage of mitochondrial counts. Generally, if the distribution is unimodal, up and down tails are removed, and if the distribution is multimodal, the top part is selected. If there is more than one sample in the dataset, these stats are plotted per sample and, depending on the differences between samples, either a common set of cutoffs is set, or different cutoffs are set per sample.
4. Basic transformations: the dataset is normalised, log1p-transformed, and genes with no counts are removed.
5. PCA, harmony integration, and kNN calculation: `sc.pp.pca` is run to calculate PCA with 30 variables. If the dataset contains different samples, harmony is used with `sce.pp.harmony_integrate` function. This function uses the PCA matrix as input, and outputs a modified matrix with the same dimensions as the PCA matrix. To calculate the kNN graph, `sc.pp.neighbors` function is used with cosine metric and the number of neighbors as $\sqrt{n_{cells}/2}$.

6. Major cell type assignment (human): UMAP–`sc.tl.umap`– and *leiden*–`sc.tl.leiden`– are run to see all major cell types. *Leiden* is run with a high resolution (3 to 10, depending on the dataset). This marker-to-population algorithm (section 5.4) is used to set different cell types in an unsupervised manner. The list of populations and markers is the following:
    - fibro: *LUM*, *PDGFRA*, *COL1A1*, *DCN*, *SFRP2*, *APOE*, *APOD*, *FN1*
    - fibro - ANGPTL7: *ANGPTL7*, *ENTPD2*, *ETV1*, *C2orf40*, *SCN7A*, *SOX8*
    - F: *B4GALT1*, *TMSB4X*, *PPP1CB*, *WTAP*, *PTPRS*, *CTNNB1*, *INSR*, *BICC1*, *CTNNB1*
    - melanocyte: *MLANA*, *PMEL*, *TRIM63*, *QPCT*, *PLP1*, *TYRP1*
    - neuro: *GPM6B*, *PLP1*, *S100B*, *SCN7A*, *NRXN1*, *GFRA3*, *MPZ*
    - secretory cell: *KRT7*, *KRT8*, *KRT18*, *KRT19*, *DCD*, *SCGB2A2*, *PPP1R1B*, *MUCL1*, *AZGP1*, *SCGB1D2*, *PDCD4*, *TSPAN8*
    - muscle: *TAGLN*, *DES*, *PCP4*, *ACTG2*, *CNN1*, *CSRP1*, *TPM1*, *SYNPO2*, *PRUNE2*, *SORBS1*, *P2RX1*
    - peri - CYCS: *TAGLN*, *CRISPLD2*, *CYCS*, *VDAC1*, *RHOB*, *SORBS2*, *PLEKHO1*, *CNN1*, *DNAJB9*, *CSRP2*
    - peri - RERGL: *TAGLN*, *ACTA2*, *CRISPLD2*, *RERGL*, *BCAM*, *ADIRF*, *NET1*, *ARPC1A*, *PLN*
    - peri - RGS5: *ACTA2*, *RGS5*, *ABCC9*, *HOPX*, *ARHGDIB*, *KCNJ8*, *FXYD6*
    - peri - ZFP36: *RGS16*, *NR2F2*, *TGFBI*, *CCL8*, *RERG*, *HOPX*
    - endo artery: *PLVAP*, *CLDN5*, *PECAM1*, *IGFBP3*, *SRGN*, *SEMA3G*, *RHOB*, *HEY1*
    - endo capillary: *PLVAP*, *CLDN5*, *PECAM1*, *SELE*, *SOCS3*, *CDKN1A*, *NFKBIA*, *DNAJB1*, *ATF3*
    - endo venule: *PLVAP*, *CLDN5*, *PECAM1*, *CYP1B1*, *CLU*, *PERP*, *VWF*, *IER3*, *TSC22D3*
    - lymph: *CCL21*, *LYVE1*, *CLDN5*, *TFF3*, *MMRN1*, *EFEMP1*, *FGL2*, *TFPI*, *MAF*
    - krt basal: *KRT14*, *COL17A1*, *KRT5*, *KRT15*, *DST*, *PDLIM1*
    - krt channel: *KRT23*, *GJB6*, *GJB2*, *CDA*, *MMP7*, *PNLIPRP3*
    - krt spinous: *KRT1*, *KRT10*, *DMKN*, *KRTDAP*, *CHP2*, *LYPD3*
    - krt gran: *FLG*, *NCCRP1*, *CNFN*, *TGM1*, *CST6*, *KLK7*
    - immune: *TPSB2*, *TPSAB1*, *HLA-DRA*, *FCER1G*, *CD74*
    - T CD4+: *CD52*, *CD3D*, *TRAC*, *TCF7*, *CD4*, *IL7R*, *CD40LG*
    - T CD8+: *CD52*, *CD3D*, *TRAC*, *CD8B*, *THEMIS*, *CD8A*, *FOXP3*, *CCR4*, *RORC*, *TIGIT*
    - plasma cell: *SDC1*, *SLAMF7*, *TNFRSF17*, *PTPRC*, *CXCR4*, *MYH9*, *PRDM1*, *CD38*, *CD27*, *IGHG1*
    - dendritic cell: *GZMB*, *MRC1*, *XCR1*, *CLEC9A*, *IRF8*, *EPCAM*, *CD1B*, *STMN1*, *IDO1*
    - APC: *HLA-DQA1*, *HLA-DRB6*, *TYROBP*, *FCER1G*, *AIF1*
    - mast cell: *IL1RL1*, *CPA3*, *HPGDS*, *TPSB2*, *HPGD*, *RGS13*, *CTSG*, *TPSAB1*, *GATA2*
    - NK cell: *NCAM1*, *XCL1*, *CD38*, *CD7*, *IL18R1*, *KLRF1*, *KLRK1*
    - mt: *MTND2P28*, *MTND4P12*, *MTCO1P40*, *ADAM33*, *RN7SL2*, *MTRNR2L6*,

- eritro: *HBB*, *HBA2*, *HBA1*, *HBD*

Most of these signatures are based on common knowledge, or extracted from several literature sources. For instance the classification of endothelial cells is based on Li et al., 2021, and the classification of keratinocytes is based on Cheng et al., 2018c. Lastly, perivascular cell classification is based on a personal classification used in a secondary unpublished project not included in this thesis, where perivascular cells could be separated into these four major types.

Regarding the characterisation of fibroblasts, *LUM*, *PDGFRA*, *COL1A1*, *DCN* genes are used to find this population across different single-cell studies. Generally, *PDGFRA*$^{+}$ cells that express the other markers are fibroblasts cells; and cells not expressing *PDGFRA* and expressing other markers are "similar" cell types (perivascular cells, endothelial cells, etc.). ANGPTL7 and F fibroblasts are described in Ascensión et al., 2021 (C_4 population) and in Vorstandlechner et al., 2020.

Once the classification is done, cell types within the fibroblast range are selected. This step is not 100% accurate though, and there might be cases of fibroblasts marked as other types (commonly neuro or melanocyte) and vice-versa. To check that the cell types selected are correct, we make use of the expression patterns of fibroblast markers, mainly *PDGFRA*. Once fibroblasts are selected, a secondary AnnData where further processing will be done is created.

7. Major cell type assignment (mouse):

   - peri: *Rgs5*, *Myl9*, *Ndufa4l2*, *Nrip2*, *Mylk*, *Rgs4*, *Acta2*, *Sncg*, *Tagln*, *Des*, *Ptp4a3*, *Myh11*
   - endo: *Pecam1*, *Cdh5*, *Egfl7*, *Cd36*, *Srgn*, *Adgrf5*, *Ptprb*, *Scarb1*, *Plvap*, *Grrp1*, *C1qtnf9*, *Mmrn2*, *Flt1*
   - kerato: *Krt14*, *Krt15*, *Perp*, *S100a14*, *Ccl27a*, *Gata3*, *Dapl1*, *Rab25*, *Ckmt1*, *Col17a1*, *Serpinb5*
   - kerato *Gjb2*: *Ucp2*, *Krt71*, *Gjb2*, *Ahcy*, *Acaa2*, *Cbs*, *Slc3a2*, *Serpina11*, *Lap3*, *Gss*, *Basp1*
   - fibro: *Dcn*, *Pdgfra*, *Lum*, *Col1a1*, *Col1a2*
   - fibro_2: *Ncam1*, *Ptch1*, *Trps1*, *Col11a1*, *Wif1*
   - fibro_acan: *Acan*, *Col2a1*, *Col11a1*, *Col9a1*, *Snorc*, *Col9a3*, *Mia*, *Cnmd*, *Ucma*, *Chad*
   - T cell: *Rac2*, *Ptprcap*, *Il2rg*, *Cd3g*, *Skap1*, *Hcst*, *Ctsw*, *Ets1*, *Cd3d*, *Ctla2a*, *Cd2*
   - APC: *Tyrobp*, *Cd74*, *H2-Aa*, *H2-Eb1*, *Ctss*, *Spi1*, *Napsa*, *Cd68*, *Lyz2*, *Csf2ra*
   - lymph: *Ccl21a*, *Egfl7*, *Mmrn1*, *Nsg1*, *Meox1*, *Gimap6*, *Kdr*
   - melano / schwann: *Syngr1*, *Pmel*, *Mlana*
   - myo: *Tnnt1*, *Tnnt2*, *Tnnt3*, *Tnnc2*, *Acta1*, *Myl1*, *Tnni2*, *Tcap*, *Eno3*, *Myoz1*
   - neural: *Itgb8*, *Plp1*, *Ptn*, *Egfl8*, *Chl1*, *Cadm4*, *Sox10*, *Cdh19*, *Snca*

   Similar to the labelling in human, the genes for each of the labels come from different sources, or as DEGs performed on the datasets to find a common set of genes that would be applicable to label these populations. In fibroblast populations, there are a set of peculiarities to be mentioned. "fibro_2" label refers to fibroblast axis w populations; and "fibro_acan" label refers to chondrogenic fibroblasts. None of the 9 mouse datasets contain these fibroblasts, but are included in the list regardless.

8. Fibroblast processing: To obtain a complete picture of all fibroblast subpopulations, secondary processing of the fibroblast datasets is performed. Many genes used for manifold learning in the general dataset (e.g. endothelial, neural or APC-related markers) will not be necessary for fibroblasts and, therefore, a secondary analysis of the subset of fibroblasts will allow finding subtler differences within its core transcriptome and, therefore, obtain a more accurate and complete picture of the dataset structure and, therefore, fibroblast subpopulations.
9. PCA, kNN and FS on fibroblasts: to do the subprocessing, genes with no expression on fibroblasts are filtered, and PCA and kNN are run as before. *triku* is used for FS, with default settings. The output from this method is a boolean array indicating if each gene is selected or not as highly variable. With this information, PCA, harmony and kNN calculation are rerun to produce the definitive set of neighbours for the following steps.

   Of note, FS is not run on the main dataset with all cell types because we are not doing a thorough analysis of major cell types, since fibroblast populations are easily separable from the rest of populations. Therefore, avoiding this step saves time frrom the initial processing, specially with big datasets.
10. Analysing fibroblast subpopulations: to obtain the different subpopulations, UMAP and *leiden* are run again, with `min_dist` and resolution values adjusted for each dataset. The goal is to obtain a high number of clusters to run the marker-to-population algorithm on them. Once populations are obtained, if there is no need to reanalyse populations or investigate new arising populations, the processing is concluded. The list of populations and markers for the detection of mouse and human subpopulations, at the time of writing this thesis, is described in section 10.1 of results.
11. Additional step - detecting new populations: sometimes the marker-to-population algorithm assigns the same label to more than two clusters, or assigns some cluster as unassigned. In these cases, it is possible that a new cluster is being discovered. To check that, DEGs of that population are calculated, and the marker-to-population algorithm is used with a set of markers. DEGs should be exclusive of the population and biologically relevant–ribosomal, IER, or basic metabolism genes should be excluded–. If the algorithm assigns the cluster correctly, it is a putative population. To assert that this putative new population is verifiable, it has to be assigned in other datasets. If not, it will not be considered as a sufficiently relevant population. In this scenario, it is possible that clusters assigned to a population are assigned to this new population for being a more likely target.
12. Additional step - removing unwanted populations: In some cases, when assigning fibroblast subpopulations, unwanted populations appear–stressed cells, non-fibroblast cells, etc.–. In these cases, these cells are removed from the AnnData, and the whole processing of fibroblast cells is redone.

5.4 MARKER-TO-POPULATION ALGORITHM

The aim of this algorithm is to, given a dictionary of populations and their respective markers, generate a mapping between the clusters of the dataset and the populations. This correspondence, schematically shown in Figure 25 is not univocal: it is possible to assign one population to several

clusters, but it is not possible to assign several populations to one cluster. Additionally, it is possible to obtain clusters that are not assigned, that is, there is no population whose markers are not sufficiently representative of that cluster. These populations are usually relevant at the initial iterations, because they might be novel populations that were unnoticed in previous datasets.

The algorithm follows these steps:

- For each category, a matrix of shape the number of cells in the AnnData by the number of markers of that category, M_0, is created.

  For each gene in M_0, the gene expression array is multiplied by the neighbour matrix, to produce a matrix similar to the kNN count matrix described in section 5.4.1. In this case, this would be the expression of the gene in each cell and in its neighbours. The aim of this product is to reinforce the expression of local genes. This kNN matrix is divided by the number of neighbours, and the result is stored in the corresponding column of M_0.

- Once M_0 has been completed, the mean is extracted, column-wise. This produces a column matrix with the mean kNN values across genes, m_1. We create the matrix M_1, resulting of concatenating m_1 matrices with all populations. Therefore, the M_1 shape is the number of cells in the AnnData by the number of populations.

- The cluster labelling from *leiden* is added to M_1, and this information is collapsed into a M_2 matrix, whose dimension is the number of clusters by the number of populations. To collapse the information of several cells from one cluster in M_1 into one number in M_2, either the percentile defined in the `quantile_gene_sel` argument, the CV, or the maximum value are computed.

- For the M_2 matrix with the selected collapsing function, the assigned population is the one with the highest value across clusters.

It is possible to select more than one population as the best for each cluster. To do that, `intermediate_states` argument is set to `True` and the populations with a smaller difference to the best population than the one defined in the argument `diff` are selected.

If none of the populations achieves a minimum value established in `min_score` the population assigned to the cluster is undefined.

5.4.1 *Calculation of kNN count distribution per gene*

Considering a scRNAseq count matrix of c cells by g genes, and given a number k of neighbours per cell (half the square root of the number of cells by default), first a k-Nearest Neighbour (KNN) matrix was created using the kNN graph from a previous step in the processing before FS. The kNN matrix contains, for each cell in the matrix, the indices of its k most similar cells.

Once the kNN matrix was calculated, the following procedure was applied for each gene. For each cell with positive read counts (read counts > 0), the number of counts of that cell and the number of counts of neighbour cells were summed. For each gene, this summarises a distribution of the counts in the kNN cells. This is the kNN count distribution. The process for the kNN count distribution construction is schematised in Figure 26.

5.5 POPULATION-TO-MARKER ALGORITHM

The aim of this algorithm is to, given a set of datasets with labelled populations, produce a list of markers that are commonly expressed in all datasets for each population. This algorithm arises because of the scenario of certain genes being markers of a specific dataset, and not in others, due to specific characteristics derived from each dataset. In order to avoid these types of genes, this algorithm achieves robustly-expressed sets of markers across all datasets instead. A scheme of the algorithm is shown in Figure 27.

For each of the populations, the algorithm follows these steps:

- First, a common set of markers across all datasets is extracted. For that, the DEGs of that population in each dataset are obtained, following a threshold–by default, p-value < 0.01–. Then, genes that follow unwanted patterns–ribosomal genes, mitochondrial genes, *MALAT1*, *S100* genes, etc.– are excluded. The non-excluded set is the set of genes that will be used for evaluation.
- For each gene and each dataset, two values are stored: the scoring value used in the DEG calculation–by default, the `score` value from `sc.tl.rank_genes` function–, and the sum of the expression of that gene in the cells within that population. This will produce two matrices of shape the number of genes from the first step by the number of datasets that will store these values.
- For each of these two matrices (M)–the one based on the DEGs, and the one based on the expression–, a mean value is computed across datasets, so that we have a value per gene. This mean is pondered (μ^*) using the sigmoid function of the value, so that expression values of datasets with greater quality–overall, genes with more localised expression and, thus, higher scores–are dampened and thus the selected markers are not completely biased towards one or two datasets:

  $$\mu^* = \frac{\sum_{dataset} a_{dataset} \cdot w_{dataset}}{\sum_{dataset} w_{dataset}} \quad , \quad w_{dataset} = 2\left(\frac{1}{1+e^{-\frac{\sum M_{dataset}}{\sum M}}} - 0.5\right)$$

  Where $a_{dataset}$ is the scoring value is in the DEG calculation, and $w_{dataset}$ is the weight computed for the pondered mean, which is based on $a_{dataset}$.
- Once mean values for genes with score based on DEGs ($\vec{\mu}^*_{DEG}$) and based on the expression ($\vec{\mu}^*_{expr}$) are computed, a score value is obtained by applying the following expression:

  $$Z = \frac{\vec{\mu}^*_{DEG}}{(\vec{\mu}^*_{expr})^{\frac{1}{4}}}$$

  That is, mean score values are divided by mean expression values to avoid selecting for genes that have higher thorough expression–which makes the DEG score higher–but which are not specifically expressed in the population of interest. Finally, in order to avoid genes selected as DEGs but with a extremely small expression, only Z values with $\mu^*_{expr} \geqslant 0.035$ are chosen. This allows, therefore, to choose from genes with middle expression values.

In the end, the population-to-marker algorithm produces a dictionary with the representative populations and a data frame of the markers–with scores and Z–.

5.6 PIPELINE TO OBTAIN A ROBUST SET OF MARKERS AND POPULATIONS

The pipeline for robust markers and populations consists of a combination of the marker-to-population and population-to-marker algorithms that tries to minimise the bias of supervised selection. This pipeline is not unsupervised, since the initial set of markers has to be manually selected and results in intermediate states can be tuned to suit the researcher's biases and thoughts about existing configurations. Nonetheless, the use of this pipeline produces reproducible and relevant results with little effort. Additionally, although individual external algorithms for the marker-to-population and population-to-marker may be found (Domínguez Conde et al., 2022; Guo et al., 2021; Zhang et al., 2019a), we were not completely satisfied with them and decided to program it from scratch to have a control on the code and the parameters. Additionally, some of the methods were developed during the development of this pipeline, and were not considered (Domínguez Conde et al., 2022; Guo et al., 2021).

The pipeline, schematised in Figure 28, consists of three steps:

1. Given a set of handpicked populations and genes, use the marker-to-population algorithm to assign clusters from each dataset their most probable population.
    - If one or more unassigned populations arise, gene markers are plotted to see if the cluster belongs *de visu* to one of the populations. If so, the threshold parameter has to be tuned.
    - If no markers are expressed in unassigned populations, DEGs from that population that are representative markers can be chosen and, if other clusters in other datasets follow the same pattern, the new population and its markers can be added to the dictionary.
    - Results of the whole pipeline are dependent on the defined populations and markers. Therefore, careful planning to create that dictionary has to be taken. One method to craft this dictionary is the one described in Materials and Methods section 4.8, where a mapping between putative clusters across datasets is performed to obtain a "consensus" graph of populations.
2. From the set of populations, applying the population-to-marker algorithm yields a "robust" set of markers for each of the populations across datasets. This set is less biased towards a specific dataset than the first set of markers used for the first market-to-population algorithm. Nonetheless, it should be noted that a change of the initial conditions of the pipeline will considerably affect the set of markers obtained after this algorithm.
3. Finally, with the "robust" set of markers we run the marker-to-population algorithm again to produce a "robust" set of populations.

5.7 UNSUPERVISED ASSIGNATION OF POPULATIONS TO AXES

The aim of this method is to confirm that a cluster belongs to a specific axis. If the method assigns a cluster to a different axis than the one expected, or to more than one axis consistently across datasets, we should consider moving that population into the corresponding axis or to a new axis, depending on the pattern of axis assignation.

This method uses the populations matched with the marker-to-population algorithm to create the axis. First, a manual set of axes is created by setting each population to the axis to which it originally belongs–i.e. populations A2 and A3 are assigned to axis A, and populations B1 and B3 are assigned to axis B–. If a population is a bridge between two axes–e.g. a/d, b/c, c/d axes on mouse–it is assigned as an independent axis. With this automatic "manual" assignation, the population-to-marker algorithm is run to obtain a dictionary of axes and markers of each axis.

Finally, the marker-to-population algorithm is run against the unsurpervisedly-assigned populations using also the marker-to-population algorithm, so that each population is mapped to an axis. Contrary to the populations, the marker-to-population algorithm is run with the `diff` parameter set to 0.05, so that populations that have a mixed transcriptome between two or more axes are not univocally assigned to one cluster. This step is crucial to see if specific populations have to be reallocated to an existing or a new axis if necessary.

5.8 PAGA AND POPULATION-RELATIONSHIP GRAPHS

One important step in the secondary analysis is to detect relationships between individual populations. This is important because fibroblasts, like any other cell type, are not independent entities, but rather interact within the tissue, either with other fibroblasts or other cell types. Therefore, fibroblast subpopulations, as well as independent axes will likely interact with each other, or share more similarities between them.

To analyse the similarities between fibroblast populations, PAGA algorithm is used (Wolf et al., 2019). The objective of the algorithm is to extend the idea behind the neighbouring concept that is applied cell-wise to clusters. Using the kNN information computed during the processing steps, and for a given set of clusters, PAGA produces an adjacency matrix indicating relationships between each cluster. With this information a graph between clusters is created where, the greater the connection between two nodes, the greater the similarity between two clusters.

To simplify the relationships within the PAGA graph a secondary tree graph is constructed, which explains the major relationships, and forces interactions to be in that tree structure. This graph comes with the limitation of "breaking" cliques or looped cluster relationships, but allows for a clearer lecture of an otherwise sometimes complex graph. Additionally, this graph tends to preserve relationships that are clearly meaningful.

Thus, for each dataset these two PAGA graphs are computed. This allows to analyse the heterogeneity of populations and interactions between populations across datasets.

In order to integrate all the information, a joined graph is built. To build this graph, to the PAGA adjacency matrix from the tree graph of each dataset, we remove interactions with values smaller than a threshold–0.6 by default–, and then compute a power of that number–between 1 and 2, depending on the graph–. The first step removes loose relationships that tend to saturate the joint graph, and the second step exacerbates differences between clearly meaningful cluster relationships and less-relevant relationships, making them stand out in the figures.

5.9 CONFIRMING THE EXPRESSION OF GENES IN FIBROBLAST SUBPOPULATIONS AND GENERAL CELL TYPES

After running the robust population and marker pipelines, we produced a table of relevant information for each gene so that it can be useful to determine the functions of each of the populations based on their gene signature. To construct this table, we need to assign the populations in which that gene is expressed first.

The chosen plots to this step are a combined dot plot and UMAP plot that show the expression pattern of each gene across datasets–the dotplots are available at Figures 43 and 49, and the UMAPs are not included within this thesis–. The size of the dot in the dot plot represents the proportion of cells within each subpopulation that expresses that gene, and the colour represents the level of expression. To better visualise the proportion, the value that is represented in the dot plot is scaled as $x := x^{\frac{2}{3}}$, which allows a better discrimination of genes expressed in fewer cells. The aim of this dot plot is to have an overall view of the expression patterns of each gene, especially when the pattern of expression is diffuse or complex.

As a complement to the dot plot, the UMAP plot with the expression of a gene in all dataset supports the findings from the dot plot, and allows to find differences of expressions across datasets.

All this information will be used to create the table with gene information, in which we will include the following information:

- Gene name
- DEG-based location: to create this column, genes appearing as the first 30-50 markers in the dictionary obtained from the pipeline described above are included in this column. If a gene is a marker of more than one population, this information is reflected. For instance, a gene with the label "A1 | B2 | C1" is a marker or A1, B2 and C1 populations at the same time, whereas a gene with the label "D1" is only a marker of the D1 population.
- UMAP-based fibroblast location: this column represents the pattern of expression of each gene within different fibroblast populations. If two populations share the same level of expression, the symbol $\sim$ is used; whereas if one population has a greater, or clearer, expression than the other, the symbol $>$ is used. For instance "A1 $\sim$ B2 > C1 $\sim$ B4" indicates that the gene is more expressed in A1 and B2 than in C1 and B4; but it is expressed at a similar level in A1 and B2, and in C1 and B4. If there is a thorough expression of the gene, but it is apparent in one or more populations, "rest" is added to the comparison. For instance, "A1 > B2 $\sim$ B3 > rest" indicates that the gene is clearly expressed at A1, then at B2 and B3, and there is a minor, basal expression in the rest of the cells.
- UMAP-based general location: similar to the previous column, a combination of dot plot- and UMAP-based visualisation are used to determine the expression patterns of each gene within the major cell types of the skin.
- Location: location of the gene product within the cell (extracellular, endoplasmic reticulum, plasma membrane, Golgi, nucleus, etc.). This information is obtained from GeneCards website (Safran et al., 2021).

- Basic gene information: in order to have a quick view of the putative functions of the gene, summaries from Entrez gene, GeneCards and Uniprot are included. All the summaries are joined, and extracted from, the GeneCards website (Safran et al., 2021).

It is important to remark that the expressions used in the UMAP-based fibroblast and general location columns are arbitrarily depicted based on the dot plot and UMAP plots. Unfortunately, this level of annotation across such heterogeneous datasets requires a subjective and arbitrary classification. Nonetheless, this notation is really helpful to have a quick view of the gene expression patterns without constantly needing to resort to visualisation methods. The list of cell types used is the one described in section 5.3 of Materials and Methods.

5.10 LIGAND-RECEPTOR ANALYSIS ON HUMAN FIBROBLASTS

Ligand-receptor (LR) analysis is a useful method to detect putative interactions between cells, and analyse underlying pathways and interaction networks. To perform the LR analysis we used CellPhoneDB v2.0 (Efremova et al., 2020), a commonly used LR analysis library. We enhanced the LR pair database using a secondary database with approximately 3400 human and mouse LR pairs, CellTalkDB (Shao et al., 2020). We used `cellphonedb database generate` command to create a common database of LR pairs from CellPhoneDB and CellTalkDB, and followed the steps as depicted in the "Generating user-specific custom database" section from their Github repository.

For this thesis, due to the workload demanding from this job, only interactions in human datasets are considered. Interactions in mouse datasets will be considered for future works.

Once CellPhoneDB is run (p-value threshold: 0.1), the output is a table where each entry contains the gene pairs and the interacting cell types, as well as the p-value. Since this analysis is run on each dataset, the final table also contains the dataset where this interaction occurs. From this table, two tables are created with the same filtering procedure: one with fibroblast-fibroblast interactions, and one with interactions between fibroblast and other cells. The last one is not used for the purposes of this thesis due to the complexity of the analysis.

The filtering procedure involves the following steps:

1. Filtering of reciprocal pairs: if pair AB and pair BA occur, their values are merged to avoid problems in further steps.
2. Same pairs from different datasets are merged and counted.
3. Dataset filter: pairs appearing in fewer datasets are removed. This is to avoid spurious pairs occurring in one or two datasets. The threshold applied is of a 30%. For instance, interaction of CXCL12 in B1 with WNT5A in C3 occurs in 4 datasets. B1 appears in 12 datasets and C3 in 8. 30% of 8 is 2.4, and since the interaction is present in 4, it is kept. Had the interaction occurred in 1 or two datasets, and it would have been discarded.
4. Gene filter: genes that are not fibroblast markers are removed. To do that, we need to first select a wide range of gene markers. For each population, from the distribution of gene rank VS Z value–from population-to-marker algorithm in section 5.5–the knee point is calculated and 6-times the number of markers given by the knee point are selected–with a minimum

number of 70 and a maximum of 350, regardless of the knee point location–. Then, any pair were one of the genes is not a marker of that population is discarded.

From these filtering steps we obtain a curated list of LR pairs. However, considering that many genes are markers of more than one population, two scenarios tend to arise–shown with examples–:

- A gene from a LR pair is shown to occur between A1 and C3. However, the marker in A1 is also a marker in A2, but the interaction between A2 and C3 of that LR pair is not recorded. We include that information in that table.
- Gene A and gene B are interacting between A1 and C3, A2 and C3, and A1 and C1. These 3 entries are simplified to genes A and B interacting between A1/A2 and C1/C3.

From these scenarios we observe that, given that many genes are markers of different populations, we cannot omit interactions that CellPhoneDB has not implicitly considered. Additionally, to avoid creating copies of the sample LR pairs in different populations, it is convenient to store each LR pair with the possible populations that can interact between them.

Therefore, the last step to build the table of LR pairs is to remove duplicate LR pairs, and condense them using the notation for marker expression patterns in section 5.9 of Materials and Methods. The table availability is disclosed in Section 7–Human L-R pairs (fb-fb) tab–, where the LR pair genes are included together with the populations they are expressed in. There are two additional columns–Clusters A/B (robust)–indicating the populations where these markers are more relevantly expressed.

5.11 COMPARISON OF HUMAN AND MOUSE FIBROBLAST POPULATIONS

In this section we are going to extend on methods used to compare mouse and human populations. To do that, we need to first apply an homology step between human and mouse genes. To do that, a human-mouse dictionary was created using a pan-organism homology table downloaded from this link (https://tinyurl.com/fc6hrv8m).

5.11.1 *Using batch effect correction methods to join human and mouse fibroblast from the same dataset*

At the time of writing this thesis, two analyses were performed with samples analysis of human and mouse dermis at the same time: Boothby et al. (2021) (Boothby et al., 2021) and Vorstandlechner et al. (2021) (Vorstandlechner et al., 2021).

To integrate the human and mouse samples from each dataset, *harmony* batch effect correction method was used, with default settings. Additionally, *bbknn* was also used but produced similar results.

5.11.2 *Comparing human and mouse markers to infer population similarities*

Another method to compare mouse and human populations is to compare their markers. By observing the overlap of markers for all pairs of populations, and analysing the sets of overlapped genes we

can infer likely comparisons between human and mouse populations. This method is also applied, as a control measurement, within human and mouse populations, as described in section 10.4.2.

The first step to do the comparisons is to select the number of markers from each population that are going to be compared. To do that we applied the following algorithm:

- For each pair of populations, given a number of markers N, calculate the Jaccard index between the first N genes in one population and N genes in the other. All Jaccard index values are stored in a matrix of indexes, M_N.
- Compute the maximum trace of M_N. To do so, we use the Hungarian method from the *munkres* package (Munkres, 1957). This algorithm finds the permutations of a matrix to satisfy that its trace will have the minimum or maximum value. Therefore, obtaining the trace of M_N after applying the Hungarian method will provide a value of wellness of overlap between the most relevant pairs of populations. This trace value is stored alongside the N value. The higher the trace, the more general overlap between the sets of markers across different populations occurs.
- The previous two steps are applied with a range of N values (10 to 300). To select the best value, we plotted N values and their respective traces and analysed the trend line. Generally, there is a knee point where trace values plateau. This was the selected N for the analysis.

Once the best N was selected, the matrix M_N was plotted and correspondences were found. In human-human and mouse-mouse comparisons, a superficial analysis was done to assert that comparisons are logical, and in line with the combined PAGA graph explained in the Materials and Methods section 5.8 and Results sections 10.2 and 10.3. In human-mouse comparisons, once the best N was selected and the M_N matrix was produced, best-matching comparisons were manually selected. Afterwards, from the list of all markers for both organisms, relevant markers were selected as follows: markers that were not relevant in humans nor mice, markers relevant in humans, markers relevant in mice, and markers relevant in humans and mice. A marker was chosen to be relevant if it was exclusive of that population, as shown in the "~ and >" notation from section 5.9 of Materials and Methods. This step is important because there are many overlapping markers that are not relevant in humans and mice, so they could be assumed as false positives. This does not mean that an overlapping marker that is not relevant is mouse and human is not expressed in those populations; rather it implies that is not sufficiently important to be considered in downstream analyses, for being too vague in its expression pattern.

5.12 ALLOCATION OF PAPILLARY AND RETICULAR SIGNATURES IN HUMAN FIBROBLAST POPULATIONS

In this section, we are going to describe different approaches used to assign human fibroblast populations to either papillary o reticular signatures, based on markers extracted from the literature. To do that, we consulted different bibliographic entries: Janson et al., 2012, Nauroy et al., 2017, Philippeos et al., 2018, Korosec et al., 2019, Haydont et al., 2019, Haydont et al., 2020. Each of those entries reflects a series of markers belonging to the papillary or reticular dermis.

Before the allocation of the populations was performed, genes from the literature were evaluated to see if they were markers of specific fibroblast populations or not, so that genes that are not population markers were not considered.

The first approach to the analysis was to create a dictionary of papillary and reticular genes. Using this dictionary, we called the marker-to-population algorithm from the Materials and Methods section 5.4 to assign each of the fibroblast populations the category of papillary and reticular, in all datasets. Populations that had a higher difference in papillary VS reticular were possibly in a papillary location and vice-versa.

The list of markers used is the following:

- Papillary: *CADM1*, *DIRAS3*, *TOX*, *ADRA2A*, *APCDD1*, *AXIN2*, *C8orf22*, *CCL14*, *CCL2*, *CCL5*, *CCL8*, *ACKR4*, *CD109*, *DPP4*, *CLEC2A*, *COL10A1*, *COL18A1*, *COL23A1*, *COL6A5*, *COL7A1*, *CTSC*, *CTSK*, *CTSS*, *CXCL1*, *DCN*, *FAP*, *FGF13*, *HSPB3*, *IL15*, *LOXL3*, *LRIG*, *MAF*, *MOXD1*, *NPTX2*, *NTN1*, *PDGFC*, *PLXNC1*, *PTGDS*, *PTGS1*, *PTK7*, *ROBO2*, *RSPO1*, *SFRP2*, *SGCA*, *SGCG*, *SPON1*, *SRPX2*, *STEAP1*, *TFAP2C*, *TNFRSF19*, *WIF1*, *WNT5A*.
- Reticular: *ACAN*, *COL11A1*, *FGF9*, *LIMCH1*, *MGST1*, *SOX11*, *VCAM1*, *A2M*, *ACTA2*, *ADAMTSL1*, *ANGPTL1*, *MCAM*, *THY1*, *CNN1*, *COL14A1*, *COMP*, *DACT1*, *EFEMP1*, *ELN*, *FBLN2*, *FGF7*, *FMO1*, *FNDC1*, *IGF1*, *KRT19*, *MFAP5*, *MGP*, *PCOLCE2*, *PCSK5*, *PDGFD*, *PPARG*, *PPP1R14A*, *SFRP4*, *SULF1*, *TAGLN*, *THBS2*, *TPM1*.

The second approach is to create a table of correspondence between the markers and the populations. To do that, we used the "~ and >" notation of the markers to assign a score of 1, 0.5 or 0.25 to each marker-population combination. For instance, if the expression for one marker is "A1 ~ A3 > B2", then A1 and A3 receive 1 point, and B2 receives 0.5 points. If the expression pattern of that marker is unclear for a population, it receives 0.25 points.

5.13 GO TERM ANALYSIS

To perform GOEA, we used the python API of the *enrichr* module (Kuleshov et al., 2016), specifically the `gseapy.enrichr()` function, and selected 6 sets of gene ontologies: *GO_Biological_Process_2021*, *GO_Cellular_Component_2021*, *GO_Molecular_Function_2021*, *Reactome_2022*, *WikiPathway_2021_Human*, *KEGG_2021_Human*. For each human fibroblast population, we selected the most relevant markers and run the GOEA, selecting $p < 0.03$ as a threshold. The markers were selected manually based on the "~ and <" notation. Each population contains between 29 and 74 terms–μ=50.33, σ=11.65–. The full list of statistically significant GO terms is available in The table disclosed in Section 7–Gene Ontology (Human) tab–.

5.14 ROBUSTNESS OF SEMI-SUPERVISED ALGORITHM

To check the robustness of the marker-to-population algorithm we applied a methodology similar to the classical jackknife resampling method (Quenouille, 1956; Tuckey, 1958). For a fixed number of iterations–30 for this analysis–, and for each dataset, we sampled a number of the cells without repetition and rerun the marker-to-population algorithm. The sampling was stratified across populations so that all populations were constantly represented. Sampling percentages were set to 50%, 70%, 80%, 95% and 99%. For the results section 10.8 we are going to present the results for the 99% resampling, which is able to discriminate any lack of robustness.

Therefore, this process was run on each dataset 30 times, producing a table per dataset with as many rows as the number of cells of the dataset, and 30 columns. For each cell, the population to which that cell is assigned in each iteration is stored. In some iterations, and for some cells, a NaN value is stored because the marker-to-population algorithm was not run in that sample of cells.

For each cell, the quotient between the number of iterations where the assigned cluster is the same as the original, and the number of assignments performed calculated. This term is called *robustness score*. For instance, if in one cell, whose original population is A2, is assigned 20 times to A2, 7 times to A1, and is not sampled in 3 iterations, the robustness score is 20/27=0.74.

To make Figures 60, S5 and S6 from section 10.8, the tables of robustness score are used. For Figures S5 and S6 the distribution is plotted directly, and for Figure 60, for each population, the mean of the robustness score across the cells in each dataset is computed, and then the distribution of means per dataset is plotted.

To make Figures 61, S7 and S8 from section 10.8, a secondary metric derived from the aforementioned table is calculated. Instead of computing only the percentage of assignment to the original cluster, assignment to other populations is also computed. To do that, an adjacency matrix is created where the rows represent the list of original populations, and the columns represent the list of populations assigned for that dataset. This differentiation is made because there might be cases of assignment of a cell to a population that is not in the set of populations from that dataset. For instance, one dataset might have populations A1, A2, A3, B1, B2, and C1; and in some iterations some cells from A1 were assigned to A4. The adjacency table would have A1, A2, A3, B1 and B2 as rows, and A1, A2, A3, A4, B1 and B2 as columns.

To create the adjacency table, for each row-column combination, the quotient between (1) the number of instances of assignment to the column population when the original labelling is the row population and (2) the number of assignments of cells labelled with the row population is calculated. Table 3 shows an example of resampling with a dataset of 7 cells and two populations (A1 and A2), and Table 4 shows the derived adjacency table from it, in which some cells are also assigned to populations A3 and A4.

Figures S7 and S8 show the adjacency tables computed for each dataset. To create Figure 61, the median value of adjacency across datasets is computed, that is, for each row-column combination, the datasets where this combination occurs are retrieved, and the median value of array of values for this combination are computed.

5.15 TABLES AND FIGURES

Table 3: **Example of resampling for robustness calculation.** Original label represents the population each cell was assigned to without any resampling, and each of the iterations show the resampling labels. Some cells are empty because they were not sampled during that iteration.

Original label	Iter. 1	Iter. 2	Iter. 3	Iter. 4	Iter. 5
A1	A1		A1	A1	A3
A1	A1	A4	A4	A1	A1
A1	A1	A1	A4		
A2	A2	A2		A2	A3
A2	A3	A2	A2	A2	
A2	A2	A2		A2	A3
A2	A2		A3	A3	A2

Table 4: **Adjacency table derived from Table 3.** The values are shown in fractional form to ease the check of the counts.

	A1	A2	A3	A4
A1	8/12	0/12	1/12	3/12
A2	0/16	11/16	5/16	0/16

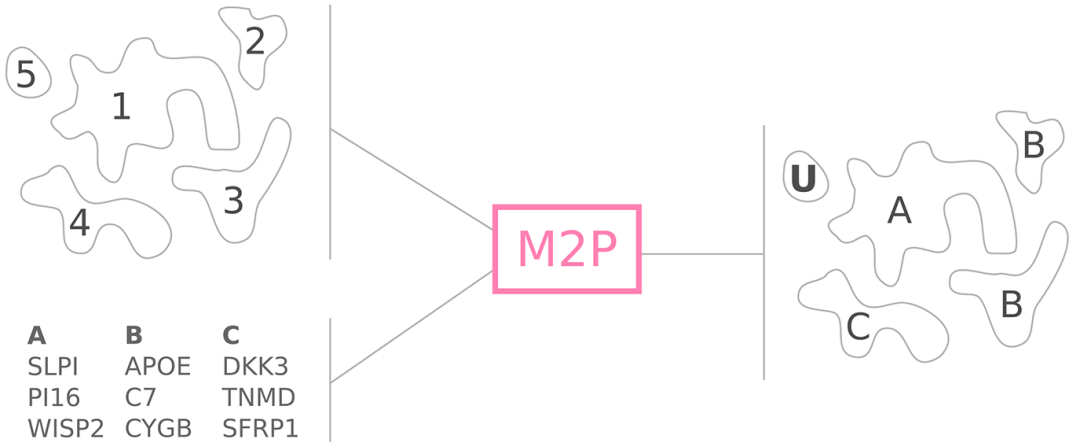


Figure 25: **Scheme of marker-to-population algorithm.** A dataset with 5 clusters, and a dictionary of three populations (A, B, C) are used in this algorithm, to assign the populations to the clusters.

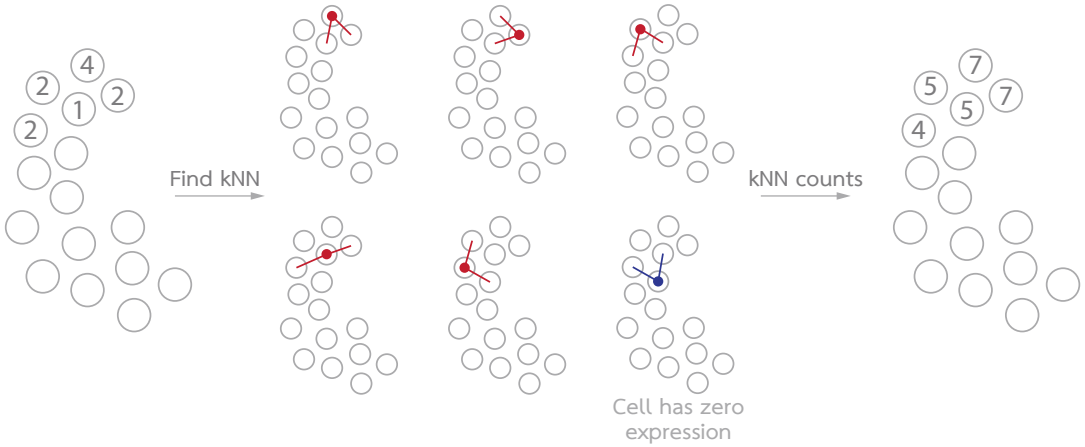


Figure 26: **Schema of construction of kNN count distribution.** In this example there are 5 cells with non-zero counts and 11 cells with zero counts. After the kNN graph is built ($k = 2$ in this dataset), for each non-zero count cell its two neighbours are considered, and the sum of counts from that cell plus its 2 neighbours is considered to build the kNN count distribution.

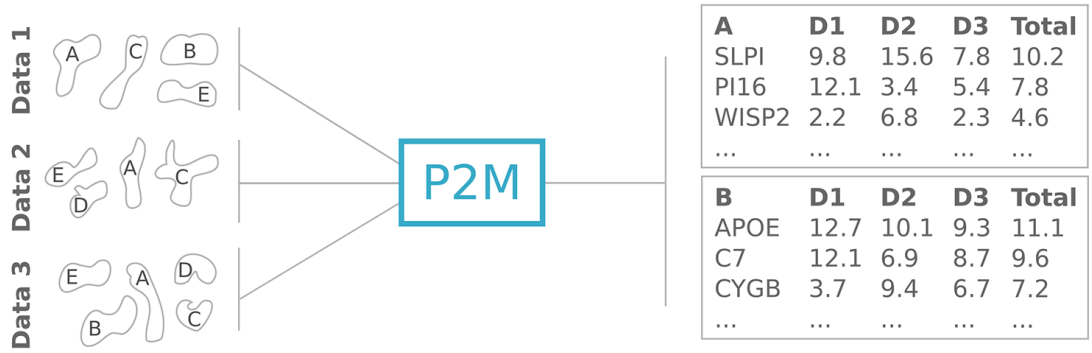


Figure 27: **Scheme of population-to-marker algorithm.** Three datasets with different populations (A - E) are used in the algorithm to output five tables of markers for each of the populations.

Figure 28: **Scheme of pipeline for robust sets of markers and populations.**

6

REANALYSIS OF REYNOLDS ET AL. DATASET

6.1 DATASET DOWNLOAD AND PREPROCESSING

Fibroblast sample data originated from five donors, as described by Reynolds et al. (Reynolds et al., 2021), and were processed from raw fastq files (E-MATB-8142). The ID numbers are 4820STDY7388991 [S1], 4820STDY7388999 [S2], 4820STDY7389007 [S3], SKN8104899 [S4], SKN8105197 [S5]. Fastq files were processed using the `loompy fromfq` pipeline. *Loompy* and *kallisto* (RRID:SCR_016582) versions are 3.0.6 and 0.46.0. Genome fasta index and annotations were based on GRCh38 Gencode v31. Additionally, for other annotations and analysis of other populations, the processed h5ad adata from Reynolds et al., 2021 was downloaded from the Zenodo repository (ID: 4536165).

6.2 PREPROCESSING OF FIBROBLAST SAMPLE DATA

Each individual sample (S1–C fibroblast types and subtypes that we had just descS5) data was processed equally using the following *scanpy* (v1.7.0rc1) (Wolf et al., 2018) procedure. To map the clusters from the original publication, cells from the processed data set were extracted and mapped to the samples. Genes with fewer than 30 counts were discarded. The sample was normalised (`sc.pp.normalize_per_cell`) and log-transformed. Then, PCA with 30 components was calculated and feature selection was performed with *triku* (Ascensión et al., 2022) (v1.3.1), and kNN with cosine metric were computed. Finally, UMAP (v0.4.6) (McInnes et al., 2018) and leiden (v0.8.3) (Traag et al., 2019) were applied to detect the fibroblast populations.

Most of the cells from the preprocessed adata were mapped to the raw dataset. However, additional unmapped cells appeared, some of them related to other cell types (e.g. keratinocytes, immune cells or perivascular cells). To assign unmapped cells to their corresponding cell types a population matching algorithm was applied, explained in section 5.4 of Materials and Methods. This algorithm requires a dictionary of cell types and markers. The markers used were the following:

- Fibroblast: *LUM*, *PDGFRA*, *COL1A1*, *SFRP2*, *CCL19*.
- Perivascular cell: *RGS5*, *MYL9*, *NDUFA4L2*.
- Erithrocyte: *HBB*, *HBA2*, *HBA1*.
- Immune cell: *TPSB2*, *TPSAB1*, *HLA-DRA*, *FCER1G*, *CD74*.

- Melanocyte: *PMEL*, *MLANA*.
- Endothelial vascular cell: *CLDN5*, *PECAM1*.
- Keratinocyte: *DMKN*, *KRT1*, *KRT5*.
- Mitochondrial content (low quality): *MTND2P8*, *MTND4P12*, *MTCO1P40*, *ADAM33*, *RN7SL2*, *MTRNR2L6*.

Once cell types have been assigned, non-fibroblast cells were discarded, and the PCA, *triku*, kNN, UMAP, *leiden* cycle was repeated to recalculate the new cell projection.

The sample S5 was discarded from the analysis due to its lack of *SFRP2* expression, a well established fibroblast marker that is expressed in the rest of samples (Ascensión et al., 2021).

Then, we separated the Fb2 population from the Fb1 and Fb3 populations for each dataset and applied the population matching algorithm to annotate them with the labels assigned from (Ascensión et al., 2021). The genes used for the population assignation were the following:

- A1: *PI16*, *QPCT*, *SLPI*, *CCN5*, *CPE*, *CTHRC1*, *MFAP5*, *PCOLCE2*, *SCARA5*, *TSPAN8*
- A2: *APCDD1*, *COL18A1*, *COMP*, *NKD2*, *F13A1*, *HSPB3*, *LEPR*, *TGFBI*
- B1: *CXCL2*, *MYC*, *C7*, *SPSB1*, *ITM2A*
- B2: *SOCS3*, *CCL19*, *CD74*, *RARRES2*, *CCDC146*, *IGFBP3*, *TNFSF13B*
- C: *CRABP1*, *PLXDC1*, *RSPO4*, *ASPN*, *F2R*, *POSTN*, *TNN*

Next, all datasets with Fb1 and Fb3, or Fb2 populations were joined. We applied the previous processing routine and, to correct for batch effects, we used *bbknn* (v1.4.0) (Polanski et al., 2020) with `metric=angular` and `neighbors_within_batch=2` parameters.

To analyse the transcriptomic profile between Fb1 and Fb3, and Fb2 populations, we joined the two datasets and applied the same processing pipeline as before. We first characterised the genes driving the differences by obtaining the DEGs between the two sets of populations, and running GOEA with the first 150 DEGs of each category. The set of ontologies used was *GO Biological Process 2018* with the module *gseapy* (v0.10.4) (Fang et al., 2021). Then, to assess that the differences were due to cellular stress in the Fb2 population, we downloaded several lists of genes, stored in the project GitHub repository (alexmascension/revisit_reynolds_fb/tree/master/papers_genes_bad_quality), and genes appearing in more than two lists were selected. This is further explained in section 11.2 of Results section. Then, the population matching algorithm was run against this list, and clusters with scores lower than 0.55 were assigned as "Non-stress" clusters.

To analyse the differences in transcriptomic profiles within Fb1 and Fb3 populations, we obtained the DEGs between the two sets of A2 populations, which were the easiest to separate in clusters. By using that list of DEGs, we applied the population matching algorithm and divided the Fb1 and Fb3 populations into two halves. We then obtained the DEGs between the two halves and ran GOEA with the first 150 DEGs of each category, which revealed a hypoxia pattern in one of the halves. To assess that the differences were due to hypoxia, we downloaded the lists of hypoxia-related genes, and genes appearing in more than two lists were selected. Since some key genes (some glycolysis genes, or important genes appearing in one list) were missing, they were manually added to obtain a more robust list. Then, the population matching algorithm was run against this list, as well as the list of stress-related genes, and clusters with scores lower than 0.5 were assigned as "Normal" clusters.

To replicate the analysis on the rest of the cell types, we used the processed h5ad file, and followed the same steps as with the fibroblast cell type.

6.3 CORRECTION OF STRESS AND HYPOXIA CELL STATES

In order to correct for stress and hypoxia cell states we used the `sc.pp.regress_out` implementation from *scanpy* on the stress and hypoxia scores. We first created two sub-datasets, one containing stress and normal cells, and another one with hypoxia and normal cells, and then the scores were regressed out. Finally, the common processing pipeline was applied. Additional correction methods tried to pursue the correction of hypoxia and stress are (1) creating the gene count matrix without the genes related to hypoxia or/and stress, (2) using *bbknn* and/or *harmony* to correct for the states, assuming they are batch-recoverable.

7

ANALYSIS REPRODUCIBILITY

When developing the projects within this thesis, all the software and analyses were conceived to be reproducible and open access for the scientific community. This section will introduce the steps applied to achieve this reproducibility.

7.1 MARKER-TO-POPULATION ALGORITHM

The source code of the marker-to-population algorithm described in Section 5.4 of Materials and Methods is available at the alexmascension/cell_asign GitHub repository.

7.2 SKIN PRIMARY ANALYSIS (HUMAN)

Skin fibroblast primary analysis–referred to in Materials and Methods chapter 4 and Results chapter 8–was developed using Jupyter notebooks, which are available at the alexmascension/fibroblast-population-detection GitHub repository (notebook 1H). Notebooks in HTML form, together with the processed `h5ad` files of the project are available at the Zenodo repository (doi: 10.5281/zenodo.4017653).

Additionally, a *CellxGene* viewer with the processed four datasets is available through fb-pop-detection.herokuapp.com Heroku application.

7.3 SKIN SECONDARY ANALYSIS

Skin fibroblast secondary analysis–referred to in Materials and Methods chapter 5 and Results chapter 10–; as well as the updated files from the primary mouse analysis were developed using Jupyter notebooks, which are available at the alexmascension/fibroblast-population-detection Github repository (notebooks 2 onwards, notebooks 1H and 1M for primary analyses). Notebooks in HTML form and the raw notebook ipybn files of the project are available at the Zenodo repository (doi: 10.5281/zenodo.7492966). Additionally, initial and processed .h5ad AnnData files, as well as the adata file for the combined datasets, can be found at the Zenodo repository for human datasets (doi: 10.5281/zenodo.7785089) and the Zenodo repository for mouse datasets (doi: 10.5281/zenodo.7785085).

The *CellxGene* viewer with the processed human and mouse datasets is available through the following link: cellxgene.cziscience.com/collections/3c4f0970-7614-43de-beb7-6128b3cb74ed (thank you Jennifer Yu-Sheng Chien for curating and preparing the datasets for the upload!). In each collection, a combined UMAP of all datasets is shown. The UMAP can display information about discrete categories such as age, author, and other categories described in the following paragraph. Expression of individual genes can be displayed, and DEGs between manually selected subpopulations can be performed. The expression values are based on the *processed* matrix counts. However, to unify value expression across datasets, mean-variance scaling from *scanpy* `sc.pp.scale()` function is performed. The arctan function is applied afterwards to remove outlier weight, whose range is [-1.5, 1.5]. This double-step transformation allows for the comparison of gene expression across all datasets.

The metadata file with the information of all human and mouse datasets is available at github.com/alexmascension/fibroblast-population-detection/blob/master/data/sample_metadata.csv GitHub repository, and the version at the time of writing this thesis is available at the Zenodo repository (doi: 10.5281/zenodo.7492966) together with the HTML and .ipybn files.

The metadata file contains the following entries:

- Author: Surname of the first author of the publication.
- Year: Registered year of last publication date.
- DOI: DOI of last registered publication.
- Accession (General): Accession number of all the samples.
- Accession (Sample): Accession number of each sample.
- Accession (SRR): If the file is a FASTQ, SRR number of the SRA archive, if existing.
- Data origin: Type of raw data–e.g. FASTQ, Raw count matrix–.
- Aligner: Alignment software used for preprocessing.
- Genome: Genome used during the alignment.
- Donor identifier: Identification number for the donor. If none, natural numbers are used.
- Sample identifier: Identification tag of the sample, as of the dataset repository.
- Internal sample identifier: Sample identifier used within the anndata file. Sometimes, the internal sample identifier differs from the sample identifier because the sample identifies is too long or missing.
- Downloaded: Raw data files were downloaded (boolean).
- Analysed: Raw data files were processed (boolean).
- Exclusion reason: If data were not downloaded or processed, indicate the reason if necessary.
- Library preparation: Library preparation method.
- Sequencer: Sequencer used after library preparation.
- Organism: mouse or human.
- Age: Individual age–in days, weeks, months or years depending on the organism–.

- Age (mean): Sometimes, due to a lack of information on individual ages, an age range is provided. For calculation measures, the mean age is computed.
- Age format (y/m): Format of the age stated before.
- Gender: Gender of the individual.
- Race: Race of individual.
- Ethnicity: Ethnicity of the individual.
- Sample location: Body location where the sample was extracted from.
- Condition: Condition of individual, either healthy/control, wounding, or any disease.
- Condition (other): Notes on the condition–e.g. severity of the disease, or days after wounding–.

Lastly, information pieces from the secondary analysis are available in a Google Docs spreadsheet: docs.google.com/spreadsheets/d/1lfI6sgjEyg37BGL7VRMfW7KgwGKwX5QrCtnKYk1DXY4–updatable version–, and at the Zenodo repository (doi: 10.5281/zenodo.7492966)–fixed version, last updated as of 2023/03/30–. These tables contain information about Gene Ontologies, human and mouse markers descriptions, and LR pairs. Some of these tables, such as "Specific marker function" were manually completed to produce certain tables from section 12 and the main text from that section.

7.4 REYNOLDS ET AL. DATASET REANALYSIS

Reynolds et al. dataset reanalysis–referred to in Materials and Methods chapter 6 and Results chapter 11–was developed using Jupyter notebooks, which are available at the github.com/alexmascension/revisit_reynolds_fb GitHub repository. Notebooks in HTML form, together with the processed `h5ad` files of the project are available at the Zenodo repository (doi: 10.5281/zenodo.4596374).

Part IV

RESULTS

8

PRIMARY OF FIBROBLAST HETEROGENEITY IN HUMAN SKIN USING SCRNASEQ

8.1 INDEPENDENT DATASETS SHOW A COMMON *DE VISU* FIBROBLAST COHERENCE PATTERN

The motivation for this thesis chapter began when researching the diversity of fibroblasts in the skin from human scRNAseq samples. At that time, four major articles were openly available: Tabib et al., 2018, Vorstandlechner et al., 2020, He et al., 2020, and Solé-Boldo et al., 2020. Additionally, Philippeos et al., 2018 published a dataset of culture-adapted fibroblasts, which were not as reliable as studies from skin biopsies. The individual papers are described in Section 3.2.1 of Introduction.

From the raw analysis of the literature, and at a first glance, we first assumed that the heterogeneity of fibroblasts across datasets was high and that more experiments would be needed to understand this phenomenon. However, doing a more thorough analysis of the Supplementary Materials of the articles, we observed that most of their predicted populations showed a clear overlap of markers, as shown in Figure 29.

We observed that there are, based only on the populations that the authors defined, 5 major types of fibroblasts that share several markers across 3 or more references. For instance, there is one population that expresses *MFAP5*, *PI16*, and *DCN*, which is thoroughly referred to in the literature as the one secreting the ECM. Also, there is one population expressing *C7*, *CCL19*, *CD74* and *CXCL12* that shows an immune profile, as stated by Tabib et al., 2018 or Solé-Boldo et al., 2020.

Therefore, from these observations, we conclude that the independent populations claimed individually by each article are already described. Therefore, a joint reanalysis of all datasets must be done to gain more insight into the skin fibroblast heterogeneity.

8.2 HUMAN DERMAL FIBROBLASTS DISTRIBUTE INTO THREE MAJOR COMPUTATIONAL AXES

To determine how the available scRNAseq datasets of human adult dermal fibroblasts presented common characteristics, we selected adult healthy cell subsets of each study for comparison, as shown in Table 5. Specifically, we analysed the following datasets: (i) 2,742 cells defined by Tabib et al., 2018 as fibroblasts through *COL1A1* and *COL1A2* expression; (ii) 2,842 young cells defined by Solé-Boldo et al., 2020 as fibroblasts through markers *LUM*, *DCN*, *VIM*, *PDGFRA*, and *COL1A2*; (iii) 1,871 cells from healthy donors defined by Vorstandlechner et al., 2020 as fibroblasts through the expression

of *FBLN1*, *LUM*, *PCOLCE*, *PLPP3*, *SDC2*, and *MXRA8*; and (iv) 6,368 cells from healthy donors defined by He et al., 2020 as fibroblasts through the expression of *COL1A1* and *DCN*. In total, 13,823 human adult dermal fibroblasts that had not been adapted to culture were used for comparison.

Each dataset was independently processed with a standard preprocessing pipeline, shown in the Materials and Methods section 4.2. In brief, *leiden* community detection with high-resolution values was used to obtain 15–20 communities per dataset. DEGs allowed merging communities with similar expression signatures, yielding 9–14 transcriptomically different populations, which show transcriptomically independent profiles. Populations expressing robust common markers, were further merged into axes. The differentiation of populations and axes is made to reflect a hierarchical ordering of similar populations, which probably similar functions.

Using this procedure, three major semi-supervised axes were robustly defined across datasets, which we termed A, B, and C (from more to less abundant percentage of cells; Figure 30 and Table 6). Axis A (Figure 30, red; *SFRP2* is shown for illustration) represents 49.5 ± 6.7% of the fibroblasts and is defined by *ELN*; *MMP2*; *QPCT*; and *SFRP2* expression. Axis B (Figure 30, green; *APOE* is shown for illustration) represents 30.5 ± 3.0% of the fibroblasts and is defined by *APOE*, *C7*, *CYGB*, and *IGFBP7* expression. Axis C (Figure 30, blue; *SFRP1* is shown for illustration) represents 16.2 ± 9.3% of the fibroblasts and is defined by *DKK3*, *TNMD*, *TNN*, and *SFRP1* expression. Other axes found were dataset-specific (such as axes D and F; Figure 30, grey) or shared by only two datasets (axis E; Figure 30, pink). However, the sum of cells included in axes A–C accounted for the vast majority of fibroblasts in all datasets, with only 7.5 ± 4.5% of the fibroblasts being left unassigned to one of the three axes (Table 6).

8.3 TYPE A FIBROBLASTS DISTRIBUTE INTO FOUR SEMI-SUPERVISED POPULATIONS

Axis A is composed of, four populations that were found across at least two datasets, which we termed A1–A4 (from more to less abundant percentage of cells; Figure 31 and Table 6). Population A1 (Figure 31, pink; *SLPI* is shown for illustration) represents 42.7 ± 5.7% of the type A fibroblasts and is defined by *IGFBP6*, *PI16*, *SLPI*, and *WISP2* expression. Population A2 (Figure 31, brown; *COMP* is shown for illustration) represents 32.3 ± 1.3% of the type A fibroblasts and is defined by *APCDD1*, *COL18A1*, *COMP*, and *NKD2* expression. Population A3 (Figure 31, orange; *WIF1* is shown for illustration) represents 19.5 ± 4.0% of the type A fibroblasts and is defined by *ELN*, *RGCC*, *SGCA*, and *WIF1* expression. Finally, population A4 (Figure 31, red; SFRP4 is shown for illustration) represents 8.2 ± 2.8% of the type A fibroblasts and is defined by *FBN1*, *PCOLCE2*, *PRG4*, and *SFRP4* expression.

At first sight, based on the marker pattern expression, there are some remarks about some of these populations. First, the population A3, despite expressing some specific genes (e.g. *WIF1*), looks like a *bridge* population between A1 and A2. Additionally, while many of the markers from A1 are jointly expressed in A4 and A3, A2 markers look more population-specific. It is possible that A1, A3 and A4 are similar populations, or the same population at different cell stages, whereas A2 could be a separate cell type within the ECM-producing fibroblasts. For instance, A2 population shows a distinct collagen pattern expression (*COL13A1*, *COL18A1*, *COL23A1*).

8.4 TYPE B FIBROBLASTS DISTRIBUTE INTO TWO SEMI-SUPERVISED POPULATIONS

Within axis B, two populations of similar abundances were robustly found in all datasets, which we termed B1 and B2 (Figure 32 and Table 6). Population B1 (Figure 32, dark green; *ITM2A* and *CCL2* are shown for illustration) represents 54.0 ± 6.2% of the type B fibroblasts and is defined by *CCL2*, *ITM2A*, *SPSB1*, and *TNFAIP6* expression. Population B2 (Figure 32, light green; *CCL19* and *CTSH* are shown for illustration) represents 46.0 ± 6.2% of the type B fibroblasts and is defined by *CCDC146*, *CCL19*, *CD74*, and *TNFSF13B* expression.

8.5 TYPE C FIBROBLASTS DISTRIBUTE INTO FOUR SEMI-SUPERVISED POPULATIONS

Within axis C, four populations were found across at least two datasets, which we termed C1–C4 (Figure 33 and Table 6). Population C1 (Figure 33, dark blue; COL11A1 is shown for illustration) represents 52.8 ± 17.5% of the type C fibroblasts and is defined by *COL11A1*, *DPEP1*, *TNMD*, and *WFDC1* expression. Population C2 (Figure 33, light violet; *COCH* is shown for illustration) represents 16.2 ± 4.7% of the type C fibroblasts and is defined by *COCH*, *CRABP1*, *FIBIN*, and *RSPO4* expression. Population C3 (Figure 33, dark violet; *POSTN* is shown for illustration) represents 26.3 ± 18.4% of the type C fibroblasts and is defined by *ASPN*, *F2R*, *GPM6B*, and *POSTN* expression. Finally, population C4 (Figure 33, light blue; *TM4SF1* is shown for illustration) represents 13.5 ± 4.2% of the type C fibroblasts and is defined by *ANGPTL7*, *APOD*, *C20rf40*, and *TM4SF1* expression.

8.6 FIBROBLAST SUBPOPULATIONS CANNOT BE FULLY RECOVERED AFTER COMPUTATIONAL INTEGRATION OF THE DATASETS

A dataset integration analysis was performed further to validate the robustness of the proposed axes and populations. Between the different common integration methods, *bbknn* (Polanski et al., 2020) and *harmony* (Korsunsky et al., 2021) were used. The integration of all four datasets yielded a non-homogeneous UMAP, where each dataset was separated from the rest (Figure 34). Because He et al., 2020 reported 43–77% cell viability within their dataset, and since low cell viability is related to artefacts in RNA capture and sequencing, which affects the biological interpretation of the results, we excluded it from this analysis. Regarding the performance of *harmony* and *bbknn*, the two showed similar integration results, including or excluding the He et al., 2020 dataset.

After concatenation of the Solé-Boldo et al., 2020; Tabib et al., 2018; Vorstandlechner et al., 2020 datasets, the combined data were normalised, and the standard pipeline of individual datasets was followed, shown in Figure 35. To evaluate the performance of data integration, the obtained UMAP were colour-coded to highlight the datasets, the proposed axes, and populations (Figures 35a to 35d). In addition, we performed *leiden* unsupervised clustering and plotted the relative proportion of cells belonging to the different datasets and fibroblast populations within each *leiden* cluster (Figures 35e and 35f).

Dataset integration showed a clear bias between, Solé-Boldo et al., 2020 and Vorstandlechner et al., 2020 on the one hand; and and Tabib et al., 2018 dataset on the other, as shown in the barplot of dataset frequency per cluster (Figure 35e). *leiden* clusters 6, 8 and 11 showed the integration of Solé-Boldo et al., 2020 and Tabib et al., 2018 datasets in axis C–the Vorstandlechner et al., 2020 dataset showed virtually no type C fibroblasts–.

Populations B1 and B2 were identified as *leiden* clusters 1 + 7 and 2, respectively. In addition, the separation between populations A1 and A2 was apparent–clusters 0 + 4 and 0 + 3, respectively–. We conclude that the structure of the most abundant populations is validated by integration analysis, although the integration is incomplete, and there are some biases between datasets. For instance, some minor populations, such as C3 or A4, were not fully recovered.

8.7 PRIMARY EXPERIMENTAL VALIDATION IN SKIN SECTIONS OF MAJOR AXES

To validate the computational findings, wet laboratory *in silico* validation of fibroblast populations was performed in the human dermis. To this end, full-thickness abdomen skin punch sections were analysed by triple immunofluorescence with antibodies detecting the expression of representative markers of type A (*SFRP2*), type B (*CXCL12*), and type C (*SFRP1*) fibroblasts, all cells being simultaneously highlighted by nuclear staining (Figure 36).

Immunofluorescence staining proves that there are no cells with colocalised expression of neither *SFRP2*, *CXCL12* nor *SFRP1* (Figures 36j and 36k). Therefore, fibroblast populations are defined as $SFRP2^{pos}CXCL12^{neg}SFRP1^{neg}$ (consistent with type A fibroblasts), $SFRP2^{neg}CXCL12^{pos}SFRP1^{neg}$ (type B fibroblasts), and $SFRP2^{neg}CXCL12^{neg}SFRP1^{pos}$ (type C fibroblasts), thus confirming that the computational stratification of fibroblast populations based on scRNAseq data was also present at the protein level *in situ*.

Some of these findings were confirmed by immunostained skin sections of the Human Protein Atlas initiative (Uhlén et al., 2015). The expression at the protein level of selected markers of type A, B, and C fibroblasts are shown in Figures S16-S18 of Ascensión et al., 2021. Several markers of fibroblast subtypes A1–A4 (Figures S19-S22), B1 and B2 (Figures S23, S24), and C1–C4 (Figures S25-S28) were also confirmed by immunostaining.

SFRP2-type fibroblasts show a dim gene expression and look scattered throughout the dermis. Immunohistochemistry location of proteins from axis A (ELN, MMP2, CD9) shows that cells expressing those markers are located throughout the dermis (CD9) or are segregated towards the reticular dermis (ELN), as shown in Figure S16 in Ascensión et al., 2021. This fibroblast type has a slightly elongated shape, although not spindle-shaped, similar to the descriptions provided by Tabib et al., 2018.

CXCL12-type cells show two levels of intensity. $CXCL12^{hi}$ cells are located within compact structures, and it is highly likely that those cells represent a non-fibroblastic population (Figure 36d). Possibly, they might be located in perivascular or near glandular structures. Additionally, there is a secondary population, $CXCL12^{dim}$, which is also scattered through the dermis. This type of fibroblast has a more spherical morphology. Immunohistochemistries of B2 markers (CD74, CTSH, RBP5) show a clear expression pattern either at vassal structures or surrounding them (Figure S24 in Ascensión et al., 2021). Moreover, some $CD74^{+}$ and $CTSH^{+}$ cells also appear with high and compacted expression outside of these structures, which may be the aforementioned $CXCL12^{dim}$ cells.

SFRP1-type cells express this gene in a dim manner too, and the cell morphology cannot be easily established. They are located in the basal epidermis and in dense cell accumulations, indicating a putative substructural role within the dermis, possibly related to hair. This is expected considering that *CRABP1*, a marker of C1-type fibroblasts, is a classical dermal papillary marker (Collins et al., 2011). Immunohistochemistry pictures of POSTN, a C3 population marker, show a clear expression within

the papillary dermis and certain structures within the reticular dermis (Figure S27 in Ascensión et al., 2021).

8.8 TABLES AND FIGURES

Table 5: **Human adult dermal fibroblast datasets used in the preliminary study** Relevant information for each of the 4 datasets used in the study.

Dataset	Donors (n)	Age (y)	Sex	Ethnicity
Tabib (Tabib et al., 2018)	6	23, 24, 54, 62, 63, 66	3M/3F	5 Caucasian, 1 Asian
Solé-Boldo (Solé-Boldo et al., 2020)	2	25, 27	M	Caucasian
Vorstandlechner (Vorstandlechner et al., 2020)	3	30, 36, 43	F	-
He (He et al., 2020)	7	38–82	4 F,	-

Anatomic Area	Time from Harvest to Processing	Number of Fibroblasts
Dorsal mid-forearm	-	2,742
Inguinoiliac (sun protected)	<1 h	2,842
Abdomen	-	1,871
Extremities	Punch biopsies at −80 °C (4–5 mo)	6,368

Original Reference Clusters	Platform and Chemistry	Cell Type Enrichment
SLM clusters 0, 3 and 4	Chromium Single Cell 3' v 1 (10X)	No
clusters 1, 2, 3, and 9	Chromium Single Cell 3' v 2 (10X)	Magnetic isolation of live cells
FB clusters	Chromium Single Cell 3' v 2 (10X)	FACS isolation of live, single cells
FB1–FB3 clusters	Chromium Single Cell 3' v 2 (10X)	No

Table 6: **Relative abundance of each axis/cluster.** Summary of computational axes and clusters including representative core genes and relative abundance of each fibroblast subpopulation in the published datasets.

Axis/Cluster	Core Genes	Tabib 2018	Solé-Boldo 2020	Vorstandlechner 2020	He 2020	Consensus %
A	*ELN MMP2 QPCT SFRP2*	1,279/2,742 (46.6%)	1,178/2,842 (41.4%)	1,117/1,871 (59.7%)	3,217/6,368 (50.5%)	49.5 ± 6.7
A1	*IGFBP6 PI16 SLPI WISP2*	472/1,279 (36.9%)	595/1,178 (50.5%)	453/1,117 (40.6%)	—	42.7 ± 5.7
A2	*APCDD1 COL18A1 COMP NKD2*	416/1,279 (32.5%)	398/1,178 (33.8%)	343/1,117 (30.7%)	—	32.3 ± 1.3
A3	*ELN RGCC SGCA WIF1*	321/1,279 (25.1%)	185/1,178 (15.7%)	198/1,117 (17.7%)	—	19.5 ± 4.0
A4	*FBN1 PCOLCE2 PRG4 SFRP4*	70/1,279 (5.5%)	—	123/1,117 (11.0%)	—	8.2 ± 2.8
B	*APOE C7 CYGB IGFBP7*	758/2,742 (27.6%)	982/2,842 (34.6%)	516/1,871 (27.6%)	2,037/6,368 (32.0%)	30.5 ± 3.0
B1	*CCL2 ITM2A SPSB1 TNFAIP6*	477/758 (62.9%)	455/982 (46.3%)	289/516 (56.0%)	1,035/2,037 (50.8%)	54.0 ± 6.2
B2	*CCDC146 CCL19 CD74 TNFSF13B*	281/758 (37.1%)	527/982 (53.7%)	227/516 (44.0%)	1,002/2,037 (49.2%)	46.0 ± 6.2
C	*DKK3 TNMD TNN SFRP1*	620/2,742 (22.6%)	682/2,842 (24.0%)	13/1871 (0.7%)	1,114/6,368 (17.5%)	16.2 ± 9.3
C1	*COL11A1 DPEP1 TNMD WFDC1*	178/620 (28.7%)	409/682 (60.0%)	—	777/1,114 (69.7%)	52.8 ± 17.5
C2	*COCH CRABP1 FIBIN RSPO4*	95/620 (15.3%)	152/682 (22.3%)	—	122/1,114 (11.0%)	16.2 ± 4.7
C3	*ASPN F2R GPM6B POSTN*	277/620 (44.7%)	54/682 (7.9%)	—	—	26.3 ± 18.4
C4	*ANGPTL7 APOD C2orf40 TM4SF1*	70/620 (11.3%)	67/682 (9.8%)	—	215/1,114 (19.3%)	13.5 ± 4.2
Un.		85/2,742 (3.2%)	—	225/1,871 (12.0%)	—	7.5 ± 4.5

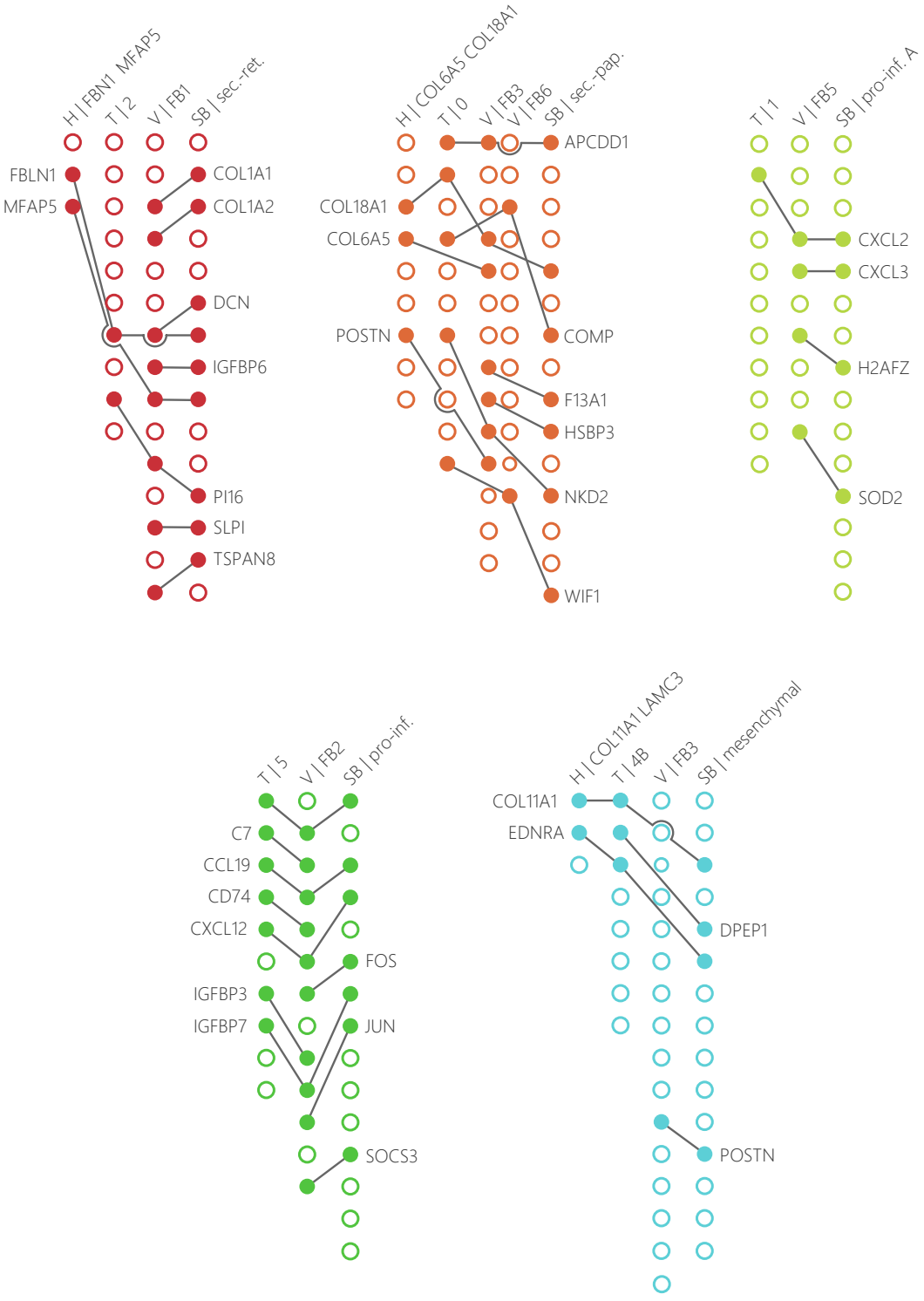


Figure 29: **Comparison of ARI for FS methods on artificial datasets.** Each diagram represents sets of markers in common across the 4 datasets. Each column represents markers obtained via the literature belonging to each dataset, sorted alphabetically. Markers in a white circle are markers that are not shared across datasets, and therefore their gene symbols are not included. Markers in a full circle are shared between datasets, using a line to connect the circles. The legend for the dataset names is: H - He et al., 2020, T - Tabib et al., 2018, V - Vorstandlechner et al., 2020, SB - Solé-Boldo et al., 2020.

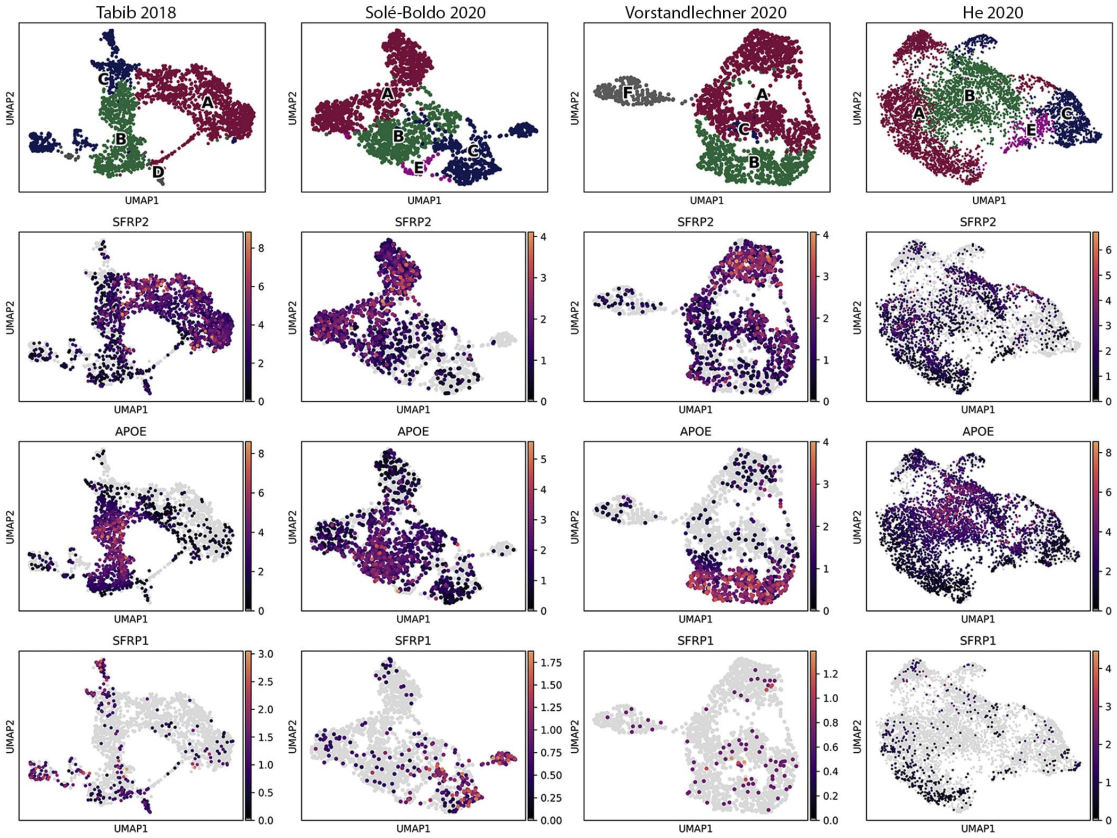


Figure 30: **Top shared markers of axes A, B, and C.** Top shared markers of axes A, B, and C. UMAP plots with expression values of *SFRP2* (axis A marker), *APOE* (axis B marker), and *SFRP1* (axis C marker). The top row shows, in different colours, the cells belonging to each of the axes for each dataset. Percentages indicate mean ± SD.

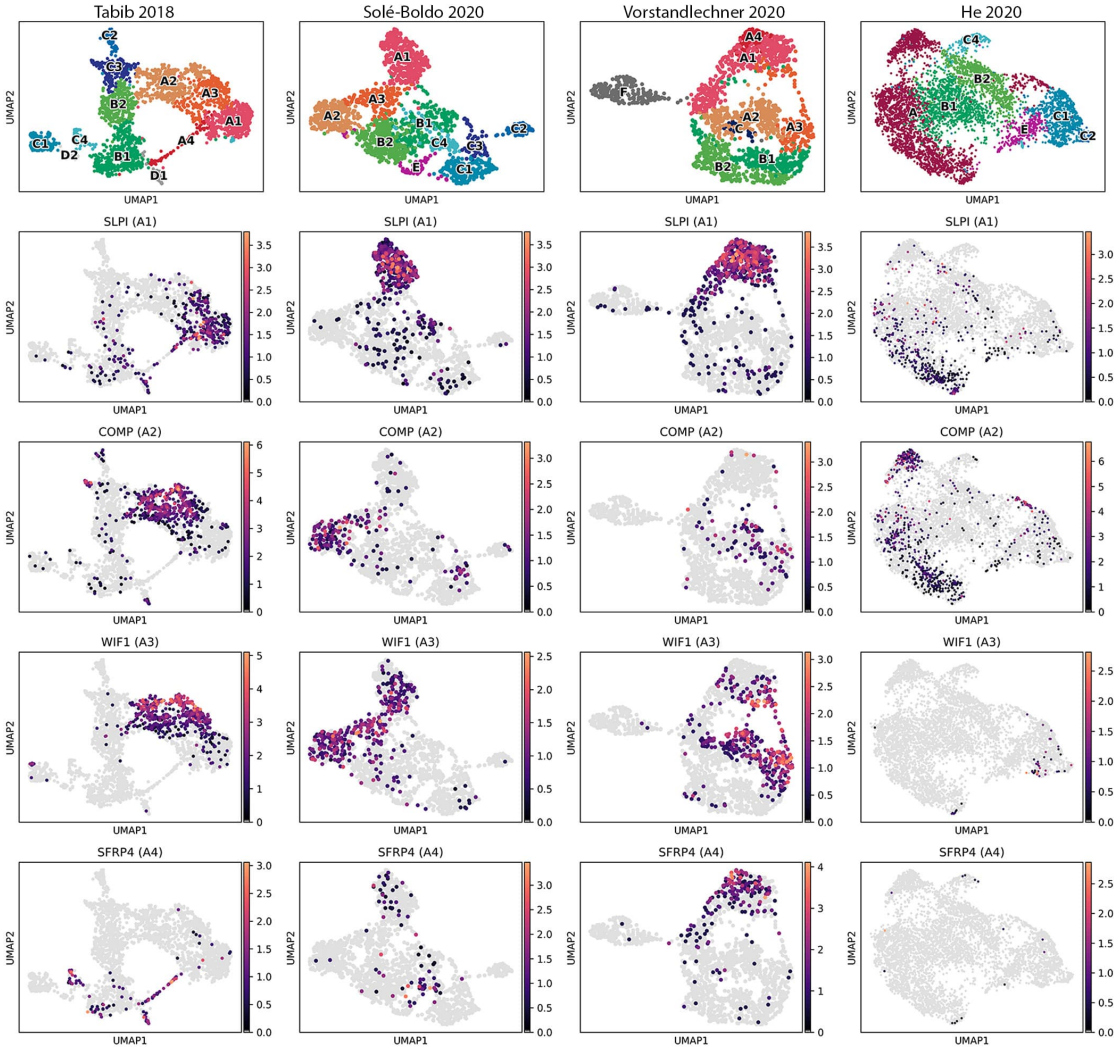


Figure 31: **Markers of axis A.** UMAP plots with expression values of the most relevant marker for each population in axis A (A1–A4). The top row shows, in different colours, the cells belonging to each of the populations for each dataset.

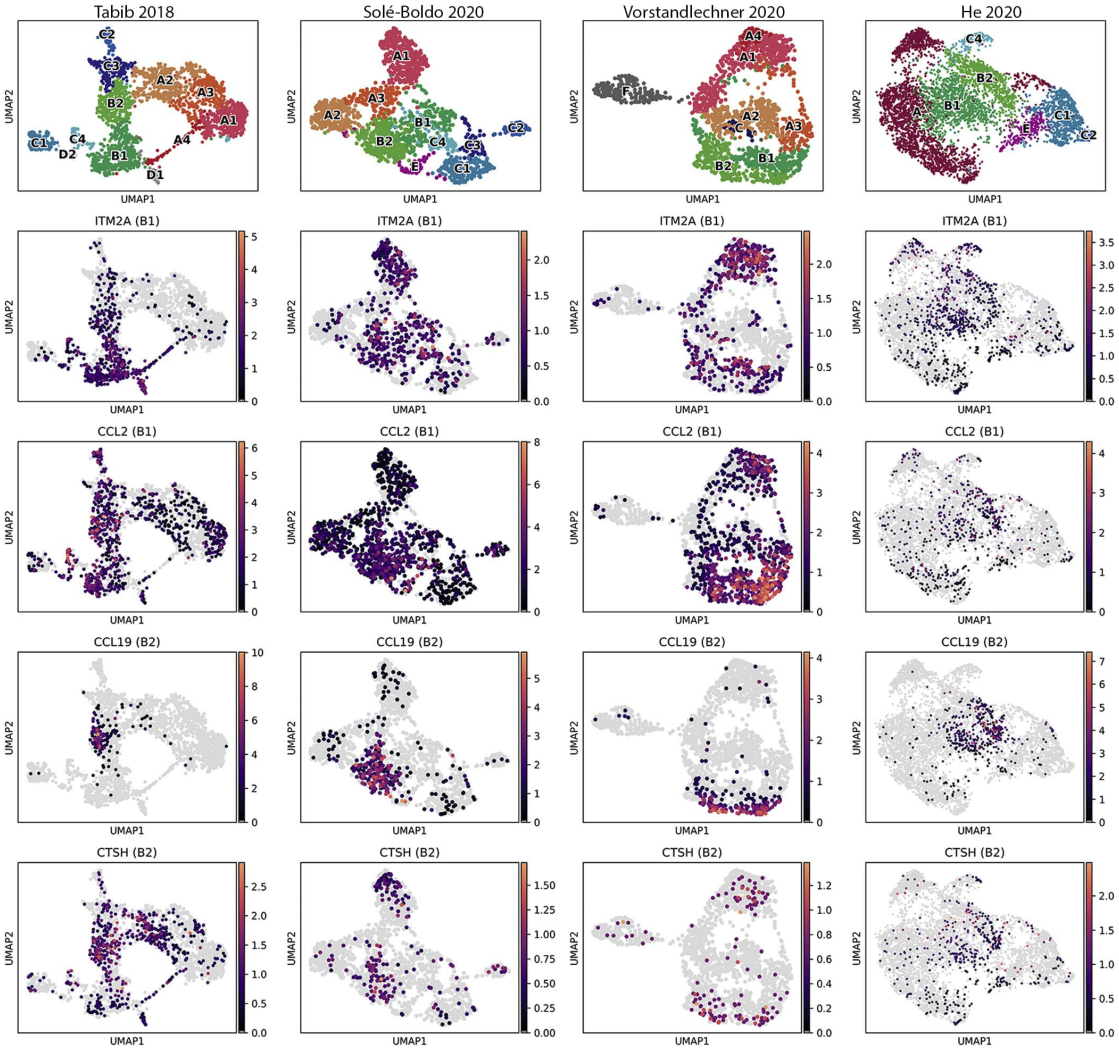


Figure 32: **Markers of axis B.** UMAP plots with expression values of the most relevant markers for each population in axis B (B1–B2). The top row shows, in different colours, the cells belonging to each of the populations for each dataset.

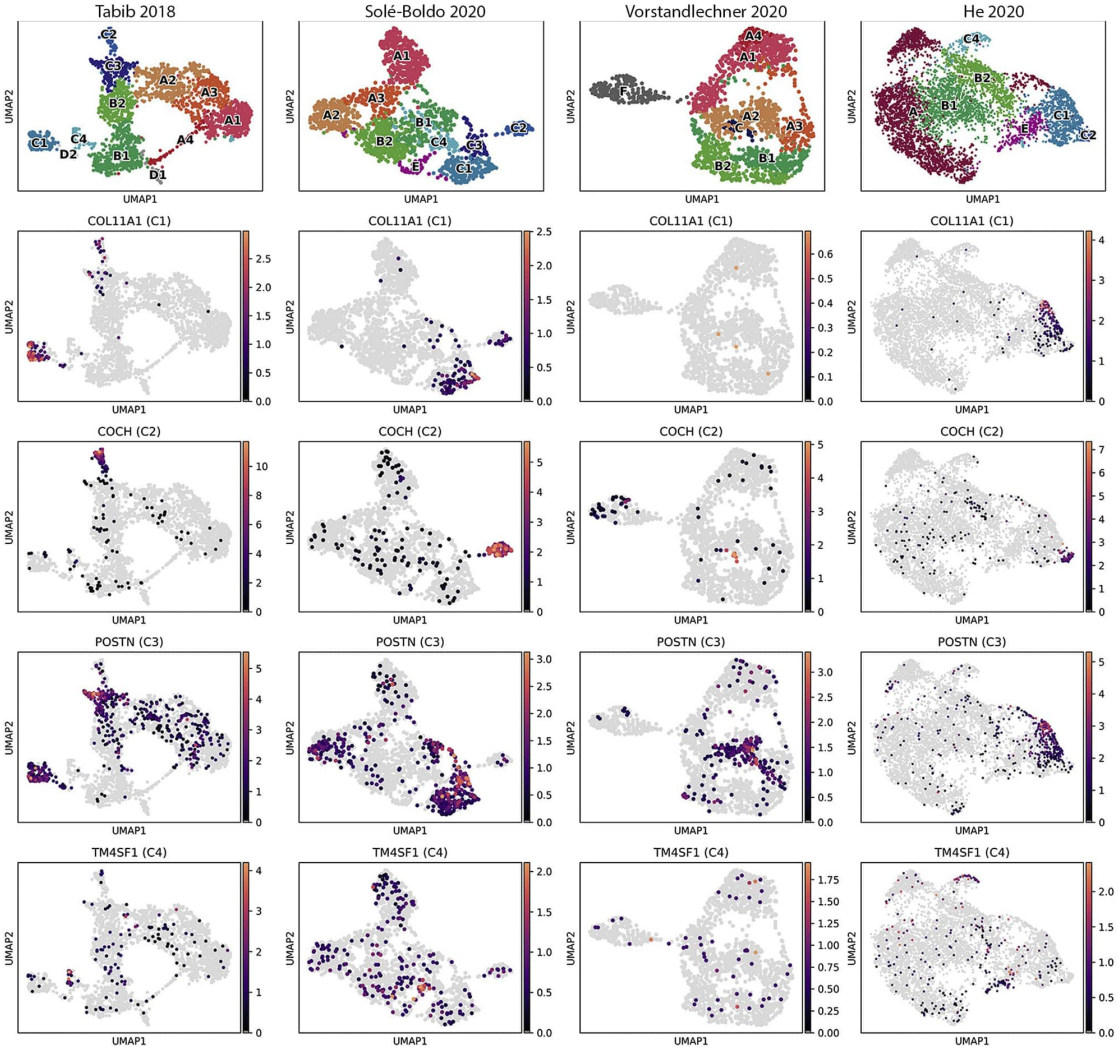


Figure 33: **Markers of axis C.** UMAP plots with expression values of the most relevant marker for each population in axis C (C1–C4). The top row shows, in different colours, the cells belonging to each of the populations for each dataset.

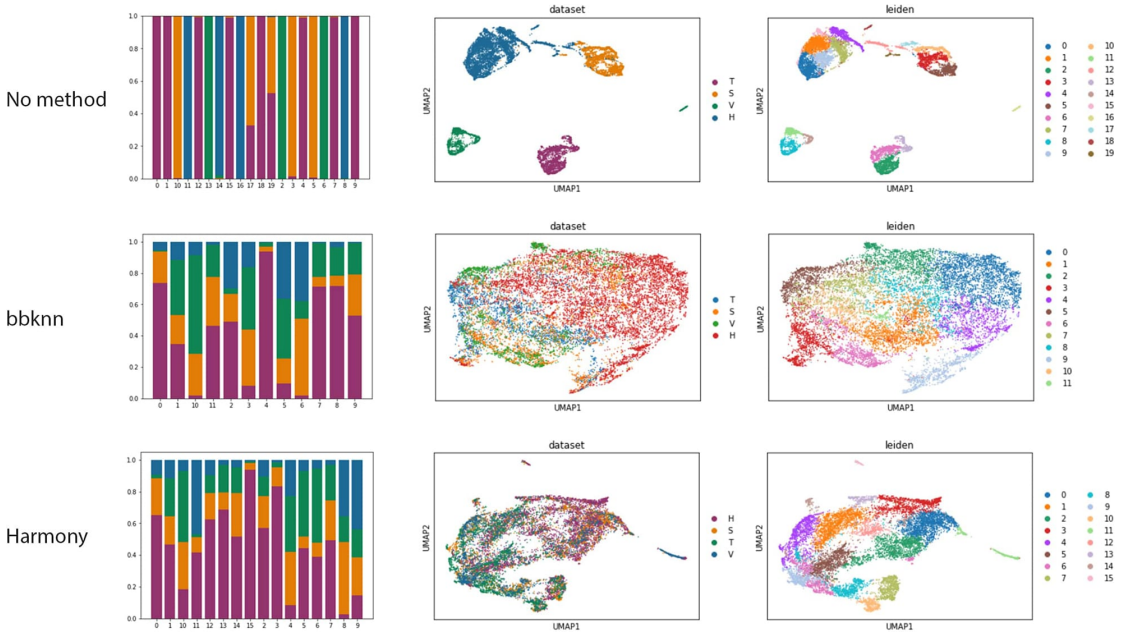


Figure 34: **Comparison of integration methods with 4 datasets.** Comparison of integration with no integration methods, *bbknn* and *harmony*. The first column shows the distribution of datasets per each of the unsupervised *leiden* clusters. The second and third columns show the UMAP plots with the dataset and *leiden* labels, respectively.

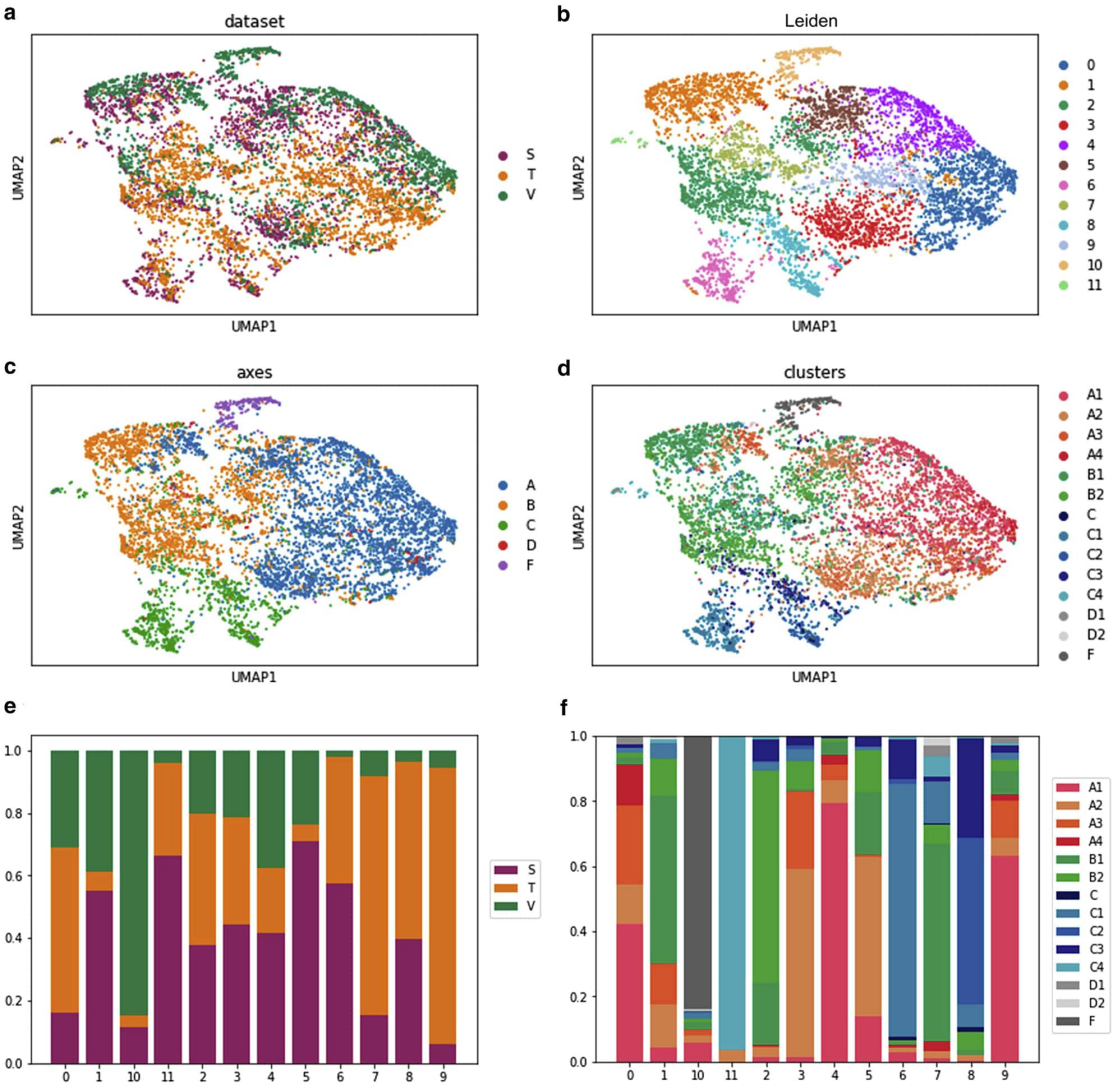


Figure 35: ***bbknn* integration of Tabib, Vorstandlechner and Solé-Boldo datasets.** To evaluate the performance of the data integration, UMAPs (colour coded by dataset, axes, and populations) are shown. (b) In addition, *leiden* unsupervised clustering is shown. The integration of the three datasets using *bbknn* yielded a fairly acceptable integration. (e) Barplots of dataset frequency per cluster and (f) fibroblast population frequency per cluster are also shown. A clear bias between S + V and T datasets was detected. Despite the integration not being 100% successful, the structure of the most important populations was recovered.

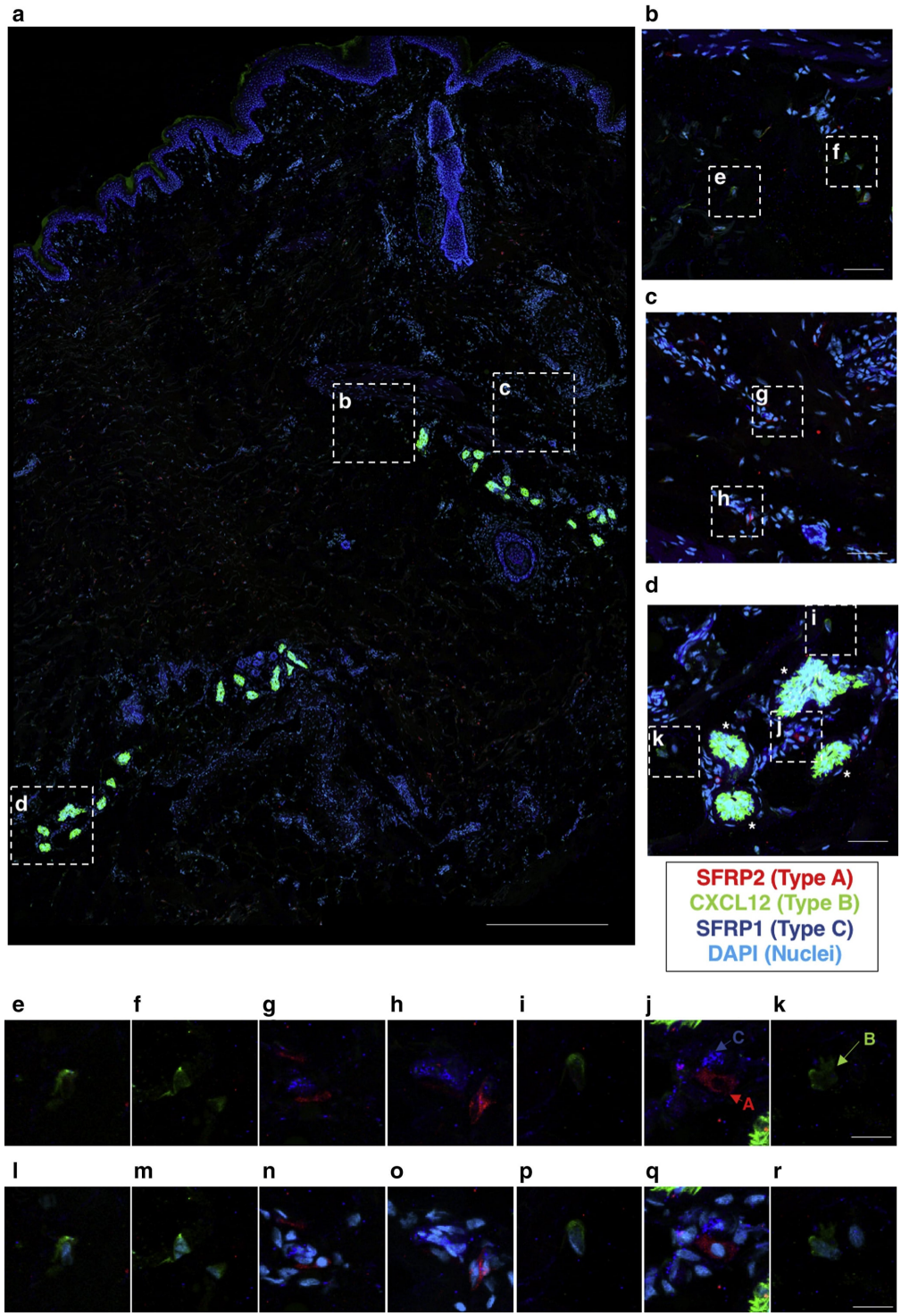


Figure 36: **Validation of major fibroblast axes human skin.** (**a**–**r**) Abdominal skin sections were tested by immunostaining with antibodies specific for the markers *SFRP2* (type A fibroblasts, red), *CXCL12* (type B fibroblasts, green), and *SFRP1* (type C fibroblasts, dark blue). Representative images at low (**a**: Bar = 500 μm), medium (**b**–**d**: Bar = 50 μm), and high (**e**–**r**: Bar = 20 μm) magnifications are shown. Asterisks in **d** denote a non-fibroblastic *CXCL12*-positive population in a perivascular location. Coloured arrows in **j** and **k** demonstrate the existence of discrete dermal cell populations expressing each of the representative markers of types A–C fibroblasts. Cell nuclei were counterstained with DAPI (light blue) in **a**–**d** and **l**–**r**. Cell nuclei in **e**–**k** are shown without DAPI staining.

9

PRIMARY OF FIBROBLAST HETEROGENEITY IN MOUSE SKIN USING SCRNASEQ

9.1 MOTIVATION FOR PRIMARY MOUSE ANALYSIS

Once Ascensión et al., 2021 was published in four human datasets, and during the development of the human secondary fibroblast analysis (Chapter 10), we found that, despite the classical papillary and reticular fibroblast classification by Driskell et al., 2013, there was no mouse fibroblast characterisation similar to the one done in human fibroblasts based on different datasets. Therefore, we began this primary mouse characterisation with 5 datasets of healthy mouse dermis available at that time: Abbasi et al., 2020, Buechler et al., 2021, Haensel et al., 2020, Phan et al., 2020, Shook et al., 2020

Furthermore, we found that the manual labelling performed in Ascensión et al., 2021 lacked robustness and was prone to be biased with clusters from datasets from personal preference. Although we reckon that population labelling and mapping across datasets will contain an inherent load of supervised execution–e.g. when deciding the clustering resolution or joining non-fully overlapping clusters–a more robust method for semi-supervised *de novo* labelling was necessary. This semi-supervised labelling method would, in turn, be applied to labelling populations in other organs, organisms, or major cell types.

9.2 FIBROBLAST HETEROGENEITY IS CAPITULATED IN AT LEAST FOUR DIFFERENT AXES

Before the combined labels were constructed, primary clustering was done in each dataset to get a set of clusters that could be mapped to each other afterwards. The results are shown in Figure 37. In Abbasi et al., 2020, 16 clusters are described in four main axes (A1a-A3a, B1a-B6a, C1a-C4a, D1a-D2a); in Buechler et al., 2021 13 clusters are described in four main axes (A1b-A5b, B1b-B4b, C1b-C2b, D1b-D2b); in Haensel et al., 2020 25 clusters are described in six main axes (A1h-A6h, B1h-B9h, C1h-C3h, D1h-D5h, Eh, Fh); in Phan et al., 2020 19 clusters are described in seven main axes (A1p-A7p, B1p-B3p, C1p-C2p, D1p-D3p, E1p-E2p, Fp, Gp); and in Shook et al., 2020 16 clusters are described in three main axes (A1s-A2s, B1s-B6s, C1s-C8s).

The clusters were classified in different axes based on the concurrent appearance in the UMAP; that is, if two or more *leiden* clusters belong to the same bigger UMAP cluster, they were merged into the same axis. In Haensel et al., 2020 and Phan et al., 2020 datasets, some axes (Eh, Fh, Fp, Gp) consist

of one cluster. It is likely that these are minor populations and do not belong to the mainstream fibroblast types. Therefore, when looking at the consistency of the axes, it is apparent that 3 to 5 axes are most surely shared between datasets.

It is important to note that the nomenclature of the axes and clusters in each dataset is arbitrary. Therefore, if there is any correspondence between axes and clusters of different datasets, this fact is not reflected in the current nomenclature but in the ones in the following steps. In other words, axis A in Haensel et al., 2020 and axis A in Shook et al., 2020 may not reflex the same axis based on transcreiptomic similarity. Thus, the next step required to build the consensus annotation of all datasets is to find the correspondences between clusters of different datasets.

9.3 CLUSTER RELATIONSHIP GRAPH UNCOVERS A JOINT FIBROBLAST COMPOSITION ACROSS DATASETS

To find the common ground for fibroblast populations, we generated a graph where relationships between fibroblasts of different datasets and relationships of fibroblasts within the same dataset are considered. The creation of this graph is explained in Section 4.8 of the Materials and Methods section. The first step was to find the markers of each cluster for each dataset and to perform a double mapping: map the markers of one population to its own dataset and the rest of the datasets. With that information, we built a plot, as depicted in Figure 38, that represents the mapping of clusters between datasets.

The interpretation of this figure is complex due to the combinatorial amount of mappings that can be done. Here we show some one-to-one and one-to-several mappings.

- C1p cluster from Phan et al., 2020 maps to D2a cluster from Abbasi et al., 2020 and C4s cluster from Shook. Each of these mappings can be done inversely, too: D2a cluster in Abbasi et al., 2020 maps to C1p in Phan et al., 2020. Interestingly, the C1p cluster in Shook maps to C4s, C5s and C6s clusters in Shook, indicating that C1p to C4s-C6s is a one-to-several mapping case.
- B1s cluster in Shook et al., 2020 maps in a one-to-one fashion to B4a in Abbasi et al., 2020 and B2b from Buechler et al., 2021
- A1s in Shook et al., 2020 maps to A2b in Buechler et al., 2021 and A2a in Abbasi et al., 2020.
- When doing the mapping, it is important to confirm that the reverse mapping is also true. For instance, the A1a cluster from Abassi maps to part of the A2b of Buechler. However, A1a cluster is not mapped to any of Buechler et al., 2021 clusters *per sé*. Therefore, it is possible that the A1a-A2b correspondence is false since A2b also maps to A2a, and thus A2b is the most similar cluster to A1a when mapping, but it is not reciprocal.

After all possible mappings are performed, and together with the relationships of clusters within the same dataset, the graph from Figure 40 is constructed. In this graph, we observe that there are many sets of mappings of 2 or more datasets in all axes (e.g. A1a-A4b-C3h, B5a-B7h-B3/B4s, C1a-C8s-C2b, C7s-A3p-A3b, etc.). Few isolated clusters are either a bridge between axes (B2s, C2b-C8s), between clusters (A4p, C4a), or are simply isolated within the axis (C2h, B8h, D3p). Of note, although it is true that some cluster mapping begins with the same letter (A2a-A1s-A2b, B1s-B2h-B1/B2b-B4a, C1a-C8s-C2d), this naming was arbitrary and coincidental.

The next step in constructing the joint populations is to merge all cluster mappings into a joint nomenclature. This nomenclature can be observed in Figures 39 and 41. From this point onwards in the analysis, and to differentiate between human and mouse axes and clusters, human axes and clusters will begin in upper case (A2, B3, C1), and mouse ones in lower case (a1, b4, c2). The simplified graph of relationships between clusters can be observed in Figure 41. Generally, all mappings have been condensed into one label, and most of the cluster labels that were only belonging to a specific dataset were removed or simplified. As a result, the preliminary common labels between datasets are divided into four axes and 18 clusters (a1-a3, b1-b7, c1-c3, d1-d5), and two additional *bridge* clusters (a/b, b/c).

During the process of condensing labels, markers shared between the labels of each dataset were found. As a result, the preliminary set of markers for these populations is the following:

- a1: Abl2, Ccl2, Ccl7, Cxcl1, Cxcl10, Dusp10, Fosl1
- a2: Ak1c18, Cotl1, Efhd1, Fez1, Il18, Lims2, Mustn1, Pts
- a3: Cdca3, Cdca8, Stmn1, Cks2, Birc5, Lockd, Top2a, Ccna2
- a/b: Ackr1, Ccn3, Cmah, Efhd1, Fbn1, Fn1, Galnt16, Has2
- b1: Adam12, Ccl19, Cd36, Col6a6, Creb3l3, Cthrc1, Cxcl12, Hmcn2
- b2: Cadm2, Angptl4, Galnt5, Mex3b
- b3: Adamtsl2, Bmp5, Cilp, Col8a1, Fbln7, Fhl2, Fibin, Gos2
- b4: Abcc9, Akr1c14, Cilp, Col12a1, Crispld2, Eln, Fibin, Igf1
- b5: C2, Cfh, Chrdl1, Cp, Cxcl16, Naalad2, Scn7a, Sned1
- b6: Acot1, Adamtsl3, Apod, Ccl11, Clec1a, Col26a1, Col8a1, Cp
- b/c: C1qtnf3, Mkx, Tnmd, Adam12, Akr1cl, Col12a1, Creb3l3, Crp
- c1: Abcc4, Cgref1, Creb3l3, Fcgr3, Wisp2, Adcy1, Ahrr, Aldh1l2
- c2: Ccdc42, Col13a1, Col7a1, Fxyd6, Grem1, Igfbp2, Sema3a, Stc1
- c3: Ccl19, Ccl2, Ccl7, Ccl8, Cpxm1, Ets2, Ltbp1, Mmp3
- d2: Cd200, Col11a1, Mafb, Robo2, Tagln, Tnmd, Actg2, Actn1
- d4: Abi3bp, Acta2, Adam12, Adamts15, Arhgdib, Bok, Cdc42ep3, Chchd10
- d5: Cilp, Cmklr1, Cpxm2, Eln, Fam180a, Il15, Mmp16, Nrep

When mapping the populations using these markers into the five datasets, we obtain a common mapping shown in the UMAP of Figure 39. However, we found some discrepancies between datasets when looking closely at the populations.

For example, there is an imbalance of axes regarding their number of cells across datasets: Haensel et al., 2020 and Phan et al., 2020 have an increased number of cells in axes *c* and *d*; and Phan et al., 2020 has, in comparison to Abbasi et al., 2020 or Shook et al., 2020 an extremely reduced amount of cells from axes *a* and *b*. Shook et al., 2020 has a low amount of cells from axis *c*, and Buechler et al., 2021 as a low amount of cells from axis *d*. We expect fluctuations in the number of cells to

happen, as is already observed in the primary analysis of human dermal fibroblasts, but some of these discrepancies seem exaggerated.

Regarding the internal coherence of populations within clusters, we observe similar results. In most datasets, we observe that axes *c* and *d* are connected by $\mathrm{d5}$ population. In all datasets, there is a coherence that axes *c* and *d* are related, $\mathrm{d5}$ being putative *bridge* population. However, while axes *a* and *b* are related in Abassi et al. and Shook et al.–probably in Buechler et al. and Phan et al. as well– in Haensel et al. these populations do not show a clear relationship.

Similar to the preliminary analysis of human datasets, in the secondary analysis, we will include new datasets that will help us achieve a better consensus on the labels and their markers.

9.4 TABLES AND FIGURES

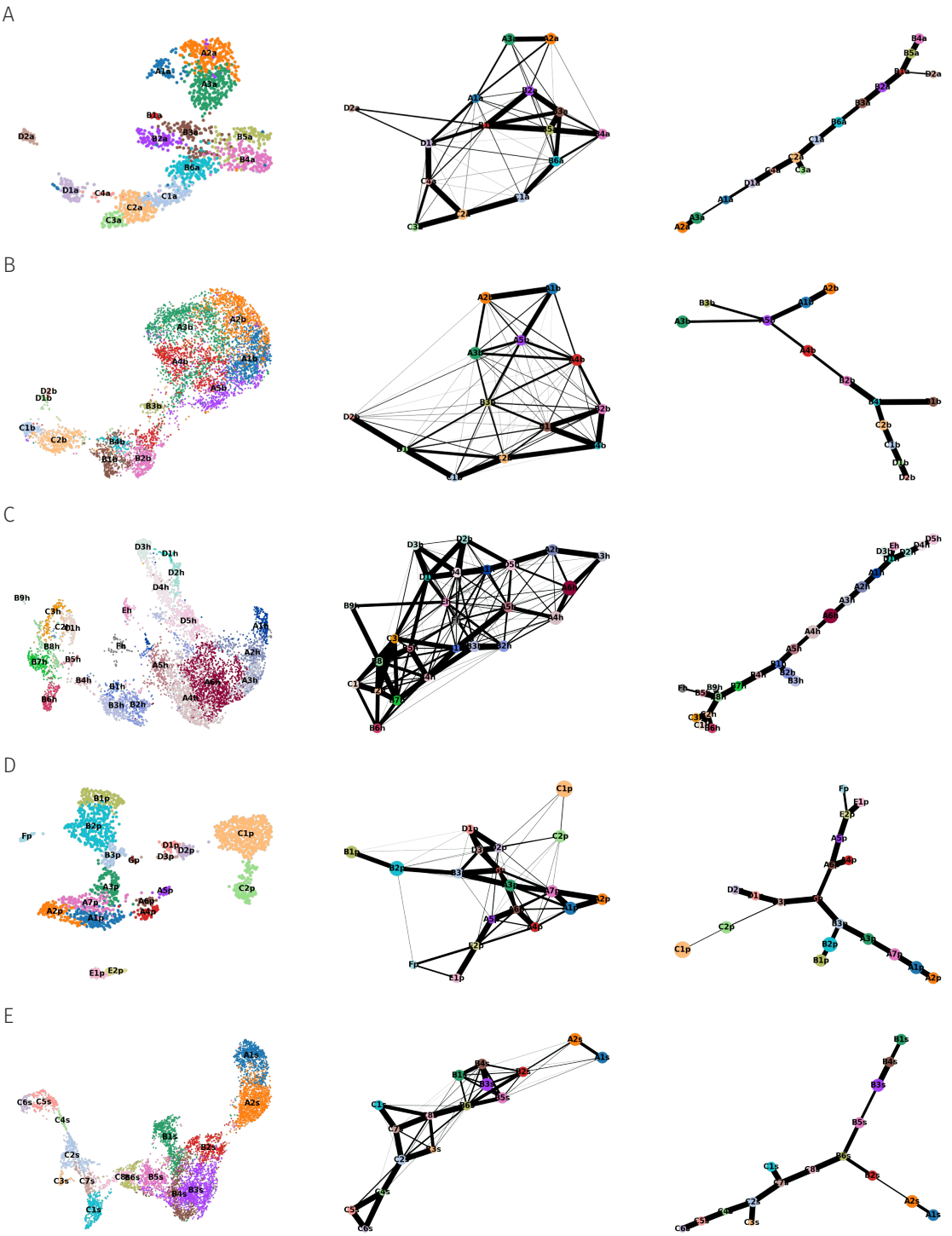


Figure 37: **Primary characterisation of clusters for five mouse datasets.** Characterisation of clusters for Abbasi et al., 2020 (A), Buechler et al., 2021 (B), Haensel et al., 2020 (C), Phan et al., 2020 (D), Shook et al., 2020 (E). The first column shows a UMAP plot with each dataset clusters, and the second and third columns show their respective PAGA graphs. The first PAGA graph is a complete graph with interactions between all clusters, and the second is a pruned graph showing the most relevant interactions in a tree structure.

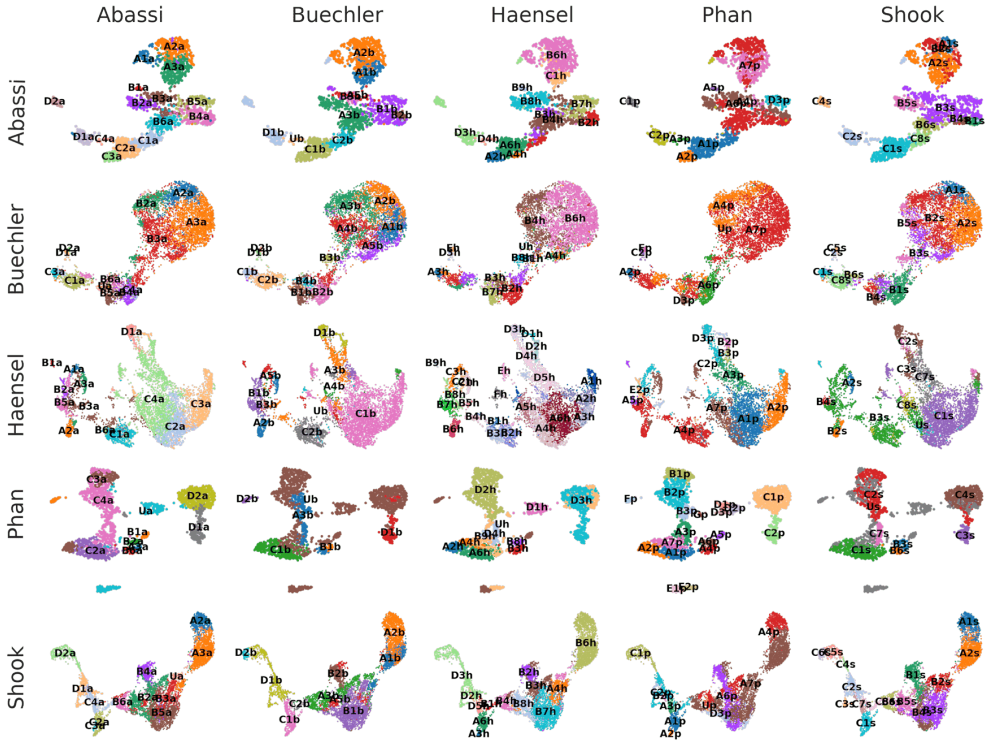


Figure 38: **Cross mapping of clusters across datasets.** For each pair of datasets, the UMAP plot shows the row dataset with the population labels and clusters of the column dataset. For example, row Abassi with column Phan shows the UMAP plot of the Abbasi et al., 2020 dataset with the labels of the Phan et al., 2020 dataset.

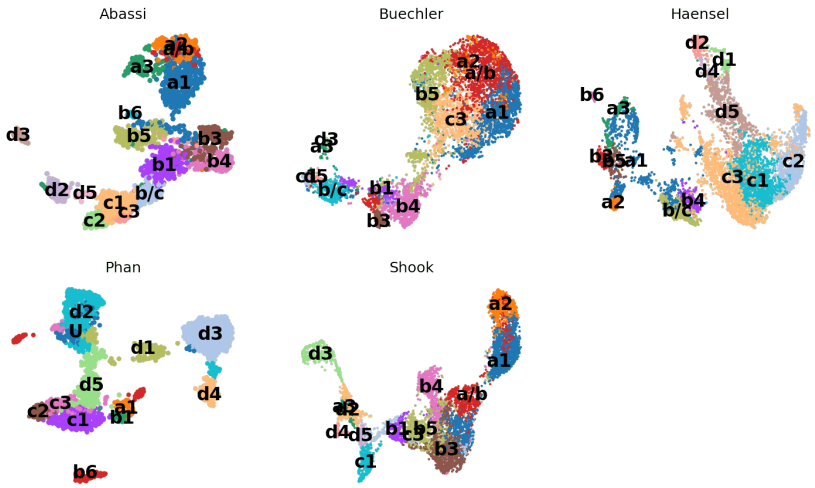


Figure 39: **Mapping of primary analysis populations into five mouse fibroblast datasets.** For each dataset, the mapping of the preliminary populations is shown.

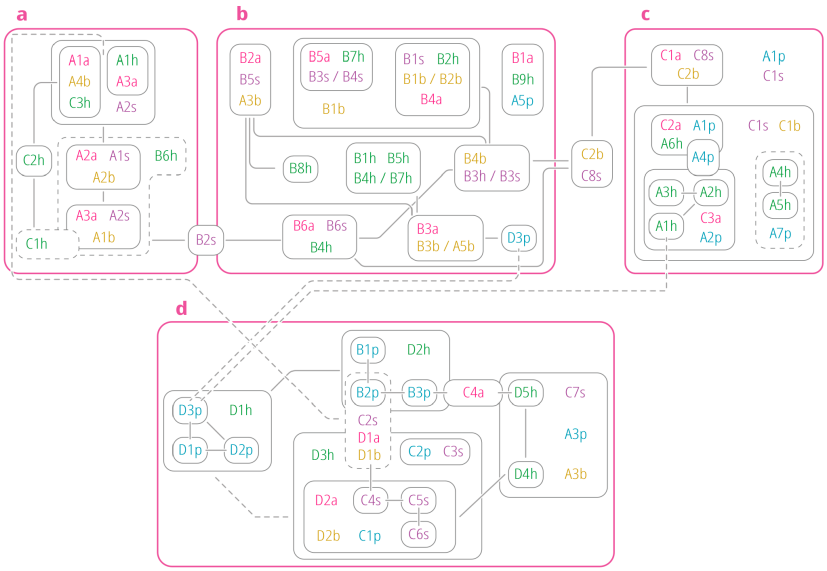


Figure 40: **Complete graph of relationships between clusters.** This graph includes the relationships from the PAGA graphs from Figure 37 and the mappings between datasets depicted in Figure 40. Groups of clusters within the same square (e.g. A1a, A4b and C3h in axis a) are the mapping of the same cluster. Mapping groups merged into one common square (e.g. A1a-A4b-C3h and A1h-A3a mappings are within the A2s cluster) indicate a one-to-several mapping case. Filled lines indicate likely relationships between clusters observed in several datasets jointly, and dotted lines indicate relationships that are likely to occur but were not directly observed in many datasets.

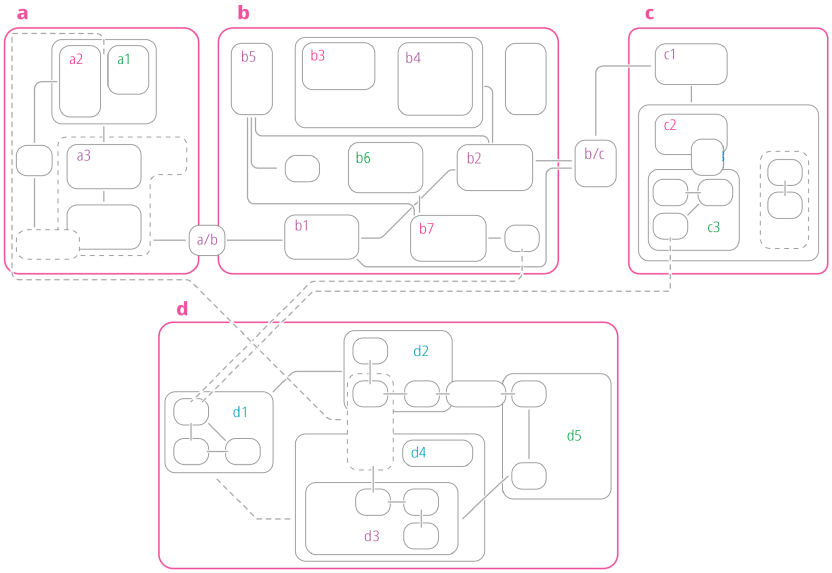


Figure 41: **Simplified graph of relationships between clusters.** This graph is created to simplify clusters of mapping into a single, combined primary notation that is present in Figure 39.

10

SECONDARY ANALYSIS OF FIBROBLAST HETEROGENEITY IN HUMAN AND MOUSE SKIN USING SCRNASEQ

10.1 MOTIVATION

This section is built upon the results from the primary analysis chapters 8–in human–and 9–in mouse–.

The primary analysis of human samples yielded 10 populations distributed in 4 axes. The main limitation of that study was that, despite the consistency of these populations across datasets, there were some populations specific to one dataset–populations D1 and D2 from Tabib et al., 2018, population E from Solé-Boldo et al., 2020 and He et al., 2020, and population F from Vorstandlechner et al., 2020–. Additionally, we could not discern the A axis in He et al., 2020 using the manual annotation. Moreover, although the consistency of the populations was promising, it was only proven in 4 datasets, and we were unsure if samples from other body parts and conditions would result in the same populations.

Regarding the mouse samples, at the time of developing the primary analysis, 5 datasets were used. The main problem from that analysis was that, although most of the populations were present across datasets, the number of cells was highly variable. For instance, axes a and b in Phan et al., 2020, or axes c and d in Buechler et al., 2021 were almost non-existent in comparison to the rest of the datasets; or the relationships between populations from axes a and b in Haensel et al., 2020 were unrelated from Abbasi et al., 2020 or Shook et al., 2020. We were not sufficiently comfortable with this approach and were unsure if the initial manual labels were correct. To gain more confidence in these results, we needed to include new datasets that were made publicly available and search for fibroblast signatures that could be compared to the existing ones in the literature.

After the inclusion of new datasets into the analysis, the number of human datasets used in this part of the analysis was incremented from 4 to 25, and the number of mouse datasets from 5 to 9. We developed a dataset integration pipeline to integrate large numbers of datasets, which outputs robust populations and markers. This pipeline is iterative; that is, after the addition of new datasets, some new populations may be detected or relabelled. This information is reflected in a new iteration of the pipeline. Therefore, because the analysis presented in this section has gone through several cycles since the development of the primary analyses, we will only show the results of the latest analysis at the time of writing this thesis.

In this chapter, we will describe the heterogeneity and characteristics of populations across datasets, similarly to the previous chapter, and the relationships and similarities between them based on marker expression and neighbouring populations. We will also try to determine the location of some relevant populations and explore the comparison between human and mouse fibroblasts. These results will be relevant for the Discussion section 12, where we will expand on the putative functions of each fibroblast population.

10.1.1 *Description of new human datasets*

A total of 25 datasets of human dermal fibroblast were used in this part of the analysis. The basic demographic characteristics of each dataset's samples are described in Table 7. Most datasets are related to samples in disease states, and some share interesting findings shaping the results from the secondary analysis and discussion.

Additionally, certain datasets originally included in the pipeline were excluded from further analysis due to several reasons.

- Alkon et al., 2022: the dataset was discarded because we found no cells expressing the common fibroblast signature at close inspection.
- Capolupo et al., 2022: the dataset was discarded because fibroblasts were cultured and showed a completely altered transcriptomic profile compared to non-cultured cells.
- Saluzzo et al., 2021: the dataset was discarded because there were too few fibroblasts (< 100).

Ahlers et al., 2022 In this study, the authors identified four dermal fibroblast populations using unsupervised clustering. These populations were labelled as previously described populations of fibroblasts (Solé-Boldo et al., 2020), including reticular, papillary, pro-inflammatory, and mesenchymal fibroblasts. The distribution of these populations varied with age, with reticular fibroblasts being more prevalent in older participants, while pro-inflammatory and mesenchymal fibroblasts being more common in younger participants. Within the mesenchymal population, two subpopulations were observed: DP (*COCH*$^{+}$ *CRABP1*$^{+}$), and a secondary, *COL11A1*$^{+}$ *DPEP1*$^{+}$ which was assigned as DS. Combining all the individual findings, the study revealed a decrease in DS cells upon ageing, significantly reducing their stemness score. The study also identified several key genes, including *COL11A1*, *TNMD*, and *SPARC*, strongly upregulated in the DS population.

Further analysis of these genes suggested that they play a role in regulating stem cell characteristics. For example, the knockdown of COL11A1 decreased adipogenic and chondrogenic differentiation capacity, while the upregulation of COL11A1 was associated with increased differentiation capacity and improved wound healing. The findings also suggest that geroprotective effects of HES1 might be mediated by its role in stem cell features of the human DS.

Deng et al., 2021 The study used single-cell RNA sequencing analysis to examine keloid, a type of fibrotic skin disease. The research found that keloid fibroblasts could be separated into four subpopulations (papillary, reticular, mesenchymal and immune, as in Solé-Boldo et al., 2020, with one subpopulation expressing mesenchymal cell markers significantly increased in keloid compared

to normal scar tissue. Unsupervised clustering identified 13 subclusters among all keloid and normal scar fibroblasts, suggesting further heterogeneity.

The study also revealed that genes associated with skeletal system development, ossification, and osteoblast differentiation–*POSTN* and *COL11A1*–significantly increased in the mesenchymal subpopulation of keloid. Furthermore, some transcription factors associated with osteogenesis, chondrogenesis, and ligament and tendon differentiation were enriched in the mesenchymal fibroblast subpopulation–*SCX*, *CREB3L1*, and *RUNX2*–, indicating the mesenchymal characteristics of this subpopulation. Myofibroblasts were increased in keloids compared to normal scars and enriched in the mesenchymal fibroblast subpopulation. Interestingly, only a portion of mesenchymal fibroblasts expressed *ACTA2*, an indicator of myofibroblast differentiation, concluding that part of mesenchymal fibroblasts were myofibroblasts, and most of the myofibroblasts were in the mesenchymal fibroblast subpopulation in keloid.

Gur et al., 2022 Systemic sclerosis (SSc) is a rare chronic autoimmune disease that primarily affects women between 30 and 50 years old, with higher morbidity and mortality rates than other rheumatic diseases. SSc is characterised by heterogeneity in clinical manifestations, autoantibody profiles, disease progression rate, treatment response, and patient survival. To further understand this disease, Gur et al., 2022 reported the first population-scale single-cell RNA sequencing profiling of skin biopsies derived from healthy controls, SSc patients at different stages of disease progression, and patients with other fibrotic skin diseases.

The study primarily focused on immune populations but also analysed stromal populations. Comparing the stromal cell compartment of SSc patients with that of healthy subjects, the study found significant perturbations in fibroblasts, vascular, and pericyte lineages. One striking finding was the reduction of LGR5+ fibroblast subpopulation, which was substantially reduced in dSSc patients compared to healthy subjects and lSSc patients. The study defined this cell population as ScAFs (SSc associated fibroblasts). The ScAFs of dSSc patients showed substantial differences in gene expression compared to the ScAFs of lSSc patients, like the upregulation of genes associated with excessive production and deposition of specific ECM components–*COL1A1*, *COL3A1*, *COL5A1*, *EFEMP1*, *BGN*, *POSTN*, and *COMP*–, while downregulating genes associated with protection against oxidative damage–*GPX3*, *MGST1*, and *SCARA5*–and adipogenesis–*ADIRF*–.

The study also detected upregulation of COL1A1 in dSSc ScAFs and other fibroblasts in the dermis, consistent with the scRNA-seq data. Additionally, the ScAFs in SSc patients were found to play a key role in aberrant L-R interactions associated with lipid metabolism–*APOD-LEPR*, textitAPOE-*SCARB1*–, matrix deposition–*CTGF-LRP1*–, and activation of profibrotic Wnt and TGF-β signalling–*LGR5-RSPO1/3/4* and *TGFBR2-TGFB1*–. Although it is not further investigated, the authors find two $MYOC^+$ populations, one overpopulated in healthy samples and the other in SSc samples.

Tabib et al., 2021 Similar to Gur et al., 2022, in Tabib et al., 2021 SSc is also studied. Compared to the study on healthy skin conducted by the same laboratory (Tabib et al., 2018), they used an updated clustering algorithm and analysed four additional discrete skin samples. The authors observed the same cell populations as previously described but also identified some additional populations of cells. For instance, SFRP2/DPP4-expressing fibroblasts were separated into $WIF1^+NKD2^+COL6A5^+APCDD1^+$ fibroblasts and $PCOLCE2^+CD55^+SLPI^+$ fibroblasts; APOE expressing fibroblasts were divided into previously described $MYOC^+FMO1^+$ fibroblasts and two additional pop-

ulations expressing *C7* and *CCL19* respectively. They also described already studied populations based on previously described murine and human fibroblast markers like the *CRABP1*$^{+}$*COCH*$^{+}$ DP and *COL11A1*$^{+}$*ACTA2*$^{+}$ DS, and two *SFRP4*$^{+}$ small populations, once expressing *ANGPTL7* and the other *LINC01133*. Finally, they discovered two new populations near DS/DP clusters: one expressing *ASPN* and *POSTN*, and the other expressing *PTGDS*.

Focusing on SSc, the authors observed that most fibroblasts from SSc patients clustered separately from the control subjects. SSc fibroblasts clustered prominently in the SFRP2hi/PRSS23+ fibroblasts and a discrete region within CCL19+ fibroblasts. The authors also noted that SFRP2hi/WIF1+ fibroblasts were largely depleted in SSc skin with the appearance of SFRP2hi/PRSS23+ fibroblasts in the adjacent subcluster. Focusing on SFRP2hi/PRSS23+ fibroblasts, the authors identified some highly upregulated genes, some of which were recognisable as genes previously shown to correlate with the severity of SSc disease. The authors broadly observed two patterns of gene expression in the SSc SFRP2hi/PRSS23+ fibroblasts; genes were either expressed by most cells in this cluster–*PRSS23*, *THBS1*, and *TNC*–or expressed by a subset of cells in this cluster–*SFRP4*, *ADAM12*, *TNFSF18*, *CTGF*, *FNDC1*, *COL10A1*, and *MATN3*–. αSMA staining of myofibroblasts showed costaining with SFRP2 and SFRP4 indicating that the *SFRP4*$^{+}$ subcluster might be myofibroblasts related to SSc.

Interestingly, and contrasting with findings from Gur et al., 2022, DS cluster present in this study was not upregulated in SSc samples. This might imply that different SSc manifestations arising from different fibroblast subtypes is possible.

Theocharidis et al., 2022 Diabetic foot ulceration (DFU) is a common complication among diabetic patients, with over 15% of them expected to develop DFUs during their lifetime. These ulcers often result from impaired wound healing, usually exclusive to the feet and associated with neuropathy and/or vascular disease. In this study, Theocharidis et al., 2022 sought to identify the differences between DFU patients whose ulcers heal (DFU-Healers) and those whose ulcers do not heal (DFU-Non-healers) within 12 weeks, with a focus on gene and protein expression profiles that affect remodelling and inflammation pathways.

The study discovered significant differences in the transcriptome profile of fibroblasts between the DFU-Healers and DFU-Non-healers, with a unique population of fibroblasts overrepresented in the samples from DFU-Healers. These were referred to as Healing Enriched-Fibroblasts (HE-Fibro). Further analysis of the HE-Fibro cluster revealed high expression of multiple ECM remodelling–*MMP1*, *MMP3*–and immune/inflammation-associated–*CHI3L1*, *TNFAIP6*–genes. Further studying identified four heterogeneous states or subtypes of HE-Fibro that were significantly enriched with cells expressing genes related to ECM remodelling, adhesion, and migration–*POSTN* and *ASPN*–with simultaneous detection of *IL6*, *TIMP1*, *PLA2G2A* and *CHI3L1* transcripts suggestive of a common inflammatory response mechanism in HE-Fibro.

Pathway analysis indicated the activation of multiple immune and inflammatory pathways in the fibroblasts from DFU-Healers, including *IL6*, *HIF1A*, and *ILK* signalling. The study further revealed that HE-Fibro were enriched in the wound bed but not in the wound edge and non-wounded samples, suggesting their association with the wound healing process. Overall, the study highlights the heterogeneity of fibroblasts across different regions of ulcers and provides insights into potential therapeutic targets for DFUs based on the identified gene and protein expression profiles in HE-Fibro.

VORSTANDLECHNER ET AL., 2021 Hypertrophic scars are a complex and significant global disease burden. The wound healing process consists of three stages, with prolonged inflammation leading to increased activity of fibroblasts and increased secretion of cytokines such as TGF-β, and insulin-like growth factor (IGF1), ultimately resulting in hypertrophic scars. In this study, Vorstandlechner et al., 2021 used scRNAseq to examine the gene expression and mechanisms involved in hypertrophic scar formation in both human and mouse skin. The study aimed to identify genes regulated in scar tissue and uncover potential targets for drug development towards scar-free wound healing.

The scRNAseq analysis of human hypertrophic scar tissue and healthy skin identified 11 clusters showing 110 significantly upregulated genes in fibroblasts derived from scar tissue compared to healthy skin. Interestingly, one cluster–FB1–was almost exclusively present in hypertrophic scars, suggesting a specific role in tissue fibrosis. The upregulated genes are related to collagen and extracellular matrix-modifying genes, including *BGN*, *COL14A1*, *COL1A1/2*, *COL3A1*, *COL5A1/2*, *FN1*, *MMP23B*, *OGN*, and *PCOLCE*. Analysis revealed a strong association of one FB cluster with TGF-β-signalling and ECM-formation, further supporting its role in skin fibrosis. Moreover, the analysis indicated the cluster's role in processes important for several other cell types, including platelets, smooth muscle cells, and cells of the skeletal system.

Similarly, when the murine scRNAseq dataset was clustered, 8-week-old mouse scars contained a higher proportion of murine FBs as well as an increased expression of matricellular and ECM-modulating proteins–*Fbln1*, *Ogn*, *Lum*, *Pcolce*, and *Tgfbi*–.

The study next focused on the identified serine proteases DPP4 and urokinase, as specific inhibitors are commercially available only for these two serine proteases. The researchers performed siRNA-mediated gene knockdown of DPP4 and PLAU in primary FBs from healthy human skin, significantly downregulated the genes' mRNA expression levels. The knockdown reduced TGF-β1-mediated expression of αSMA, accompanied by a reduced ability to contract a collagen matrix in vitro. The specific inhibitors for DPP4 (Sitagliptin) and PLAU (BC-11) abolished TGF-β1-induced αSMA production to a comparable degree as the specific gene knockdown. Sitagliptin and BC-11 also attenuated TGF-β1-induced overproduction of ECM-proteins Col1a1 and fibronectin by FBs. The study suggests that Sitagliptin and BC-11 interfere with matrix deposition *in vivo*, representing promising candidates for improving hypertrophic scarring.

10.1.2 *Description of new mouse datasets*

Nine datasets of mouse dermal fibroblast were used in this part of the analysis. The basic demographic characteristics of each dataset samples are described in Table 8. Most of the studies use single-cell as a side element to the main results from the experiments and therefore lack significant relevance in shedding light on the fibroblast heterogeneity.

Additionally, certain datasets were excluded from further analysis due to several reasons.

- Efremova et al., 2020: the dataset was discarded because the number of cells was small, and the transcriptome of the populations failed to match any of the consensus populations.
- Lin et al., 2022: the dataset was discarded because the number of cells was too low to be suitable for the analysis.

- Leyva-Castillo et al., 2022: the dataset was discarded because of large differences in the transcriptomic profiles of cells. This is probably because the mouse strain is BALB/c, while the strain of most datasets is C57BL/6J.
- Salzer et al., 2018: the dataset was discarded because the number of cells was too low to be suitable for the analysis. Upon closer inspection, the results showed two clusters that did not resemble transcriptomically the main fibroblast axes in mice.

Boothby et al., 2021 In the study by Boothby et al., 2021 the authors explore the impact of a reduction in neonatal regulatory T cells (neoTreg cells) on developing the immune system in murine skin. The researchers found that a temporary reduction in Treg cells caused dysfunction in subcutaneous tissue and expansion of an unknown stromal cell type. Using single-cell RNA sequencing, the researchers identified two clusters of stromal cells highly increased in the skin with reduced neoTreg cells and expressed Th2 cytokine receptor genes *Il4ra* and *Il13ra*. These stromal cells were identified as Th2-interacting fascial fibroblasts (TIFFs) and were responsible for forming fibrous bands seen in the skin of mice with reduced neoTreg cells.

The study also found that the expansion of TIFFs was induced by subcutaneous injection of Th2 effector cytokines IL-4 and IL-13, and IL-33 in young wild-type mice. Furthermore, the researchers isolated stromal cells from healthy human skin and identified two clusters of potential fascial fibroblasts (fFB1 and fFB2) that expressed murine TIFF markers–*PI16*, *SFRP2* and *OGN*–and immunological genes–*IL13RA1*, *CXCL12* and *CXCL14*–. The human fFB clusters were found to have a strong transcriptional similarity to mouse TIFFs and THBS4+ fibroblasts.

Buechler et al., 2021 In this paper, Buechler et al., 2021 create a fibroblast-specific single-cell atlas by collecting mouse scRNAseq datasets from their own lab and public repositories (~120,000 cells). After cleaning the data and correcting for batch effects, they identified ten clusters based on differential gene expression. Each cluster was annotated based on its dominant cluster-specific gene–*Pi16*, *Col15a1*, *Ccl19*, *Coch*, *Comp*, *Cxcl12*, *Fbln1*, *Bmp4*, *Npnt* and *Hhip*–, and the functional identity of most clusters was ascribed using tissue-distribution patterns and hallmark genes. From these populations, it was observed that the $Pi16^{+}$ and the $Col15a1^{+}$ populations were present in all tissues and were termed as "pan-tissue" fibroblasts.

Based on their gene signatures and surface markers–SCA1 for $Pi16^{+}$ and LY6C and $Col15a1^{+}$–$Pi16^{+}$ were suggested an adventitial stromal cell role, and $Col15a1^{+}$ fibroblasts a basement membrane role. The ubiquity of both populations was suggested to be due to the elevated stemness potential.

The authors investigated how fibroblasts were affected by infection, injury, cancer, fibrosis, metabolic changes, and arthritis. They created a perturbed-state fibroblast atlas by integrating 17 publicly available scRNA-seq datasets across 13 tissues. The analysis showed ten clusters of fibroblasts in the perturbed-state atlas. Most of these clusters were present at the healthy atlas, like the universal $Pi16^{+}$ and $Col15a1^{+}$ fibroblasts, as well as tissue-specific ones like $Cxcl12^{+}$, $Ccl19^{+}$, $Comp^{+}$,$Npnt^{+}$, and $Hhip^{+}$; and other populations like the $Cxcl5^{+}$,$Adamdec1^{+}$, and $Lrrc15^{+}$ appeared to be perturbation-specific. Thus, the authors provided evidence that the $Pi16^{+}$ and $Col15a1^{+}$ clusters were universal, not only in healthy but also in diseased tissue.

Finally, the authors also investigated fibroblast heterogeneity in human tissues by selecting three tissues–pancreas, intestine and lung–and sequencing cells in diseases and normal adjacent tissue. It

was observed, for instance, that mouse *Lrrc15*$^{+}$ population seemed to have its human *LRRC15*$^{+}$ analogue and represented a myofibroblastic state. Another myofibroblast population expressing *COL3A1* was found.

10.2 HUMAN DERMAL FIBROBLASTS ARE DIVIDED INTO 5 MAIN AXES AND 15 POPULATIONS

Human dermal fibroblasts are divided into 5 axes, A to E. UMAP plots of human fibroblast populations across datasets (Figure 42) reveal a consistent pattern of division in two main axes: axis A and axis B. In most, but not all datasets–Reynolds et al., 2021, Theocharidis et al., 2020, Vorstandlechner et al., 2020 do not show–there is a third major axis, C. The presence of this axis is also variable since there are some datasets–Theocharidis et al., 2022, Vorstandlechner et al., 2021–where there is a vestigial number of cells. Lastly, there are two other axes, D and E, which have fewer cells than other axes. For instance, the E axis is only clearly visible in 10 datasets, whereas in 8, it is integrated within other cell types, or is composed of a minority of cells.

10.2.1 *Axis A*

Axis A is divided into four main populations: A1 to A4. Population A1 is defined by the expression of *WISP2*, *SEMA3B*, *LGR5* and *ANGPTL5*; A2 by *COL18A1*, *NKD2*, *COL6A5* and *HSPB3*; A3 by *WIF1*, *SGCA*, *CORIN* and *SOSTDC1*; and A4 by *C1QNTF3*, *SCARA5*, *PRG4*, and *TRAC* (Figure 43).

Based on UMAP populations, as well of individual and combined PAGA graphs–Figures 44 and S2, and 45 respectively–there is a common agreement that all four A axis populations form a continuum. A1 and A2 are originally the most prominent and archetypal populations; if the A axis had to be divided into 2 populations, it would be these two. A3 is a bridge population between A1 and A2. In fact, most of the A3 markers are also expressed in A1 and, to a lesser extent, in A2–with a few exceptions such as *WIF1* or *SOSTDC1*, which are mainly expressed in A3–. Lastly, A4 is a population close to A1–most A4 and A1 markers are cross-expressed–, but its own marker expression–e.g. *PRG4* or *TRAC*–separates it from A1. Based on UMAP plots and PAGA graphs, A4 seems like a *differentiated* state, or a subtype, of A1. Generally, the A4 population comprises fewer cells and is in a terminal UMAP location of A axis cells.

Interestingly, since the A1 population is transcriptomically related to A3 and A4, it is complicated to define a core transcriptomic profile for this population since most of the A1 markers will be expressed either in A3, A4 or both. The exception to this rule are a few markers with low expressions, such as *EYA2* or *MTCL1*. Regarding the location of the populations within individual PAGA graphs (Figure 44), the relationships are generally consistent across datasets–when A2-A3-A1 appear together, A3 goes between A1 and A2, and when A4 is present, it goes next to A1–. Interestingly, in some datasets, we observe that A2 is sometimes linked to clusters from the B axis in Figure 44, although this might be, in part, an artefact derived from the constraint of building a tree graph.

Lastly, all four populations are assigned to axis A using the unsupervised population assigning method (Table 9). This indicates that axis A has a clear and robust transcriptomic profile across datasets. Additionally, these populations have remained unchanged since the primary analysis (Figure 46).

10.2.2 *Axis B*

Axis B was originally divided into two populations: B1 and B2 (Figure 46). During the iterations in the secondary analysis, we realised that there was an independent population similar to B1, but with its own transcriptomic profile–especially the expression of *ITM2A* and *MYOC*–, which is now termed B4. This population was originally present in Tabib et al., 2018 but was merged with B1 because there was not enough evidence to consider it a separate population, based on the other three datasets. Secondly, in some datasets–especially in Kim et al., 2020a and Solé-Boldo et al., 2020–, we observed a bridge population between B1 and B2, termed B3. Interestingly, we originally thought that *CCL19* was a clear marker of the B2 population, but in many datasets, we observed that there was a $CCL19^{+}$ population that also expressed markers of B1–e.g. *IL6* or *CXCL2*–, which moved us to assign this division with a bridge cluster, in a similar fashion to what occurs in axis A. Like population A3, the limitation of establishing bridge clusters is that their assignment might be inaccurate some times, as we will see later, in Section 10.8.

Therefore, axis B is divided into 4 populations: B1 to B4. Population B1 is defined by the expression of *GEM*, *CXCL2*, *TNFAIP6* and *CXCL1*; B2 by *GGT5*, *IL33*, *C7* and *SCN4B*; B3 by *CCL19*, *CTSH*, *RBP5* and *ACHE*; and B4 by *EFEMP1*, *ITM2A*, *MYOC*, and *GDF10*. Based on marker expression, despite B3 being a bridge between B1 and B2, B3 markers tend to be more expressed in B2–surprisingly, even "canonical" B2 markers such as *CCL19* or *CD74* show a high expression in B3–. B3 is termed as the bridge and not B2 because B1 markers such as *CXCL2* and *CXCL3* are also expressed in B3, not in B2.

When looking at the PAGA graphs and UMAPs, the relationships between populations are not as clear as with axis A. In UMAPs we generally see that B3, when present, appears next to B1 and B2, or as a *mixture* of both, and not as an independent population–e.g. in Deng et al., 2021; Gao et al., 2021; Kim et al., 2021; Liu et al., 2021a; Mirizio et al., 2020; Rindler et al., 2021; Solé-Boldo et al., 2020; Theocharidis et al., 2020, 2022; Vorstandlechner et al., 2020, Figure 42–. This effect is replicated in individual PAGA graphs from the mentioned datasets (Figure S2). Interestingly, in some individual PAGA graphs–Hughes et al., 2020; Tabib et al., 2018; Theocharidis et al., 2022; Vorstandlechner et al., 2021–the B1 population looks separated from the rest of the B axis populations–although this effect is not apparent in UMAPs, or partially replicated in complete individual graphs (Figure S2). This is probably an artefact of branching a complexly interlaced graph, as with the A2 population. Regarding the B4 population, it seems to interact with the rest of the axis B populations–especially with B2 and, to a lesser extent, with B1–, as well as with the E1 population, as clearly seen in joint PAGA graphs from Figure 45. This effect is also seen in individual PAGA graphs and in UMAPs–e.g. in Ahlers et al., 2022; Gao et al., 2021; Gur et al., 2022; Rindler et al., 2021; Solé-Boldo et al., 2020; Theocharidis et al., 2020; Vorstandlechner et al., 2020, 2021–, where B4 and E1 appear close together.

Lastly, unsupervised axis assignation in Table 9 reveals that axis B populations are overall assigned to axis B, although some cases, like in B1 and B4, are assigned both to A and B in certain datasets.

10.2.3 *Axis C*

Axis C was originally divided into four populations: C1 to C4. However, during the addition of new datasets, we observed that the C4 population did not express most of the typical C axis markers–e.g. *TNN*, *ASPN*–and, indeed, the unsupervised classification algorithm categorised C4 into an Unknown

axis (data not shown). As a result, we decided to separate C4 into a new axis, D. Addition of further datasets–Deng et al., 2021; Gaydosik et al., 2019; Kim et al., 2021; Tabib et al., 2021; Theocharidis et al., 2022–showed that the C4 population could be divided into two populations, which were not observed initially in the primary analysis with Tabib et al., 2018 or Solé-Boldo et al., 2020 datasets; now termed as D1 and D2. Additionally, in some datasets–e.g. Deng et al., 2021; Gao et al., 2021; Hughes et al., 2020; Kim et al., 2021; Liu et al., 2021a–we observed a small population that expressed *INHBA* and *LUZP2* that could be separated from the rest of populations and was termed as C5. Currently, there is no C4 population to keep a mapping consistency between the populations from the primary and secondary analyses.

Therefore, axis C is divided into 4 populations: C1, C2, C3, and C5. C1 is defined by the expression of *COL11A1*, *MEF2C*, *DPEP1*, and *WFDC1*; C2 by *COCH*, *CRABP1*, *NDNF*, and *SLITRK6*; C3 by *POSTN*, *BGN*, *LRRC15*, and *LTBP2*; and C5 by *IGFBP3*, *SLC5A3*, *WNT5A* and *LUZP2*. When looking at the expression patterns, we see that C1 and C2 have their own transcriptomic profile, but C3 does not–with the slight exception of *POSTN*, which is more expressed in C3 than in C1 or C2–, and is instead defined as a population that is of the axis C, but which does not express markers from C1, C2 or C5. Unlike A3 or B3, the C3 population is not a mixture of C1 and C2 *per se*, because it does not express their markers, or they are expressed marginally. This effect is partially observed in PAGA graphs from several datasets–Deng et al., 2021; Gao et al., 2021; Gur et al., 2022; Hughes et al., 2020; Kim et al., 2021; Solé-Boldo et al., 2020; Tabib et al., 2018; Xu et al., 2021b, where C3 is a node between C2 and C1, and sometimes with more populations. This effect is replicated in joint graphs, especially in the one created from the tree PAGA graphs. C5, when present, appears to branch from C1, C2 or C3 without a clear pattern. This effect is partially replicated in UMAPs, although this population tends to appear isolated within the axis C vicinity.

Unsupervised axis assignation reveals that axis C populations have a more diffuse assignment pattern, being also assigned to the A and E axes. This may be due to a lack of clear axis markers or interaction with those axes, like C1 and C5 and E1. Regardless, we still assign these populations to the C axis, first to keep the consistency with the primary analysis and second because, as further discussed in the Discussion section, populations from the axis might have a clear biological function related to HF.

10.2.4 *Axes D and E*

As explained before, D1 and D2 populations arise from the branching of the C4 population from the primary analysis. Additionally, with the addition of new datasets–e.g. Deng et al., 2021; Gur et al., 2022; Hughes et al., 2020; Liu et al., 2021a; Theocharidis et al., 2022– we observed a new population expressing *RAMP1* and *OLFML2A* that differed from any other population from the rest of the datasets, and was thus assigned to a new axis, E. This is the only axis consisting of one population, E1.

D1 population is defined by the expression of *ANGPTL7*, *ENTPD2*, *CDH19* and *ATP1A2*; and D2 population is defined by the expression of *BNC2*, *ITGA6*, *ITGB4* and *TNNC1*. Although these two populations are different, there are not many clear markers that separate one from the other, and instead, most markers are co-expressed by the two populations. Lastly, E1 population is defined by the expression of *IGFBP2*, *RAMP1*, *COL26A1* and *WNT2*. Since there are few populations within each axis, unsupervised

axis assignation shows a clear assignation pattern of D1 and D2 to axis D and E1 population to axis E.

Regarding the PAGA graphs, the joint graphs show that D1 and D2 populations are quite independent of the rest. It may be argued that they are related to a certain extent to B axi based on the graph in Figure 45B; however, this links are not well preserved in the combined graph derived from PAGA trees (Figure 45A).

On the other hand, the E1 population shares a high degree of similarity with the axes B, C and D–especially with populations B4 and C3–. This probably indicates that the E1 population shares common functions with other axes and might be a *multipurpose* population.

10.2.5 *T1 population*

The T1 population arose during the secondary analysis as a population that failed to be classified with the marker-to-population algorithm and did not show a clear expression pattern. When looking at the UMAPs–Deng et al., 2021; Gaydosik et al., 2019; He et al., 2020; Kim et al., 2021; Mariottoni et al., 2021; Rindler et al., 2021; Solé-Boldo et al., 2020; Tabib et al., 2021–T1 population seems to be a mixture, or bridge, between B and C axes. In fact, marker expression consists mainly of markers from the populations C1 and C3–e.g. *ADAM12* (C3), *ASPN* (C), *DKK3* (C1), *DPEP1* (C1), *F2R* (C3), *LGR4* (C1), *MDK* (C3), *MME1* (C1), *UGT3A2* (C1)–, which might indicate that these cells are either a bridge between axes–like an *undifferentiated* population–, or a processing artefact. Although this is an interesting population, it is not prioritised for this thesis.

10.3 MURINE DERMAL FIBROBLASTS ARE DIVIDED INTO 5 MAIN AXES, 3 BRIDGE AXES, AND 17 POPULATIONS

In this section, we will explain the evolution of mouse populations from the primary to the secondary analysis and describe each population and its relationship to the others, as we have already done in human datasets. Mouse dermal fibroblasts are divided into 5 major axes (v to z) and 2 bridge axes (w/x and x/y), consisting of 17 populations, a similar number to the 15 human populations.

Axis nomenclature in the secondary analysis is changed from a to e, to v to z. Most prominent axes–c, b and a–are renamed to x, y, and z, respectively; the d axis is renamed to w and e to v. This is because of two main reasons. First, we observed that, despite the differentiation of lower casing in mouse populations (a1, a2, a3) and upper casing in human populations (D1, D2), this nomenclature was sometimes confusing and could induce indirectly associating axis A in humans to axis a in mice, and similarly to the rest of the axes, when there is no direct relationship between them. Therefore, renaming axes with a new set of letters while keeping the difference in casing, would allow less biased comparisons.

Additionally, the naming of the nomenclature was inspired by the findings of Joost et al., 2020. In their work, the authors analyse the composition of skin cells of hair follicles in anagen and telogen phases, including dermal fibroblasts. Regarding the fibroblast subpopulations, they separate them into 7 populations: FIB1 to FIB4, DS1, DS2 and tDP. FIB populations are dermal fibroblasts, DS1 and DS2 are dermal sheath cells –in anagen and telogen phases–, and tDP are cells of the DP in telogen phase (Figure 47A). Looking specifically at the location of the FIB populations using immunofluorescence,

they observe FIB1 and FIB2 to be located within the dermis, where FIB1 has more presence during anagen and FIB2 during telogen; FIB3 is located within the hypodermis, and FIB4 in the adventitial layer beneath the *panniculus carnosus* (Figure 47C)–Figures 6F to 6H in the publication (Joost et al., 2020)–.

Based on this characterisation, we renamed the fibroblast populations from these axes to match the layering observed by Joost et al., 2020. Axis x matches the expression profile of FIB1 and FIB2 fibroblasts–*Col1a1*, *Sparc Dcn*–, axis y matches the expression profile of FIB3–*Cxcl12*, *Gpx3*–, axis x/y is a mixture of markers from FIB1/2 and FIB3, and axis z matches the expression of FIB4–*Mfap5*, *Plac8*– Figures 47B and 47E. Populations v and w, since they are not directly associated with this layering–axis w maps to DS and DP cells–, their letters were arbitrarily assigned. Lastly, population y5, based on the marker expression, is unlikely to be associated with FIB3 and FIB4 populations, and it is possibly associated with FIB2.

10.3.1 *Axis x*

Axis c, now renamed as x, consisted of 3 populations–c1 to c3–and a mixed axis b/c. After the addition and reanalysis of new datasets, we observed that c3 and d5 populations shared a common transcriptomic profile, so they were merged into the mixed axis c/d, now w/x. No alterations were observed after adding new datasets for the rest of the populations, so they are kept with the same labels–c1 to x1, c2 to x2, and b/c to x/y– (Figure 52).

Axis x populations are defined as follows (Figure 49): population x1 is defined by the expression of *Fgfr4*, *Gpha2*, *Cib3* and *Serpina3m*; population x2 is defined by the expression of *Igfbp2*, *Stc1*, *Sema3a* and *Enho*; population x/y is defined by the expression of *Crp*, *Akr1cl*, *Lgr5* and *Mup20*; and population w/x is defined by the expression of *Coch*, *Emid1*, *Kera* and *Ntn5*.

Regarding the structure and relationship with populations from axis x, it is clear that x2 is a terminal state of x1; that is, x1 and x2 are always related, and x2 is in terminal points of UMAPs and in individual tree graphs. Although x2 is related to other populations, x1 shows stronger relationships, especially with x/y and w/x.

Unsupervised axis assignation of populations from axis x reveals that both populations are unanimously assigned. w/x mixed population seems to be assigned to all axes, and x/y population seems to be assigned to either x or y.

10.3.2 *Axes y and v*

Axis b (now axis y) was originally divided into six populations–b1 to b6–and a mixed axis a/b. After a reanalysis of markers, we observed that markers from the b2 population were inconsistent across datasets in newer iterations. The marker-to-population algorithm failed to assign that population accurately. Therefore, it was removed from the dictionary. Consequently, clusters b3, b4 and b5 were renamed b2, b3 and b4; and now are set as y2, y3 and y4. After adding Joost et al., 2020 and Phan et al., 2020 datasets, we observed that the b6 population was assigned to two different clusters. Therefore, a reanalysis of the different clusters revealed that there was a primary cluster–b5, now termed as y5–which shared an axis y-like transcriptomic profile and a secondary and less-frequent cluster that was not unsupervisedly assigned to axis y and was therefore termed as v1. Lastly, the mixed a/b axis,

originally termed like that for being a bridge between axes a and b, was often assigned to b1, now y1, after adding the new datasets. This merging also occurs because, originally, a/b was only assigned in Shook et al., 2020 dataset. Therefore, a/b and b1 were joined into the y1 population.

Axis y populations are defined as follows: population y1 is defined by the expression of *Postn*, *Fabp4*, *Cd36* and *Pparg*; population y2 is defined by the expression of *Hmcn2*, *Col6a6*, *Fbln7* and *Bmp5*; population y3 is defined by the expression of *Ccn5*, *Ccn2*, *Ecrg4* and *Fgf9*; population y4 is defined by the expression of *C2*, *C4b*, *Chrdl1* and *Gdf10*; population y5 is defined by the expression of *Vwa1*, *Vit*, *P2ry14* and *Kcnk2*. Lastly, population v1 is defined by the expression of *Cldn1*, *Moxd1*, *Tenmd2* and *Krt19*.

The net of interactions within and with other axes is complex, almost behaving like a clique. Based on the joint PAGA graphs, individual PAGA graphs, and UMAP plots, y2 seems to be a nexus of interactions between y3, y4, and, to a lesser extent, y1; and y1 also interacts with y4. y5 is closer in interaction to y4. Therefore, axis y is highly connected, probably due to the large number of populations within it.

Lastly, population v1 seems to interact with w3, w5 and y5; however, the interactions are less confident due to this population's presence in only two datasets, and it is not reflected in the combined PAGA graphs, where it interacts only with w5. In section 10.4, we observe that y5 and v1 populations have similar correlates in the human axis D so that this relationship could make sense. Except for y5, all axis y populations are unanimously assigned to that axis based on unsupervised assignation. Population b5 is either assigned to y or v—in line with what was explained previously–.

10.3.3 *Axis z*

Axis z (previous axis a) was divided into three populations: a1, a2 and a3. Population a1 and a2 are now renamed z1 and z2, with no changes. Changes in population a3 will be described on axis w.

Axis z populations are defined as follows: population z1 is defined by the expression of *Ptx3*, *Ptgs2*, *Has1* and *Il16*; and population z2 is defined by the expression of *Akr1c18*, *Aldh1a3*, *Chst1* and *Itgb7*.

There is some variation between datasets regarding the topology and relationships between populations. In the UMAPs of Abbasi et al., 2020; Shook et al., 2020; Vorstandlechner et al., 2020 datasets, we see that these two clusters are independent, whereas in the rest of datasets–Boothby et al., 2021; Buechler et al., 2021; Haensel et al., 2020; Joost et al., 2020–axis z is integrated next to axis y and tends to be close to populations y2 and y3. The joint tree PAGA graph in Figure 51A–and individual PAGA graphs from some datasets show that z1 and z2 share a high resemblance and that axis z links to axis y mainly via the interaction between z1 and y1 or y4.

Based on the unsupervised axis assignation, z1 and z2 are mainly assigned to the z axis, although the z1 population is sometimes assigned to the y axis, maybe due to the z1-y1 interaction described before.

This discrepancy between z being a separate entity in certain datasets and being integrated within y axis in others is intriguing and may underlie a biological basis. Considering z axis to represent fibroblasts below the *panniculus carnosus*–and immediately above–and axis y to represent fibroblasts within the hypodermis, a *mixture* may represent that these two layers come in close contact and may interact with each other.

10.3.4 *Axis w*

During the primary analysis, axis d was divided into 5 populations–d1 to d5–. As previously explained, d5 was merged with c3 into the c/d, now w/x mixed axis. After adding new datasets, especially Joost et al., 2020, and Shin et al., 2020, we observed that the d1 population was not consistently assigned, so it was removed during the secondary analysis. The unsupervised analysis of Shin et al., 2020, which contains a large set of cells from this axis, revealed 4 main clusters. These clusters, which were translatable to the rest of the datasets, showed that previous population d3 could be consistently divided into two populations–w1 and w2–and d4 and d2 populations could be merged into w3. d2 population was also assigned to w1, although not that frequently. Lastly, population w4 was not assigned to any previous cluster, so it is set as it is.

Lastly, regarding the a3 population from the primary analysis, now w5, it was originally observed only in Abbasi et al., 2020 dataset, and due to its UMAP location near axis a, it was assigned to it. Later reanalyses with other datasets–Boothby et al., 2021; Haensel et al., 2020; Shook et al., 2020–showed that this population was located within axis a and b–Haensel et al., 2020–but also within axis d–Boothby et al., 2021; Shook et al., 2020–. Additionally, unsupervised axis assignation revealed that the a3 population was assigned to either the a, b or d axes. considering all this information, the population was renamed to a/d. However, based on the analysis developed by Shin et al. (2020) (Shin et al., 2020), this population is associated with hair stem cells–based on the shared expression of the *Cdk1* gene, shown in Figure 49 of this thesis and Figure 5E of the original publication–, and therefore is renamed as w5 due to this association.

Axis w populations are defined as follows: population w1 is defined by the expression of *Chodl*, *Slc26a7*, *Edn3* and *S100b*; population w2 is defined by the expression of *Hhip*, *Ddx26b*, *Rspo4* and *Rhbdf2*; population w3 is defined by the expression of *Cox4i2*, *Pdlim3*, *Sema3g* and *Heyl*; population w4 is defined by the expression of *Actg2*, *Hck*, *Hapln3* and *Fam65b*; and population w5 is defined by the expression of *Birc5*, *Diaph3*, *Cdk1* and *Cdkn3*.

Regarding the relationships between populations, it is clear that w1 and w2 populations share similar transcriptomic profiles, based on either the UMAPs or graphs. Additionally, w3 and w4 are also connected, although w3 seems connected to w1/w2 and w/x, as observed in the condensed graph from Figure 51. Lastly, the w5 population shows an inconsistent pattern related to axes z, w and v in the PAGA graphs and UMAP plots.

10.4 COMPARISON BETWEEN MOUSE AND HUMAN POPULATIONS

In this section, we will compare mouse and human populations. Despite the differences in the skin structure between the two organisms, we assume that some degree of overlap between functions and transcriptomic profiles of fibroblasts is feasible.

To perform that comparison, we used two methods. First, we merged the samples from the two organisms in datasets that contain information from human and mouse cells–Boothby et al., 2021; Vorstandlechner et al., 2021–. Second, we established an overlap of populations based on the transcriptomic profiles of all datasets, assuming that populations with similar functions should have a higher degree of overlap in relevant markers than populations with dissimilar functions.

There are advantages and disadvantages to each method. With dataset merging, the main benefit is that samples are more likely to merge based on their transcriptomic profile; therefore, an overlap between two populations is more likely to be more reliable. On the other hand, the number of datasets where this is feasible is currently 2, so the information from the rest of the datasets is actually lost. Considering the inter-dataset heterogeneity, and possible artefacts from datasets integration algorithms, this method may not be sufficient to have a clear image of human-mouse interactions. Regarding the overlap of markers, the main advantage is that all human and mouse datasets can be used in the process, so inter-dataset heterogeneity is considered. On the other hand, the overlap is based on a sub-selection of markers and not in the full transcriptomic range, so the nuances of the expression of specific subsets of cells may be lost.

10.4.1 *Traditional batch effect correction methods fail to integrate murine and human datasets on a large subset of genes*

We aimed to merge the populations from Boothby et al., 2021; Vorstandlechner et al., 2021. We applied an ortholog mapping between mouse and human genes and a common processing scheme for each merged dataset. Then, we batch-effect correction by setting the organism as a batch to merge the datasets from both organisms. The results of the batch-effect correction method are shown in Figures 53A and 54A for Boothby et al., 2021; Vorstandlechner et al., 2021 datasets, respectively. UMAP plots show that the method failed to integrate the organism variable. UMAPs clearly show that human cells are separated from the ones in mice.

We expected that, if this is the result of the integration, some underlying transcriptomic variability was causing this effect. To check that, we obtained the list of DEGs between human and mouse cells, as shown in Figures 53B and 54B. We observe a clear transcriptomic bias in both datasets. Interestingly, this bias is partly dataset-specific: the biggest changes are due to differential expression of *RPL* and *RPS* genes, *B2M*, *MALAT1*, etc., but different across organisms and datasets. Additionally, in Boothby et al., 2021, we see the overexpression of *SOD2* or *IER3* in the human dataset, genes associated with stress response, and we also observe an increase of some fibroblast-associated markers such as *COL3A1*, *COL1A2* or *BGN* in the mouse dataset, which were reported as a possible false-positive signal due to contamination of samples ((Rojahn et al., 2020; Vorstandlechner et al., 2021)).

After this observation, we used a linear regression method to correct unwanted sources of variation from *scanpy* (`sc.pp.regress_out`). Usually, this method is not used due to the existence of more modern and sophisticated batch-effect correction methods, since it may introduce artefacts into the dataset due to overcorrection (Tran et al., 2020), and because it implicitly assumes identical cell type composition across batches (Andrews, 2020; Haghverdi et al., 2018). After the application of this method, the cells from human and mouse organisms were nearer to each other in the UMAP plot but did not have enough overlap to do the analysis (Figures 53C and 54C). Therefore, we discarded this method to analyse overlap between populations across datasets, not only because it does not work but also because applying further processing steps to merge the samples may introduce such a large amount of artefacts that hinder the certainty of the overlap between samples.

These results were initially intriguing since both Boothby et al., 2021 and Vorstandlechner et al., 2021 show a merged UMAP within their figures. To try to replicate their method, we looked at their references to see how they performed the integration. Vorstandlechner et al., 2021 showed no processing

method to integrate the samples, and Boothby et al., 2021 stated the following: "[...] genes within the resulting Seurat object were renamed with their most probable human orthologues. [...] An initial list of 313 mTIFF signature genes was generated [...]. This TIFF signature was then filtered to remove genes that were not variably expressed in human data, which resulted in a final set of 144 genes." From this sentence, the understanding is that they used a core set of genes commonly expressed in human and mouse fibroblasts. Although this approach is feasible, it is not what we seek since we lose information relevant to integrating such a reduced set of genes. Additionally, this set might bias the integration and produce false positive results.

10.4.2 *Control results in human-human and mouse-mouse marker-based comparisons*

In addition to dataset integration, human-mouse population comparisons based on the expression of common markers is also a feasible option. Before we established the comparison between human and mouse populations, we wanted to assess the comparisons within human and mouse populations. This is due to two main reasons: 1) to check that the overlap of markers between populations follows a logic pattern–e.g. A1 and A4 populations, or B1 and B2 populations should share more markers than between A1 and B2–, and 2) to assess the amount of overlap–if a common overlap in human-human and mouse-mouse dataset is of 15%, then observing a 12% overlap in human-mouse populations might be expected–. For the comparisons and their corresponding heatmaps, overlaps smaller than 0.03 are not considered.

10.4.2.1 *Human-human comparisons*

Figure 55 shows the heatmap of Jaccard index values with the top 125 markers of each population. Most of the conclusions from the heatmap are shared based on PAGA graphs, and UMAPs commented before. These results are expected and confirmatory to some extent since both UMAP and PAGA information is based on the transcriptomic profiles of the populations.

- A axis is highly isolated from the rest of the clusters, except for A2 population, which shares some markers with C and E axis (~0.05), indicating a possible interacting functionality between them. However, its relatively low similarity profile with C1, C3 and E1 makes it a somewhat "independent" population.
  - A3 population is a bridge between A1 and A2, being more *similar* to A1 (0.27) than to A2 (0.05). The most relevant overlap markers between A1 and A3 are *AOX1*, *ARFGEF3*, *COL14A1*, *CORIN*, *GLRB*, *OMD* and *SVEP1*.
  - A4 overlaps both with A1 and A3, but the overlap is more apparent with A1 (0.19) than with A4 (0.08). The most relevant markers between A1 and A4 are *ACKR3*, *CD12*, *CD55*, *CD70*, *CLEC3B*, *DBN1*, *FBN1*, *GPX3*, *IGFBP6*, *ISLR*, *LGR5*, *SLPI*.
  - No overlap between A1 and A2 is found, and some overlap between A2 and A3 is found (0.05).
  - Relevant markers overlaping between A2 and C1/C3 are *COL21A1*, *COL6A1/3*, *COL7A1*, *COMP* *LAMC3*, *LOXL2*, *PTK7*, and *SPON1*.
- B axis, similar to A axis, looks independent from most of the other axes.

  - It is clear again that B3 acts as a bridge cluster between B1 and B2, with a more apparent overlap with B2 (0.36) than with B1 (0.08). Markers shared between B2 and B3 are *ADRA2A*, *C3*, *CCL19*, *CTSC*, *CX3CL1*, *ICAM2*, *IGFBP3*, *IL34*, *NLGN4X*, *PTGDS*, *RBP5*, and *TNFSF13B*; and markers shared between B1 and B3 are *CCL2*, *CXCL1/2/3*, *GEM*, *HAS2*, and *TNFAIP3*.
  - B1 and B2 clusters are not related–similarly to A1 with A2–, and B2 is related to B4 as well (0.11): *ABCA8/10*, *APOC1*, *C7*, *EPHX1*, *GGT5*.
  - Despite the relative independence of the B axis, B1 has a partial overlap with D1 (0.04), and B4 has partial overlaps with D1 (0.07), D2 (0.06) and E1 (0.07)
- Regarding C, D and E axes, there are instances of a high correlation–e.g. C1 and C3, D1 and D2–, and, as a whole, there is an underlying basal correlation between all of these populations. It is likely that all three axes have a common interaction due to a similar location or a similar function.
  - C1 and C3 are specifically related (0.19), and also C2 with C3 (0.16). This may indicate the partial relevance of C3 as a bridge cluster between C1 and C2.
  - C5 cluster is related to C1 (0.14).
  - As observed from UMAPs and graphs too, D1 and D2 are highly related (0.23).
  - Similar with D axis, E1 population overlaps with several populations (C and D axes, as well as A2 and B4), which makes it a good candidate for further analysis.

10.4.2.2 *Mouse-mouse comparisons*

We calculated the Jaccard index with the top 150 markers to perform this comparison, similar to human-human populations. Figure 56 shows the heatmap of Jaccard index values. Overall, there is a high overlap of clusters with each axis and a low overlap of clusters between axes. Also, as expected, bridge clusters show some overlap between their respective clusters.

- z2 and z1 are highly similar (0.25), as expected.
- Surprisingly, contrary to what was observed in UMAPs and PAGA graphs, there is no relevant marker overlap between z1 and y1.
- Populations from axis x have a certain degree of similarity–x2 with x1 (0.15) and x1 with x/y (0.13)–. The bridge axis x/y overlaps with y3 (0.10) and y1 (0.08).
- y axis shows a general overlap across all populations in that axis. The population interactions with a higher degree are between y2 and y3 (0.12), y4 (0.15) and y5 (0.12); and between y4 and y5 (0.11). These observations align with what was previously commented before, based on UMAPs and PAGA graphs.
- Within w axis, there are two pairs of interactions with a high overlap–between w1 and w1 (0.39) and between w3 and w4 (0.17)–. Interestingly, and as it was previously stated, w3 also interacts with other populations, mainly with w/x (0.08) and v1 (0.05). w/x interacts with w3, w4, and with x2 (0.06), but not with w1, which is observed in UMAPs and PAGA graphs.

10.4.3 *Human-mouse marker-based comparison*

In human-mouse comparisons, we have used the first 250 markers from each population to compute their Jaccard index. The results are displayed in Figure 57. At a first glance, overlapping values range between 0.08 to 0.14 in overlaps considered "relevant". Those values are in range with the ones from human-human and mouse-mouse comparisons, taking into account that values higher than 0.2 are considered for populations that are extremely similar–e.g. z1 and z2, or B2 and B3–.

In this heatmap, the possible pairings are more diffuse; thus, we will select the most relevant overlaps. Additionally, for each pair of populations, we will show their overlapping markers, divided into four categories: ***markers relevant in human and mouse***, **markers relevant in human**, *markers relevant in mouse* and the rest of markers. We consider markers relevant if they are expressed mainly in that population. This criterion is arbitrary, so we try to make it conservative; we select relevant markers that are only expressed visibly in that population or, at most, in closely similar populations–like A1 and A4–. Genes expressed in more populations or have mild expression are not considered.

This secondary criterion, additional to the overlap, is relevant to help us choose markers in downstream parts of the analysis, instead of using sets of markers that we know beforehand are not accurate of the depicted populations.

The heatmap in Figure 57 does not show the markers in their original order. Instead, they are changed to reflect some of the interactions better visually. We will use human populations as the reference for comparison with mouse populations.

Beginning with axis A, there is an overlap of populations from axes x, y and z in mouse; which might imply that the functionality of this axis might be distributed in mouse.

- A2: although it shares some markers with some populations from axes y and w, the highest overlap is with population x2. Interestingly, the interaction of A2 with w axis populations (w1 especially), is something that we observe within human-human and mouse-mouse comparisons–A2 with C1/C3 and x2 with w/x–, assuming that C and w are homologous populations. Although there are slight differences between populations, this pattern is recurrent, and could be of interest for further analysis.
  - A2 - x2: CD109, COL3A1, COL5A1, LOXL2, NFATC2, NOTUM, PAM, RSPO3, SCARF2, TCF4, ZNF608, **COL13A1**, **CYP26B1**, **DAAM2**, **ISM1**, **NKD1**, **NKD2**, **PAPPA**, **PREX1**, **PTGS1**, **TGFBI**, **THBD**, **THSD4**, *C10orf105*, *CAV1*, *CCBE1*, *EMX2*, *ENHO*, *LSAMP*, *MAMDC2*, *SPRY1*, *TWIST2*, *WNT5A*, ***AHRR***, ***AXIN2***, ***CD9***, ***COL7A1***, ***F13A1***, ***GREM2***, ***IGFBP2***, ***KCNK2***, ***PTK7***, ***PTPRE***, ***RSPO1***, ***SMIM3***, ***STC1***, ***TNFRSF19*** [11, **12**, *10*, ***14***] - 47
  - A2 - w1: C4orf48, EMB, ETV1, PAPPA, PLCB1, SFRP2, WIF1, WNT5A, **APCDD1**, **AXIN2**, **FGFR2**, **NKD1**, **RAMP3**, **RGS2**, **SMIM3**, *LAMC3*, *NOTUM*, *NRN1*, *RSPO1*, *SPON1*, *TFAP2C*, ***APELA***, ***COL13A1***, ***COL23A1***, ***DAAM2***, ***F5***, ***NKD2***, ***PTK7***, ***RSPO3*** [8, **7**, *6*, ***8***] - 29
    Some of these markers, such as *DAAM2*, *NKD2*, *PTK7* or *RSPO3* are coexpressed between A2 and C2 or other C members.
- A1: this population shows a relevant overlap in several populations, although we will focus on x1, x/y, z2 and y3. Among the combinations, the mouse population that best overlaps A1 is z2, with 12/40 relevant markers in mouse and human. Interestingly, none of these markers overlap with relevant markers in the A1-x1 and A1-y3 comparisons, which may imply that A1 population establishes functions that might be of separate fibroblast types in mice.

  - A1 - x1: C1QTNF3, CD34, CHPF, CLEC3B, COL1A1, COL1A2, COL3A1, CREB5, CTSK, DCN, ELN, FKBP9, HTRA1, IGFBP5, ITGBL1, KDELR3, LOXL4, MFAP4, MMP14, P4HA2, PCOLCE2, PDGFRL, PLPP1, RCN3, SLC38A10, SPARC, TSPAN4, VKORC1, **AEBP1**, **ANGPTL1**, **CADM3**, **GDF15**, **GPNMB**, **HPGD**, **MMP2**, **PODN**, **SCARA5**, **TIMP2**, **TNXB**, *ADAMTS2*, *PPIC*, ***CGREF1***, ***CPZ***, ***CYBRD1***, ***CYP4B1***, ***MMP27***, ***SEMA3B*** [28, **11**, *2*, ***6***] - 47
  - A1 - x/y: ADAMTSL1, ANGPTL1, BASP1, CCDC80, CD34, CILP, COL12A1, COL1A1, COL1A2, CTSB, CTSK, DCN, FKBP9, IGSF10, ITGBL1, NUCB2, OGN, P4HA2, MGST1, PCOLCE2, PPIB, RCN3, SERF2, SERPINF1, VKORC1, **ABCC9**, **ADGRD1**, **AEBP1**, **AGTR1**, **CADM3**, **CYP4B1**, **GALNT15**, **HPGD**, **LOX**, **PCOLCE**, **PDGFRL**, **SEMA3B**, **SMOC2**, **SVEP1**, **THBS2**, **THBS3**, **TNXB**, *C1QTNF3*, *GPX3*, *GREB1L*, *PLTP*, ***CGREF1***, ***CPZ***, ***LGR5***, ***MMP27*** [25, **17**, *4*, ***4***] - 50
  - A1 - z2: CREB5, ECM1, FBN1, IGFBP5, IGFBP6, METRNL, PAMR1, PLXDC2, SEMA3C, TIMP3, VGLL3, **CHRDL1**, **CLEC3B**, **MFAP5**, **QPCT**, **SCARA5**, **TIMP2**, *ADAMTSL4*, *BASP1*, *CD248*, *CD34*, *COL14A1*, *DBN1*, *EMILIN2*, *LRRN4CL*, *PTGIS*, *SFRP2*, *UCHL1*, ***ACKR3***, ***ADGRD1***, ***CD55***, ***DPP4***, ***ISLR***, ***LIMS2***, ***MTCL1***, ***NPR1***, ***PI16***, ***PRKG2***, ***SEMA3E***, ***TUBB4A*** [11, **6**, *11*, ***12***] - 40
  - A1 - y3: ANGPTL1, COL12A1, COL14A1, CRYAB, CYB5R3, DHRS3, ELN, FGL2, GAS1, GPX3, IGFBP6, ITGBL1, MFAP4, OGN, PRELP, SERPINF1, SFRP2, TIMP3, **CLU**, **FBLN1**, **FBLN2**, **GALNT15**, **LGR5**, **LOX**, **CD151**, **PCOLCE**, **PDGFRL**, *C1QTNF3*, *CILP*, *MGP*, *PTGIS*, ***ABCC9***, ***OMD***, ***PODN***, ***SMOC2*** [18, **9**, *4*, ***4***] - 35

- A4: the most overlapping mouse populations in mouse are z1 and z2. Interestingly, although y1 shows a degree of overlap comparable to z1 (0.09), it does not have relevant enough markers to be focusing on it. Between z1 and z2, there seems to be more overlap and relevant markers from z2 (0.12) than from z1 (0.09), although many of them are shared–e.g. *AIFL*, *CD248*, *CD55*, *NPR1*, *SEMA3C*, *SEMA3E*, *SFRP4* or *WNT2*–.
  - A4 - z1: ADAMTS5, FSTL1, GFPT2, MGLL, RAMP2, SDK1, TIMP2, TNFAIP6, VASN, ZYX, **ACE**, **CLEC3B**, **GPX3**, **IGFBP6**, **LOXL1**, **MFAP5**, **PTGIS**, **RAB32**, **SCARA5**, *ACKR3*, *ADGRD1*, *AXL*, *CHRDL1*, *FLNC*, *GAP43*, *HAS1*, *HEG1*, *PROCR*, *PRSS23*, *UGDH*, ***AIF1L***, ***CD248***, ***CD55***, ***DBN1***, ***DPP4***, ***EMILIN2***, ***NPR1***, ***SEMA3C***, ***SEMA3E***, ***SFRP4***, ***WNT2*** [10, **9**, *11*, ***11***] - 41
  - A4 - z2: BMP7, CD34, DDAH2, GFPT2, IGFBP5, MGLL, NHSL1, PCSK6, PPP1R14B, PXN, RAMP2, SCARA3, SDK1, SMURF2, TRIO, ZNF385A, ZYX, **ACE**, **ACKR3**, **CLEC3B**, **DBN1**, **FBN1**, **IGFBP6**, **LOXL1**, **MFAP5**, **SCARA5**, **TIMP2**, **TIMP3**, *ACKR2*, *ADAMTSL4*, *ADGRG2*, *AXL*, *CHRDL1*, *DACT2*, *GAP43*, *HEG1*, *PROCR*, *PRSS23*, *UGDH*, ***ADGRD1***, ***AIF1L***, ***CD248***, ***CD55***, ***DPP4***, ***EMILIN2***, ***ISLR***, ***LIMS2***, ***NPR1***, ***PAMR1***, ***PTGIS***, ***RAB32***, ***SEMA3C***, ***SEMA3E***, ***SFRP4***, ***WNT2*** [17, **11**, *11*, ***16***] - 55
  - A4 - y1: ADAMTS5, ADGRD1, AHNAK2, APBB1IP, AXL, CAPG, CDH13, FAP, FSTL1, GFPT2, GLIPR2, HEG1, LGR4, LSP1, MEDAG, MGST1, MMP2, PRSS23, RAMP2, TAGLN2, THBS3, TMSB10, TNFAIP6, UGDH, VCAN, ZNF385A, **CD248**, **CD55**, **CLEC3B**, **EMILIN2**, **FBN1**, **GPX3**, **IGFBP6**, **LOXL1**, **MFAP5**, **TNXB**, **TPPP3**, *HMCN2*, *NOVA1*, *PLTP*, *SSC5D*, ***ACE***, ***CTHRC1*** [26, **11**, *4*, ***2***] - 43

  Axis B is related to several populations in mouse, mainly from axis y; although for some populations, like B4, there is an array of interaction from several axes.

- B1: B1 interacts mainly with z1 population (0.08). Interestingly, z1 also interacts with A4, which might imply that they have similar functions, split in human.
  - z1 - B1: ADAMTS1, BCL3, C3, CSRNP1, CSRP2, GFPT2, HMOX1, IFI16, MYC, NFE2L2, NFKBIZ, NOCT, PHLDA1, PIM1, PLSCR1, PNP, SOD2, **ERRFI1**, **KDM6B**, **MAFF**, **NFKB1**, **NFKBIA**, **NR4A3**, **REL**, **TIPARP**, **TNFAIP3**, **TNFAIP6**, **UAP1**, **ZC3H12A**, *HAS2*, *PTGS2*, *PTX3*, *TNFAIP2*, ***CCL2***, ***CXCL2***, ***FOSL1***, ***GCH1***, ***IL6*** [17, **12**, *4*, ***5***] - 38

- B2 and B3: these two populations interact mainly with y4–0.07 for B3 and 0.08 for B2–. Interestingly, y4 also interacts with B4, and with a higher overlap (0.14), but it is more likely that B2/B3 show relevant markers–3/61 from B4 and 7/38 from B2–.
  - y4 - B2: ABCA8, ABI3BP, C1S, FRMD6, HGF, ID4, IL11RA, LDB2, NRP1, OSMR, P2RY14, PLCXD3, TRIM47, TSPAN11, VEGFA, **APOC1**, **C7**, **CXCL12**, **CYGB**, **GGT5**, **IGFBP3**, **IGFBP7**, *ADCYAP1R1*, *AVPR1A*, *COL4A4*, *NDRG2*, *PTCH2*, *RBP1*, *SNED1*, *TMEM176A*, *TMEM176B*, ***APOE***, ***C3***, ***IL33***, ***MGP***, ***NFIB***, ***SLCO2B1***, ***TNFSF13B*** [15, **7**, *9*, **7**] - 38
  - y4 - B4: ARHGDIB, COL4A2, CYGB, DACT1, EGFR, EPHA3, EPS8, F3, FAM13A, FOXP1, GSN, HGF, IGFBP3, IGFBP7, ITM2B, MGST1, NGFR, NID2, OLFML2B, SERPING1, SRPX, TMEM176B, TNFSF13B, TSHZ2, **ABCA8**, **APOC1**, **APOD**, **C7**, **CXCL12**, **FGF7**, **FMO1**, **FZD4**, **GGT5**, **GPX3**, **HHIP**, **IGF1**, **NFIB**, **NTRK2**, **PODN**, **VIT**, **ZFHX4**, *ABCC9*, *ADAMTSL3*, *APOE*, *BMPER*, *BNC2*, *C3*, *COL4A4*, *FMO2*, *IL33*, *INMT*, *NR2F2*, *NRP1*, *PEAR1*, *PPL*, *PTCH2*, *SFRP4*, *TMEM176A*, ***CHRDL1***, ***GDF10***, ***MGP*** [24, **17**, *17*, **3**] - 61
- B4: B4 shows a diffuse pattern of overlap throughout y axis (y1 to y5). Although the overlap grade is enough to be considered relevant (0.07 to 0.14), none of the overlaps have enough relevant genes (2 or 3 in most cases). Most of the genes are relevant in human or mouse only, but not in both. Therefore, it is likely that this population is not present in mouse and its putative functions might be shared across the existing mouse populations. Additionally, we observe that y2 and y3 populations also map to populations from A, C, D and E axes, so it is likely that these populations directly cannot be comparable to any human fibroblast population.
  - B4 - y1: APBB1IP, ARPC1B, COL4A2, EBF1, EGFR, GPC3, HIC1, IGFBP3, LGALS3BP, LGMN, MEDAG, NFIB, NR1H3, PRSS23, SLIT2, TGFBR2, THY1, TMEM135, **CXCL12**, **CYGB**, **EFEMP1**, **FMO1**, **FZD4**, **GGT5**, **GPX3**, **IGFBP6**, **ITM2A**, **LSP1**, **MGST1**, **RARRES2**, **ZFHX4**, *BMPER*, *FABP4*, *NOVA1*, *PLAT*, ***PPARG*** [18, **13**, *4*, **1**] - 36
  - B4 - y2: ARPC1B, CAPN6, COL15A1, COL4A2, COL4A4, CRLF1, EGFR, EPS8, F3, FHL2, HIC1, HSPG2, HTRA3, IGFBP7, ITM2A, ITM2B, NID2, SPRY1, VWA1, **ABCA9**, **APOD**, **CYGB**, **FMO1**, **FZD4**, **GGT5**, **GPX3**, **GSN**, **IGF1**, **LSP1**, **MGP**, **MYOC**, **NFIB**, **PODN**, **TSHZ2**, **ZFHX4**, *ADAMTSL3*, *BMPER*, *INMT*, *PEAR1*, ***ABCA8*** [19, **16**, *4*, **1**] - 40
  - B4 - y3: ARHGDIB, DEPTOR, DHRS3, EPHA3, F3, FHL2, GHR, GPC3, IGFBP3, IGFBP6, IGFBP7, KCNJ8, NPY1R, PEAR1, SFRP4, SLIT2, SUSD2, TIMP3, **ABCA8**, **APOD**, **CYGB**, **FGF7**, **FMO1**, **GPX3**, **GSN**, **IGF1**, **ITM2A**, **MGP**, **MYOC**, **NFIB**, **PHLDA3**, **PODN**, **TXNIP**, **ZFHX4**, *ABCC9*, *FMO2* [18, 16, 2, 0] - 36
  - B4 - y4: ARHGDIB, COL4A2, CYGB, DACT1, EGFR, EPHA3, EPS8, F3, FAM13A, FOXP1, GSN, HGF, IGFBP3, IGFBP7, ITM2B, MGST1, NGFR, NID2, OLFML2B, SERPING1, SRPX, TMEM176B, TNFSF13B, TSHZ2, **ABCA8**, **APOC1**, **APOD**, **C7**, **CXCL12**, **FGF7**, **FMO1**, **FZD4**, **GGT5**, **GPX3**, **HHIP**, **IGF1**, **NFIB**, **NTRK2**, **PODN**, **VIT**, **ZFHX4**, *ABCC9*, *ADAMTSL3*, *APOE*, *BMPER*, *BNC2*, *C3*, COL4A4, *FMO2*, *IL33*, *INMT*, *NR2F2*, *NRP1*, *PEAR1*, *PPL*, *PTCH2*, *SFRP4*, *TMEM176A*, ***CHRDL1***, ***GDF10***, ***MGP*** [24, **17**, *17*, **3**] - 61
  - B4 - y5: ABCA8, F3, GPC3, HSPG2, IL33, ITM2A, LXN, MGP, RARRES1, **COL15A1**, **GSN**, **NTRK2**, *ABCA6*, *APOD*, *FOXS1*, *NR2F2*, *PHGDH*, *VIT*, *VWA1*, ***MYOC***, ***TSHZ2*** [9, **3**, *7*, **2**] - 21

  Axis C in human mainly overlaps with axis w in mouse. This is expected due to the similarity in functions of both axes, which are related to hair follicle cells.
- C2 population is related to several mouse populations, but mainly w/x, with whom it shares 12/42 relevant markers–based on an overlap value of 0.09–. Secondly, w1/w2 populations share 6/34 relevant markers–overlap of 0.06–. Lastly, y3 only shows overlap with 2 relevant markers,

despite showing a similar overlap with w1/w2–0.05–. Therefore, we do not consider the C2-y3 interaction as useful.

  - C2 - w/x: ADAMTS9, CBFA2T3, COL11A1, EDNRA, ENHO, HTRA1, SLC48A1, SRPX, TBXA2R, TENM3, **CPNE5**, **CRABP1**, **FZD1**, **GPM6B**, **MEOX2**, **NCAM1**, **NFATC2**, **PLXDC1**, **PTH1R**, **RSPO4**, **SLC40A1**, **TBX15**, **TCF4**, **TRIB2**, **TRPS1**, *HS3ST6*, *KIF26B*, *MEGF6*, *NR2F1*, *TSHZ3*, ***CHST15***, ***COCH***, ***CYP1B1***, ***DKK2***, ***EMID1***, ***FIBIN***, ***FMOD***, ***MAFB***, ***MKX***, ***NRP2***, ***PTGFR***, ***TNMD*** [10, **15**, *5*, ***12***] - 42
  - C2 - w1|w2: CNTN1, CTTNBP2, PRLR, PTGER3, PTPRD, SFRP1, TSPAN7, **ATP1B1**, **BTBD11**, **CLEC14A**, **DKK2**, **EMB**, **GPM6B**, **IGFBP5**, **MEIS2**, **NCAM1**, **SPARCL1**, *ALX4*, *CRABP2*, *DUSP10*, *LEF1*, *NDP*, *PDE1A*, *PTK7*, *RSPO4*, *RUNX3*, *SDC1*, *WNT5A*, ***CRABP1***, ***DAAM2***, ***NDNF***, ***NOTUM***, ***TRPM3***, ***TRPS1*** [7, **10**, *11*, ***6***] - 34
  - C2 - y3: CALM2, MDFIC, OGN, PDE4B, PMEPA1, PRELP, RERG, SERTAD4, SUSD2, **EMB**, **FIBIN**, **FMOD**, **FZD1**, **KCNAB1**, **PTPRD**, **PTPRK**, **TBX15**, *CAVIN2*, *F3*, *FGF9*, *LTBP4*, *PAQR6*, *SFRP1*, ***ASPN***, ***CADM1*** [9, **8**, *6*, ***2***] - 25

- C1 population interacts mainly with w4 and w3 mouse populations–0.12 and 0.11 overlap respectively–. When looking at the population marker overlap, we observe a much larger overlap of relevant markers with w4–18/53– than with w3–2/48–.
  - C1 - w4: ARHGAP28, ATP10A, CDH11, COL27A1, COL4A1, KIAA1217, KLF5, LMO7, MGLL, NPNT, NUAK1, NXN, PALLD, PMEPA1, POSTN, RNF152, SPARC, **CNN2**, **MME**, **SEMA5A**, *ADAMTS6*, *CD200*, *COL7A1*, *FOXD2*, *LMO4*, *LRRC15*, *NID2*, *NTRK3*, *PARD6G*, *PAWR*, *PRR5L*, *SATB2*, *TMEM119*, *TNMD*, *TPM2*, ***ACAN***, ***ACTA2***, ***ADAMTS18***, ***ADAMTS9***, ***BCL11B***, ***CALD1***, ***CCND1***, ***COL11A1***, ***COL12A1***, ***COL8A2***, ***CPXM2***, ***EDNRA***, ***EDNRB***, ***EGFL6***, ***MEF2C***, ***RAMP1***, ***TAGLN***, ***TENM3*** [17, **3**, *15*, ***18***] - 53
  - C1 - w3: AFAP1L2, COL6A3, FNBP1L, FOXD2, JAG1, KIF26B, LMO4, MDK, MGLL, MICAL2, NPM1, PALLD, PARD6G, PAWR, PMEPA1, SHOX2, SOX18, SOX4, TMEM119, TNS3, TPM2, **ACAN**, **ACTA2**, **ADAMTS18**, **ALX4**, **BCL11B**, **CALD1**, **CD200**, **CDH11**, **CNN2**, **COL11A1**, **DKK3**, **EDNRA**, **EVA1A**, **F2R**, **KIAA1217**, **LAMC3**, **MEF2C**, **PTCH1**, **ROBO2**, **RUNX2**, **STMN1**, *LRRC15*, *MMP11*, *NTRK3*, *TAGLN*, ***CCDC3***, ***EGFL6*** [21, **21**, *4*, ***2***] - 48

- C3 population, being established as a putative "bridge" between C1 and C2, is expected to overlap with mouse populations with C1 and C2. As such, C3 overlaps with w/x–0.11–and w4–0.10– populations, as expected. Looking at the significative markers, both w/x and w4 show a slight overlap–4/51 and 5/44– with C3. Of note, the number of relevant markers of C3 is reduced; therefore, we would expect this lower level of relevant marker overlap between human and mouse. Additionally, C3 also shows some degree of overlap with x1–0.08–, but there is only 1 relevant marker between human and mouse, so this comparison should not be considered.
  - C3 - w/x: ADAMTS9, AQP1, CDH11, CHCHD10, CNN2, COL11A1, COL16A1, CRABP1, DKK3, EDNRA, EGFLAM, EMP2, FIBIN, FZD1, GPM6B, HMCN1, MEF2C, MFAP4, MMP11, MXRA8, PLXDC1, PTMA, RSPO4, SESN3, SLC40A1, TBX15, TCF4, THY1, TMEM119, TMEM204, TPM2, TRPS1, **F2R**, **HTRA1**, **MFAP2**, **POSTN**, **TENM3**, *CPXM2*, *DKK2*, *EMID1*, *KIF26B*, *MEGF6*, *NR2F1*, *NREP*, *NRP2*, *RFLNB*, *TSHZ3*, ***COL7A1***, ***MAFB***, ***MMP16***, ***RASL11B*** [32, **5**, *10*, ***4***] - 51
  - C3 - w4: BGN, BHLHE41, CALD1, CDH11, CNN2, EMP2, FABP5, ITGB1, ITGBL1, MGLL, MYH9, NPNT, PALLD, PPIC, PRELP, SORCS2, SPARC, **KIAA1217**, **LMO7**, **PMEPA1**, **SEMA5A**, *ACAN*, *ACTN1*, *ADAMTS9*, *AQP1*, *CGNL1*, *COL11A1*, *COL12A1*, *CPXM2*, *EDNRA*, *EGFLAM*, *LBH*, *MEF2C*, *MME*, *SLC40A1*, *TAGLN*, *THBS4*, *TMEM119*, *TPM2*, ***COL7A1***, ***COL8A2***, ***LRRC15***, ***POSTN***, ***TENM3*** [17, **4**, *18*, **5**] - 44
  - C3 - x1: BMP1, CHPF, COL3A1, COPZ2, GLT8D2, ITGBL1, LUM, MARVELD1, MFAP4, MMP14, MMP2, P4HA2, PYCR1, RCN3, RRBP1, SLC16A3, SPARC, SPHK1, SPON2, **BGN**, **C1QTNF6**, **COL5A1**, **COL5A2**, **ELN**, **HTRA1**, *ADAMTS2*, *AEBP1*, *COL16A1*, *COL1A1*, *COL1A2*, *EMID1*, *PPIC*, *PPP1R14A*, *SULF2*, **LTBP2** [19, **6**, *9*, **1**] - 35

- C5 population shows some overlap with population w3–0.08–and w1|w2–0.07–. When looking at the relevant markers, we observe a better overlap with w1|w2 than with w3–8/29 vs 2/26–. Interestingly, the amount of markers overlapping between C2 and w1|w2 is similar to the ones with C5 and w1|w2, although higher in proportion–6/34 vs 8/29–; and the markers that are relevant in human and mouse between the two comparisons do not overlap between them. Therefore, C2 and C5 functions may be shared between w1 and w2, without a clear pattern.
  - C5 - w3: CFL1, COL11A1, DKK3, EDNRA, EEF1A1, F2R, PFN2, PMEPA1, PTMA, SMS, SOX4, STMN1, **ALX4**, **CXCR4**, **HEY2**, **JAG1**, **LMO4**, **MARCKSL1**, **PRDM1**, **PREX2**, **PTCH1**, **SOX18**, *KIF26B*, *LAMC3*, ***CDH11***, ***ROBO2*** [12, **10**, *2*, **2**] - 26
  - C5 - w1|w2: C4orf48, FST, LAMC3, MARCKSL1, PLK2, PTCH1, PTMA, RAB34, SPARCL1, **CXCR4**, **IGFBP3**, **PDE3A**, **PRDM1**, **PREX2**, **ROBO1**, *EDN3*, *HEY2*, *SDC1*, *SLC26A7*, *SPON1*, *TRPS1*, ***ALX4***, ***BMP7***, ***INHBA***, ***LEF1***, ***PGM2L1***, ***SOX18***, ***TFAP2A***, ***WNT5A*** [9, **6**, *6*, ***8***] - 29

  Axis D has three possible populations with considerable overlap: y4, y5 and v1. To see which populations are the most likely fit for D1 and D2 independently, we will analyse each combination's relevant marker overlap. We observe that the best fits are D1 with y5 and D2 with v1. y5 shares 9/24 relevant markers with D1, whereas only 2/13 with D2, which makes D1 more suitable than D2 to be associated to y5. Regarding v1, there is a 7/43 overlap with D1, whereas there is a 19/60 overlap with D2, implying that D2 is likely to be more associated with v1. Lastly, regarding population y4, we see that the relevant overlaps are scarce–2/29 with D1 and 1/29 with D2–compared to y5 and v1, so this mouse population is far less likely to be comparable to D1 or D2 in human.

- D1 vs D2 - y5
  - D1 - y5: ARL4A, CNN3, CSPG4, ITM2A, LUM, NID1, **EBF2**, **ENTPD2**, **ETV1**, **GPC3**, **PHLDA1**, **SPARCL1**, **TM4SF1**, *CDKN2B*, *MATN2*, ***ABCA8***, ***APOD***, ***COL8A1***, ***FOXS1***, ***NR2F2***, ***P2RY14***, ***SOX9***, ***TGFBI***, ***VIT*** [6, **7**, *2*, ***9***] - 24
  - D2 - y5: ABCA8, CNN3, EBF2, ETV1, ITM2A, NR2F2, SMOC2, TGFBI, TM4SF1, VIT, *FOXS1*, ***MATN2***, ***P2RY14*** [10, **0**, *1*, ***2***] - 13
- D1 vs D2 - v1
  - D1 - v1: CAVIN2, CSPG4, CSRP1, CTNNAL1, DOCK9, FRMD4B, JAG1, MTUS1, NDRG1, PEAR1, PLEKHA4, PTCH1, STMN1, SYNE2, TLN2, TUBA4A, **CD200**, **DUSP5**, **EFNA1**, **EGR3**, **ETV1**, **MRAS**, **NR2F2**, **OLFML2A**, **SLC12A2**, **SOX9**, **TENM2**, **TGFBI**, **VIT**, *BNC2*, *EZR*, *ITGA6*, *ITGB4*, *KLF5*, *SBSPON*, *SLC2A1*, ***AKAP12***, ***CLDN1***, ***EBF2***, ***EFNB1***, ***ETV4***, ***MTSS1***, ***NDRG2*** [16, **13**, *7*, **7**] - 43
  - D2 - v1: v1 - D2: BHLHE40, DMD, ENDOD1, ETV1, FRMD4B, JAG1, KTN1, MDFIC, MFAP5, NDRG1, NR2F2, PHLDA3, PLSCR4, PTCH1, SFRP1, STMN1, STXBP6, SYNE2, TLN2, **ADAMTSL3**, **AKAP12**, **CAVIN2**, **CSRP1**, **DACT1**, **DUSP5**, **IGFBP6**, **ISYNA1**, **MRAS**, **OLFML2A**, **PEAR1**, **PLEKHA4**, **SORBS1**, **TGFBI**, **TJP1**, **VIT**, *CCDC3*, *DDIT4*, *EZR*, *PERP*, *TPD52*, *TRIB2*, ***AQP3***, ***BNC2***, ***CAV1***, ***CAV2***, ***CLDN1***, ***DOCK9***, ***EBF2***, ***EFNB1***, ***GAB1***, ***GPC1***, ***ITGA6***, ***ITGB4***, ***KLF5***, ***KRT19***, ***MTSS1***, ***NDRG2***, ***SBSPON***, ***SLC2A1***, ***TENM2*** [19, **16**, *6*, ***19***] - 60
- D1 vs D2 - y4
  - D1 - y4: APOD, BNC2, EBF2, EGFR, EPS8, FMO1, IGFBP7, ITGA6, KLF5, LTBP4, MEOX2, NDRG2, NGFR, NR2F2, PEAR1, S100B, SCN7A, VIT, **CYP1B1**, **P2RY14**, **PHLDA1**, **SPARCL1**, **WFDC1**, *ABCA8*, *CYGB*, *INMT*, *SFRP4*, ***ENTPD2***, ***FMO2*** [18, **5**, *4*, **2**] - 29

  - D2 - y4: y4 - D2: ABCA8, ADAMTSL3, CCDC3, COL4A2, EBF2, EGFR, CYP1B1, IGFBP7, ITM2B, LTBP4, MEOX2, MFAP5, NDRG2, NR2F1, NR2F2, P2RY14, SCN7A, VIT, **BNC2**, **CAV1**, **DACT1**, **GPC6**, **ITGA6**, **KLF5**, **NGFR**, **PEAR1**, *CHRDL1*, *SFRP4*, ***INMT*** [18, **8**, *2*, ***1***] - 29

- E1 population shows more reduced levels of overlap–0.05 to 0.07– with several mouse populations: y2, y3, y4, w3, w4, and v1. Some of these populations, like w3, w4, and v1, have a high chance of being related to other human populations like C1 or D2, and it is less likely, based on the smaller overlaps of relevant markers, to be associated with E1. On the other hand, based on the heatmap from Figure 57, and on the overlapping markers, populations y2, y3 and y4 are not one-to-one correspondence candidates, since they have marker overlap with a wide range of human populations. Therefore, based on this analysis, we cannot assign a clear human-to-mouse comparison candidate to E1.
  - E1 - y2: APOD, CAPG, COL15A1, FHL2, GOS2, GSN, IGF1, ITGA11, ITM2A, ITM2B, JUP, LAMA2, LPL, LSP1, MFAP5, SCN7A, SFRP1, SPRY1, TMEM204, **CMKLR1**, **MGP**, **PLEKHA6**, **RGMA**, *COL14A1*, *CRLF1*, *GPM6B*, *INMT*, *PEAR1*, *TGFBI*, *TSHZ2*, *VWA1*, ***MEOX1*** [19, **4**, *8*, ***1***] - 32
  - E1 - w3|w4: ACOT7, ALX4, BHLHE40, BHLHE41, C1QTNF7, CALD1, CCDC34, CDC42EP3, CDK6, DKK3, EFNB1, EMP2, EVA1A, FHL2, HES1, HEY1, ITGB1, JUP, KLF5, LAMC3, LMO7, MAF, MFGE8, MSI2, RASGRP2, RGCC, RUNX3, SEMA3G, SEMA5A, SHOX2, TBX18, **EGR2**, **KIAA1217**, **OLFML2A**, **RAMP1**, **TCF7L2**, *AQP1*, *ARHGDIB*, *CDH11*, *COL8A2*, *CSRP1*, *EDNRA*, *F2R*, *FOXD2*, *PARD6G*, *RUNX2*, *STMN2*, ***IGFBP2***, ***NTRK3*** [31, **5**, *11*, ***2***] - 49

Therefore, as a summary of all the comparisons, we conclude that human and mouse populations share a certain degree of overlap, with some cases of merely a one-to-one or one-to-two combinations–e.g. A2-x2, C1-w, A4-z1/z2–; some cases where a set of human and mouse populations are comparable as a set and not individually–e.g. A3 and A1 with x1 and x/y, or D1 and D2 with y5 and v1–; and some cases where the overlap may be confounded by that population being similar or interacting with other populations–e.g. E1 in human or y2 and y3 in the mouse–, and therefore not having putative comparable populations from the other species.

10.5 DEPICTION OF PAPILLARY AND RETICULAR PROGRAMS ON HUMAN DATASETS

For many years, fibroblast-related literature has focused on the differences between dermal fibroblasts from the papillary dermis and the reticular dermis. In this section, we will analyse the most relevant markers based on the literature and use them to classify the human fibroblast populations into papillary-like or reticular-like.

To check the relevance of each population to the papillary-reticular location, we obtained a list of markers from 6 different references: Janson et al., 2012, Nauroy et al., 2017, Philippeos et al., 2018, Korosec et al., 2019, Haydont et al., 2019, and Haydont et al., 2020. The full list of markers is available in the Materials and Methods section 5.12. Based on that list, we applied the marker-to-population algorithm to see which fibroblast populations matched more with papillary or reticular signatures. Figure 58 shows the results of that score.

The results from the figure indicate that population A2 is the most likely to be assigned to the papillary dermis based on the signatures and, on the other hand, A1, A4, B4, C1, D2 and E1 are more likely to be assigned to the reticular signature. The rest of the populations tend to be equally assigned to both signatures.

Another way to check the belonging of the populations to each type of fibroblast is to assess, for each marker, which population is best assigned. We used the manual scoring of markers and applied that scoring to each fibroblast signature. Therefore, for each marker and population, we have the information of whether the marker is assigned as papillary or reticular based on the signature and if it is a good marker for that population. The results are available at Figure 59.

For population A2, we observe that there is a large set of genes almost specific to that population–e.g. *COL23A1*, *RSPO1*, *PTGS1*, *COL6A5*, *COL18A1*–that belong to the papillary signature. For the A axis in general, there is a larger set of genes–e.g. *FAP*, *SGCG*, *PDGFC*, *NTN1*, *DPP4*–that belong to the papillary signature, as well as a smaller set of genes–e.g. *ANGPTL1*, *THBS2*–which belong to the reticular signature. Additionally, for A1 and A4 populations, there are more genes–*MFAP5*, *FNDC1*, *TMP1*, *PCOLCE2*, *FBLN2*–which further contribute to the reticular signature of these populations.

Regarding the rest of the axes, the evidence of a clear signature is more complex. For instance, except B3, which has a core set of papillary genes–e.g. *CCL2*, *CXCL1*, *ADRA2A*, *CCL8*, *IL15*–; B4, which has a set of reticular genes–e.g. *MGP*, *IGF1*, *FMO1*, *EFEMP1*, *PPARG*, *CTSK*–; or C1 with a set of papillary genes–e.g. *CTSK*, *LRIG*, *CADM1*, *FGF13*, *ROBO2*, *DIRAS3*, *PTK7*–, the rest of populations do not share a core set of papillary or reticular genes.

The main limitation from these analyses comes from the selection of papillary and reticular markers based on the literature. These markers are probably biased to the mostly expressed markers of the populations at each dermal layer, or markers known to differ between each layer beforehand. The fact that most markers from the papillary layer are associated with population A2 reflects that this population may be the main contributor by cell number to the papillary dermis. On the other hand, the reticular dermis may be populated with a more diverse range of cell types, and therefore the markers of less predominant cell types may not be discovered, or may be underreported, and therefore, are not suitable for this analysis. This is more prevalent with the populations from axes B and C, where only a few markers are expressed by them.

10.6 LIGAND-RECEPTOR ANALYSIS ON HUMAN DATASETS

The interaction between fibroblast populations is an important component in exploring the underlying biology of the skin. In this section, we will perform a basic quantitative analysis of the results from the ligand-receptor (LR) algorithm CellPhoneDB. The biological relevance of these results will be further explored in the discussion section.

Typical LR analyses yield a set of LR pairs with a p-value associated with the significance of the interaction. The LR pairs are set for specific cluster combinations, that is, for each pair of clusters, a set of LR pairs is computed. This analysis type is unsuitable for our fibroblast characterisation since we want to summarise all LR pairs across datasets. Additionally, since we already know that most fibroblast markers are not expressed exclusively in one population, most of the LR pairs set for a pair of populations can be extended for other pairs, and this information has to be summarised as well.

Therefore, to integrate results from several datasets and to retain only the most relevant interactions, we applied a two-step filtering, firstly on datasets–i.e. LR interactions occurring in at least a minimum number of datasets are considered–and secondly on genes–i.e. genes that are good pop-

ulation markers are considered–so that we obtain a "good-quality" list of putative LR interactions between different populations. Once the table is obtained, the additional putative populations that can be targets of the LR pair are included. For example, the pair ACKR3-CXCL12, which is statistically significant for populations A1 and B2, is also extended for populations A4 and B4 because ACKR3 is also a marker of A4 and CXCL12 is also a marker of B4.

The list of interactions is defined in Table 11. A total of 69 LR pairs have been filtered between 84 different genes. From these, some appear recurrently: *WNT5A* 9 times, *CD44* and *EGFR* 4 times, and *SLIT2*, *CADM1*, *NRP1*, *NRP2*, *CXCL12*, *LGR5*, *LGR4*, *LDLR*, *DPP4*, *FZD2*, *FGFR2* and *EGFR* 3 times.

It is interesting to observe that, although there is a variety of interactions between different families of fibroblasts, interactions between certain axes are more persistent.

- There are 6 interactions between axes A and B–e.g. *ACKR3-CXCL12*, *CCL19/CCL2*-ACKR4, *CXCL2-DPP4*, *FGF7-FGFR2*–. Most of them have a relationship with immune signalling and chemokine-based communication.
- There are 14 interactions between axes D and the rest of axes–e.g. *APOD-LEPR*, *EGFR-ICAM1/FGF13*, *LDLR-APOE/F3*, *VEGFD-NRP2*–. These interactions belong to different categories.
- There are 14 interactions between C5 or E1 populations and the rest of axes–e.g. *BMP7-PTPRK*, *LDLR-WNT5A* and most of them of the type *FZD1/2/4/6/7-WNT5A/11*–. Although the interactions belong to different categories, the most prominent one is related to the Wnt pathway.
- There are also a large set of 6 interactions, mainly between C1, C2, A1 and A2, that belong to the family of R-spondins and they Lgr family receptors, which potentiate the effects of Wnt signalling (Carmon et al., 2011; Lau et al., 2011; Ruffner et al., 2012).

When we look at the interactions, they can be categorised into different categories.

The most prominent one, stated before, is Wnt signalling, either via canonical–*FZD1* receptor and *WNT2* ligand–, non canonical–*FZD2* receptor and *WNT5A* ligand–, or ambivalent patways–*FZD6* and *FZD7* receptors, *PTPRK* and *PTK7* correceptors–. Interestingly, R-sponding signalling between *LGR4/5* and *RSPO1/3/4* is apparent Niehrs, 2012. In total, more than 18 interactions involve Wnt signalling.

Chemokine-mediated communication involving more than 7 interactions *ACKR3/4*, *CCL2/13/19*, *CXCR4*, and*CXCL2/12* is relevant for signalling between populations from axes A and B predominantly. Interestingly, this communication may not be driven by an inflammatory context, but rather for proper dermal homeostasis.

Ephrin and neuropilin-mediated signalling is highly relevant, with more than 7 interactions involving *NRP1/2*, *EFNA5/B1* and *EPHA3/B6*, sometimes between neuropilins and ephrins, sometimes with other interacting proteins. These interactions are mediated by populations from all axes, without any major population being more relevant.

With similar properties to neuropilins and ephrins, more than 7 interactions involve growth factors or their receptors, including *EGFR* with *GRN*, *ICAM1*, *FGF13* and *LRIG1*; or *FGF7/13-FGFR2* and *FGF7-NRP1*. *EGFR* is mainly expressed by D axis, but it binds to populations from the rest of axes.

These categories and interactions may be relevant later on in the Discussion section to understand the biological functions of each cell type.

It should be noted that the LR analysis from this section contains two biases: on the one hand, although these interactions are curated from bibliographic records based on biological experiments, these are putative fibroblast interactions in the sense that they may not occur specifically in that cell type. Additionally, the list of interactions is biased towards already studied pathways and interactions, therefore, many other unregistered interactions may not have been captured.

10.7 GOEA EXPLAINS THE DIFFERENCE OF FUNCTIONS BETWEEN FIBROBLAST AXES, BUT LACKS GRANULARITY TO DO SO WITH POPULATIONS

A common step during the analysis of the function of cell types is Gene Ontology Enrichment Analysis (GOEA). The aim of this step is to obtain sets of different ontologies for each fibroblast population to shed some light on their putative functions. To do that, we used the python API of the *enrichr* module (Kuleshov et al., 2016), and selected 6 sets of gene ontologies–. The full list of statistically significant GO terms is available at this link.

Although GO terms may be functional to indicate major putative roles exerted by cells, we came across several findings that may compromise the characterisation of fibroblast populations:

- There is a large group of genes that is repeated across ontologies and the number of genes used in the ontologies is reduced compared to the number of genes used as input.
- This fact is in part supported because there are sets of ontologies within each population that are similar or have the same set of genes associated to them.
- There are sets of ontologies shared across different populations.
- Most sets of ontologies are vaguely informative. Moreover, many repetitive sets are based on a few markers and may give the false sensation of that population doing specifically one function based only on a few markers.

The first point is addressed in Table 12. Across populations, we observe a tendency that only half of the genes–μ=0.56, σ=0.13–are selected in GO terms, and the rest are not considered. Therefore, much of the biological information of the population is lost, and the GO terms are biased towards the few genes associated with them.

To prove the second point, we are going to show several examples from different populations:

- In population A1 we observe the terms *carboxypeptidase activity (GO:0004180)*, *metallocarboxypeptidase activity (GO:0004181)* associated with similar sets of genes; *positive regulation of cytosolic calcium ion concentration (GO:0007204)* and *regulation of cytosolic calcium ion concentration (GO:0051480)* associated to the same set of genes (*AGTR1*, *ACKR4*, *ACKR3*, *CD24* and *CD55*); or *dystroglycan complex (GO:0016011)* and *sarcoglycan complex (GO:0016012)* GO terms associated with *SGCA* adn *SGCG* genes.
- In population A2 we observe the 4 GO sets *collagen-containing extracellular matrix (GO:0062023)*, *external encapsulating structure organization (GO:0045229)*, *extracellular structure organization (GO:0043062)* and *extracellular matrix organization (GO:0030198)* associated with the same set of 10 genes; or *regulation of Wnt signaling pathway (GO:0030111)*, *positive regulation of Wnt signaling pathway (GO:0030177)*, and *regulation of canonical Wnt signaling pathway (GO:0060828)* associated with similar sets of genes.

- In population B1 we observe the 4 GO sets *response to lipopolysaccharide (GO:0032496)*, *cellular response to molecule of bacterial origin (GO:0071219)*, *cellular response to lipopolysaccharide (GO:0071222)*, and *NOD-like receptor signaling pathway* associated with a set of 7 genes; or the gene sets *Vitamin D in inflammatory diseases WP4482*, *TLR4 Signaling and Tolerance WP3851*, *Resistin as a regulator of inflammation WP4481*, and *Signal transduction through IL1R WP4496* associated with the genes *NFKBIA*, *IL6* and *NFKB1*.
- In population C3 there are three similar terms *collagen-containing extracellular matrix (GO:0062023)*, *external encapsulating structure organization (GO:0045229)*, and *extracellular structure organization (GO:0043062)* linked to a similar set of 8 genes; or sets *lysosomal lumen (GO:0043202)*, *glycosaminoglycan catabolic process (GO:0006027)*, *glycosaminoglycan biosynthetic process (GO:0006024)*, and *sulfur compound catabolic process (GO:0044273)* associated with *ACAN*, *BGN* and *SDC1* genes.
- In population D1 we see a set of 8 different GO terms such as *camera-type eye photoreceptor cell differentiation (GO:0060219)*, *negative regulation of myoblast differentiation (GO:0045662)* or *morphogenesis of an epithelium (GO:0002009)* linked with genes *SOX8* and *SOX9*; or GO terms *negative regulation of cell activation (GO:0050866)*, *negative regulation of neuroinflammatory response (GO:0150079)*, or *negative regulation of macrophage activation (GO:0043031)* associated with *LDLR* and *CD200* genes.

This effect does not resolve by adding more GO terms by lowering the p-value. In fact, many of the following GO terms for most populations only contribute with the same markers. In some cases, additional markers are added, but sparsely. In fact, the redundancy rate, that is, the mean number of different GO terms in which a gene appears, is more than five–$\mu = 5.19$, $\sigma = 2.45$–. Therefore, most GO terms will also tend to be redundant.

This effect is expected and exacerbated since many families of terms are repeated for one population and across populations, as we mentioned in the third point. We show some examples.

- The most repeated family of terms is related to the ECM. The term *collagen-containing extracellular matrix (GO:0062023)* is shared by populations A1, A2, A3, A4, B4, C1, C2, and C3. Other GO terms from the same family are repeated across populations–mainly across A1, A2, A3, C1, C3 and D2–: *extracellular matrix organization (GO:0030198)*, *extracellular structure organization (GO:0043062)*, *collagen fibril organization (GO:0030199)* and *collagen-containing extracellular matrix (GO:0062023)*.
- Several immune-related populations–e.g. B1, B2 and B3–share cytokine- and chemokine-related GO terms, such as *cytokine-mediated signalling pathway (GO:0019221)*, or *cellular response to cytokine stimulus (GO:0071345)*.

Unexpectedly, the markers of each population for the same terms vary because, although useful at first sight, these terms are extremely vague in their formulation and do not provide a detailed picture.

GO term redundancy is partly explained by the pooling of GO terms from different ontology set–e.g. KEGG, Wikipathways and, GO biological process–which, despite increase the diversity of GO terms, most of them may be redundant. Despite that, even within each GOs there is considerable redundancy by including vertically-related GO terms–branches or parent GO terms–.

All these ideas can be summarised into one observation: most GO terms are biased towards existing and general cell functions, and therefore many genes and functions are not accurately complied. Therefore, GOEA will be severely biased towards a few GO terms based on a subset of genes and thus can only explain some bold functions of the fibroblast populations. Thus, despite GO terms being interesting to have a general view on the putative functions of a population based on a set of markers, which we will be supporting during the discussion, we believe that this analysis falls short of providing a relevant and detailed view of the functions of each population, which may require a more profound literature analysis.

10.8 SEMI-SUPERVISED CLASSIFICATION ALGORITHM ROBUSTNESS

In this section, we are going to discuss the robustness of the unsupervised classification algorithm. Since most of the results from this chapter are derived from the gene-to-population algorithm, it is necessary to check that the categories provided are accurate and replicable so that the downstream inferences we performed are also valid.

We applied a methodology similar to the classical Jackknife resampling method to check the robustness. For a fixed number of iterations–30 for this analysis–, and each dataset, we sampled 99% of the cells and reran the marker-to-population algorithm. The sampling was stratified across populations so that all populations were consistently represented.

The result of this resampling is, for each dataset, a table with as many rows as cells and 30 columns. For each cell, the population to which that cell is assigned in each iteration is shown. To show the results numerically, we calculate the robustness score per cell, that is, the quotient between the number of iterations where the assigned cluster is the same as the original and the number of times the assignment was done–because for each iteration 1% of the cells is not considered, and therefore the total number of times might be lower than the theoretical number–.

10.8.1 *Robustness across datasets and populations*

Since the sampling percentage is high, we should expect that the assignment is robust and that the robustness score values are high. Figure 60 shows the distribution of the mean robustness score of each population across datasets.

In human datasets, we observe a variation in scores between 0.6 and 0.9, with a mean value of about 0.75, which is acceptable. Some populations, such as B3, B4 or D2, show lower values. Interestingly, the boxplots from Figure 60 show a wide range of variation across datasets, so that the robustness might be dataset dependent. In fact, when we look at the distribution of scores across datasets, we observe a wide disparity of values, with some datasets like Reynolds et al. (2021), Hughes et al. (2020) or Vorstandlechner et al. (2020) having lower mean scores with high deviations.

If we look at the violin plots from Figure S5, which show the distribution of scores per population and dataset, we confirm some previous observations. For instance, regarding population B3, we observe that in some datasets–Liu et al., 2021a; Rindler et al., 2021; Solé-Boldo et al., 2020; Theocharidis et al., 2020–the scores are high, whereas in other datasets–Ahlers et al., 2022; Burja et al., 2022; Reynolds et al., 2021; Tabib et al., 2021– the scores are extremely low. In Section 10.2.2, we mentioned that, in some datasets, there was a clear separation between B3 and B2 and B1. Interestingly, most of these

datasets show high robustness scores, whereas, in the datasets with very low scores, the location of B3 in the UMAP is far more diffuse or a very small cluster (Figure 42).

Similar effects can be observed for populations such as E1. The robustness score of this population is generally approaching 1, except in a couple of datasets–Gao et al., 2021; Gaydosik et al., 2019; Vorstandlechner et al., 2020–, where the scores are near 0. In these datasets, E1 population was primarily assigned to B1 or B4 (Figure S7).

Regarding A1 population, we observe that, unlike other populations, it does show lower scores–between 0.4 and 0.7–in many datasets, which is explained due to the similarity with A4. A similar effect is observed with A3 population, which is also explainable by its similarity with A1 and A2.

Regarding mouse populations, we observe that robustness score values are higher compared to humans, with many of them reaching values higher than 0.85 (Figure 60). Despite that, there are 5 notorious exceptions with lower score values or wider deviations: w/x, w2, x1, x2 and z1 populations. When looking at the violin plots from Figure S6 we see that, in all cases, the general scores for these populations are high, although, for a few datasets, it is either low or a mixed combination or high and low. For instance, w/x is low in Joost et al., 2020 and variable in Shook et al., 2020; x2 is a variable in Boothby et al., 2021 and low in Phan et al., 2020; z1 is variable in Phan et al., 2020. In fact, most of this variability comes from the Phan et al., 2020 dataset, as we observe in Figure 60. This may indicate that assignment of populations in Phan et al., 2020 dataset lacks robustness, probably due to higher transcriptomic noisiness related to sample processing.

10.8.2 *Non-robust populations "sublimate" to similar populations*

From the previous section, we can easily argue that a certain degree of inter-dataset variability explains the differences in robustness scores. However, another variable that can be discussed is which populations are assigned by the algorithm when the assigned population is not the original.

This is a relevant point to analyse since, as we have discussed earlier, fibroblast populations have certain degrees of similarity between them, and therefore, we should expect the classification algorithm to assign the markers to the most similar population. In fact, if two populations are highly similar, it is understandable the algorithm fails to assign these populations due to the populations and not due to the algorithm itself. If "less similar" populations were assigned instead, this would justify a lack of robustness.

To check this, we calculated the adjacency matrix of the resampling results per dataset, as shown in Figures S7 and S8. The matrix indicates the proportion of assignments for each cell type (row) to the rest of the populations (columns). The median value across datasets was calculated to produce the combined adjacency matrix in Figure 61.

Regarding human datasets, there is a generally high proportion of cases where the assigned cluster is the same as the original one, with some examples such as B1, C2, C5, D1 or E5, with values higher than 0.92. From the previous, we discussed some populations, such as A1, A4 or B3, which have lower values–0.73, 0.69 and 0.59, respectively–. When looking at the assigned clusters, we observe that for A1, the most assigned cluster is A3–21% of the times–, for A4 is A1–21% of the times– and for B3 are B2 and B1–25% and 16% of the times, respectively–. All cases respond similarly to what we discussed previously in Section 10.2.

This pattern is kept for individual datasets (Figure S7). For instance, for datasets which have lower proportions of A1 population assigned to A1, the second most assigned population is A3 (Dataset: proportion originally assigned/proportion of secondary assignation)–Ahlers et al., 2022: 0.49/0.20, Gur et al., 2022: 0.58/0.30, Reynolds et al., 2021: 0.47/0.47, Rindler et al., 2021: 0.56/0.38, Tabib et al., 2018: 0.60/0.32, Tabib et al., 2021: 0.38/0.56, Vorstandlechner et al., 2020: 0.58/0.35, Vorstandlechner et al., 2021: 0.59/0.35–. In some datasets, the second most assigned population is not A3–Kim et al., 2020a: 0.45/B4 (0.25), Mirizio 2020: 0.50/A4 (0.21)–. This pattern is repeated for A4 and B3 as well.

For the rest of the populations, we observe similar assignment patterns according to previously commented similarities. For instance, B4 population (0.75) is assigned similarly to B1 and B2 (0.12 and 0.09); or D2 (0.89) is assigned to D1 (0.03) but also to B4 (0.06)–as observed in Figure 55–.

Regarding mouse populations, the assignment rate to the original population is much higher–more than 0.85–and assignment to other populations follows similar patterns to those discussed in Section 10.3. For instance, w/x (0.87) was primarily assigned to x1 (0.06); w2 (0.88) is assigned secondarily to w1 (0.12); w4 (0.94) is assigned to w3 (0.06); y3 (0.94) is assigned to y2 (0.04); y2 (0.86) is mainly assigned to y3 (0.09); or z1 and z2 (0.89, 0.92) are assigned to z2 and z1 respectively (0.06, 0.08).

Therefore, in conclusion, the population assignment algorithm shows good robustness levels in human datasets and very good robustness levels in mouse datasets. Moreover, the cases of lower robustness are either due to dataset-specific reasons and are not generalised to the population or because the population is similar transcriptomically to other populations and thus follows similarity patterns already described in sections 10.2 and 10.3.

10.9 TABLES AND FIGURES

Table 7: **Human dermal fibroblast dataset information.**

Reference	Donors (n)	Age (y)	Sex	Ethnicity	# of fbs.
Ahlers et al., 2022	3	22, 25, 29	F	Caucasian	20743
Billi et al., 2022	14	-	-	-	5142
Boothby et al., 2021	3	64, 66, 67	M	-	4915
Burja et al., 2022	2	46, 79	F/M	Caucasian	2587
Deng et al., 2021	3	26-39	1F/2M	Han chinese	5983
Gao et al., 2021	3	23, 32, 47	1F/2M	-	2099
Gaydosik et al., 2019	1	64	M	-	2640
Gur et al., 2022	60	28-75	49F/11M	-	15393
He et al., 2020	7	38-82	4F/3M	-	3443
Hughes et al., 2020	2	58,68	1F/1M	-	450
Kim et al., 2020a	4	-	2F/2M	-	2266
Liu et al., 2021a	4	26-32	F	Han chinese	1270
Mariottoni et al., 2021	1	48	M	Af. American	894
Mirizio et al., 2020	6	-	-	-	1074
Reynolds et al., 2021	5	25-60	F	-	7327
Rindler et al., 2021	4	44-57	3F/1M	Caucasian	2885
Solé-Boldo et al., 2020	2	25, 27	M	-	2739
Tabib et al., 2018	6	23-66	3F/3M	White	2742
Tabib et al., 2021	10	-	-	-	4814
T. Sapiens Cons. et al., 2022	2	33, 59	M	Hispanic	2315
Theocharidis et al., 2020	7	49-75	F	-	3285
Theocharidis et al., 2022	10	34-75	6F/4M	White	10441
Vorstandlechner et al., 2020	3	30, 36, 43	F	-	1161
Vorstandlechner et al., 2021	2	36, 43	F	Caucasian	1338
Xu et al., 2021b	4	18-41	3F/1M	Han chinese	494

Table 8: **Mouse dermal fibroblast dataset information.**

Ref	Donors (n)	Age (y)	Sex	Ethnicity	Number of fbs.
Abbasi et al., 2020	1	-	-	C57BL/6J	1718
Boothby et al., 2021	1	22 (d)	M	C57BL/6J	9280
Buechler et al., 2021	1	9 (w)	F	C57BL/6J	6621
Haensel et al., 2020	5	7 (w)	F	C57BL/6J	6327
Joost et al., 2020	10	5(6), 9(4) (w)	F	C57BL/6J	949
Phan et al., 2020	2	21 (d)	-	C57BL/6J	1493
Shin et al., 2020	3	2, 12, 18 (m)	F/M	C57BL/6J	8090
Shook et al., 2020	1	7 (w)	M	C57BL/6J	7728
Vorstandlechner et al., 2021	1	7 (w)	F	BALB/c	4343

Table 9: **Unsupervised axis assignation of human dermal fibroblast populations.** The assignation was done by applying the method in Materials and Methods section 5.7.

Dataset|Population	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C5	D1	D2	E1	T1
Ahlers 2022	A	A	A	A	B	B	B	B	C	C			D		B/E	B
Billi 2022	A	A	A		A/B	B	B	B	C/E	C	C	C	D	D	E	E
Boothby 2021	A		A		B			A/B	C/E	C	A	C	D		E	
Burja 2022	A	A	A		A/B		B	B							E	
Deng 2021	A	A	A		B	B	B	B	C	C	C	C	D	D	E	A/C
Gao 2021	A	A	A		B	B	B	B	C	C	C	C/E	D	D	E	A/C/E
Gaydosik 2019	A	A	A		A/B	B		B	A/C/E	C	C		D	D	E	B
Gur 2022	A	A	A	A		B		B	C	C	C				E	A/B
He 2020	A		A	A		B	B	A	C	C		C		D		B
Hughes 2020	A	A	A		A	A/B	B	B	C	C	U	U				
Kim 2020	A/B/E	A	A		B	B	B		C	C	U	C/E	D	D	E	A/C/E
Liu 2021	A	A	A		B	B	B		A/C	C	A	C/E	D		E	
Mariottoni 2021	A	A	A	A	B	B		B	C/E	C				B/D	E	E
Mirizio 2020	A	A	A	A	B	B	B	B	C	C				D		U
Reynolds 2021	A	A	A		A/B	B	B									
Rindler 2021	A	A	A	A	B	B	B	A/B	C	C	A/C			D		
Solé-Boldo 2020	A	A	A		B	B	B	B	C	C	A/C/E		D		E	A/E
Tabib 2018	A	A	A	A	B	B	B	B	C	C	C		D	D	E	
Tabib 2021	A	A	A	A	B	B	B	B	C/E	C	A/C		D	D	E	U
Tabula Sapiens 2021	A		A		A/B					C			D			
Theocarditis 2020	A	A	A	A	B	B	B	B					D		E	
Theocarditis 2021	A	A	A		B/E	B	B	B			E			D	E	
Vorstandlechner 2020	A	A	A		A/B	B	B	A/B							E	
Vorstandlechner 2021	A	A	A	A	B	B		B		C						
Xu 2021	A	A	A		A/B	B	B	B		C		C				

Table 10: **Unsupervised axis assignation of mouse dermal fibroblast populations.** The assignation was done by applying the method in Materials and Methods section 5.7.

Dataset|Population	x1	x2	x/y	y1	y2	y3	y4	y5	z1	z2	w/x	w1	w2	w3	w4	w5	v1
Abbasi 2020	x		x	y	y	y	y	y	z	z	U		w	w		z	
Boothby 2021	x	x	x	y	y	y	y	y	y	z	U	w		w	w	w	
Buechler 2021	x		x/y	y	y	y	y		z	z			w				
Haensel 2021	x	x	y	y	y		y	y/v	z	z	y/v			w		y	
Joost 2020	x		x/y	y	y	y	y	y	z	z	y/w/v	w	w	w		x	v
Phan 2020	x	x	x/y			y		y	y		U	w	w	w			v
Shin 2020	x	x		y			v		y		x/y	w	w	w/v	w/v	w	
Shook 2020	x			y	y	y	y		z	z	w	w	w	w	w	y	
Vorstandlechner 2021	x	x	U	y	y	y	y		z	z	U		U				

Table 12: **GOEA gene statistics.** Statistics from GOEA analysis on human population markers. The proportion of genes is the ratio between the number of genes appearing associated with GO terms and the number of genes submitted to the GOEA. The redundancy rate is the mean number of times the genes associated to GO terms from a population appear.

	A1	**A2**	**A3**	**A4**	**B1**	**B2**	**B3**	**B4**	**C1**	**C2**	**C3**	**C5**	**D1**	**D2**	**E1**
# genes submitted	74	71	31	57	54	47	45	46	53	46	29	49	53	56	44
# GO terms (p <0.03)	12	37	15	54	44	51	58	55	8	4	32	48	41	25	56
# genes across GO terms	48	39	16	37	37	29	28	28	14	13	20	27	32	22	29
% genes in GO terms	.65	.55	.52	.65	.69	.62	.62	.61	.26	.28	.69	.55	.60	.39	.66
redundancy rate	1.5	5.0	3.0	4.9	8.7	5.1	10.1	7.2	2.6	1.5	5.6	7.9	4.2	4.2	6.3

Table 11: **LR pairs in human fibroblast populations.** For each LR pair, the interacting genes, as well as the most relevant populations where that gene is expressed are shown.

Gene A	Gene B	Population A	Population B
ACKR3	CXCL12	A1, A4	B2, B4
ACKR4	CCL13	A1, A3	D2
ACKR4	CCL19	A1, A2, A3	B3
ACKR4	CCL2	A1, A2, A3	B1, B3, D1
ADA	DPP4	A1, B4	A1, A4
ANTXR1	WNT5A	A2	C5
APOD	LEPR	D1	A2
APOE	LDLR	B2, B3	D1
BAMBI	INHBA	D1	C5
BMP7	PTPRK	C5	C2
BST1	CAV1	B4	A2, A3, D2
CADM1	CADM1	C2	C2
CADM1	CADM3	C2	A1
CAV1	ICAM1	D2	B1
CSF1	SIRPA	B2, B3	A3
CD40	TNFSF13B	B1, B2, B3, B4	B2, B3
CD44	PRG4	A2, A3, B1	A4
CD44	TIMP2	A2, A3, B1	A1, A4, C2
CD44	TIMP3	A2, A3, B1	E1
CD44	FGFR2	A2, A3, B1	A2
CMKLR1	RARRES2	B2, B4	B2, B4
CXCL12	CXCR4	B2, B3, B4	C5, D2
CXCL12	DPP4	B2, B3, B4	A1, A3, A4
CXCL2	DPP4	B1,B2	A1, A4
EDN3	EDNRA	A2	C1
EFNA5	EPHA3	A2	C1
EFNA5	EPHB6	A2	A2, A3, B3
EFNB1	EPHB6	D1, D2	A2, A3, B3
EGFR	GRN	D1, D2	A1, A2, A3, A4
EGFR	ICAM1	D1, D2	B1
EGFR	FGF13	D1, D2	C2
EGFR	LRIG1	D1, D2	C2
F3	LDLR	D2	D1
FGF13	FGFR2	C2	A2
FGF7	FGFR2	B4	A2
FGF7	NRP1	B4	B1, B2, B3, B4
FZD1	WNT2	C2, C3, D2	E1
FZD1	WNT5A	C2	C5
FZD2	SFRP1	D1	C2, E1
FZD2	WNT2	D1	E1
FZD2	WNT5A	C1, D1, D2	C5
FZD4	WNT5A	B4	C5
FZD6	WNT5A	A1	C5
FZD7	WNT11	E1	A2
FZD7	WNT5A	E1	C5
GPC1	SLIT2	C1, C2, D2	B2, B4
HGF	NRP1	B2	B2
HLA-DRB1	OGN	B1, B2, B3	C2, D1
HRH1	IL6	D2	B1
JAG1	NOTCH3	C1, C5, D2	B2, D2
JAM3	JAM3	D2	D2
LDLR	WNT5A	D1	C5
LGR4	RSPO1	C1	A2
LGR4	RSPO3	C1	C5
LGR4	RSPO4	C1	C2
LGR5	RSPO1	A1	A2
LGR5	RSPO3	A1	A2
LGR5	RSPO4	A1	C2
NCAM1	ROBO1	C2	C3, C5
NGFR	TTR	D2	A2
NRP1	PGF	B2, B4	C2, E1
NRP2	PGF	C2, C3, C5, D1, D2	C2, E1
NRP2	SEMA3C	D1	A4
NRP2	VEGFD	D1	C2
PTK7	WNT5A	A2	C5
PTN	SDC1	C3, E1	C3
PTPRK	WNT5A	C2	C5
ROBO1	SLIT2	C3, C5	B2, B4
ROBO2	SLIT2	A2, C1, C5	B2, B4

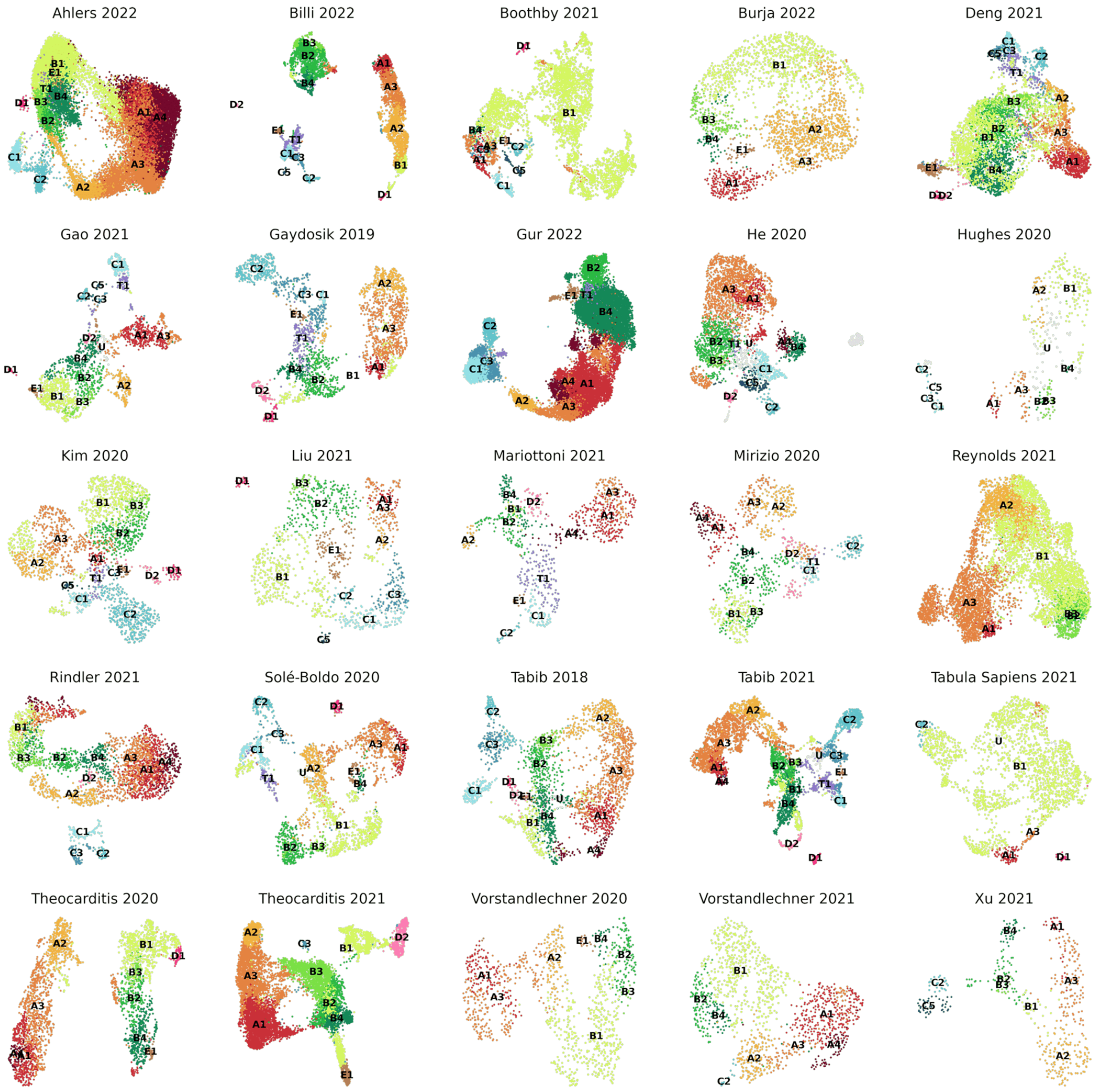


Figure 42: **UMAP plots of all human fibroblast datasets from secondary analysis.** Colours of populations are shared across datasets.

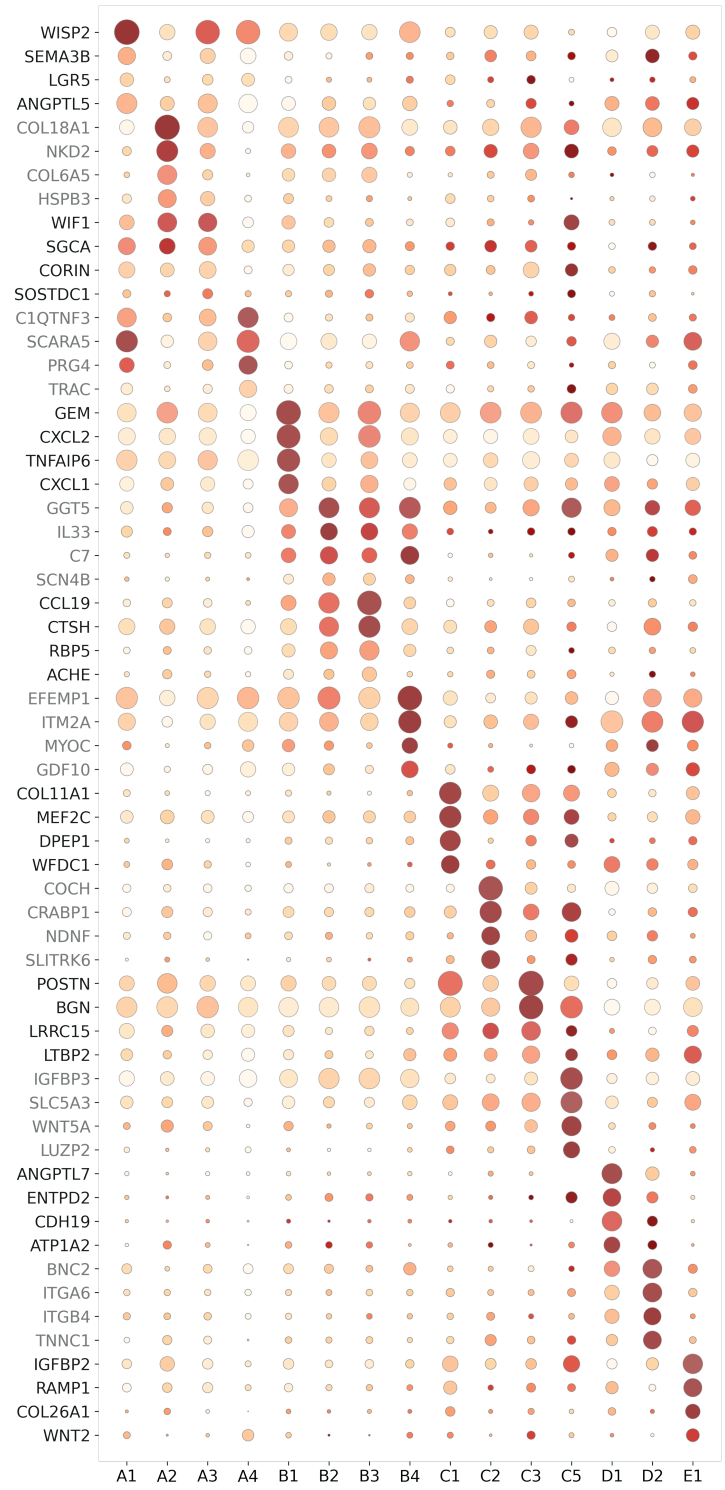


Figure 43: **Dot-plot of relevant markers of human fibroblast populations.** 4 markers are chosen for each population. The size of the circle represents the proportion of cells in that population expressing the marker–the larger the circle the larger the proportion–; and the colour represents the mean expression of the gene in that population–the redder/browner the colour the higher the expression–.

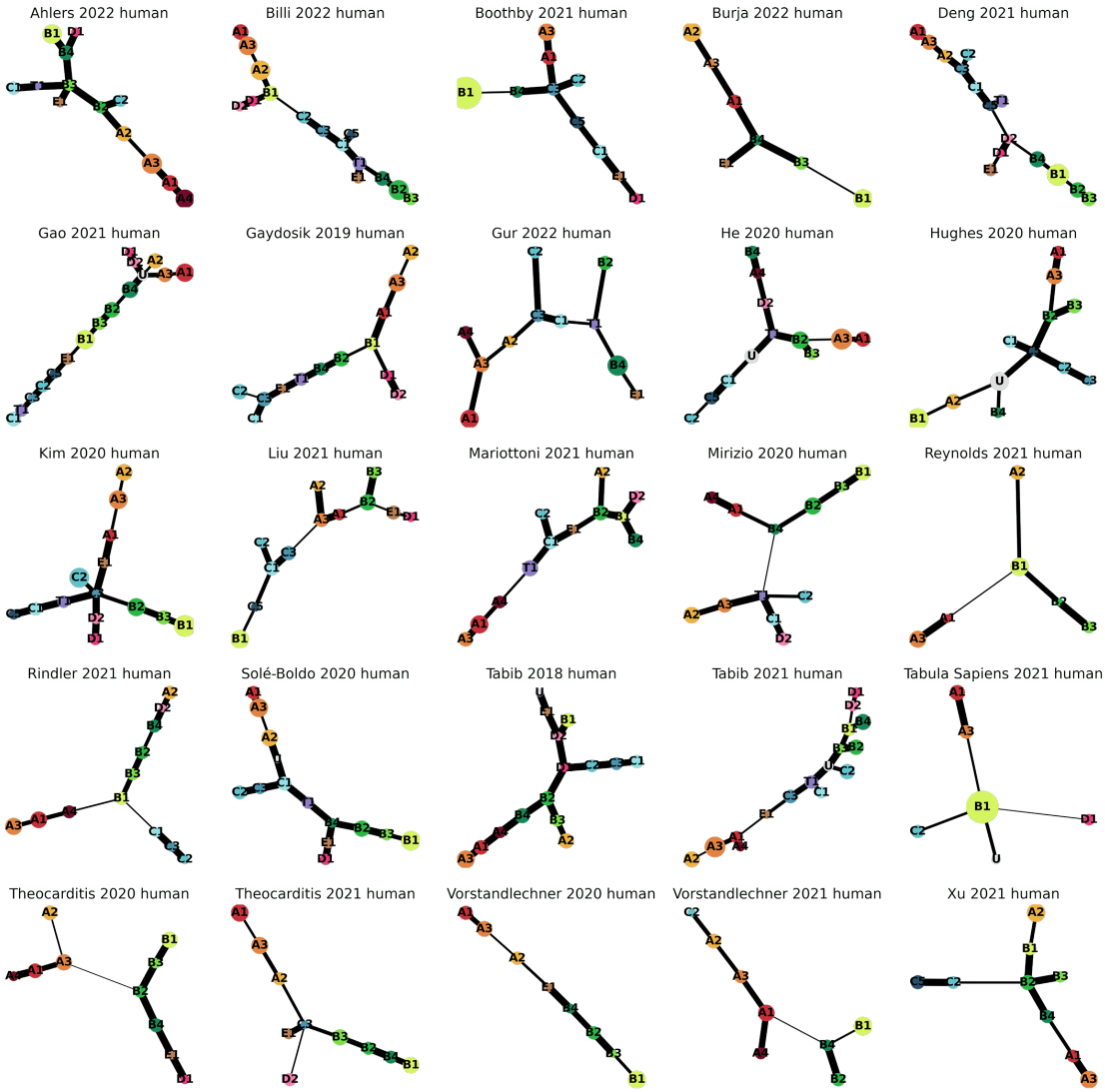


Figure 44: **PAGA tree graph of human dermal populations.** Colours of populations are shared across datasets. Thicker lines indicate greater connectivity between nodes.

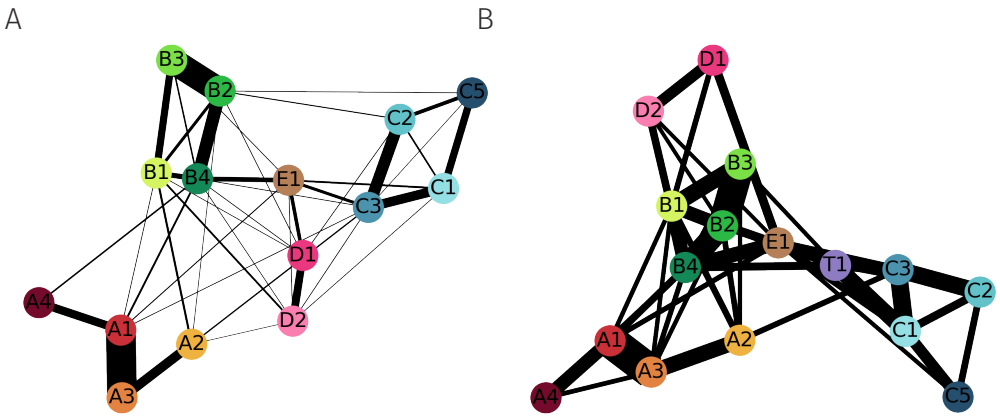


Figure 45: **Joint PAGA graphs of human dermal populations.** Graph in A is constructed from merging all PAGA trees, and graph in B from merging all PAGA graphs.

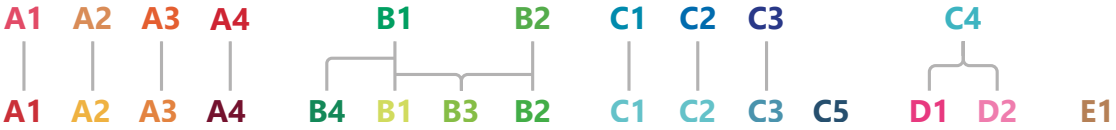


Figure 46: **Changes in human populations between primary and secondary analysis.** Populations in the upper row are the ones depicted in the primary analysis from chapter 8; and the ones in the lower row are the ones from the secondary analysis, at the time of writing this thesis.

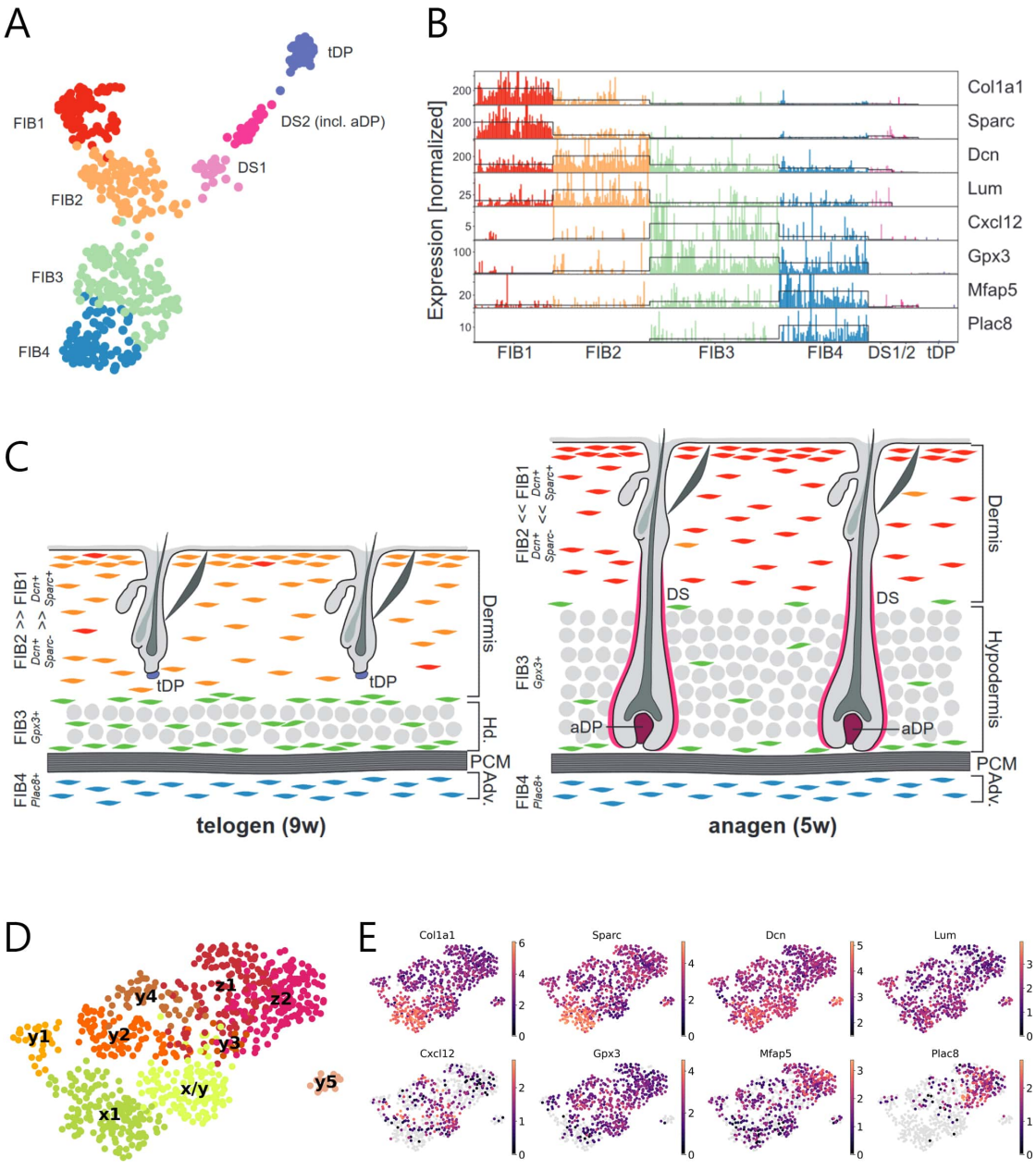


Figure 47: **Mapping of Joost et al. (2020) populations.** (A) UMAP plot of mouse dermal fibroblasts from Joost et al. (2020), adapted from Figure 6A of the publication. (B) Tracksplot of markers genes of populations FIB1 to FIB4, adapted from Figure 6B of the publication. (C) Scheme of the location of FIB1 to FIB4 populations in telogen and anagen phases, adapted from Figure 6I of the publication. (D) UMAP of dermal fibroblasts from Joost et al. (2020) after being reanalysed. Only cells from axes x, y and z are shown. (E) UMAP plots showing the expression levels of markers from subfigure (B).

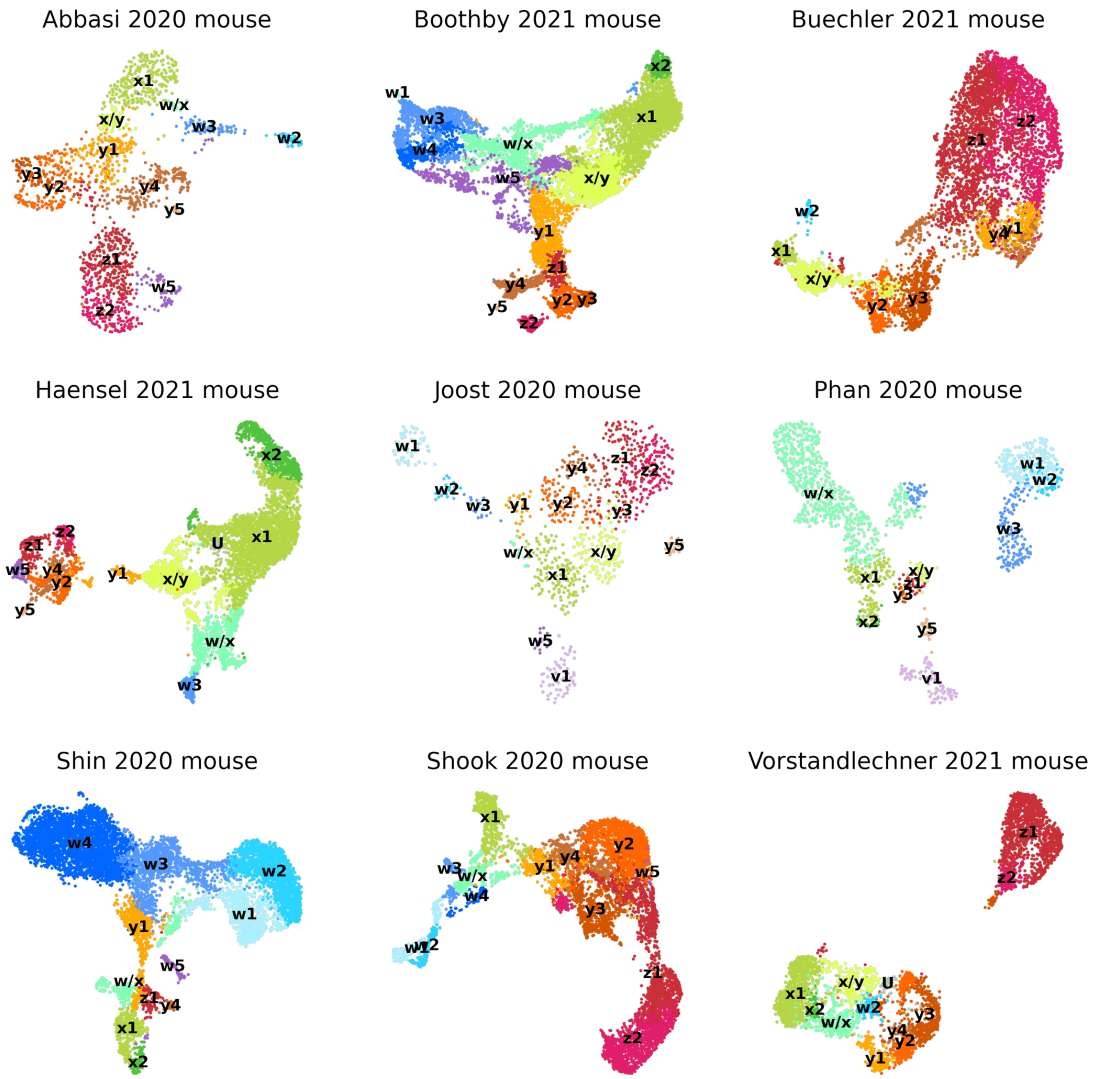


Figure 48: **UMAP plots of all mouse fibroblast datasets from the secondary analysis.** Colours of populations are shared across datasets.

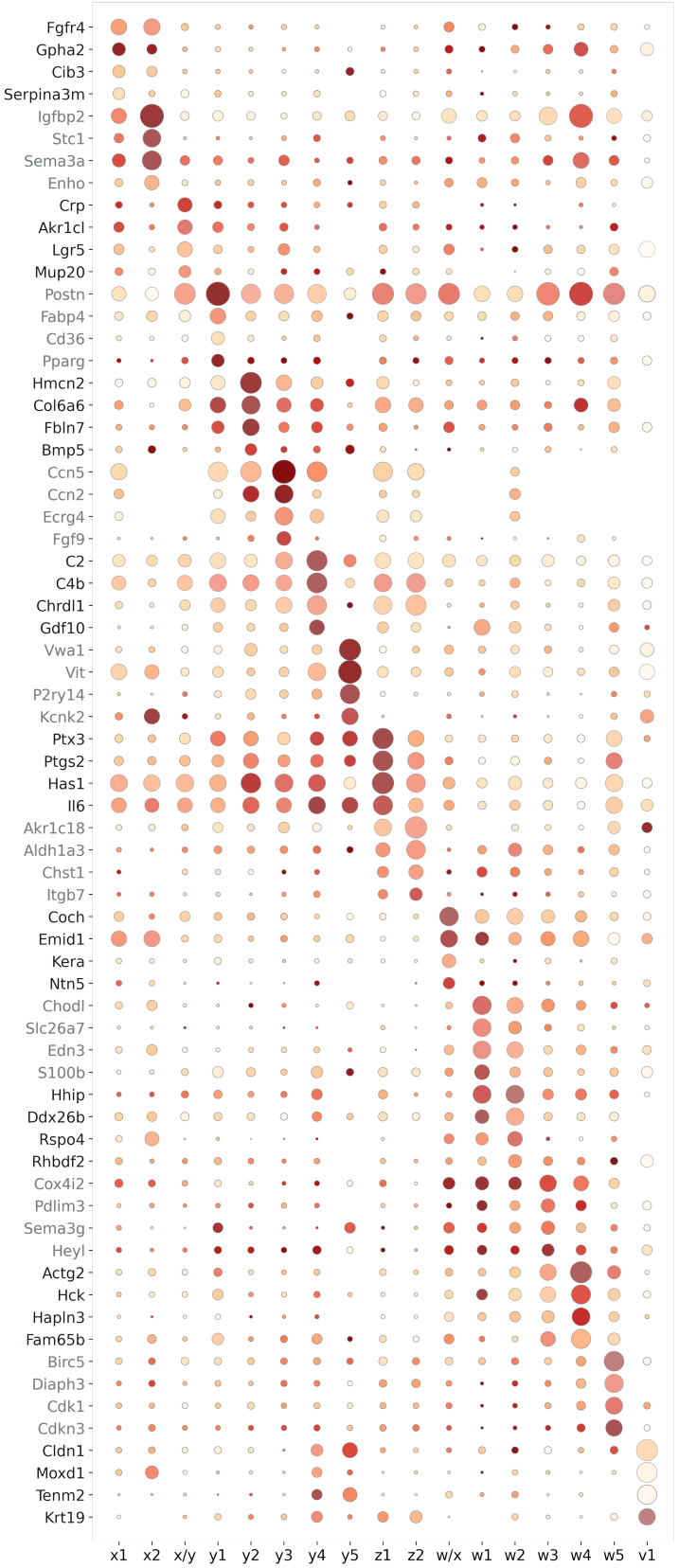


Figure 49: **Dot-plot of relevant markers of mouse fibroblast populations.** 3 markers are chosen for each population. The size of the circle represents the proportion of cells in that population expressing the marker–the larger the circle the larger the proportion–; and the colour represents the mean expression of the gene in that population–the redder/browner the colour the higher the expression–.

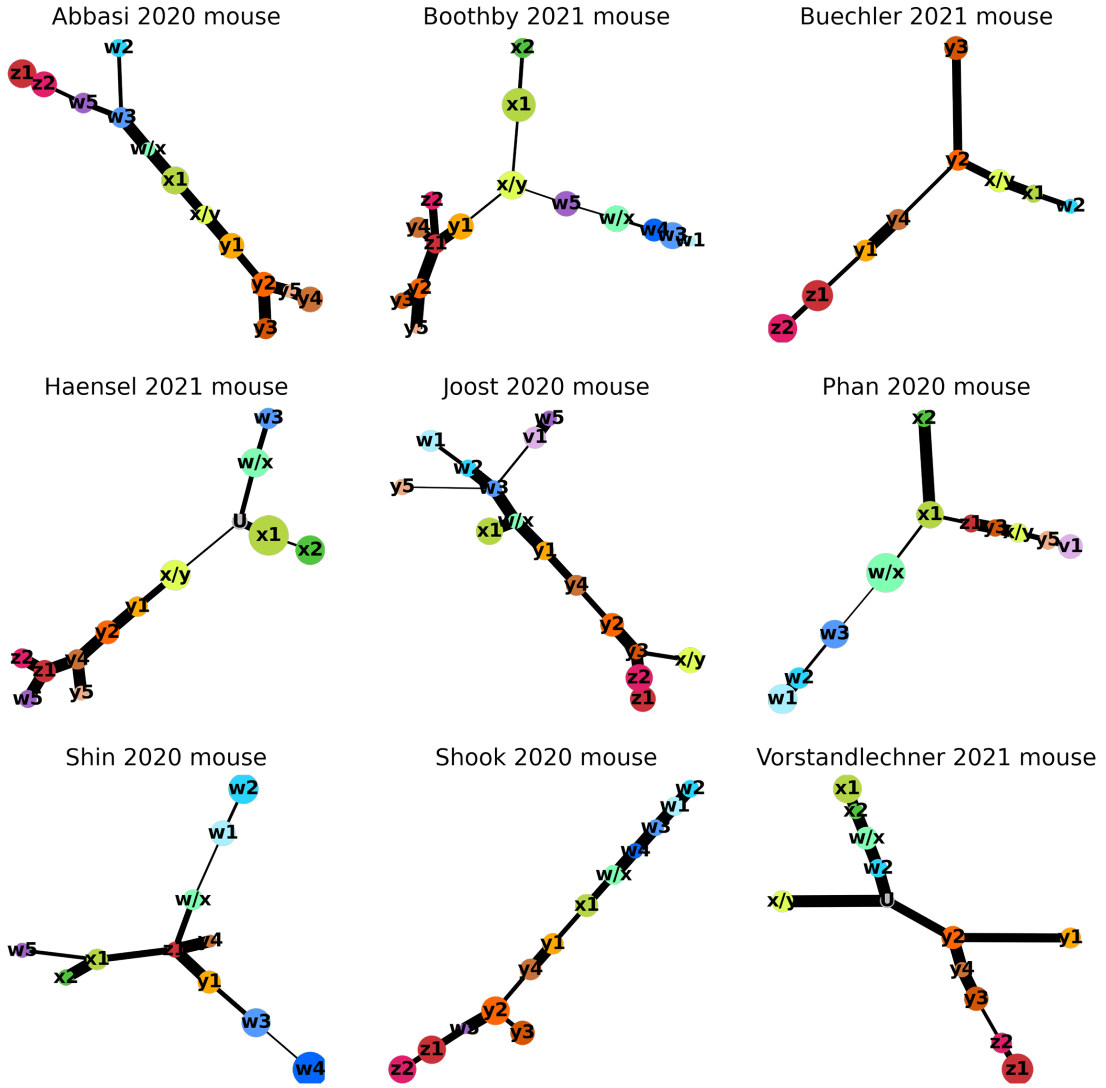


Figure 50: **PAGA tree graph of mouse dermal populations.** Colours of populations are shared across datasets. Thicker lines indicate greater connectivity between nodes.

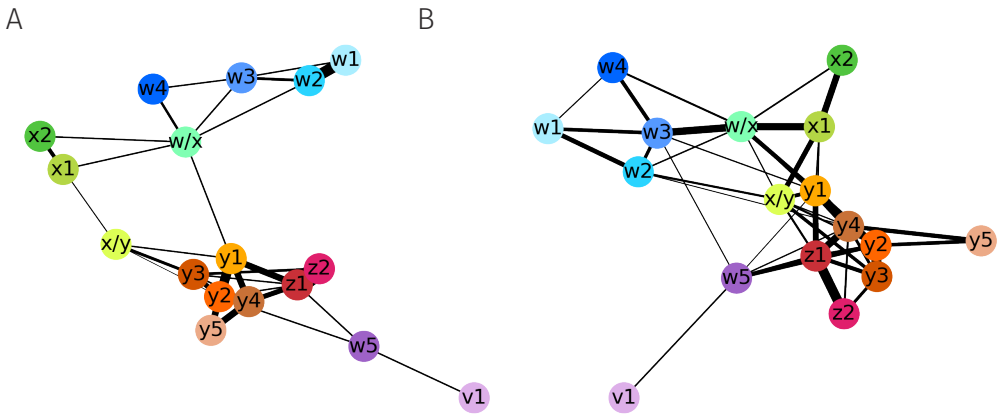


Figure 51: **Joint PAGA graphs of mouse dermal populations.** Graph in A is constructed from merging all PAGA trees, and graph in B from merging all PAGA graphs.

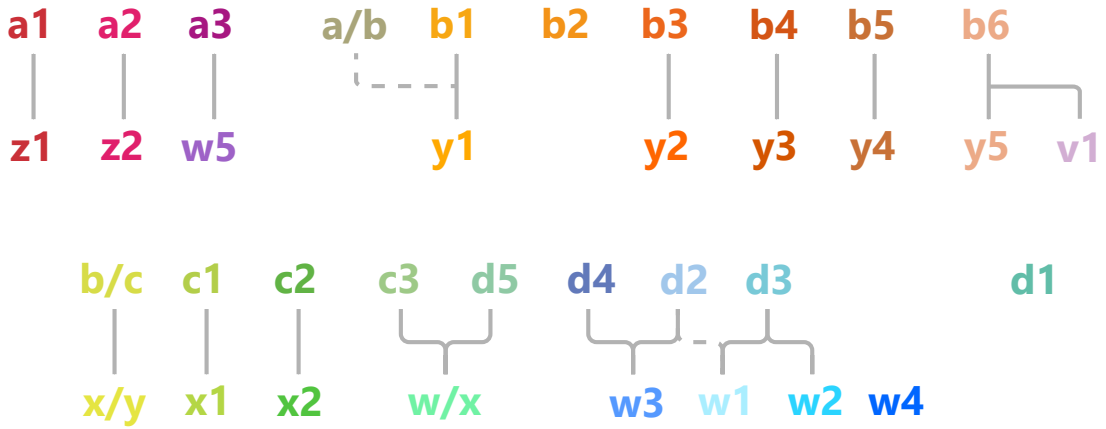


Figure 52: **Changes in mouse populations between primary and secondary analysis.** Populations in the upper row are the ones depicted in the primary analysis from chapter 9; and the ones in the lower row are the ones from the secondary analysis, at the time of writing this thesis. Since there are major changes between populations, compared to human ones, the primary populations are depicted with an asterisk to separate them in instances where primary and secondary populations are mentioned at the same time.

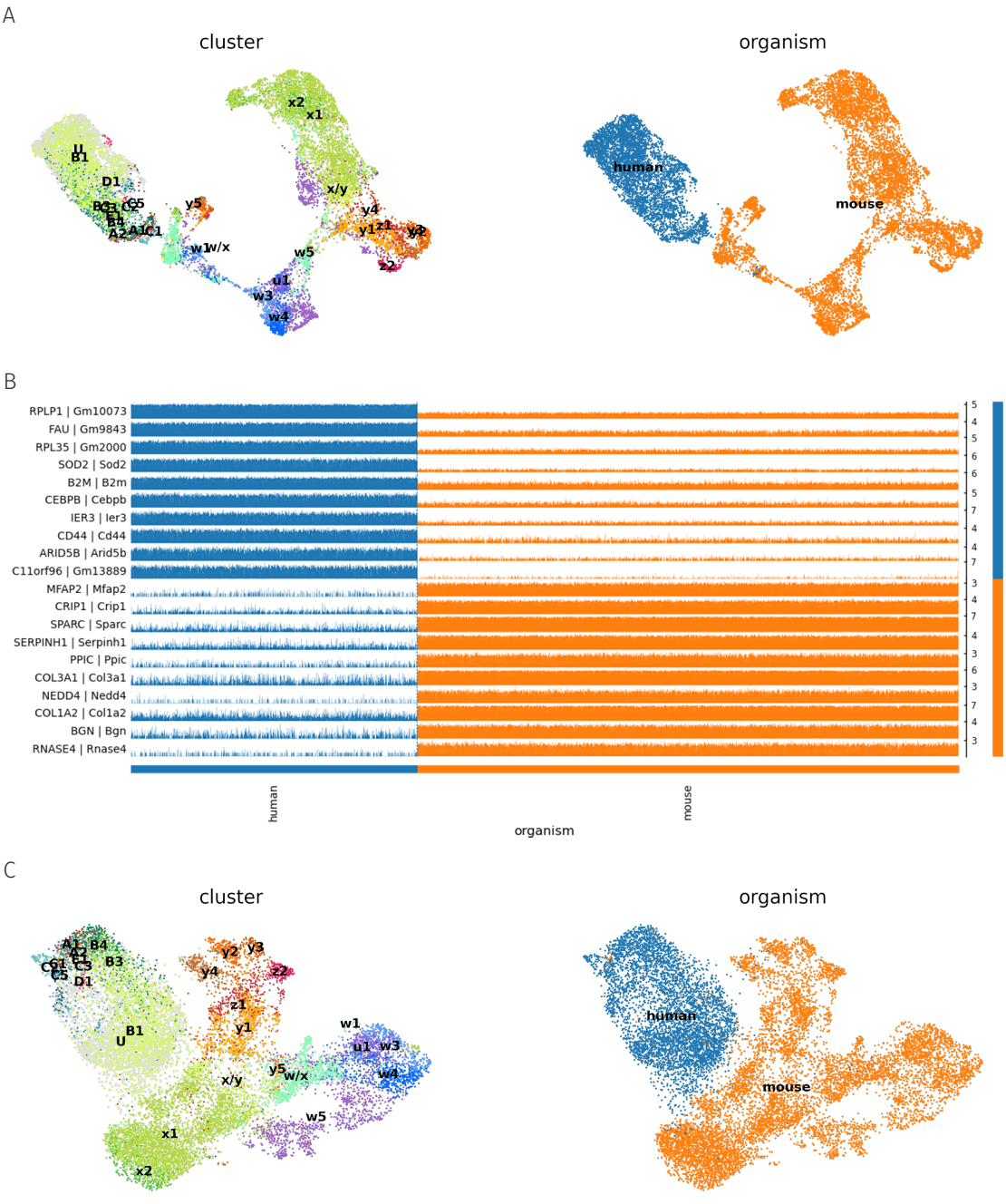


Figure 53: **Analysis of mouse/human integration on Boothby et al..** (A) UMAP plots of clusters and organism of Boothby et al. dataset, using only `harmony_integrate` function. (B) First 10 DEGs between human and mouse populations. (C) UMAP plots of clusters and organisms of Boothby et al., 2021 dataset, using `regress_out` and `harmony_integrate` functions.

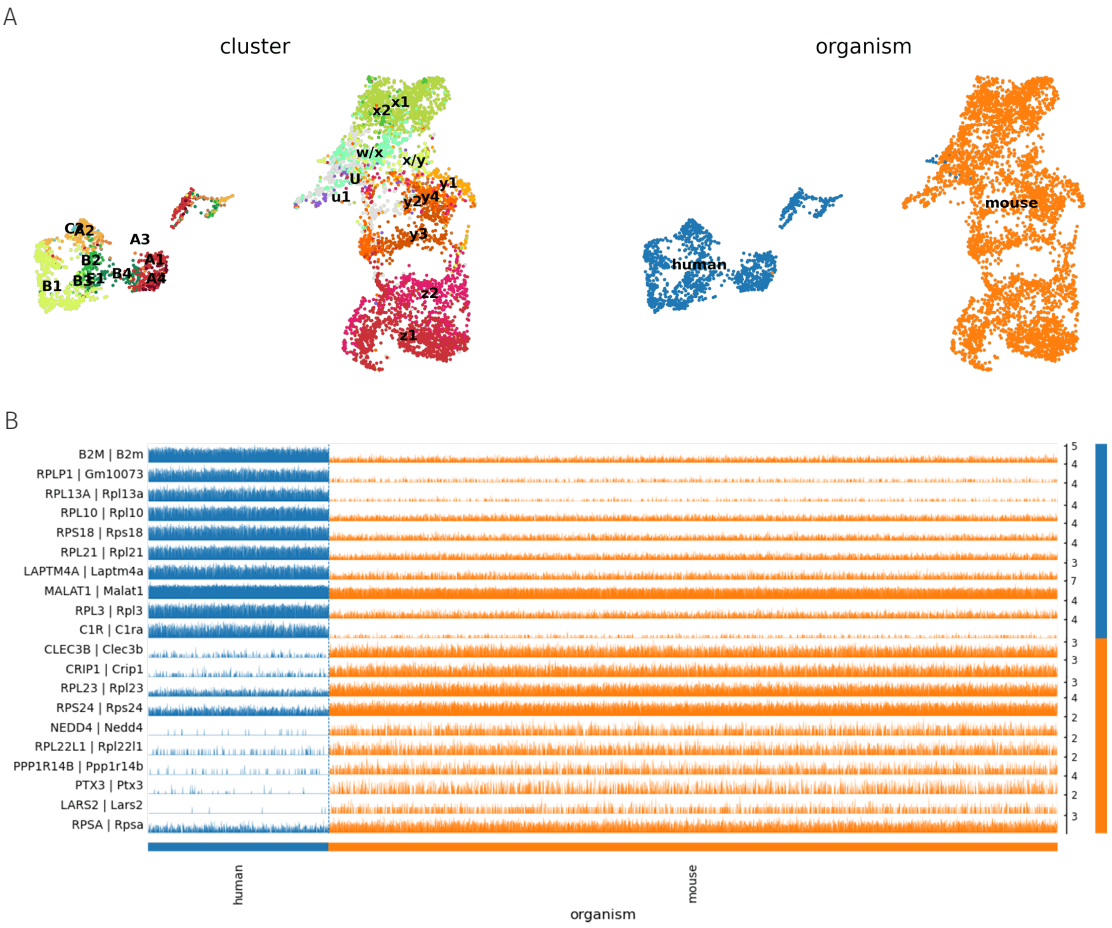


Figure 54: **Analysis of mouse/human integration on Vorstandlechner et al., 2021.** (A) UMAP plots of clusters and organism of Vorstandlechner et al., 2021 dataset, using only `harmony_integrate` function. (B) First 10 DEGs between human and mouse populations. (C) UMAP plots of clusters and organism of Vorstandlechner et al., 2021 dataset, using `regress_out` and `harmony_integrate` functions.

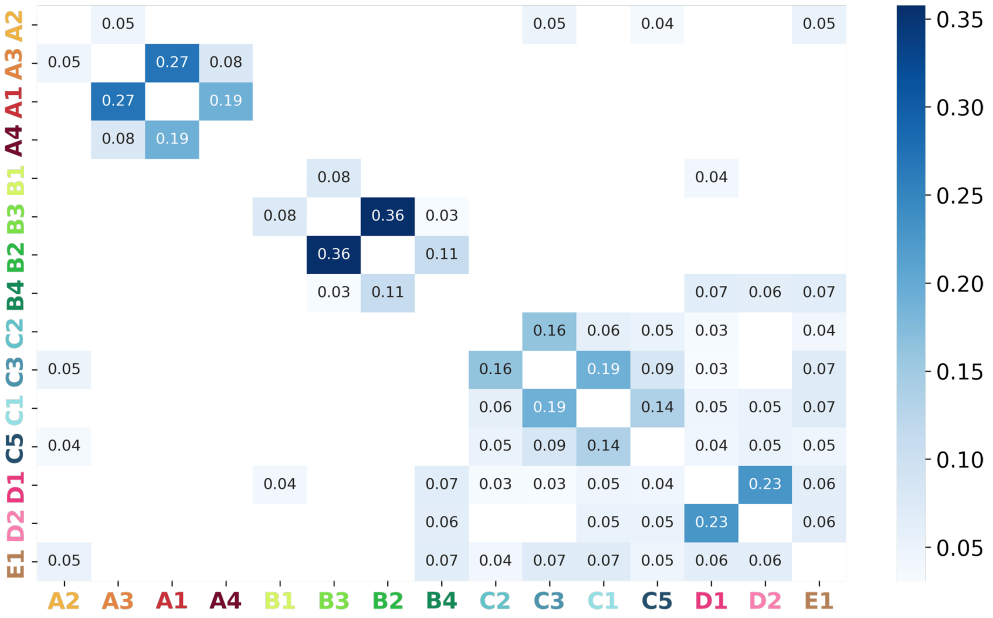


Figure 55: **Heatmap of human marker overlap across populations.** For each pair of populations, the number in the cell represents the proportion of overlapping markers.

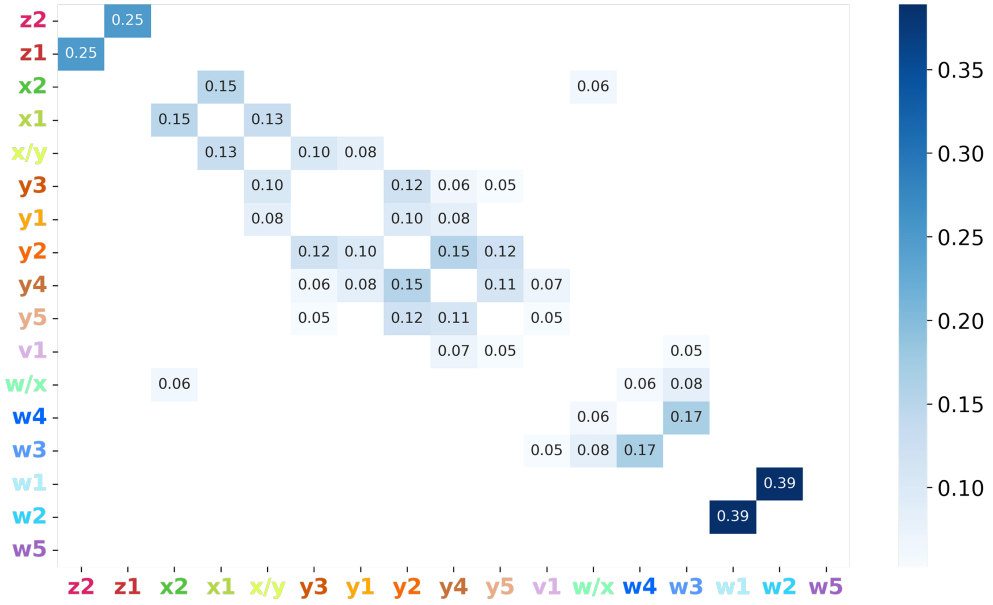


Figure 56: **Heatmap of mouse marker overlap across populations.** For each pair of populations, the number in the cell represents the proportion of overlapping markers.

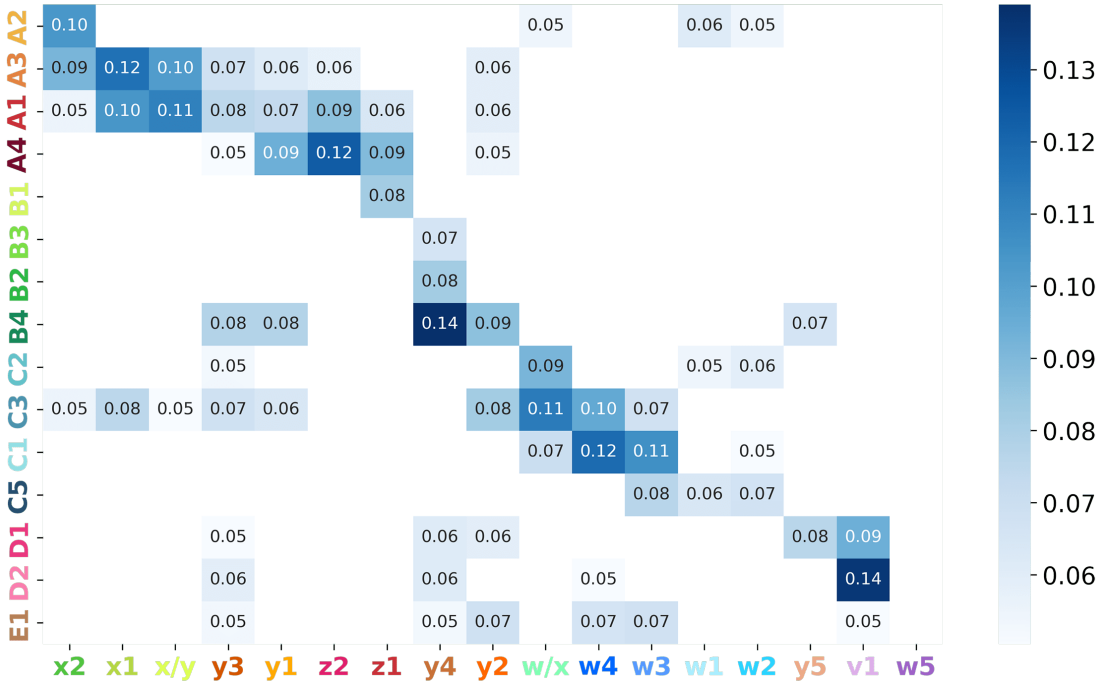


Figure 57: **Heatmap of human and mouse marker overlap across populations.** For each pair of populations, the number in the cell represents the proportion of overlapping markers (N=150).

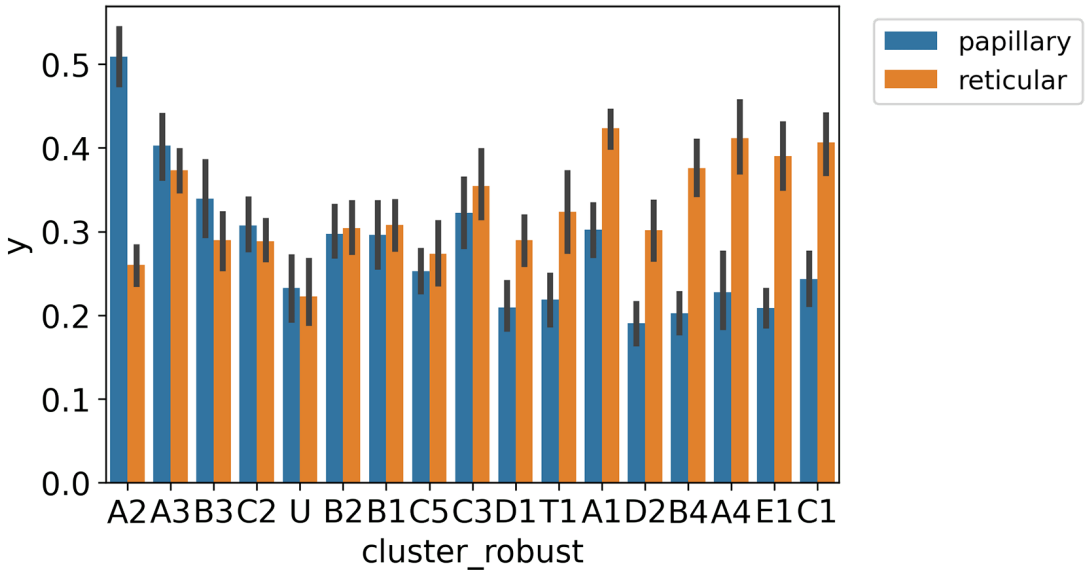


Figure 58: **Comparison of scores for papillary and reticular signatures.** Barplot of marker-to-population algorithm scores for each population. Barplot error bars represent the std of the score across datasets.

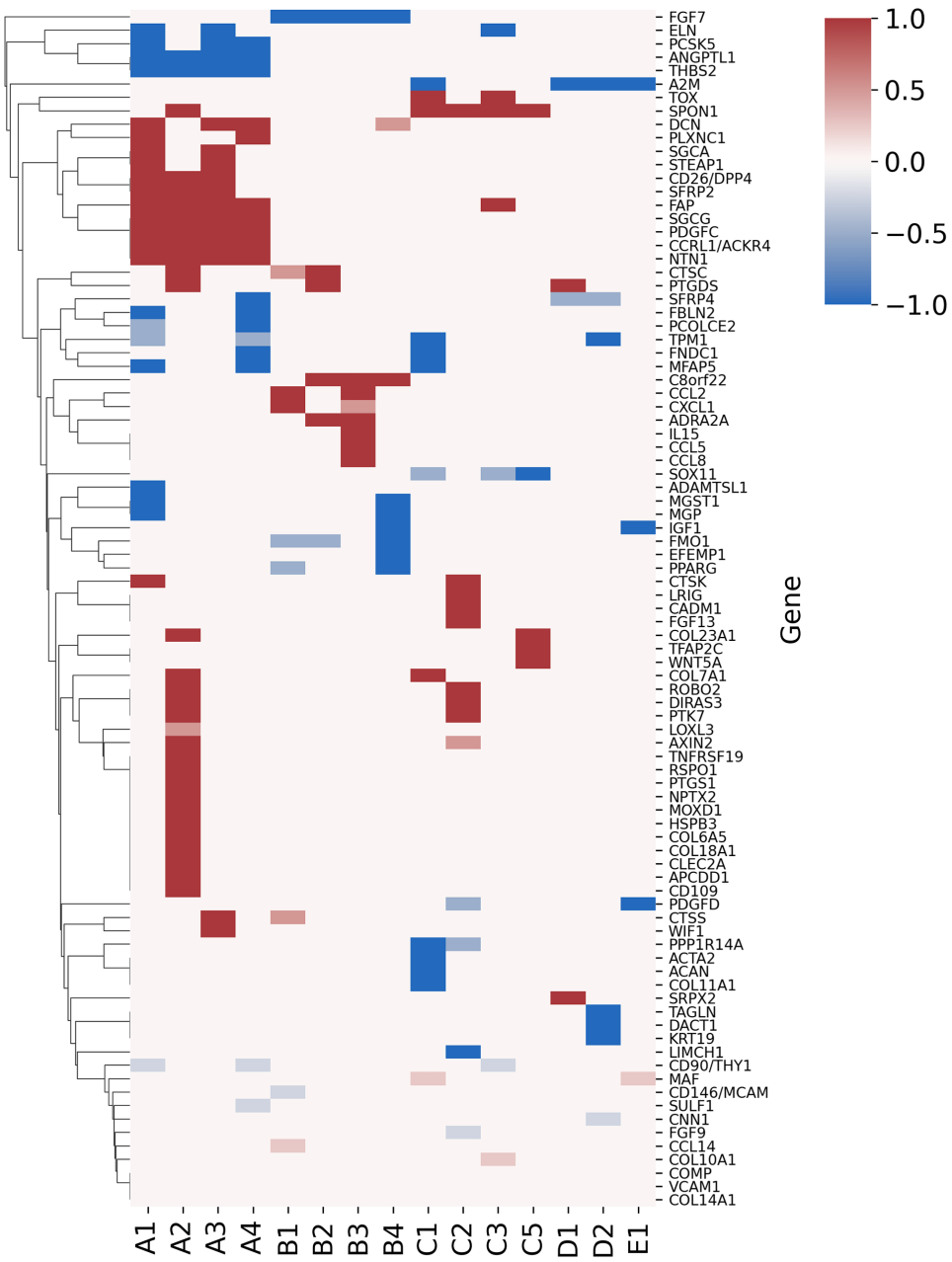


Figure 59: **Heatmap of papillary/reticular markers across human populations.** For each gene/population pair, a darker colour represents that the gene is a better marker of that population.

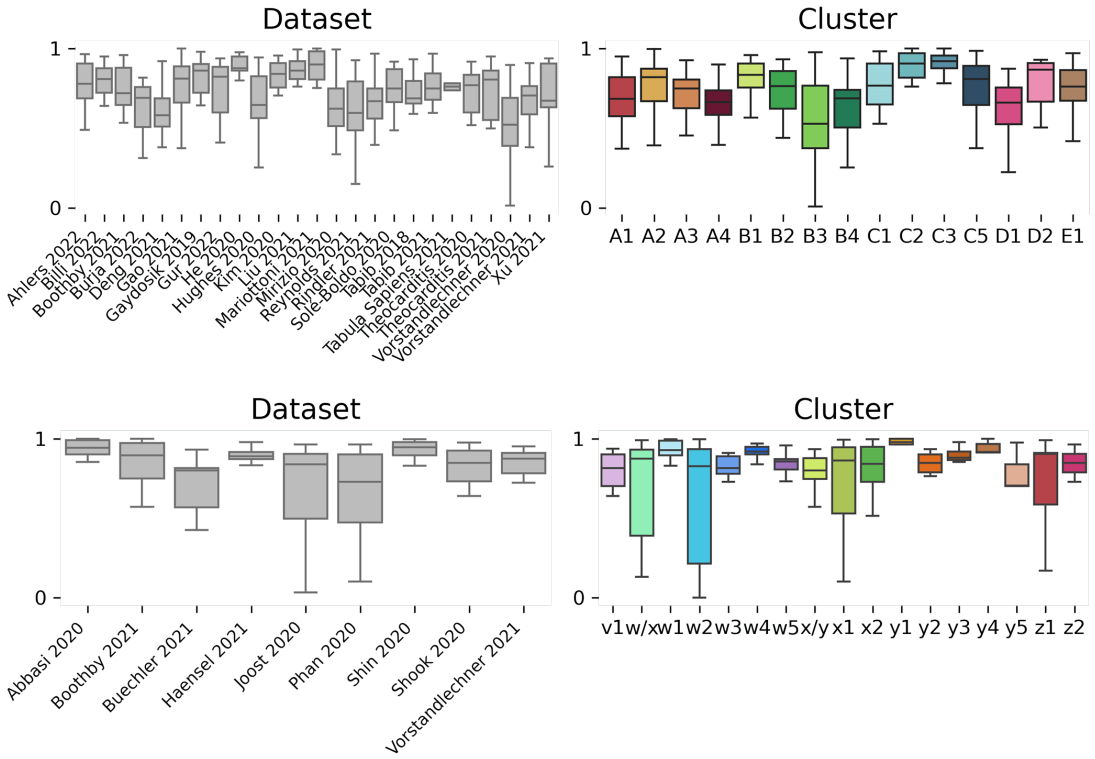


Figure 60: **Boxplots of general robustness score.** Each boxplot represents the mean robustness score of each dataset, across datasets. Each box represents the quartiles of the distribution, centre bar represents the median, and whiskers extend to Q1 - 1.5 * IQR and Q3 + 1.5 * IQR.

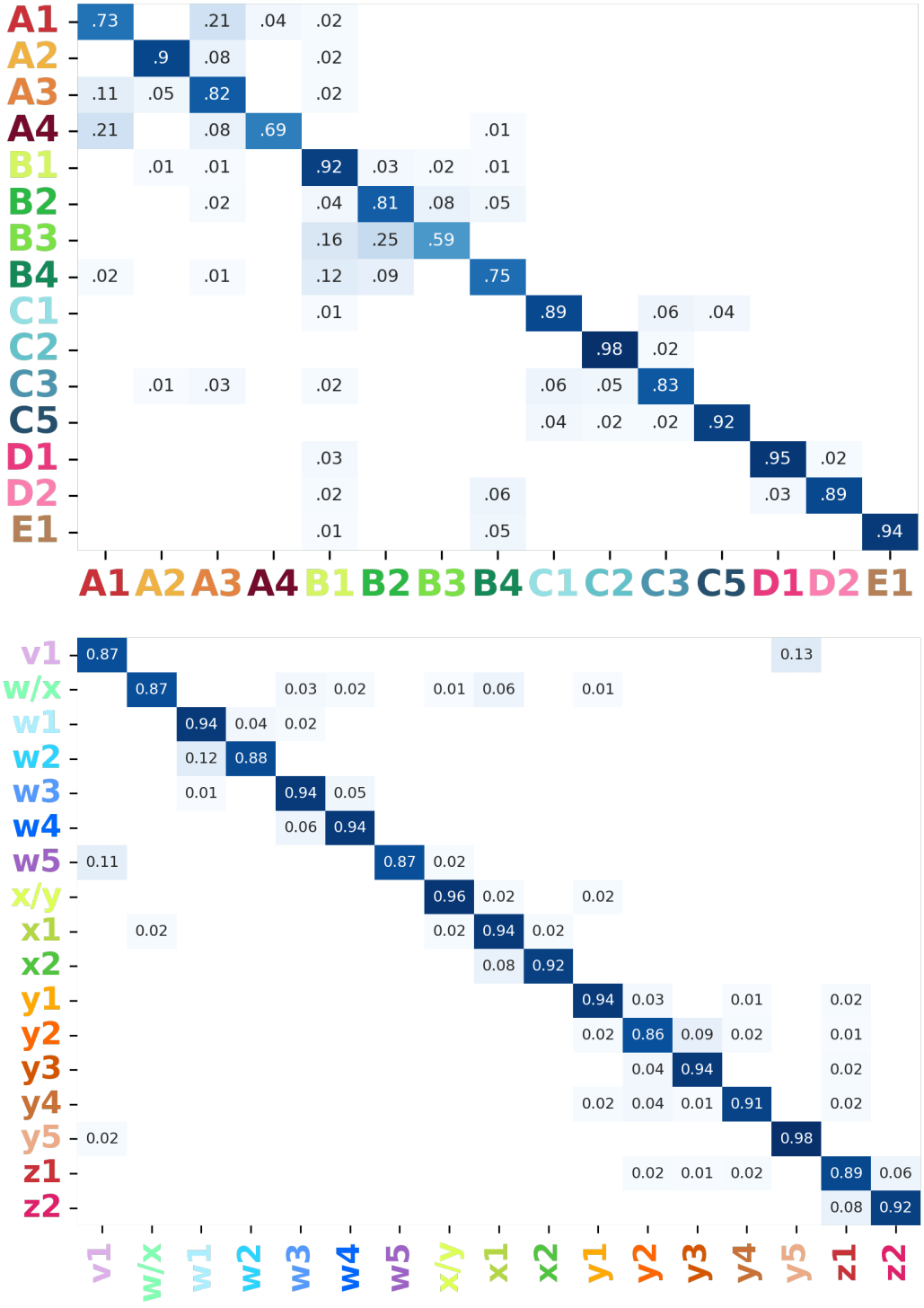


Figure 61: **Adjacency matrix of robust cluster assignation.** Each row-column combination is the proportion of times the row cluster is assigned to the column cluster. In this heatmap, numbers represent the median of the proportions across datasets.

11

REANALYSIS OF REYNOLDS ET AL. DATASET

11.1 MOTIVATION

The analysis presented in this section is motivated by the extensive gathering and analysis performed in the secondary analysis performed in chapter 10. One of the datasets in this analysis is from a landmark paper published in Science, where Reynolds et al., 2021 produced a dataset of 528,253 sequenced cells obtained from healthy adult skin (five female patients undergoing mammoplasty surgery) and fetal samples, as well as inflamed skin from atopic dermatitis and psoriasis patients.

In healthy dermal fibroblasts, the authors described three populations: the main cluster termed Fb1, and two minor subpopulations, Fb2 and Fb3. Fb2 was additionally described as enriched in fetal and inflamed skin samples. We aimed to analyse whether the Fb1, Fb2 and Fb3 populations were consistent with the A–C fibroblast types and subtypes that we had just described in Ascensión et al., 2021, and which were used in chapter 8 and in chapter 10 before the minor axis and cluster reassignment. More specifically, we reasoned that at least the most abundant subpopulations we had defined, A1, A2, B1 and B2, should be clearly detected in a >500k cell dataset, thus further validating our previous scRNAseq study. In contrast, we found that a substantial proportion of the Reynolds et al., 2021 scRNAseq dataset shows a predominance of differential expression of stress and hypoxia-related genes. Thus, data extracted from this source should be interpreted in light of this bias. It is possible that other existing large datasets suffer from similar methodological problems, which might be due to insufficient oversight.

11.2 REASSESSMENT OF THE MAIN CELL POPULATIONS IN A LARGE SKIN DATASET REVEALS THE PRESENCE OF CLUSTERS WITH STRESS- AND HYPOXIA-RELATED GENE SIGNATURES

Using an unsupervised population-matching algorithm, we observed that in each of the healthy donors analysed by Reynolds et al., 2021, at least two independent fibroblast clusters expressed signature markers of the A1, A2, B1 and B2 populations. One set of cells corresponded to the Fb2 population, and the second set corresponded to the Fb1 and Fb3 populations. After batch effect correction, a joint analysis of all donors showed that the cluster duplication observed in each donor could be replicated jointly (Figure 62).

We, therefore, assumed that some global effect should be affecting the cells, i.e. Fb2 might be "a copy" of Fb1+Fb3 cells, although perhaps affected by some alteration. DEG analysis between Fb2 and Fb1+Fb3 revealed enrichment in ontology terms associated with cell stress–e.g. *unfolded protein response*, *regulation of apoptotic process*, *mRNA catabolic process*–. We then designed a signature gene list composed of 50 DEGs commonly associated with stress in very different scRNAseq settings–e.g. *ATF3*, *BTG2*, *FOS*, *FOSB*, *GADD45B*, *HSPA1A/B*, *IER2/3*, *JUN*, *JUNB*, *NFKBIA*, *NR4A1/2*, *PPP1R15A*, *RHOB*– (Adam et al., 2017; Brink et al., 2017; Denisenko et al., 2020; O'Flanagan et al., 2019; Waise et al., 2019). Using this signature, the Fb2 population over-expressed *BTG2*, *EGR1*, *FOSB*, *IER2*, *SOCS3*, and *ZFP36*, among others, indicating that these cells clustered together mainly due to cellular stress.

Further analysing the Fb1 and Fb3 cells independently, we observed that the A1, A2, B1 and B2 populations appeared twice again. A DEG analysis between each pair of duplicated populations disclosed genes in one of the split populations that were related to glycolysis (*ALDOC*, *ENO2*, *GAPDH*, *PGK1*, *PDK1*, *PFKFB4*, *PYGL*), cell integrity, hypoxia and apoptosis (*BNIP3*, *BNIP3L*, *ANGPTL4*, *LOX*, *HILPDA*); whereas the second split population over-expressed units of the mitochondrial ATPase and complex I, indicating an active oxidative metabolism. It is well known that cells under hypoxic conditions switch from aerobic to anaerobic metabolism to keep energy homeostasis within the cell (Mohyeldin et al., 2010; Simon et al., 2008; Xiao et al., 2019).

We, therefore, generated a curated list of hypoxia-related genes and managed to separate the non-hypoxic from the hypoxic group with the population-matching algorithm. Once stressed or hypoxic cells were removed based on a set threshold of expression of signature genes, we mapped the main types of fibroblasts in what we termed the "normal" cell subset of Reynolds et al., 2021 (Figure 63A). Fibroblast A1, A2 and B2 populations were independently mapped, and we also found clusters which seemingly were mixtures of previously defined populations e.g. B1/B2, A1/A2, or A2/B2. No type C fibroblasts were detected.

To understand whether the stress and hypoxic signatures were only present in fibroblast subsets or could also be traced to other populations within the dataset, we mapped the stress and hypoxia gene signatures to perivascular cell, keratinocyte, vascular endothelial cell, lymphoid cell, and APC clusters. Our reanalysis of healthy donors, fibroblasts, perivascular cells, keratinocytes, and vascular endothelial cells showed clear hypoxia and stress-related clusters (Figure 63B). For instance, the VE3 population, described by Reynolds et al., 2021 as increasing in patients suffering from inflammatory conditions, presented a clear stress-related transcriptomic profile. On the other hand, most of the VE2 population over-expressed hypoxia-related genes. On lymphoid cells, we did observe a sub-cluster of stressed Tc/Th cells but no clear hypoxic profiles. On APCs, an inflammatory macrophage cluster showed hypoxia, and the M2 and DC2 clusters showed stress-related profiles. Some of these results may be expected in physiological conditions for immune cells, but others could be attributed to sample handling.

Finally, we tested if the aforementioned stress and hypoxia-related signatures were present in the previously published scRNAseq datasets of human skin (He et al., 2020; Solé-Boldo et al., 2020; Tabib et al., 2018; Vorstandlechner et al., 2020). The levels of expression of these genes were clearly superior in the Reynolds et al. dataset as compared to other available resources (Figure 64). For instance, regarding the hypoxia profile, some of the clusters in Reynolds et al., 2021 dataset show values higher than 0.6, whereas none of the clusters from the rest of the datasets achieve these values and, instead, show predominantly values lower than 0.3.

11.3 CORRECTION OF STRESS AND HYPOXIA SIGNATURES SHOWS THAT STRESSED CELLS SHOW A NON-RECOVERABLE GENE SIGNATURE

Since the stress and hypoxia-related expression profiles are apparent, we were interested in studying the "reversibility" of the transcriptomic signatures, creating a normalised dataset where hypoxic and stressed cells could merge with the normal cells, and classifying the whole dataset into the original cell types described in (Ascensión et al., 2021). To this end, we applied two approaches with similar results. On the one hand, we considered cell states as batches and applied batch effect correction with *bbknn* and *harmony*. On the other hand, we applied regression on the stress and hypoxia scores shown in Figure 64 based on the *Seurat*'s linear regression function implemented in *scanpy*. Since both approaches showed similar results (Figure 65), we show the results of the latter case in Figure 66.

To further study if stress and hypoxia transcriptomic profiles are "recoverable", we generated two types of datasets, one with the stress or hypoxia cells and another containing normal cells. When applying the correction to the stress + normal dataset, we observed that there was no integration between the two states (Figure 66A). On the other hand, there was a good integration between the hypoxia and normal cell states (Figure 66B), and the main fibroblast populations could be correctly mapped (Figure 66C). From these results, we infer that the transcriptome from stressed cells is much more altered than the one from hypoxic cells, to the extent that stress cells are in a computationally non-reversible state.

11.4 HARSH SAMPLE PROCESSING CAN BE THE CAUSE FOR STRESS AND HYPOXIA SIGNATURE

After concluding the stress and hypoxia analysis, we questioned why this phenomenon occurred to the Reynolds et al., 2021 dataset. We believe the reason for this could be the harsh sample processing of skin. As stated in the article's Materials and Methods section: "The top 200 μm-thick layer of the skin was cut with a dermatome and digested with dispase (1h at 37ºC) to separate dermal and epidermal layers. Both layers were digested in collagenase for 12h at 37ºC, cells were filtered and subjected to FACS sorting before library generation and sequencing".

While this strategy warrants high purity of the obtained cell populations, the long processing times (>16h) and the use of warm dissociation for a long period might have significantly affected gene expression patterns of relevant numbers of cells in this setting. In this sense, aiming to process large numbers of cells involves longer processing times. High processing times (even > 60') have previously been reported to generate significant transcriptomic alterations (Brink et al., 2017; Waise et al., 2019) and, in particular, warm dissociation is associated with stress response (Denisenko et al., 2020), which is apparent in the transcriptomic profiles of part of the cells. An alternative to warm dissociation may be the use of cold-active psychrophilic proteases (Potter et al., 2019).

A recent methodological study developed by Burja et al., 2022 states that digestion times vary from 2h to overnight digestion, although long times are completely discouraged due to their impact on cell viability and transcriptome alteration. Additionally, QC performance in scRNAseq is rarely reported, which hinders standardisation and result comparison. In their paper, the authors apply skin digestion for 2 hours.

11.5 HIGH COMPUTATIONAL TIMES ARE A SECOND KEY FACTOR TO HINDER SCRNASEQ DATASET ANALYSIS QUALITY

TheReynolds et al., 2021 dataset contains, after some basic filterings, approximately 450k cells. We became interested in analysing the run times of a standard single-cell pipeline procedure–consisting of quality control, PCA, graph neighbour construction, dimensionality reduction, clustering and DEG calculation–using different cell numbers, to see how this analysis is scaled. The analysis results are observed in Figure 67 and detailed in Table S1. The analysis shows that running the pipeline a single time in the whole dataset in a working station takes approximately 1 hour. The parts with the longest-running times are the batch, clustering and DEG calculation. Additionally, when analysing trends in the processing times, we observe an inflexion point at around 30,000 cells, marking two clear run time trends. For higher numbers of cells, the processing times are further increased–2.1 times per doubling of cells– compared with a lower number of cells–1.5 times per doubling of cells–. For good measure, a single run of this pipeline analysis on an extended dataset with 1M cells would take 2 hours and 25 minutes.

Apart from the sample quality and processing, there are computational and analytic challenges with such complex datasets. As shown in Figure 67, the run times of analytic pipelines vastly increase with the number of cells. Thus, if the processing time is expected to be the same for a small dataset and a big dataset, due to the low time to perform a beginning-to-end analysis adapted to the current fast-paced times of publication, this leads to a more shallow exploration of the initial stages.

Before a pipeline is run on a single-cell dataset, researchers usually have to spend some time doing an exploratory analysis, where they select the cutoff values for quality control, explore different batch effect removal methods, or tune the parameters for clustering, neighbour graph calculation, and other steps in the pipeline.

These decisions are made on the basis of the output of the differently-preprocessed versions on downstream analyses: how the datasets look on UMAP plots, how robust their DEGs are, etc. Usually, this part of the analysis requires several reruns of the same pipeline to find the best parameters and obtain an overall view of the limitations of the dataset and the general information elements that will be obtained from it. This means that single-cell pipelines are not linear but rather an iterative process where researchers have to make decisions based on the output of previous steps.

As a consequence, if the results from the initial stages of the analysis are overlooked, and biases go unnoticed, these effects propagate downstream the pipeline–e.g. observing differences in healthy vs diseased samples, finding rare populations, pathway/ontology analysis, or RNA velocity analysis–and artefacts can be presented as genuine results, hindering the dissemination of quality results to the scientific community.

11.6 TABLES AND FIGURES

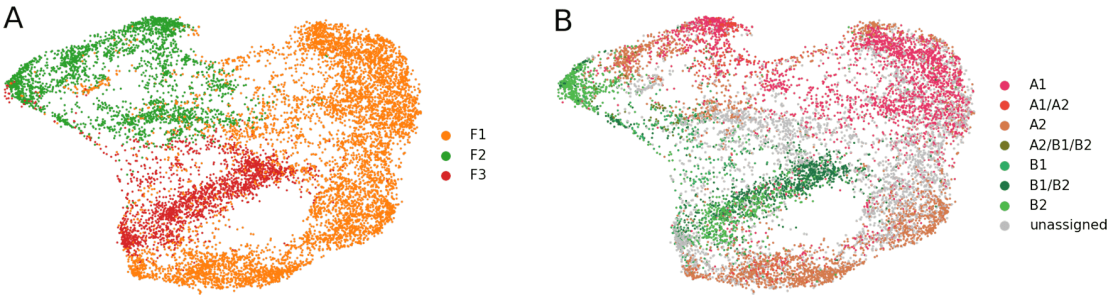


Figure 62: **Duplicated fibroblast populations can be observed in Reynolds et al. clusters.** (A) UMAP plot of Reynolds et al. fibroblast populations where each original cluster (F1=Fb1, F2=Fb2, F3=Fb3) is marked. (B) Same UMAP as in (A) showing main fibroblast populations depicted in Chapter 8.

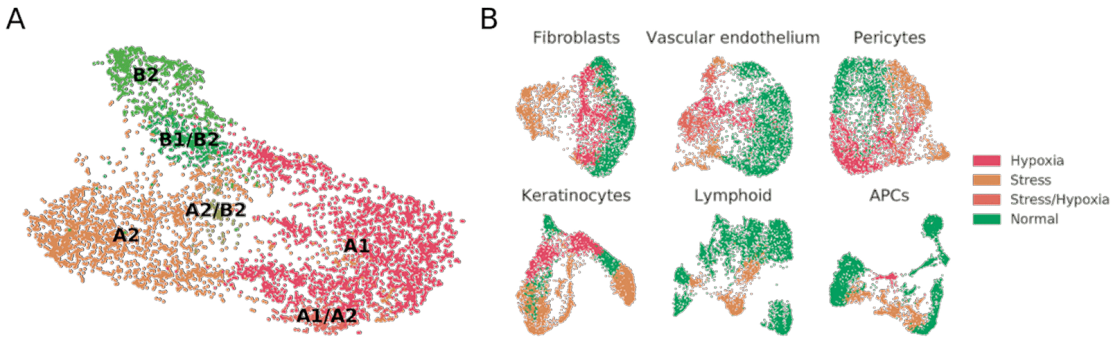


Figure 63: **Fibroblast subpopulations reveal the presence of substantial proportions of stressed and hypoxic cells**. (A) UMAP plot of "normal" fibroblasts (after removal of hypoxic and stressed cell subsets) reveals conservation of some, but not all, cell types previously described in independent datasets (Ascensión et al., 2021). (B) UMAP plots of fibroblast, vascular endothelium, pericyte, keratinocyte, lymphoid and APC cell populations from healthy donors, labelled to highlight hypoxic and stressed cell subpopulations as characterized by overexpression of defined gene signatures.

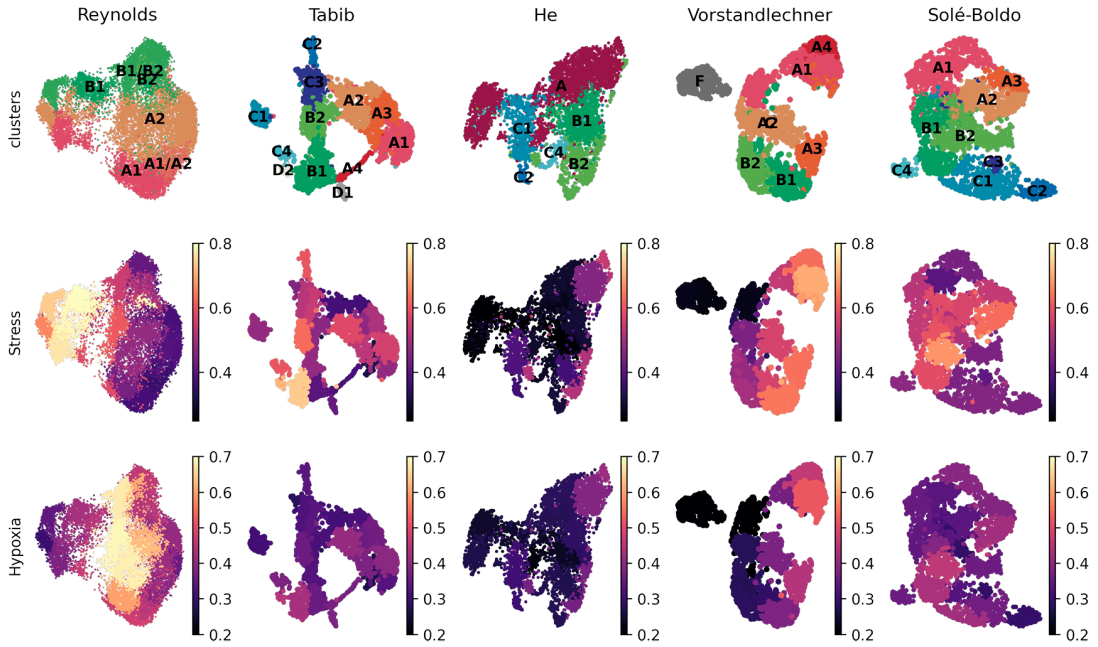


Figure 64: **Stress and hypoxia-related signatures in published human dermal fibroblast datasets**. (A) UMAP plot of "normal" fibroblasts (after removal of hypoxic and stressed cell subsets) reveals conservation of some, but not all, cell types previously described in independent datasets (1). (B) UMAP plots of human dermal fibroblast subsets as defined in (Ascensión et al., 2021) are shown here for five published datasets (He et al., 2020; Reynolds et al., 2021; Solé-Boldo et al., 2020; Tabib et al., 2018; Vorstandlechner et al., 2020), and depicted by the average levels of expression of stress and hypoxia gene signatures.

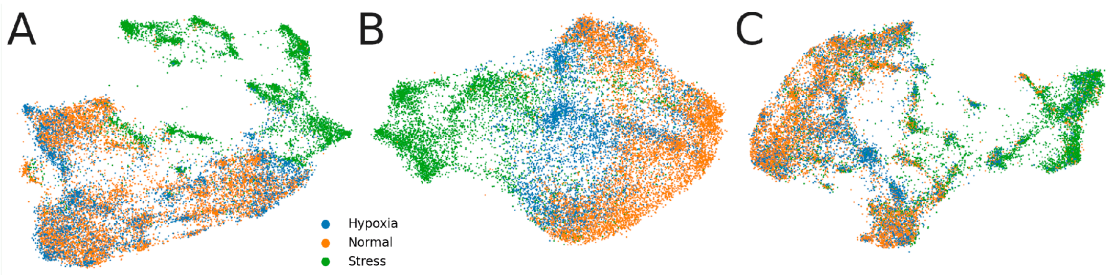


Figure 65: **Comparison of different correction methods for stress and hypoxia.** UMAP plots of (A) regression, (B) *bbknn* and (C) *harmony*.

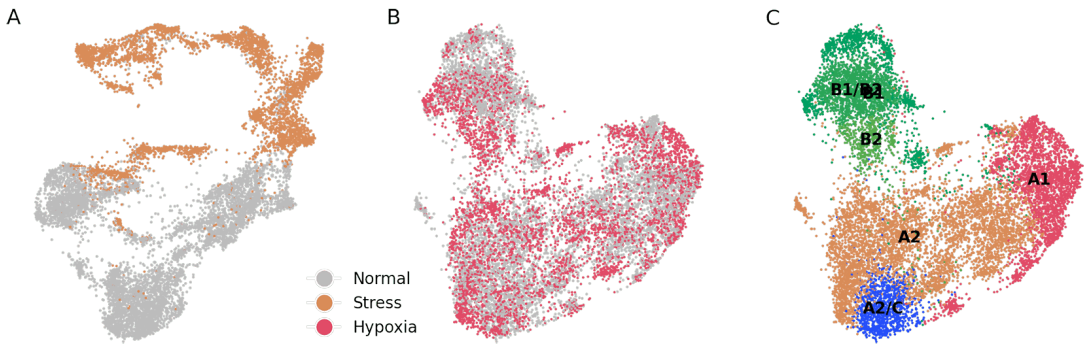


Figure 66: **Dataset merging of stress and hypoxia populations show mixed degrees of integration with the "normal" dataset**. (A) UMAP plot of merged "stress" and "normal" cells. There is a low degree of integration between both cell types. (B) UMAP plot of merged "hypoxia" and "normal" cells. There is a high degree of integration between both cell types. (C) Unsupervised assignation of fibroblast types from (B) reveals, similar to results from Fig. 63A, major fibroblast types.

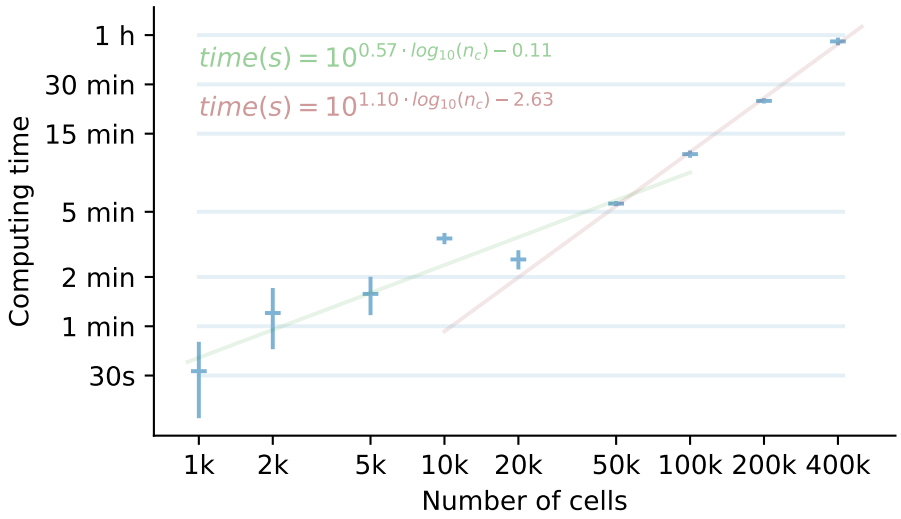


Figure 67: **Running times of a basic single-cell pipeline on Reynolds dataset**. For each number of cells, three running times are collected, and the mean (horizontal bar) and standard deviation (vertical bar) are shown. For the [1k - 20k] and [50k - 400k] intervals, two linear regressions were trained and extended outside of these intervals to show that there is a change in the processing time rate at ~ 30k cells. A doubling in the cell number implies 1.48 times more run-time in the [1k - 20k] interval, and 2.14 times in the [50k - 400k] interval.

Part V

DISCUSSION

12

DISCUSSION ABOUT FIBROBLAST HETEROGENEITY

In chapters 8, 9, and 10, we analysed the heterogeneity of dermal fibroblast populations in human and mouse samples. Throughout this discussion section we are going to use all the previous information–including gene markers, ontology terms, interactions with other populations and relationships across species–to generate a detailed and justified view of the putative roles of each human fibroblast population. Due to time constraints, analysis of mouse populations is not feasible, although human-mouse population similarities may be useful to extend this knowledge.

To explain the main functions of fibroblasts, we used the most relevant markers of each population and performed an extensive literature review of each marker. The information of all markers is recompiled in the *Specific marker function* table described in the Materials and Methods section 7. During the literature review, we observed that many genes were classified based on common themes and decided to describe them in depth each of the themes first. The selected themes are (1) ECM, (2) ECM modulation, (3) complement system, (4) immune response, (5) Wnt signalling, (6) TGF-β signalling, (7) lipid metabolism, and (8) vitamin A metabolism.

Hence, prior to analyzing each fibroblast population, we will first provide a description of each classification. This will help identify any shared patterns among the populations, making it to simpler explain the roles of the markers in each population afterwards.

12.1 PATHWAYS OF RELEVANCE

12.1.1 *ECM*

Due to the component diversity within the ECM, the gene markers corresponding to this section have been divided into four categories: (1) collagens, (2) elastic fibres, (3) small leucine-rich proteoglycans (SLRPs) and (4) other elements. The descriptions of individual genes are available in Supplementary Table S2.

COLLAGENS Classical collagens are collagens type I and III. Each collagen is composed of different fibres encoded by individual genes–e.g. *COL1A1*, *COL1A3*–. Both *COL1* and *COL3* families of genes are expressed by A1, A3, A4 reticular fibroblasts, and C1, C2, C3 HF-associated fibroblasts. Additional collagens–types IV, V, XII and XV–are expressed mainly by these populations as well. The principal

function of these fibroblasts is to provide structural support to the dermis, either by interacting with their own fibrils or interacting with other fibrils from this category.

There are collagens that are principally restricted to one or a few populations. One clear example is A2, which almost exclusively expresses collagens *COL4A2*, *COL6A1/2/3*, *COL6A5*, *COL7A1*, *COL13A1*, *COL18A1* and *COL23A1*. Most of these collagens are located in BM or BM-like structures, surrounding blood vessels, and in the DEJ, like COL13A1 or COL18A1.

Additionally, populations C1 and C2 express *COL11A1*, *COL21A1* and *COL24A1*; populations D1 and D2 express *COL8A1/2*, *COL9A3* and *COL28A1*; and E1 expresses *COL26A1*. Some of the functions of these collages will be extended later. The expression of these collagens may be relevant for the future identification of the populations in histological samples.

Lastly, there are two additional types of collagens with specific expression patterns: nucleating and FACIT collagens. Regarding nucleating collagens, traditionally, type V is located more in the papillary dermis and type XI more in the reticular dermis (Nauroy et al., 2017). These observations are not fully replicated in our single-cell analysis: we observed that *COL5A1/2* are expressed by A3, A4, C1, C2 and C3 populations, with a putative reticular location. *COL9A2/A3* chains are expressed predominantly by D1 population. Interestingly, these essential collagens are not produced by the *traditional* ECM-producing A1/A4 populations. Regarding FACIT collagens, *COL12A1*, reported as papillary (Nauroy et al., 2017) is not expressed by A2, but by the rest of A populations. *COL14A1* is also expressed by A1/3/4, and is reported as reticular.

Therefore, the expression of collagens is extremely diverse based on the different populations that expressed it–almost all of them, with the exception of B axis–and, therefore, their functions too. Interestingly, mRNA expression from scRNAseq analyses does not correlate with protein deposition for certain papillary collagens. Considering a mean turnover rate of 15 years for different collagens (Verzijl et al., 2000), it is possible that some of the putative reticular populations are also located in papillary regions in specific moments to produce these collagens.

ELASTIC FIBERS Elastic fibre genes expressed by dermal fibroblasts are *ELN*, *EMILIN2* and *FBN1*; as well as the associated glycoproteins *MFAP2* and *MFAP5*. All of these fibres are primarily expressed by A4, and by A1 to a lesser extent. This indicates that one of the differentiating functions of A4 with respect to A1 is the production of elastic fibres.

SLRPS SLRPs are a family of proteoglycans that contain leucine-rich repeats (LRRs) flanked by cysteine clusters that bind different GAGs. This structure allows them (1) to bind and regulate collagen and elastic fibre expression, (2) participate in immune functions by binding TLRs, TNFα and C1q, (3) bind different growth factor receptors like IGFRs or EGFR and some of their ligands like BMPs and WISP1, and (4) modulate fibrogenesis by binding TGF-β (Merline et al., 2009).

Similar to collagens, SLRPs are primarily expressed by A1, A4 from A axis; and C2 and C3 from C axis. The most common SLRPs expressed by A and C axes–e.g. *BGN*, *DCN*, *OGN*, *POSTN*, *ASPN*–are associated with binding to ECM components (collagens, fibronectin, tenascin C). Interestingly, ASPN, mostly expressed in C axis; competes with DCN, mostly expressed in A axis, to bind collagen (Kalamajski et al., 2009).

Therefore, as expected with collagens, it is apparent that A axis populations are involved in the expression of the majority of ECM components. Additionally, many of these SLRPs are partially related to some of their *non-ECM* functions that we will extend on, such as Wnt and TGF-β signalling, or certain immune responses.

OTHER ECM COMPONENTS For the elements of this category, we do not observe a clear pattern of expression by fibroblast populations. However, we do find some genes that are principally expressed in C axis populations–*ACAN*, *COCH*, *NPNT*, *SDC1*–; and some by B4 population–*EFEMP1*, *HSPG2*–.

Interestingly, we do observe *TNC* and *TNXB*, two tenascins with a complementary expression pattern (Valcourt et al., 2015), be expressed by complementary populations too: *TNC* is expressed by A2, B and C; whereas A1/A3/A4 express TNXB. Moreover, *TNXB* interacts with type I, III, V, XII, and XIV collagens, which A1/A3/A4 mainly expresses. Therefore, combined with ASPN/DCN pattern mentioned before; *TNXB* and *DCN* expression may be restricted to the dermal surroundings of A1/A3/A4 fibroblasts, and *TNC* and *ASPN* expression to A2/C fibroblasts.

In general, due to the large range of putative binding molecules described in the literature, the net of interactions of ECM proteins is highly complex; and specific analysis should be performed to elucidate specific binding partners and putative roles of each gene.

12.1.2 *ECM modulators*

ECM modulators are proteins that change the structure or composition of the ECM, usually replenishing some of the components, or during injury processes, allowing immune and other cells to reach the injury site. ECM modulators have been divided into three main categories: MMPs and TIMPs, ADAM family metalloproteases, and other elements. The descriptions of individual genes are available in Supplementary Table S3.

MMPS AND TIMPS There is a wide range of MMPs in dermal samples, and in this analysis, dermal fibroblast express mainly 6 of them: *MMP1*, *MMP2*, *MMP3*, *MMP11*, *MMP16* and *MMP27*. There is a large variability of expression within the populations and between MMPs.

Immune populations, concretely B1 population, expresses *MMP1* and *MMP3*, which show a wide range of collagen degradation patterns–MMP1 degrades primarily collagens I, II, III and VIII; and MMP3 degrades collagens IV, V, IX, X and XI–. Interestingly, MMP9, commonly found in immune responses, is not expressed by any fibroblast, but in immune cells (Jiang et al., 1998).

On the other hand, non-immune populations from axis A mainly express *MMP2* and *MMP27*. While MMP2 is well characterised to degrade a wide range of collagens, as well as vitronectin, fibronectin and laminin (Cabral-Pacheco et al., 2020), the functions of MMP27 are unknown (Cominelli et al., 2014). Lastly, MMP11 and MMP16 are populations expressed by C axis populations, and therefore are likely to be located in HF. These MMPs degrade mainly collagen III, but also BM-associated and nucleating collagens IV, V and IX; or aggrecan expressed by axis C populations (Cabral-Pacheco et al., 2020).

Regarding TIMPs, inhibitors of MMPs, most of them are expressed by A1 and A4–*TIMP1/2/3*–; but other populations express them differentially. *TIMP2* is expressed by axis C populations and is possibly regulating the presence of MMP11 and MMP16. On the other hand, *TIMP1* and *TIMP3* are expressed by

E1 and B4. Interestingly, the CD44 molecule interacts with TIMP2 and TIMP3 based on Table 11, which might be useful since CD44 binds a vast array of ECM components and might protect them from MMP-mediated degradation.

ADAM FAMILY This family is composed of ADAMTS and ADAMTS-like (ADAMTSL) proteins. Markers extracted for this analysis reveal the expression of three ADAMTS members–*ADAMTS4/9/18*–as well as 4 ADAMTSL members–*ADAMTSL1/3/4/5*–.

ADAMTS9 and *ADAMTS18* are expressed by C axis populations. ADAMTS9 acts on ACAN and VCAN. While VCAN is more or less thoroughly expressed in the skin, ACAN is almost exclusive of C populations, thus its homeostatic degradation may be regulated by ADAMTS9/18. ADAMTS4, expressed by B axis populations, also degrades ACAN and VCAN.

Regarding ADAMTSL members, there is only one member of ADAM expressed in A axis: ADAMTSL1, with an unknown role in the skin. On the other hand, three ADAMTSLs, including *ADAMTSL3*, which bind fibrillin-1 (FBN1, expressed by A4 > A1), are expressed by D2 population. FBN1 is associated with LTBP–a binding peptide that sequesters TGF-β–, and ADAMTSL3 is related to the inhibition of TGFB signalling. Thus, TGF-β signalling may be directly mediated by FBN1 interactions, and its activity may be regulated by ADAMTSL proteins.

OTHER One of the key ECM regulators from this category are proteins that are required for ECM maturation. These components are again majorly expressed by axes A and C. Elements secreted by axis A are (1) members of LOX family (*LOX*, *LOXL1/4*) and *PCOLCE1/2* expressed by A1/A3/A4, and *LOXL2* expressed by A2 and C, which crosslink collagen and elastin chains to strengthen the ECM; and (2) proline hydroxylases–*P3H2*/*P4HA3* in C and *P4HA2* in A–, which are necessary for correct collagen synthesis.

The rest of the elements tend to be "negative regulators" of ECM synthesis. For instance, *ANGPTL7* (D1), *CHADL* (C2), *CHST15* (C2) *CTSK* (A/C), *WISP1/2* (A/C) may inhibit expression of different ECM components or their binding to integrins or other anchoring elements. The complementary function tends to be related to inflammatory processes or ECM degradation. Thus, these enzymes may be expressed either to regulate ECM in homeostasis or may be also responsible for the initiation of ECM degradation and remodelling during inflammatory processes.

12.1.3 *Complement*

The complement pathway is a complex mechanism of the innate immune response. This pathway triggers a set of reactions between different proteins that culminate in the formation of the membrane attack complex, a set of complement proteins (C5b, C6, C7, C8, C9) that form a pore in the target cell's membrane, resulting in lysis.

Supplementary Table S4 displays the expression of complement members, showing that the B2/B3 and B4 populations are involved in the effector components of the complement and support the immune function of these populations. Interestingly, complement activation is not active in population B1.

Additionally, Supplementary Table S4 shows three other components: *CD55*, *CFH*, and *CLU*, which belong to the complement inhibitory regulation. *CD55* and *CFH* bind different upstream C proteins and hinder subsequent steps in the pathway, while *CLU* directly inhibits MAC formation. These genes are expressed mainly by A1 and A4, ECM-secretory populations that, in homeostatic conditions, may ensure that no immune activation occurs. *CLU* is expressed, to a lesser extent, by B2/B3; and *CFH* is also expressed by B4. This may indicate a partial immunomodulatory function of B4, which will be discussed later.

12.1.4 *Immune*

Although frequently overlooked, immune signalling is a very relevant function of fibroblasts. The descriptions of individual genes are available in Supplementary Table S5.

Looking at the related gene expression, it is apparent that most of the genes expressed in the immune environment belong to the axis B fibroblasts, especially B2 and B3. Interestingly, B4 fibroblasts do not express many immune-related genes as B1, B2 or B3 do, except for *CXCL12* and, with secondary expression levels, *IL11RA*, *IL33* and *RARRES2*.

Thus, considering that most of the genes in this table are expressed in pro-inflammatory environments, B1, B2, and B3 are undoubtedly pro-inflammatory fibroblasts. In this section, we are going to extend on the immune response markers, which have been divided into five main categories: (1) CCL - CXCL - CX3C - CXCR cytokines, (2) ACKRs, (3) interleukins, (4) TNF family members, and (5) other elements.

CCL - CXCL - CX3C - CXCR CYTOKINES Within the subset of chemokines and their receptors, we observed a slight difference between B1 and B2/B3. B1 expresses principally *CXCL1/2/3*, which attract several immune cells, but show an affinity for neutrophils implicated in primary immune responses. On the other hand, B2/B3 express *CCL19*, *CX3CL1* and *CXCL12*, which show higher affinity towards cells related to adaptive immune responses, specially T cells, but also DCs, B cells, and NK cells.

Interestingly some cytokines are expressed in populations other than "canonical" axis B immune fibroblasts, such as CCL13 or CCL2, expressed by D axis populations. Both CCL13 and CCL2 bind to CCR2, and attract a range of immune cells, including T helper and NK cells, mast cells, DCs, LCs, monocytes and macrophages (Craig et al., 2006; Mendez-Enriquez et al., 2013; Ouwehand et al., 2010). Therefore, D axis populations may show some immune action, as it will be explained later.

ACKRS ACKRs are receptors for certain common chemokines. The most common registered interactions are ACKR3-CXCL12 and ACKR4-CCL2/19 Table 11. ACKR3 and ACKR4 are principally expressed by populations from axis A. Interestingly, CXCR4, expressed by C5 and D axis populations, binds CXCL12 as well. All these interactions imply that either these populations attract immune-related fibroblasts for signalling, or vice-versa; indicating a vast net of interactions related to migration of immune cells and fibroblasts.

ACKRs are essential immune components since they act as scavengers of their respective ligands to (1) remove and clean the environment from these cytokines and (2) to generate a gradient for immune cell migration into the tissue. One example is the ACKR3-CXCL12-CXCR4 interaction. While CXCL12 binds to CXCR4, ACKR3 sequesters CXCL12 to generate a gradient necessary for the migration

of $CXCR4^+$ cells. Therefore, axis A fibroblasts may be responsible, indirectly, for immune migration and may also act as "cleaners" of the ECM they are producing from immune response byproducts (Berahovich et al., 2013; Donà et al., 2013; Lipfert et al., 2013).

INTERLEUKINS Interleukins were first described as cytokines produced by immune cells that act upon immune cells (Giovine et al., 1990). However, this definition has expanded since many interleukins are also produced, and may act upon, other non-immune cell types.

With the exception of *IL6*, which is exclusively expressed by B1 population, the majority of interleukins are expressed by B3. The function of most of these cytokines is the activation of B and T cells, NK cells and other immune types (IL15/IL15RA, IL33, IL34). There are some relevant family members like *IL11RA*, a soluble form of *IL11* receptor, which may induce IL11 signalling on cells with no receptors (Lamertz et al., 2018; Lokau et al., 2016); and has been observed to induce fibroblast infiltration and TGF-β signalling (Elshabrawy et al., 2018).

TNFS Contrary to chemokines, TNF signalling is not exclusive of B axis populations, but it is expressed also in populations from other axes. This may be because TNF signalling is not "restricted" to the immune response but is present in a large range of signalling pathways. For instance, A2 population expresses *TNFRSF19*, which acts as a co-receptor of TGFBR and LGR5 to inhibit TGF-β and activate Wnt signalling, respectively (Fafilek et al., 2013). The population A2 also expresses *TNFRSF21*, which acts as a negative regulator of B and T cell proliferation and also induces axonal pruning (Nikolaev et al., 2009).

On the other hand, ligands expressed by immune cells show a positive regulatory function. *TNFSF14*, expressed by the population B1, may induce DC and T cell maturation and proliferation (Albarbar et al., 2015); and *TNFSF13B*, expressed by the B2/B3 populations, acts similarly to the CD40/CD40L system necessary for immune cell maturation, such as DC and B cells (So et al., 2013).

OTHER Regarding the rest of immune-signalling elements, the most relevant system is conformed by the RARRES2-CMKLR1 interaction. *RARRES2* ligand is expressed by B2 and B4, but also by A, D and E axes populations, and is consistently linked to immune response, probably more in early than in late phases. For instance, *RARRES2* activation induces chemotaxis of immature APCs, but reduces recruitment of macrophages; or is involved in early psoriatic lesions, but it is downregulated in IL17/IL22-mediated advance stage psoriasis (Cash et al., 2008; Luangsay et al., 2009). Interestingly, RARRES2 receptor, CMKLR1, besides being expressed on immune types, is also expressed by E1 and A2 populations, which may imply further putative roles in the skin.

Other chemokines, such as *SOCS3* or *VCAM1* are usually expressed after the binding of pro-inflammatory interleukins or factors, although with contrary effects: VCAM1 mediates the adhesion and migration of immune and other cell types, whereas SOCS3 has a principal immunomodulatory effect by toning down pro-inflammatory signalling. Although *SOCS3* is expressed by A2, D and B fibroblasts, both *SOCS3* and *VCAM1* are predominantly expressed by B3 population, despite their *contrary* effect, showcasing the complexity of immune modulation by fibroblasts.

12.1.5 *Wnt*

Wnt signalling consists of a complex interaction of ligands and receptors, which activate β-catenin canonical pathway and other non-canonical pathways. More information about Wnt signalling is available in the Introduction section 2.2.2.1, and descriptions of individual genes are available in Supplementary Table S6.

Wnt signalling is highly relevant in the skin, principally during development and for HF homeostasis, where it serves as a key component of the hair cycle (Chen et al., 2012; Ouji et al., 2012; Romanowska et al., 2009). This effect is reflected in the LR interactions shown in Table 11, where we observe a large range of interactions, including LGR4-RSPO1, FZD6-WNT5A, PTPRK-WNT5A or FZD7-WNT11, mainly exerted between A and C axes; as well as other interactions between D and E axes, such as FZD2-WNT2 or FZD7-WNT5A.

Looking at the genes involved in Wnt signalling from Table 11, 4 of them are associated with canonical signalling–*APCDD1*, *FZD1*, *WNT2/10B*–; 6 with non-canonical signalling–*DAAM1/2*, *FZD2*, *WNT5A/10A/11*–; and 4 with mixed signalling–*CTHRC1*, *FZD6/7* and *PTK7*–.

Regarding the patterns of populations present in each type of signalling, there are no clear differences between axes for each type of signalling. In fact, A1/A2/A4, C2/C5, D1/D2, and E1 populations are involved in Wnt signalling; but there are no remarkable differences with each type of signalling. Also, both Wnt ligands and Fzd receptors are present in all axes, showing that Wnt expression patterns are complex, diffuse, and possibly redundant.

Analysis of positive and negative modulation of Wnt signalling shows similar results as with the signalling itself. There are several families of repressors and activators, such as DKKs, NKDs, LGRs or RSPOs. Each one of these families has at least one member being expressed by populations from axis A, and another member by populations from axis C–DKK1 (A) vs DKK2/3 (C), NKD1 (C) vs NKD2 (A), LGR4 (C) vs LGR5 (A), RSPO1 (A) vs RSPO3/4 (C)–. Thus, A and C populations are complementary in Wnt signalling.

Therefore, from the analysis of Wnt-related genes, we highlight these points: (1) assuming assuming that Wnt is highly related to HF homeostasis, D and E axis are likely to be related to it; however, (2) many other relevant pathways are dependent on Wnt, and therefore Wnt expression cannot be trusted as a proxy for HF development. (3) All in all, Wnt signalling and regulation is a complex network that may require a profound knowledge of the details of the pathway members to elucidate the details within the populations.

12.1.6 *TGF-β*

TGF-β signalling is a complex pathway, principally mediated by binding of TGF-β dimer to its receptor, to induce fibrogenic processes in fibroblasts by induction of collagen I and III production and increased proliferation. This topic is introduced in sections 2.2.2.2 and 2.2.4 of the Introduction, and descriptions of individual genes are available in Supplementary Table S7.

When looking at the gene markers, we observed that most of the markers belonged to the regulation of TGFβ processing instead of the pathway itself, whose members–e.g. *TGFB1*, *TGFB2*, *TGFBR1*, *TGFBR2*–show a decreased or scarce expression among all fibroblast populations.

We divided TGF-β regulators into positive or negative/ambivalent regulators. While the literature on positive regulators shows a pronounced increase in TGF-β signalling, the effects are mixed regarding negative/ambivalent modulators. In several cases, proteins encoded tend to sequester TGF-β in the environment for release under the proper conditions or interfere with the binding of TGF-β to its receptor until in homeostasis. Interestingly, regarding the markers from Table S7, the vast majority of modulators are negative/ambivalent (19) compared to positive modulators (3); which probably implies that TGF-β signalling is active constitutively, and it is regulated in homeostatic conditions to allow its immediate activation when the conditions are appropriate.

Regarding the positive modulators, the most common modulator is *BMP7*, expressed by C5 and D1 populations. A similar modulator, *GDF10*, is expressed by the B4 population. Interestingly, other modulators, like *SLPI* or *ADAMTSL3* may also be expressed by B4 immune fibroblast population, implying that either TGF-β signalling partially mediates immune processes or that these genes are pleiotropic and act on other pathways.

Looking at the heterogeneity of populations expressing negative/ambivalent modulators, results are similar to Wnt signalling: there is a range of genes expressed by axis C populations–*ASPN*, *FMOD*, *HTRA1*, *INHBA*, *LTBP2*, *MFAP2*–, genes expressed by axis A populations–*CILP*, *GDF15*, *SLPI*, *SOSTDC1*, *TNXB*–, and genes expressed by axis D populations–*ADAMTSL3*, *BAMBI*, *CAV1/2*–. Some of these genes have multiple functions, but regardless, the presence of diverse markers in different populations suggests that TGF-β signaling is also diverse and may have distinct functions for each population. This could be influenced by the environment or additional functions of the population, which requires further research.

12.1.7 *Lipid metabolism*

Lipids were traditionally framed within biochemistry research as "energy molecules", with negligible secondary functions. Nowadays, lipid metabolism is known to be more than energy-related, and several lipid families–e.g. eicosanoids, sterols, sphingolipids, fatty acids–have a very relevant function in signalling. In this section, we are going to divide lipid-related genes into cholesterol/lipoprotein metabolism and eicosanoids. The descriptions of individual genes are available in Supplementary Table S8.

Cholesterol and lipoproteins, besides their traditional role of "energy storage", are key immunomodulatory components. Oxidised lipoproteins have long been known to induce proinflammatory processes, primarily in macrophages, that lead to atherosclerosis (Parthasarathy et al., 2009). More recently, the range of immunomodulatory functions of cholesterol-related molecules and lipoproteins has been expanded to (1) activation of neutrophils, (2) pro- or anti-inflammatory macrophage lineage phating, (3) activation of T cells by gathering of receptors into cholesterol-rich lipid rafts, (4) control of hematopoiesis, (5) activation of TLR signals by lipoproteins (Aguilar-Ballester et al., 2020; Tall et al., 2015).

On the other hand, eicosanoids are traditionally known as signalling molecules derived from the metabolism of arachidonic acid, which acts upon a large range of processes, including inflammation (pro- and anti-inflammatory action), allergy, fever, pain, cell growth, blood pressure, platelet aggregation, or vascular permeability. (Dennis et al., 2015).

In line with the role of cholesterol metabolism in immune function, we observe that many related genes are expressed by axis B fibroblasts–*APOC1*, *APOE*, *CH25H*, *CYP7B1*, and less predominantly *ABCA8*, *APOD* and *CYP4B1*–. Most of these genes are expressed concretely by B2/B3 populations and are associated with pro-inflammatory responses–CXCL1 expression (APOE), induction of TNF and IL6 (CH25H), activation of innate immune cells and production of immunoglobulin (CYP7B1)–. Interestingly, 3 genes–*ABCA8*, *APOD*, *LDLR*–are expressed by D1 and B4 populations. Thus D axis may also be relevant in cholesterol metabolism.

Based on interactions from Table 11, APOD (D1 > B4 ~ E1) interacts with LEPR (A2 > A1 ~ A3); and APOE (B2 > B3 ~ B4 > B1) interacts with LDLR (D1). We see that in both interactions, population D1 is a key element.

Regarding eicosanoid metabolism, fibroblast population patterns are more diverse. For instance, IL1- and IL6-inducing *PTGS2* is expressed by B1, and its family member *PTGS1* by A2; and other prostaglandin synthesis/degradation enzymes such as *HPGD*, *MGST1*, *PLA2G2A*, *PTGDS* or *PTGIS* are expressed by A and B axis populations, as well as D1, D2 and E1. The only exception are C axis populations which, despite *PTGFR* expressed by C2, show no remarkable gene expression.

12.1.8 *Vitamin A metabolism*

Vitamin A is thoroughly used in the body, including in the skin and HF. Vitamin A can exist in different forms, including retinol, retinal or retinoic acid, depending on the context of action. Vitamin A is transported in plasma in its retinol form by being bound to Retinol Binding Proteins (RBPs), and is translocated to the nucleus by binding to CRABPs in the retinoic acid form (VanBuren et al., 2022).

Vitamin A is necessary for the regulation of HF, although its effects are contradictory, depending on dosing and dietary changes. In mouse studies, it is observed to act on anagen, catagen and telogen; some studies indicate that it prolongs anagen, while others indicate that it favours telogen effluvium and prolongs this phase (Everts et al., 2004, 2007; Suo et al., 2021). Vitamin A metabolism is also involved in wound-induced HF neogenesis (WIHN), a process whereby HFs are created *de novo* in large murine wounds (Bhoopalam et al., 2020). This process has been observed to be mediated by the activation of *ALDH1A3* and *CRABP1*, expressed by populations D1/D2 and C2 respectively (Abbasi et al., 2020; Kim et al., 2019). Lastly, vitamin A participates in the production of melanin by melanocytes, and also favours melanocyte differentiation (Hu et al., 2017; Inoue et al., 2012). The descriptions of individual genes are available in Supplementary Table S9.

Based on this, we observe that the majority of genes associated with vitamin A metabolism are expressed by axis C (*CRABP1*, *RBP4* and *CYP26B1*), although we also observe *CYB26B1* and *TTR* expression in A2, possibly indicating a paracrine role in melanin production; or any other unidentified signalling in the papillary dermis. Interestingly, we see that *RBP5* is expressed by B2 and B3 populations, which are related to immune signalling. It has been previously reported that immune cells express *CYP26* and *ALD1A* proteins (Stevison et al., 2015) and therefore, retinol import into cells via RBP5 in B type fibroblasts might be necessary for part of their functioning.

12.2 PUTATIVE FUNCTIONALITY OF HUMAN FIBROBLASTS POPULATIONS

12.2.1 *C axis*

12.2.1.1 *C1*

C1 population can be described by three main markers used in other single-cell publications: *ACTA2*, *COL11A1* and *DPEP1*. This population, which shares these markers with mouse w4 population, as well as *EDNRA* and *TENM3*, is described by Ahlers et al., 2022–*COL11A1*$^{+}$ *DPEP1*$^{+}$–and Tabib et al., 2021–*COL11A1*$^{+}$ *ACTA2*$^{+}$–as DS cells.

One of the most relevant markers studied from C1 populations is *ACTA2*, the gene encoding for α-SMA. *ACTA2* is already described as a marker of DS cells (Mistriotis et al., 2013), and is traditionally described as a myofibroblast marker. In fact, during wound healing, *ACTA2*$^{+}$ fibroblasts also express *COL1A1* and *ITGB1*, although neither of them is strictly required for proper wound closure (McAndrews et al., 2022). Another report indicates that, although *ACTA2*$^{+}$ myofibroblasts contribute to wound closure, they are not strictly necessary, and other actins (*ACTB*, *ACTG1*) may also be involved in the process (Ibrahim et al., 2015). Additionally, at least in myocardial fibroblasts, they may still differentiate into myofibroblasts regardless of ACTA2 expression (Li et al., 2022b). All these studies may be relevant to postulate whether *ACTA2*$^{+}$ DS fibroblasts already exist in a primed pro-fibrotic state, or if myofibroblasts already originate from another population. This point will be further discussed in the Discussion chapter 13.

A set of markers relevant to DS expression are related to its location. Some markers descriptive of this population are also expressed in the BM surrounding the CTS, such as *KRT17* (Antonini et al., 2013), *COL5A1/2* (Chanoki et al., 1988), MME (Morisaki et al., 2013), and *MDK* (Woo et al., 2022). Interestingly, some of these markers are also detected in DP as well as in DS: *ADAMTS18* (Hagner et al., 2020), *COL7A1* (Tsutsui et al., 2021), *COL12A1* (Sasaki et al., 1996), *LEF1* (Sun et al., 2022), *MDK* (Rendl et al., 2005).

C1 has a clear link to ECM. For instance, *TNMD* is expressed probably to control the width of collagen fibres (Docheva et al., 2005); and *ACAN*, *ASPN*, *COL11A1* and also *TNMD* are usually expressed in cartilage, tendon and ligaments (Kalamajski et al., 2009; Luo et al., 2019; Roughley et al., 2014; Sun et al., 2019).

C1 may also be related to the blood vessels surrounding the CTS, due to the expression of COL15A1 (Manon-Jensen et al., 2019a) and COL21A1 (Kehlet et al., 2019a), which are usually found surrounding blood vessels. Additionally, C1 also expresses *EDNRA*, which regulates blood pressure and blood vessel constriction (Liu et al., 2019), and which, in individuals with mutations in this gene, may lead to alopecia (Gordon et al., 2015); indicating the putative relevance of this gene, and C1 population in general, in HF vasculature.

Also related to HF homeostasis, C1 expresses genes that are related to hair cycle regulation. For instance, *MME* is more active in early anagen (Morisaki et al., 2013); KRT17 binds CDKN1B to prolong the anagen phase (Panteleyev et al., 1997; Tong et al., 2006); and *LEF1* is associated with lengthening of anagen and shortening of telogen and catagen (Zhang et al., 2013).

Therefore C1 DS cells that are active in the anagen phase express several genes contributing to that state; also express HF-related specific ECM genes, and have a wide control of CTS microvessels, probably to supply the IRS and ORS cells within the HF.

12.2.1.2 *C2*

C2 population shares common robust markers both with mouse w/x–*CHST15*, *COCH*, *DKK2*, *FMOD*, *POSTN*–and w1/w2–*CRABP1*, *DAAM2*, *NOTUM*, *TRPM3*, *TRPS1*–populations. One of these markers, *CRABP1*, is a classical DP marker described by Collins et al., 2008. Additionally, Ahlers et al., 2022 described a DP population expressing both *COCH* and *CRABP1*, C2 population exclusive markers.

One of the most relevant markers of C2 is COL24A1, an specific collagen of this population. Although its function is not well known, it has been observed to be highly expressed in bone and has 3 non-collagenous domains that may confer additional, *non-traditional*, functions (Nielsen et al., 2019b). The relationship of this collagen with bone is relevant since we observe that C2 shows highly active participation within the ECM homeostasis, and expressed other markers, like *ASPN*, which has been shown to actively participate in calcium binding of collagen to induce ECM mineralisation (Kalamajski et al., 2009).

Other relevant ECM regulatory components are *CHADL*, which regulates collagen and aggrecan (*ACAN*) fibrillogenesis (Tillgren et al., 2015), *F13A1*, which binds fibronectin and vitronectin (Huttenlocher et al., 1996; Winnemöller et al., 1991); *P3H2*, which binds collagens I and IV and is necessary for their maturation (Tiainen et al., 2008); and *SDC1*, a promiscuous HSPG that binds collagens I, II and IV, fibrin, fibrillin, TNC and vitronectin, as well as MMPs, ADAMTSs, BMPs, EGFs, cytokines and complement proteins (Stepp et al., 2015). Other relevant ECM modulators expressed by C2 are *ADAMTS9* and *CHST15*, which degrade CSPGs such as aggrecan (*ACAN*) and versican (*VCAN*) (Kai et al., 2017; Kelwick et al., 2015).

This ECM regulatory activity, as well as the active expression of ECM components that are more "cartilage-like", similar to markers from C1 DS population, may confer C2 a unique ECM microenvironment necessary for DP cells to dwell within the HF.

One of the hallmarks of HF signalling is the Wnt pathway. From the analysis of markers, we observe that C2 shows a high Wnt activity. Interestingly, this signalling is *mixed*, that is, both pro- and inhibitory markers are expressed. For instance, *DAAM2* enhances Wnt activity by receptor aggregation, which potentiates the downstream cascade (Lee et al., 2012a, 2015), but also expresses *NOTUM*, which inhibits this aggregation (Zhang et al., 2015). C1 population also expresses *FZD1* and *RSPO4*, receptor and co-activator of canonical signalling (Gazit et al., 1999; Szenker-Ravi et al., 2018); together with *DKK2* and *PTK7*, which inhibit canonic signalling (Ahn et al., 2011; Lhoumeau et al., 2011). PTK7, additionally, is a non-canonic signalling co-receptor, which may interact with C1-expressed WNT10, a non-canonic ligand (Chen et al., 2012).

Wnt signalling in DP was observed to be mainly pro-anagenic since many of these markers are not expressed during telogen phase (Lim et al., 2012; Reddy et al., 2001). Nonetheless, some markers like *SFRP1*, also observed in cuboidal ORS at the base of the HF, are potentially active in regressing telogenic HFs (Geyfman et al., 2014).

12.2.1.3 *C3*

C3 population, based on the PAGA graphs and other population analyses from chapter 10, is highly related to C1 and C2, since many markers expressed by C3 are also expressed by C1 and C2, but not vice-versa. Based on human-mouse population similarity, C3 is most related with w/x and w4, with shared markers such as *COL7A1*, *MMP16*, *LRRC15* and *POSTN*. w4 is also similar to C1, so probably C3 is more related to C1 than C2.

TNMD, expressed by all C populations–Ahlers et al., 2022 classifies it as a DS marker–, reduces the amount of *BGN* (biglycan), *COMP* and *FN1* (fibronectin). Therefore, C3 may be, at least ECM-wise, a spatially complementary population to C1 and C2, with a separate ECM niche (Lin et al., 2017).

Despite being a "mixture" of C1 and C2, C3 expresses certain genes more than the other two populations. Some of these genes, such as *MMP11*, *BGN*, or *POSTN*, are tightly related to the functioning of the ECM. *BGN* is an SLRP that binds collagens I and VI, chondroitin and dermatan sulfate (Roughley et al., 1989; Schönherr et al., 1995); and *POSTN*, localised around the HF, binds several components, including collagen I, fibronectin and tenascin C (Yamaguchi, 2014). Interestingly, Deng et al., 2021 observed a *POSTN*$^{+}$ population in keloid samples, which increases collagen production (Zhang et al., 2014); and the authors show that blockade of POSTN function led to a decrease in collagen production. Therefore, C3 might be probably related to ECM homeostasis of HF cells.

Regarding its location, one of the few exclusive markers expressed by C3 is *KLK4*, a member of the kallikrein serine proteases family. It does not have a clearly defined function, although it may participate in certain immune interactions (Filippou et al., 2020), androgens may regulate its expression in certain cancers (Korkmaz et al., 2001), and may also regulate the degradation of ECM components (Obiezu et al., 2006). Regarding its location, KLK4 has been found in IRS, but also in ORS, SG and eccrine glands; similar to other KLKs such as KLK6/10/11 (Komatsu et al., 2003), which are described in the literature but are not expressed in our analysis.

Based on the expression of these markers, the population C3 may be involved in both the secretion of HF-specific ECM, as well as in its degradation, based on the expression of specific metalloproteases and kallikreins.

12.2.1.4 *C5*

Similar to C3, C5 function is not a cell type categorised based on bibliographic records. Moreover, the fact that, compared to the rest of C axis populations, has comparatively fewer cells, indicates that it is probably a cell population not present in all HF stages, or that it represents a specific cell type with a reduced number of cells, like stem cells.

Looking at the PAGA graphs in human, C5 shares a higher transcriptomic similarity with C1 DS compared to C2 DP, coexpressing genes such as *CDH11*, *INHBA*, *KRT17*, *LEF1*, *MME*, *SOX18*. However, it also shares specific key genes with C2 DP population–*CRABP1*, *PTK7*, *DAAM2*, *CYP26B1*–. Regarding mouse-human similarity, C5 is transcriptomically more similar to w1/w2 populations, with shared genes such as *BMP7*, *INHBA*, *LEF1*, *SOX18*, *WNT5A*; mostly expressed by C1.

Some of these markers, apart from being associated by similarity, are reported to be expressed by DP cells, like *WNT5A* (Lim et al., 2012), *CRABP1* (Collins et al., 2008), *TPD52* (Rubin et al., 2004), *CXCR4* (Zheng et al., 2022), *LEF1* (Sun et al., 2022), or *SOX18* (Villani et al., 2017). Another relevant marker,

RSPO3, has been described in DP cells to induce proliferation of HF stem cells and DS (Hagner et al., 2020). In fact, considering that some of these markers are also expressed in other parts of the HF, like *LEF1* in DS, *WNT5A* in IRS (Lim et al., 2012), *CXCR4* in ORS (Zheng et al., 2022), *KRT17* in ORS and SG (Antonini et al., 2013) and *MME* in bulge DS (Morisaki et al., 2013), it is possible that the expression of certain DP markers in additional cell types is due to C5 population being part of these cell types.

Despite differences and similarities with other HF populations, C5 expresses a few exclusive markers, such as *TPD52*, *KRT9*, *WNT5A*. There are other markers which do not have a clear role, either in general or in HF, such as *LMO3*, *LUZP2* or *PKP4*. This last one is expressed in desmosomal plaques (Berika et al., 2014).

Similarly to C1, C5 shows a high and, at the same time, diffuse Wnt activity. For instance, *APCDD1* and *NKD1* expression is linked to negative Wnt signalling (Angonin et al., 2013; Mazzoni et al., 2017), whereas *DAAM2* is an activator (Lee et al., 2012a), and *RSPO3* and *DKK2* are canonical Wnt activators and inhibitors respectively (Ahn et al., 2011; Szenker-Ravi et al., 2018). Lastly, *WNT5A* is a non-canonical Wnt ligand (Romanowska et al., 2009), which can be expressed by *SOX18*–another marker of C5– (Villani et al., 2017), and WNT5A induces the expression of *KRT9* (Rinn et al., 2008); all 3 markers are expressed by C5. Surprisingly, *KRT9* is a classical palmoplantar keratin (Fu et al., 2014), with no bibliographic records indicating its expression if HF and other skin appendages.

Similar to C1, there are certain markers involved in cell cycle regulation, directly or indirectly, such as *IRF1*, which regulates *CDKN1A* (Dornan et al., 2004); as well as in HF cycle, like *MME*, which shows increased activity in anagen (Morisaki et al., 2013), or *CXCR4* which, besides its classic CXCL12 receptor involved in chemotaxis, delays the transition from telogen to anagen (Zheng et al., 2022).

Another function of C5, not described previously in C1 or C2, is TGF-β signalling. Two of the markers of C5 are BMP7, a key ligand of the non-canonical TGF-β pathway (Meng et al., 2016); and INHBA, a common TGF-β inhibitor that binds to the ligand and inhibits receptor coupling (Ferdous et al., 2007). It is possible that both genes act synergistically to promote only non-canonical TGF-β signalling.

Although it might not be its main function, C5 expresses *S100B* and *SLC5A3*, two key genes in neurogenesis and vasculogenesis. *S100B*, also expressed by D1/D2, is a classical neuron marker, although it is also present in melanocytes, adipocytes, DCs and lymphocytes (Donato et al., 2009). S100B has been found to participate in neuron myelinisation (Fujiwara et al., 2014), neurite growth, (Donato et al., 2009), satellite cell activation (Sorci, 2013) and production of iNOS and VEGF by endothelial cells (Donato et al., 2009). On the other hand, *SLC5A3*, an inositol-Na^+ transporter expressed in nerve, pancreas, lung and muscle (Ma et al., 2012); besides being a key cellular and intercellular osmolarity regulator (Pizzagalli et al., 2020), it forms complexes with other channels in nerve and blood vessels to sense neuronal excitability (Dai et al., 2016) and vessel contractibility (Barrese et al., 2020).

Lastly, an interesting remark about C5 population is the expression of certain B axis markers, like *CD74* (B3), *IER1* (B1) and *IRF1* (B1 and B3). The functions of these genes will be discussed later but, at first sight, it is possible that either the C5 population carries some level of immune response or these genes have additional functions apart from the ones exerted by immune cells and immune-related fibroblasts.

Therefore, based on transcriptomic similarity, markers and putative functions, there are several hypotheses regarding the cell type that C5 population could be assigned to. It is possible that C5 represents a population that shares its location with the DS, but also lies within the DP environ-

ment. Firstly, C5 cells could either be matrix cells adjacent to DP, implicated in HF proliferation, or precortex cells, implicated in differentiation. However, both cell types have a marked expression of *KRT15* o *KRT6* (Geyfman et al., 2014), which are not specifically expressed within the fibroblast cells in our analysis. Secondly, C5 cells could also be dermal cup cells, that is, DS cells located below the dermal papilla. Hagner et al., 2020 analyse dermal cup cells in mice and show that these cells express specifically *Epha3*, *Hic1*, *Igfbp2* and *Tnnt1*, *Mcam1* and *Pcp4* come of which are expressed by E1 or C1 sparsely–*Epha3*, *Igfbp2*, *Tnnt1*–and the rest of genes do not show any expression pattern. This effect is already expected due to the gene expression differences that are commonly observed across species. Lastly, C5 population is a candidate of telogenic DP or DS cells. Considering that the HF cycle is asynchronous in humans and the fact that around 5% of HF cells are found to telogen (Serrano-Falcón et al., 2013), the reduced amount of C5 cells may be an indicator. Additionally, despite telogen being popularly considered as an inactive HF state, in reality, telogen DP is very active and necessary to maintain the club hair properly (Geyfman et al., 2014). The main drawback to this assumption is that C5 population expresses both pro-anagenic and pro-telogenic markers, and the number of markers expressed is reduced, which decreases the conclusiveness of the results.

12.2.1.5 *Consistency of C axis populations*

Results from this analysis, as well as other studies that characterise HF cells using scRNAseq (Chovatiya et al., 2021; Wu et al., 2022) show that the HF is an incredibly heterogeneous structure and, focusing on fibroblasts, this analysis shows that, besides DP and DS cells, other new fibroblastic populations are relevant for HF biology, and are yet to be characterised. Even more, the characterisation of DP and DS, at least based on single-cell and other studies, is yet to be 100% defined as well.

In fact, not only differences between human and mouse markers are to blame, but also the differences across studies. For instance, the comparison of DS and DP populations described by Shin et al., 2020 and Joost et al., 2020 performed in Table 13 shows that the overlap of markers is still extremely small. For instance, we see that CTS1 from Shin et al., 2020 is more similar to DS2 from Joost et al., 2020, although the Jaccard index is only 11.11% (10 shared out of 99 compared between both datasets); and CTS2 is more similar to DS1, but the similarity to DS2 cannot be discarded–with Jaccard indexes of 12.36 (11/89) and 7.53 (7/93)–. Regarding DP populations, the only "relevant" overlap is between aDP and DP3–Jaccard index of 12.36 (11/89)–.

The same effect occurs comparing Shin et al., 2020 or Joost et al., 2020 markers to the merged populations from this analysis. Both tDP and aDP populations reported by Joost et al., 2020 match w1 and w2 indistinctively, and DS1 and DS2 populations match w3 and w4. However, the differences between w1/w2 and w3/w4 populations, despite being highly similar, are consistent across datasets, which implies that the characterisation of individual datasets may not be consistent and that, at least in mouse, annotation of w axis does not reflect differences between anagen and telogen states. However, considering the small number of cells within Joost et al., 2020 population, it is possible that anagen/telogen differentiation was not relevant enough to be considered for the marker-to-population algorithm. If further studies including more anagen/telogen cells are introduced in the analysis, probably new w populations reflecting this heterogeneity will appear.

A similar phenomenon can be described for the mouse population w5. This mouse population, characterised by its increased expression of cell cycle markers–e.g. *Mki67*, *Cdk1*, *Cdkn3*–is described by

Shin et al., 2020 as HF dermal stem cells (hfDSCs). However, although this population is reported to exist in human HF (Wu et al., 2022), based on the expression of *Col17a1* and *Krt15*, it is not located within the fibroblasts clusters. Therefore, it is possible that hfDSCs do not share the same lineage in both organisms.

With all this in mind, even a *proper* mouse HF fibroblast characterisation may not lead to new human populations, firstly, due to the fundamental limitations between human-mouse comparisons; and secondly, due to the fact that, even within the same organism, fundamental differences between mRNA and protein levels arise. Tsutsui et al., 2021 comment that several genes, including *Col4a1*, *Tnc* or *Smoc1* show significative differences between expression levels of mRNA and protein. They state that these differences may be due to posttranslational mechanisms or due to protein expression by adjacent populations affecting the measurement levels.

This last effect is also sound in human samples. In fact, many of the fibroblast markers are also expressed by many cell types, including keratinocytes, endothelial cells, perivascular cells and immune cells; so the putative fibroblast functions inferred during this discussion section may be affected by additional interactions with adjacent populations.

12.2.2 *A axis*

12.2.2.1 *A2*

A2 population is one of the key archetypes of the A axis, together with A1, and shows both a differentiated marker profile and location. A2 population is highly likely to be associated with the papillary dermis, based on the results from Section 10.5, which we will extend on this section.

ECM AND ECM MODULATION A2 population shows a rich expression of collagens that are known to be bound to DEJ structures, located in BM around vessels or other structures, or are located in the papillary dermis. The most relevant collagens in this aspect are the following: (1) *COL7A1* is a component of the anchoring fibrils in the BM, which bind collagen I and III fibres (Barbieri et al., 2014); (2) *COL6A1/2/3*, which can be found in the DEJ, but also around blood vessels, reticular dermis and hypodermis (Sabatelli et al., 2011; Theocharidis et al., 2017), in DEJ it assembles anchor cells to the ECM (Kaur et al., 2015); similarly, (3) *COL6A5* is located in the papillary area, and attaches several growth factors and MMPs (Freise et al., 2009); (4) *COL13A1* and *COL18A1* are also found in the DEJ, as well as in vascular and epithelial BMs (Bonnet et al., 2017; Pehrsson et al., 2019; Peltonen et al., 1999); (5) *COL4A2/4* is similarly found in the lamina densa of BMs (Barbieri et al., 2014). There are three additional collagens: *COL14A1* and *COL21A1* FACIT collagens; and transmembrane *COL23A1*, which is mainly located in epithelia (Kehlet et al., 2019b). Interestingly, some collagens, like *COL14A1* or *COL6A5*, are also found in reticular dermis (Nauroy et al., 2017; Theocharidis et al., 2017). This may be because these genes are also expressed by other populations that are likely to be reticular–e.g. *COL14A1* is also expressed by A1 and A3, and *COL6A5* is also expressed by B3–or because this population, aside from being located in the papillary area and DEJ, is also necessary for proper endothelial BM functioning, regardless of the location of the blood vessels.

Other relevant ECM components expressed by A2 are *COMP*, which binds collagens I, II, IX, XII and XIV (Agarwal et al., 2012; Farina et al., 2006); and *POSTN*, which binds collagen I, fibronectin and tenascin C, and is also found at the DEJ (Yamaguchi, 2014). Specifically with collagens XII and XIV, COMP acts

as a bridge between these collagens located in the anchoring plaques and collagen I, to stabilise the structure (Agarwal et al., 2012). Regarding *POSTN*, it is interesting to recall that it binds tenascin C, expressed by A2, C and B populations, in contraposition to tenascin XB, expressed by A1/A3/A4 populations.

Besides ECM components, A2 population also expresses a large array of ECM modulatory molecules, that are known to either bind ECM components (1) without knowing their function, (2) to change the properties of the ECM, or (3) to have a secondary function not involved with ECM modification. For instance, *APCDD1* acts like *PCOLCE*, expressed by A1, which cleaves pro-collagen molecules into their active form, and also binds BMP1, another similar procollagen proteinase, to enhance its activity (Baicu et al., 2012); and *LOXL2*, a member of the lysyl oxidase family, crosslinks collagen chains to make them more resistant to insults (Sarrias et al., 2004). Interestingly, *APCDD1* is reported to be expressed by reticular dermis too (Solé-Boldo et al., 2020), possibly due to being expressed by C2 and C5 populations. On the other hand, *ADAMTS9* and *MMP11* are known to degrade, respectively CSPG such as ACAN or VCAN (Kelwick et al., 2015), collagens IV, V, IX, X, XI, vitronectin, fibronectin and laminins, as well as activate several proMMPs (Kahari et al., 1997).

Additionally, A2 population also expresses (1) *CD44*, a pleiotropic molecule involved in fibrosis and TGF-β signalling, immune activation and cell proliferation, which binds HA, fibronectin, proteoglycans, and collagens I, IV and XIV (Bennett et al., 1995; Fujimoto et al., 2001; Ishii et al., 1993; Jalkanen et al., 1992; Kawashima et al., 2000). It may maintain the levels of collagen I, N-cadherin and fibronectin in homeostasis (Tsuneki et al., 2015) and wounding, avoiding its accumulation (Govindaraju et al., 2019).(2) *TGFBI*, which codifies a multidomain protein that may bind integrin αvβ3, fibronectin, vitronectin, collagens–including XII–, fibrinogen and VWF (Runager et al., 2013; Ruoslahti et al., 1987). (3) *F13A1*, the A subunit of the coagulation factor XIII, binds fibrin, fibronectin, vitronectin, collagen VI, and other ECM components, crosslinking them (Muszbek et al., 2011).

Lastly, *ANTXR1* can bind collagen VI (Nanda et al., 2004), and might be a link between ECM and cytoskeleton, necessary for ECM-sensing and migration (Gu et al., 2010). A similar function may be exerted by sarcoglycans *SGCA* and *SGCG*–commonly expressed by muscle, and also expressed by A1/A3–, which act as a link between the cytoskeleton and the ECM (Noguchi et al., 1995).

NERVE AND VASCULAR The A2 population, together with other populations like C3, or D axis, expresses several markers related to nerve and/or blood vessel physiology. For instance, *KCNQ3*, which encodes a potassium channel protein (Singh, 2003), is observed at the endings of the lanceolate complex in mice (Schütze et al., 2016). Similarly, *SEMA5A* is observed to act both as an attractor and a repulsor in neural guidance (Hilario et al., 2009); and probably also in vasculogenesis (Purohit et al., 2014). *TGFBI*, besides binding ECM components, is shown to have antiangiogenic properties (Son et al., 2013); similarly to *EDN3*, an endothelial cell vasoconstrictor (Perrine, 2007). Interestingly, *EDN3* was shown to induce pigmentation in melanocyte *in vitro* (Garcia et al., 2008), and hair plucking led to melanocytic stem cell activation via *EDN3*/*EDNRB* activation (Li et al., 2017).

HF Despite several genes expressed by A2 showing evidence that it may be expressed in the papillary dermis, certain exclusive markers show that the presence of A2 population may be also extended to HF. We previously mentioned the expression of *COL7A1* and *COL13A1*, related to DEJ and BMs. Additionally, these collagens are located at the BMs surrounding DP and ORS (Tsutsui et al., 2021). We also mentioned *KCNQ3*, a K^+ channel observed at the endings of the lanceolate complex, which

surrounds and innervates the HF. Additionally, *TNFRSF19*, which in humans is located at the stratum basale, also marks IFE and infundibulum in mice HFs in telogen and hair bulb in anagen. *EDN3* is also expressed by DP cells in mouse, although not in human (Hagner et al., 2020). *WNT11* is expressed in DS and ORS cells (Lim et al., 2012). Lastly, *FGFR2*–which binds FGF7 and FGF10 expressed by B4–is observed to be expressed by keratinocytes, SG (Katoh, 2009) and, in mouse HF, by matrix cells near DP (Rosenquist et al., 1996).

WNT AND TGF-β SIGNALLING Linking to the putative association of the A2 population with HF, another classical pathway involved with HF homeostasis and hair cycle is Wnt signalling. A2 expresses more than 10 Wnt-related markers, which are also expressed by HF populations, mainly C2, C3 and C5. Although this evidence of Wnt signalling by A2 does not directly imply that A2 fibroblasts have to belong in HF, it is highly likely that Wnt signalling is not only focused on HF cells but also by surrounding populations–including A2, B4, or D axis–by paracrine signalling.

Among the Wnt pathway markers, we observe a minority of markers related to canonical signaling, specifically *RSPO1/3/4* canonical activators (Szenker-Ravi et al., 2018) and *DAAM2* potentiator, which aggregates Fzr/Axin/Dvl complexes (Lee et al., 2012a, 2015). The rest of the markers are involved with non-canonical signalling, either by (1) inhibition of canonical signalling by *AXIN2*, a member of APC|Axin|GSK3β complex that inhibits canonical Wnt through β-catenin inhibition (Jho et al., 2002); *DKK2/3*, *PTK7* and *TNFRSF19*, which sequester LRP5/6 co-receptors and promote their internalisation (Ahn et al., 2011; Lhoumeau et al., 2011); and *NKD1/2*, inhibitors of the translocation of β-catenin to the nucleus (Angonin et al., 2013; Zhao et al., 2015); (2) direct Wnt inhibition by *WIF1*, which sequesters several Wnt ligands (Angonin et al., 2013), or (3) non-canonical markers, like *WNT11* and *PTK7*.

Interestingly, although the vast majority of markers are either (1) inhibitors of the canonical pathway, (2) general Wnt inhibitors, or (3) ligand or co-receptors of either the canonical or non-canonical pathways, we do not see any receptor being expressed by A2 population. Therefore, although some markers are located in the cell membrane, the lack of proper receptors may imply that A2 is a candidate to be a paracrine Wnt canonical inhibitor.

Regarding TGF-β signalling, A2 only expresses two related markers, also expressed by C populations: *MFAP2*, which sequesters TGF-β1 in the ECM microfibrils (Weinbaum et al., 2008), and *NOG*, a BMP7 inhibitor (Blazquez-Medela et al., 2019).

IMMUNE AND LIPID SIGNALLING The A2 population also shows the expression of certain markers involved in side, but relevant functions, including immune, and lipid metabolism.

We have previously mentioned the expression of *FGFR2*, which binds B4-expressed *FGF7* and *FGF10*. The binding of these two factors induces the synthesis of IL-1α (Melnik et al., 2008). IL-1α is highly relevant, acting synergistically with TNFα and also acting in a positive feedback loop to activate NF-kB signalling (Laberge et al., 2015). Another gene, *CLEC2A*, is involved in NK cell-mediated cytotoxicity by binding to its receptor (Steinle et al., 2009). *ANTXR1* is a transmembrane protein that, besides from binding collagen VI (Nanda et al., 2004), may have an immunosuppressive role, modulating M2 polarisation and T cell exhaustion (Huang et al., 2020).

Another interesting protein that binds the immune with lipid signalling is *OSBP2*. As we will see in B axis signalling, oxysterols–cholesterol-derived molecules–are relevant to mediate oxidative stress and inflammatory processes (Gargiulo et al., 2016), but also in physiological aspects of cholesterol

metabolism and membrane homeostasis (Liu et al., 2020a). In that sense, *OSBP2* may bind oxysterols and be related in the regulation of their synthesis and toxicity (Gargiulo et al., 2016).

Other two genes related to lipid signalling are *PTGS1* and *AHRR*. *PTGS1* is involved in the synthesis of thromboxanes from arachidonic acid, such as TXA2 or PGI2, involved in homeostatic and inflammatory functions (Heide et al., 2006). *AHRR*, on the other hand, is a TF that binds a wide list of ligands, which includes eicosanoids, bilirubin, indoles, as well as external pollutants and dietary ligands; and is related to the transcription of (1) several metabolism-related genes, including many members of the CYP450 family; (2) endocrine disruptors, and (3) proinflammatory factors such as IL-1β, TNFα, IL8, MMP1 or TNFSF13B (Larigot et al., 2022; Vogel et al., 2017).

Lastly, A2 population also expresses two genes involved in vitamin A transport and metabolism: *TTR*, which binds RBPs to co-transport retinol into the cells, and *CYP26B1*, which degrades all-trans retinoic acid by hydroxylation (Isoherranen et al., 2019). In HF, retinoic acid contributes to the refractory telogen phase, and CYP26B1 may take part in the modulation of this step (Hovland et al., 2020).

12.2.2.2 *A1/A3*

In this section, we will discuss the main functions of A1 markers. However, we remind that many of these markers are also shared by A3 and A4. Thus, these populations, although not explicitly mentioned, may also be active in some of these functions.

ECM AND ECM MODULATION A1 is one of the key fibroblast subpopulations due to its collagen and ECM component-producing capacity. The most important collagen genes expressed by this population are *COL1A2* and *COL3A1*, the main chains from collagens I and III, the most abundant collagens in dermis (Henriksen et al., 2019; Nielsen et al., 2019a). Interestingly, although collagen III is said to be more predominant in the papillary dermis (Barbieri et al., 2014; Stunova et al., 2018), A2 population does not express it, so it is likely, and also expected, that either A1/A3/A4 or C axis populations, fulfil this task in the papillary dermis. The other two relevant collagens in skin are *COL12A1* and *COL14A1*, which belong to the FACIT family. These collagens show interruptions in their chains that allow the interaction and binding of other collagens, GAG chains, DCN, COMP, TNC and other ECM components (Mortensen et al., 2019b). Interestingly, and similarly to collagen III, *COL12A1* is more expressed in the papillary dermis (Nauroy et al., 2017), but it is not expressed by A2.

Compared to other populations, like A2 or C axis, which express a wide range of collagens, A1 only expresses the aforementioned 4 collagens. Even nucleating collagens, relevant for the proper structure of the ECM, are not expressed by A1. However, despite that, these are the most abundant, and therefore, functionally relevant collagen components of the ECM.

Besides collagens, the A1 population also expresses other relevant ECM components like the SLRPs *PODN*, *DCN* and *OGN*, which bind fibronectin and other elements (Winnemöller et al., 1991). *DCN* also regulates the expression of collagenases (Huttenlocher et al., 1996), which play an active part in ECM regulation. Other two relevant ECM components are *HSPG2* and *TNXB*. *HSPG2*–perlecan–, binds a large array of targets such as laminin, collagen IV, V, VI, XI, elastin and FBLN2; and also binds lipids with an LDLR-like domain, as well as Wnt morphogens (Hayes et al., 2022). *TNXB*, on the other hand, interacts with type I, III, V, XII and XIV collagens, as well as decorin and integrins, and is antagonistic to the location of *TNC*, mainly in the papillary dermis and HF (Lethias et al., 2006).

Other proteins belonging to the matrisome, and with high binding capacity are *WISP1/2* and *MFAP5*. WISP proteins contain several domains, including IGFBP, von Willebrand type C repeats, thrombospondin type 1 repeat and cysteine knot motif. *WISP1* in particular has been observed to bind DCN and BGN (Desnoyers et al., 2001). *WISP2* is known to inhibit the binding of fibrinogen to integrin receptors (Janjanam et al., 2021). Additionally, it is upregulated in hypertrophic scars (Chaudet et al., 2020) and is involved in *PPARG* activation in adipocytes. *PPARG* despite its *clasical* role in lipid metabolism, suppresses the expression of TNFα, IL-1β and IL6 as well as MMP9 in many immune cells (Jiang et al., 1998), an activity shared by certain A1 markers such as *PI16*, or *SLPI*. Lastly, *MFAP5* is a key component in the organisation of elastic fibres by interacting with fibrillin-1/2 (Penner et al., 2002).

Certain genes are involved in the maintenance, protection and maturation of the ECM fibres. One of these genes is *CA12*, a carbonic anhydrase that catalyses the production of $HCO3^{-}$ and H^{+}, thus regulating pH and CO2 homeostasis (Supuran, 2008). This anhydrase is also present in eccrine sweat glands, regulating Cl^{-} levels (Na et al., 2019). The other two families of ECM regulators are LOX and PCOLCE. LOX family, composed of *LOX* and *LOXL1*–although *LOXL2* is expressed by A2 and C axis–is necessary to crosslink collagen and elastin chains–*LOXL1* favours elastin (Liu et al., 2004)–a process necessary for the stabilisation of the ECM. However, excessive activity of *LOX* may be related to stiffened ECM in ageing (Langton et al., 2012). On the other hand, *PCOLCE* and *PCOLCE2* are necessary to cleave pro-collagen molecules, activating them and allowing them to polymerise into the final helical fibres (Steiglitz et al., 2002; Takahara et al., 1994).

Another aspect of ECM homeostasis is not only its synthesis but also its degradation. ECM degradation is carried out by MMPs, like *MMP2*, which can degrade collagens I, IV, V, VII, X, XI, fibronectin, elastin, laminin, vitronectin and activate proMMPs (Cabral-Pacheco et al., 2020); and *CTSK*, a cathepsin that activates MMP9 (Christensen et al., 2015), decreases *COL1A1* expression (Soundararajan et al., 2021) and catabolises collagens and elastin (Kondo et al., 2022).

This effect is primarily reverted by TIMPs, such as *TIMP2* or *TIMP3*, which inhibit different MMPs and ADAMs (Cabral-Pacheco et al., 2020). Interestingly, another inhibitor of MMP2 is *PI16*, a peptidase that is activated under ECM stress or inflammation (Hazell et al., 2016). *PI16* also cleaves B2-produced RARRES2 into chemerin, which binds to CMKLR1 (E1) to exert its functions, including immunomodulation, adipogenesis and angiogenesis (Regn et al., 2016).

A similar function is exerted by *SLPI*, an inhibitor of serine proteases, to control that ECM is not over-degraded (Nugteren et al., 2021). It is also an inhibitor of monocyte-produced MMP1 and MMP9 (Zhang et al., 1997) and inhibits the activity of IL-1β (Zakrzewicz et al., 2019). *DPP4* is a pleiotropic dipeptidase that may regulate CXCL12, MMP1, and MMP3 levels (Ospelt et al., 2010) and its binding to certain ECM components may activate or inhibit fibroblast proliferation. For instance, binding to fibronectin induces mobilisation (Cheng et al., 2003), whereas binding to glypican 3 leads to DPP4 inhibition and reduced cell proliferation (Davoodi et al., 2007). Moreover, *DPP4* is often complexed with *ADA*, also expressed by A1, which may show an antifibrotic potential (Fernández et al., 2013) and regulates the collagen production (Marucci et al., 2022).

DPP4 has been used in several studies as a marker for different fibroblasts, even with contradictory findings. For instance, Philippeos et al., 2018 used *Cd26/Dpp4* to isolate papillary cells, and Korosec et al., 2019 also stated its presence in the papillary dermis. On the other hand, *Dpp4* is a clear marker of z1/z2 populations from adventitia cells, as stated by Joost et al., 2020; and Tabib et al., 2018 located an $SFRP2^{+}DPP4^{+}$ population throughout the dermis.

This anti-inflammatory potential from *SLPI* and *DPP4-ADA* is also shared by markers involved in the negative regulation of the complement pathway. A1, among other populations, expresses *CD55*, which inhibits the complement by binding to C4 and C3b fragments and inhibiting the creation of C2aC4b and C3bBb complex respectively (Dho et al., 2018). Additionally, *CLU* directly inhibits the formation of the MAC (Shinjyo et al., 2021); and *DCN*, aside from its SLRP component function, binds to C1q to inhibit the initiation of the classical pathway (Krumdieck et al., 1992).

Related to immune signalling, with a possible role in immune regulation, are the chemokine receptors *ACKR3* and *ACKR4*. These receptors bind, respectively, CXCL12–which binds CXCR4–and CCL19/CCL21–CCL19 binds CCR7–, ligands produced not only by immune cells but also by B2/B3 populations. Both ACKRs act as scavenger receptors for these chemokines. One of the proposed and studied functions of this scavenging is to create a gradient that favours immune cell migration (Berahovich et al., 2013; Donà et al., 2013). For instance, in lymphatic vessels, *ACKR4* is expressed in afferent lymphatic collectors. There, T cells migrate by adhesion from $CCL21^{+}$ lymphatic capillaries into $CCL21^{dim}ACKR4^{+}$ collectors, where they lift and go with the lymph flow (Friess et al., 2022). However, considering the additional functions of other immune-related markers, *ACKR3* and *ACKR4* may also simply just scavenge these chemokines from the ECM to avoid unnecessary immune infiltration.

Two last markers that may be related to immune and vascular function are *AGTR1* and *CD248*. *AGTR1* is a receptor for angiotensin II, a potent vasoconstrictor (Barnes et al., 2005; Bergsma et al., 1992). However, in cardiac fibroblasts, *AGTR1* was observed to regulate fibroblast proliferation and collagen expression (Tadevosyan et al., 2017). On the other hand, *CD248*–endosialin–is a transmembrane glycoprotein related to stromal cell proliferation and microvascularisation during tissue remodelling (Wu et al., 2021), and may also act as a receptor of CCL17–expressed by macrophages–to induce collagen expression in fibrotic environments (Pai et al., 2020). It should be noted that the two genes show a secondary function related to proliferation, in line with the ECM production function of these fibroblast populations.

METABOLISM Fibroblasts are not only in charge of synthesising ECM but are also active scavengers of hormones, xenobiotics and other small molecules that may be necessary for paracrine or endocrine signalling, or which are harmful to the environment.

SLC47A2 is a H^{+}/organic cation antiporter expressed in the kidney for excretion of toxic endo or exocompounds (Yonezawa et al., 2011). In skin, it may have a role in skin drug absorption (Alriquet et al., 2015). Among the molecules that it transports there are endogenous molecules such as creatinine, guanidine or thiamine, and exogenous molecules such as tetraethylammonium–TEA, a potent channel inhibitor–, acyclovir, ganciclovir, paraquat, metformin or DAPI (Avsar, 2022; Yonezawa et al., 2011).

CES1 metabolises a wide range of molecules–heroin, cocain, etc.–into their active forms; and also converts monoacylglycerides to free fatty acids and glycerol, and cholesterol ester to free cholesterol (Markey, 2010). Similarly, *MGST1* belongs to the MAPEG family, involved in eicosanoid and glutathione metabolism, and catalyses the reduction of many reactive intermediates and drugs, and aids them to be further metabolised (Morgenstern et al., 2011).

Belonging the CYP450 family, the A1 population also expresses *CYP4B1*, a member that oxidises a large set of substrates, including fatty acids, arachidonic acid and cholesterol, which is activated by hypoxia, androgens, and other factors (Röder et al., 2023). Lastly, we have *HPGD*, a member of the

alcohol dehydrogenases which, similar to *CYP4B1*, metabolises a wide range of prostaglandins, mainly to regulate their levels (Yan et al., 2004), (Cho et al., 2006). For instance, it degrades the IL1-derived proinflammatory cytokine PGE2 (Arai et al., 2014). It also activates resolvins, enzymes that promote restoration of normal cellular function following the inflammation (Arita et al., 2006). The functions of some of these *metabolic* genes–e.g. *CYP4B1*, *MGST1*–are compatible with the immunoregulatory function of this fibroblast population.

WNT AND TGF-β SIGNALLING Wnt signalling is partially present in A1 population, with the expression of *FZD6*, a receptor of canonical and non-canonical signalling (Corda et al., 2017); *LGR5*, a co-receptor of RSPOs to sequester ZNFR3 and allow Wnt binding to the receptor; and *CTHRC1*, a pleiotropic protein that binds FZRs and induces their activation (Mei et al., 2020). On the other hand, we also observe the expression of *DKK1*, a canonical Wnt inhibitor (Ahn et al., 2011), and *WIF1*, a Wnt regulator that sequesters Wnt ligands in the ECM (Gajos-Michniewicz et al., 2020).

Contrary to the possible non-canonical Wnt signalling in this population, genes involved in TGF-β signalling expressed by A1 show a marked inhibitory effect, in line with the anti-fibrotic potential from *DPP4* and *ADA*. This effect could be contrary to the proliferative effect of *AGTR1* and *CD248*, although this effect could be pro-fibrogenic and thus compatible with *ADA* and *DPP4*, or could not be active in this fibroblast population.

CTHRC1 expression does not only control Wnt signalling but also attenuates TGF-β pathway by induction of proteasomal degradation of Smad2/3 complex (Myngbay et al., 2021). *CILP* shows a similar inhibition mechanism by impeding the phosphorilation of Smad2/3, and also by binding to TGFBR (Liu et al., 2020b). Similarly, *SLPI* has an additional function of suppressing TGF-β (Ashcroft et al., 2000).

This effect is mainly associated with canonical TGF-β signalling. Regarding non-canonical pathways, we see that A1 expressed *GDF10/15*, which act similarly to BMP proteins (Wang et al., 2021b). However, the A3 population expresses *SOSTDC3*, an antagonist of BMP protein (Faraahi et al., 2019), which may counteract GDF proteins.

12.2.2.3 *A4*

In this section, we will refer to the markers of A4 and their main functions. However, most of these markers are already described in A1 so that we will focus on the new markers. Regarding these markers, most of them are only more expressed in the A4 population that in the A1 population, and only a few of them are exclusive of this population. Thus, newly attributed functions in A4 may also likely be performed by the A1 population.

ECM AND ECM MODULATION Most ECM components expressed by A4 have already been described: *COL1A2*, *COL3A1*, *COL12A1*, *DCN*, and *MFAP5*. There are additional elements expressed that are relevant for the ECM, such as *EFEMP1*, a matrisome protein containing several EGF-like repeats and a fibulin-type C-terminal domain (Sun et al., 1998); or *PRG4*, a proteoglycan usually found in cartilage due to its high viscosity, which binds to hyaluronan and fibronectin (Eguiluz et al., 2015; Rhee et al., 2005).

Three additional genes are related to the synthesis of elastic fibres and microfibrils: *FBN1*, *ELN* and *EMILIN2*. *FBN1*–fibrillin 1–, is a glicoprotein that constitutes a structural component of microfibrils

(Lee et al., 2004). These fibres can be associated with elastic fibres and elastin-independent networks (Jensen et al., 2016), and can also bind other matrisome proteins, including BMPs or LTBPs, playing an indirect role in TGF-β response (Jensen et al., 2016). *ELN*–elastin–is one of the key components of elastic fibres, which also needs the binding of GAGs, fibrillin, heparan sulfate and other ECM components (Gheduzzi et al., 2005). Elastin, being more than 1000 times more flexible than collagen, is key to the flexibility of the ECM (Kristensen et al., 2016). Lastly, *EMILIN2* is another component of the microfibrils, which binds fibrillin to other structures (Doliana et al., 2001; Schiavinato et al., 2016).

In order for fibroblasts to secrete ECM and, more generally, move alongside the ECM, they express certain markers involved in cellular motility. Some of these markers, like *SGCA/G* are also key markers of the rest of A axis populations. In this case, A1 and A4 express additional markers related to sensing and motility: *PIEZO2*, *DBN1* and *TPPP3*. *PIEZO2*, a mechanosensory channel located in Merkel cells (Wu et al., 2017), could potentially sense variations of pressure and tension in ECM to modify its structure. *DBN1* and *TPPP3* bind to F-actin in growth cones and filopodia and to microtubules, respectively (Butkevich et al., 2015; Vincze et al., 2006)

Finally, regarding ECM modulators, all the genes expressed by A4 have already been described for A1.

METABOLISM Metabolism of complex molecules is a hallmark of A1 population. In the A4 population, in addition to *MGST1* and *HPGD*, two more genes are expressed: *PLA2G2A* and *PTGIS*. *PLA2G2A* is a phospholipase that transforms different phospholipids into arachidonic acid, which can later be transformed into eicosanoids by enzymes like *HPGD*. Another enzyme is *PTGIS*, which transforms PGG2 into PGI2–prostacyclin–, which is involved in the regulation of HF regeneration. In fact, higher levels of PGI2 and related to androgenic alopecia (Chovarda et al., 2021), and minoxidil inhibits the action of PGI2 (Messenger et al., 2004). Thus, A1 and A4 fibroblasts are active members of arachidonic acid metabolism due to the expression of *MGST1*, *CYP4B1*, *HPGD*, *PLA2G2A* and *PITGIS*.

WNT SIGNALLING Compared to A1, TGF-β signalling is not present as the expression of specific markers, whereas Wnt signalling is apparent. Besides the expression of *CTHRC1* and *DKK1* regulators, A4 population also expresses *APCDD1L*, which is predicted to be involved in the negative regulation of Wnt pathway (Otsuki, 2005); *SFRP4*, which acts similarly by sequestering Wnt ligands; and Wnt canonical ligands *WNT2* and *WNT10B*. *WNT2* may promote fibre deposition by fibroblasts (Cai et al., 2017), and *WNT10B* has been studied in the context of DP proliferation and maintenance (Ouji et al., 2012).

VASCULATURE Although A1 population was indirectly related to vascular homeostasis with the expression of genes such as *AGTR1* of *CD248*, the A4 population may show a greater implication in this matter. PGI2, synthesised by *PTGIS*, has been shown to activate *PPARD* and induce VEGF production (Wang et al., 2013). Another relevant marker is *MGP*, a protein with affinity for Ca^{2+} cations, which, aside from decalcifying elastic fibres, is also necessary to avoid the calcification of vascular cells.

Two other markers largely studied in the vasculature and cell migration context are *SEMA3C* and *SEMA3E*, belonging to the semaphorin family. SEMA3C binds NRP1 (A1/A4), NRP2 (D and C) and Plexin D1 molecules, and shows a mixed angiogenic profile, which is pro or antiangiogenic dependent on the tumour type (Neufeld et al., 2008). Similarly, SEMA3E binds Plexin D1 only, with a similar mixed profile (Neufeld et al., 2008).

IMMUNE RESPONSE Although certain genes with some links to immune signalling have been discussed for A1 population–complement inhibitors *CD55* and *CLU*–a large set of similar markers is expressed by the A4 population.

One of the mentioned markers, *SLPI*, aside from its preotease function associated with ECM, also regulates immune signalling by inhibiting NF-kB and TLR signalling (Greene et al., 2004; Nugteren et al., 2021), and also inhibits the maturation of IL-1β (Zakrzewicz et al., 2019) and IL6 (Zakrzewicz et al., 2019). *PRG4* proteoglycan expression can also target to immune-related CD44. Binding of PRG4 to CD44 internalises it and targets the inflammasome, reducing IL1β expression, and affecting TLR-mediated cascades, including NF-kB (Richendrfer et al., 2020). Similarly, DPP4 may truncate T chemoattractant CXCL12–produced by B2/B3–into an inactive form that only binds ACKR3, and not CXCR4 (Elmansi et al., 2022).

Four additional molecules with pleiotropic functions involved in immune signalling are *C1QTNF3*, *PDPN*, *HSD3B7* and *MGP*. (1) *C1QTFN3* inhibits IL1 and TNFα (Guo et al., 2020); but also TLR4, IL6, VCAM1, ICAM1, and E/P selectins (Schmid et al., 2021); (2) *PDPN* may favour the motility of NK cells, neutrophils and DCs (Kerrigan et al., 2009; Seymour et al., 2016; Sobanov et al., 2001), and a subset of $PDPN^{+}$ cells form a reticular network near lymph nodes was observed to facilitate leukocyte migration and antigen presentation via CCL19, CCL21 and IL7 (Fletcher et al., 2015); (3) *HSD3B7* participates in the degradation of the proinflammatory molecule 7α,25-dihydroxycholesterol–oxysterol–synthesised by CH25H and CYP7B1 from cholesterol by B2/B3 fibroblasts; and (4) *MGP* has additionally been involved in the downregulation of TNFα, IL-1β and NF-kB; as well as inhibition of Ca^{2+}-dependent inflammatory processes (Viegas et al., 2017).

Lastly, there are two markers that are related with T cells: *CD70* and *TRAC*. *CD70* is a cytokine that belongs to the TNF ligand family. It induces the proliferation of costimulated T cells, enhances the generation of cytolytic T cells, and contributes to T cell activation. For instance, in the skin, LCs present CD70 to augment $CD8^{+}$ T cell presence in the epidermis (Polak et al., 2012). On the other hand, *TRAC* encodes part of the TCR molecule, which binds to MHC molecules and activates the adaptive immune response in the T cell (Brownlie et al., 2013). The presence of TRAC in fibroblasts is surprising since the presence of the TCR is restricted to T cells, and its expression in fibroblasts has been recorded in the literature. Interestingly, *TRAC* only encodes the α subunit of the receptor, and since the β subunit is not expressed in fibroblasts, it is likely that TRAC is constitutively express without exerting a relevant function, or that it exerts an unknown function in the A4 population.

12.2.3 *B axis*

12.2.3.1 *B1*

B1, together with B2, is one of the initially discovered immune-related fibroblast populations. There is a marked difference between B1 and B2 markers and therefore, it is likely that immune response in the dermis, as with many other organs, requires different types of responses.

B1 shares some relevant markers with the z1 population, including *CCL2*, *CXCL2*, *FOSL1*, *IL6*, or *GCH1*. Interestingly, this mouse population also shares relevant markers with A4 population, such as *CD248*, *DPP4*, *EMILIN2*, *NPR1*, *PTGIS*, *SEMA3C/E*, *SFRP4* and *WNT2*. Thus, the z1 population, due to its restricted location in mice, may serve the double function of these two populations to act as an ECM-producing

and immune-related fibroblast. This is not surprising considering the array of immune-related markers expressed by the A4 population.

IMMEDIATE EARLY GENES Immediate Early Genes are genes that are transcribed within minutes after stimulation in response to both intrinsic and extrinsic signals. Many of these responses involve cell differentiation, signalling related to the immune system and stress response (Bahrami et al., 2016). Identification of IEGs tends to be complex and depends on the system of study. However, some "universal" lists of these genes have been obtained for mammalian cells (Tullai et al., 2007), which were used during this analysis.

Among the immune populations, B1 is the one expressing more IEGs (8) compared to others like B2/B3 (4), which B1 also expresses. These genes are *IER3*, *JUNB*, *SLC2A3*, *TNFAIP3*, *ZFP36*, *IL6*, *FOSL1*, and *NFKBIA*. Some of these genes, like *FOSL1* and *JUN* belong to the AP-1 TF complex, which binds JUN family proteins (Baines et al., 2017). Some of these markers belong to the acute phase immune response, which involves the production of cytokines secreted to attract the innate immune response, among other roles. These markers are discussed below.

ACUTE PHASE CHEMOATTRACTION (CHEMOKINES) There is a family of chemoattractants expressed by B1, and to a lesser extent by B3: *CXCL1*, *CXCL2* and *CXCL3*. All of these CXCLs bind to CXCR2, and the most studied one is CXCL1. This chemokine, and probably CXCL2 and CXCL3 as well, binds to heparan, dermatan and chondroitin sulfate GAGs to function, after being liberated by MMPs (Wang et al., 2003). Also, GAGs seem necessary for full activation of CXCR2 as well (Wang et al., 2003). CXCL1 has a marked angiogenic potential (Murphy, 2007), which may facilitate the chemoattraction of eosinophils, basophils, macrophages, immature DCs and naïve T cells; and neutrophils, towards which shows the most chemoattractant activity (Bautista-Hernandez et al., 2017; Murphy, 2007). CXCL2 and CXCL3 both act similarly to CXCL1 (Murphy, 2007), although we note that binding of CXCL2 to ACKR1 in blood vessels is necessary for neutrophils to perform diapedesis (Girbl et al., 2018).

Another factor necessary for immune cell chemotaxis is *ICAM1*, expressed by endothelial cells to induce leukocyte trafficking (Bui et al., 2020). Although this effect is studied in endothelial cells, a similar process may be necessary in fibroblasts to facilitate leukocyte motility within the ECM. For instance, certain HF fibroblasts express *ICAM1* to attract perifollicular macrophages, which seem necessary for proper HF functioning since *ICAM1* KO mice suffer from hair regression (Müller-Röver et al., 2000).

ICAM1 expression is induced by *NFKB1*, one of the effector members of the NF-kB pathway, highly active in immune processes, and also expressed by B1. Key chemoattractant targets induced by *NFKB1* are *ICAM1* and *VCAM1* adhesion molecules, MMPs, *COX2*/*PTGS2* and *NOS2* (iNOS), these last two genes involved in angiogenesis (Liu et al., 2017). Additionally, *NFKB1* expression has been observed to correlate with M1 (pro-inflammatory) macrophage polarisation and neutrophil recruitment (Liu et al., 2017). Interestingly, B1 primarily also expresses *NFKBIA* the IkBa inhibitor. This protein sequesters NFKB1 in the cytoplasm until a pro-inflammatory signal–e.g., TNFs, TCR–targets it for degradation and liberates NFKB1 (Yu et al., 2020). Therefore, the B1 population seems to be the first to react by being prepared to activate NF-kB signalling whenever a stimulatory signal is triggered.

ACUTE PHASE CYTOKINES Another key aspect of immune response is the production of cytokines that can mediate the inflammatory response. The cytokines produced by the B1 population, or in-

duced by genes expressed by it, are mainly acute phase cytokines, that is, cytokines that induce the chemotaxis of innate immune cells, during early immune responses.

Most of these cytokines, like IL-1β, IL6, TNFα, TGFβ, are induced by many genes expressed by B1, including *NFKB1*, *IL32*, *TNFSF14*, *CD44*, and *FOSL1* (Alsaleh et al., 2010; He et al., 2022; Liu et al., 2017; Pierer et al., 2007). Unsurprisingly, these genes were also expressed as primary chemoattractants. Interestingly, one of the target chemokine genes, *IL6*, is also expressed by B1. Its expression is induced by LPS and other external agents that bind TLRs, as well as IL1, TNFα, TGFβ and other factors. In fact, as an IEG, it is one of the most readily inducible cytokines and activates the differentiation of B cells to plasma cells, peripheral T cells–Th2, Th17, Treg–, as well as ECM remodelling–production of collagen and GAGs–and fibrosis (Duncan et al., 1991; Paquet et al., 1996; West, 2019). It also induces the secretion of more IL6, as well as different chemokines (CCL2, CCL11) and adhesion molecules (ICAM-1, VCAM-1), which mediate its chemotactic properties (West, 2019).

ECM REGULATION One family of target genes of NF-kB signalling are MMPs. ECM plays a key role in immune response, which is observed by the expression of other genes whose targets are also ECM regulatory components. For instance, *CD44*–previously mentioned in A2 population to maintain collagen and fibronectin levels and to bind many ECM molecules–plays an active role in immune activation too. For instance, low molecular weight hyaluronic acid mediates CD44-induced IL6, CXCL1 and CXCL2 expression in fibroblasts (Vistejnova et al., 2014). Another member involved in ECM production under immune responses is *FOSL1* which, besides regulating Th17 commitment (Shetty et al., 2022), induces the expression of *TGFB1*, *FN1*, *VIM* and *MMP1/9/14* (Sobolev et al., 2022).

Some of these MMPs, like MMP1 and MMP3 are also expressed by B1 population. MMP1 degrades mostly collagen III, but also I, II and BM collagen IV, as well as aggrecan and entactin (Cabral-Pacheco et al., 2020). It also activates proMMP2 (Kahari et al., 1997). MMP3 can degrade collagens IV, V, IX, X, and XI, aggrecan, vitronectin, fibronectin and laminins; and also activates proMMPs 1, 8, 9, and 13 (Cabral-Pacheco et al., 2020), (Kahari et al., 1997). Similarly, *ADAMTS4* metalloprotease is also expressed and can degrade CSPGs such as aggregan or versican (Kelwick et al., 2015).

The last member of the ECM regulators is *TNFSF14*, a member of the lymphotoxin system, a network of LR pairs expressed by T cells activated during viral infection (Dostert et al., 2019). TNFSF14 induced the expression of MMP9 and IL6 in synovial fibroblasts in rheumatoid arthritis (Pierer et al., 2007), which are also expressed in the skin.

TGF-β SIGNALLING Pro-inflammatory immune signalling goes unsurprisingly hand in hand with the pro-fibrotic TGF-β pathway. However, contrary to expected association, pro-inflammatory signalling induces an antifibrotic response, as observed by B1 markers. In fact, this signalling is not only unique to the B1 population but also to B2/B3 populations, with shared and independent markers.

One of the previously mentioned markers, *CXCL1*, is also related to TGF-β signalling. It is observed that TGF-β reduces its expression (Fang et al., 2015); although the inverse might not be true, since CXCR2 KO mice resulted in decreased TGF-β1 and collagen I expression (Zhang et al., 2020a), and this reduction might be a result of indirect processes. Another previously mentioned marker is *CD44*, which is observed to act as a negative regulator of TGF-β and PDGFRB (Porsch et al., 2014).

Lastly, B1 expresses *PPP1R15A*, which belongs to a family of regulatory phosphatases expressed under stress conditions. PPP1R15A is recruited to prevent the excessive phosphorylation of the translation initiation factor eIF-2A/EIF2S1. This, in turn, reverses the shut-off of protein synthesis that is initiated by stress-inducible kinases and facilitates the recovery of cells from stress (Choy et al., 2015; Santos et al., 2016). Additionally, PP1R15A down-regulates the TGF-β signalling pathway by promoting the dephosphorylation of TGFB1 (Santos et al., 2016), and also by recruiting Smad7 inhibitor and by dephosphorylating TGFBR (Shi et al., 2004).

OTHER FUNCTIONS B1 expresses additional markers with mixed functions that are not directly classified into the previous categories. One example of a marker is *PTGS2* (COX2), mentioned to be induced by *NFKB1*. In fact, *PTGS2* is induced by upstream IL1 and IL6. PTGS2 belongs to the family of arachidonic acid metabolism and is involved in the synthesis of PGE2, related to fever and pain signalling, as well as other immune signalling processes (Heide et al., 2006).

Another marker is *GCH1*, the rate-limiting enzyme in the synthesis of tetrahydrobiopterin (BH_4) (Zhang et al., 2007), a cofactor necessary for NOSs and tyrosine hydroxylase–e.g., to make compounds such as dopamine–(Lewthwaite et al., 2015). Additionally, BH_4 is a potent, diffusable antioxidant that resists oxidative stress and enables cancer cell survival (Kraft et al., 2019).

A similar oxidative stress protector is *SOD2*, a protective enzyme against ROS that transforms superoxide from the mitochondrial electron transport chain into H_2O_2 and O_2 (Pias et al., 2003). In skin–and presumably other tissues–SOD2 is necessary for age-related damage protection and, in fact, SOD2 expression decreases with age (Treiber et al., 2012); and SOD2 deficiency promotes aged phenotype in mice (Weyemi et al., 2012). However, despite its protective role, SOD2 is necessary for an active innate response. Lack of SOD2 impairs IFN-I and cytokine production, as well as dysregulation of NF-kB; all of which are related to active ROS production (Wang et al., 2017). Additionally, at least in lung cancers, *SOD2* expression is associated with MMP upregulation (Yi et al., 2017).

12.2.3.2 *B2/B3*

B2 and B3 populations are what would traditionally be understood as adaptive response immune fibroblasts, that is, their main function is to attract immune cells focused on the adaptive response against specific antigens, or becoming cells that produce responses against such antigens for future infections. This response is usually preceded by the innate/acute inflammatory response although both responses may coexist and be complementary.

Comparing it with mouse populations, B2 shares the highest homology with the y4 population, based on the co-expression of some of the following markers: *APOE*, *C3*, *IL33*, *MGP*, *SLCO2B1*, and *TNFSF13B*; most of which have defined functions that will be discussed throughout this section. Similarly to the B1 population, B2 also expresses IEGs such as *JUNB*, *ZFP36*, and *TNFAIP*; as well as more specific IEGs such as *SLC2A3* and *CCL2*. For instance, *SLC2A3*–a.k.a. *GLUT3*–is a monosaccharide transporter, including glucose and galactose, to supply the cell with metabolic precursors for glycolysis (Deng et al., 2015; Seatter et al., 1998).

COMPLEMENT B2/B3 populations and, to some extent B4, expressed key components of the complement pathway, namely C3, C6 and C7. C3 is a cornerstone of the classical and alternative pathways; and can initiate the alternative cascade on its own. It activates into C3a anaphylatoxin and C3b, which

binds to C2a and C4b to activate C5. (Rutkowski et al., 2010). C5 is also activated by CTSH, produced by B2/B3 (Bhakdi et al., 2004). Additionally, C3 can also be an immune primer: it is observed that in the synovial space, C3 exposure primes fibroblasts for future immune responses, which may lead to chronic inflammation (Afzali et al., 2021).

The other two members of the complement cascade expressed by dermal fibroblasts are C6 and C7. These proteins, together with the end-of-cascade protein C5b, as well as C8 and C9 form the MAC (Rutkowski et al., 2010). Interestingly, some components, like the pore protein C9, are not expressed by these fibroblasts. In fact, C9–and other complement terminal proteins–is produced by monocytes and mature DCs as well as hepatocytes (Lubbers et al., 2017), a target suppressed by hepatitis C virus to impair MAC formation and hinder immune response in the host (Kim et al., 2013a).

Therefore, B2/B3 fibroblasts express complement cascade initiators and some terminal components as part of the innate response, but a proper response is reliant on C9-expressing immune populations.

CHEMOATTRACTANT One of the main functions of B2/B3 populations is the chemotaxis of adaptive and innate immune cells. A clear marker is *CCL19*, a chemoattractant of $CCR7^{+}$ cells, such as DCs, B cells, NK cells and several types of T cells (Laufer et al., 2019; Ohl et al., 2004; Reif et al., 2002; Robbiani et al., 2000). In APCs, it is observed that migration occurs after CCR7 phosphorylation and internalisation (Anderson et al., 2015; Tian et al., 2013). Moreover, CCL19-induced chemoattraction is also observed in certain tumours, which favours T $CD8^{+}$ infiltration and antitumorigenic response (Cheng et al., 2018c).

Another member of the CCL family expressed by B2/B3 is *CCL2*, which binds CCR2 and CCR4, expressed by bone-marrow-derived monocytes (Craig et al., 2006; Vanbervliet et al., 2002). Together with CCL5 it is involved in epidermal-to-dermal migration of LCs after injury (Ouwehand et al., 2010). Additionally *CCL2* is also expressed as a regenerative signal in the context of HF. It was observed that hair plucking led to the release of CCL2 to recruit macrophages that help in adjacent hair creation by activation of regenerative signals (Chen et al., 2015b; Rahmani et al., 2020).

Two additional expressed chemokines are CX3CL1 and CXCL12. CX3CL1 binds CX3CR1 to attract T cells, NK cells, monocytes and DCs (Johnson et al., 2013; Limatola et al., 2014). In fact, during wound healing, *CX3CL1* leads to the promotion of macrophage and fibroblast accumulation (Ishida et al., 2008); and in atopic dermatitis, it induces the retention of T cells in the skin (Staumont-Salle et al., 2014). Interestingly, *CX3CL1* promotes leukocyte attachment in physiological conditions in its anchored form, whereas cleavage by ADAMTS10 and ADAMTS17 in inflammatory conditions leads to its soluble form, which increases its activity (Thelen et al., 2016). Regarding CXCL12, it binds CXCR4 expressed by T cells and Langerhans cells (Bautista-Hernandez et al., 2017; Ouwehand et al., 2008).

B2/B3 populations also produce several chemotactic interleukins, namely, IL15, IL32, IL33 and IL34; of which IL15 works as a chemokine. IL15 shows roles in innate and adaptive responses, mainly activating B, T and NK cells (Lodolce et al., 2002), and also induces the production of IL8 and CCL2–in monocytes and fibroblasts–to attract neutrophils and monocytes (Badolato et al., 1997). IL32 has been observed to switch between pro- and anti-inflammatory programs (Heinhuis et al., 2015) and, in the pro-inflammatory program, it induces expression of pro-inflammatory IL1β, IL18 and TNFα cytokines (Alsaleh et al., 2010).

A relevant chemokine that has been previously mentioned is RARRES2, that binds CMKLR1 in the context of angiogenesis. Additionally, RARRES2 shows pro- and anti-inflammatory properties depending on cleavage by proteases (Mattern et al., 2014). For instance, it was observed to enhance the chemotaxis of immature DCs and monocytes but also reduce the recruitment of neutrophils and macrophages into inflamed tissues (Cash et al., 2008; Luangsay et al., 2009).

Other relevant molecules for proper immune response are adhesion molecules. B2/B3 populations express *ICAM1*, *ICAM2* and *VCAM1*. *ICAM2* works similarly to *ICAM1*, commented on B1 population section, although it may be necessary for neutrophil crawling between endothelial cells during diapedesis (Halai et al., 2013). Additionally, and related to this previous role, CXCL2 regulates endothelial barrier function and permeability (Amsellem et al., 2014). Regarding *VCAM1*, it is usually expressed in endothelial cells after cytokine stimulation–i.e. IL1, TNFα, IL4, IL3–, which promotes the adhesion of lymphocytes, monocytes, eosinophils and basophils (Broide et al., 2014; Mantovani et al., 1998). In other contexts, like muscle, VCAM1 is necessary for satellite cells to communicate with other satellite cells as well as immune cells and may affect myofibril growth (Choo et al., 2017). Therefore, adhesion molecules may also be used by B axis fibroblasts to gather other fibroblasts into the inflammation site.

Lastly, *HAS2* may also work as an indirect chemoattractant, since large hyaluronic chains are necessary for protection against stressors and tissue repair, but also for proper leukocyte homing (Sussmann et al., 2004), possibly by binding chemoattractants like CXCL1 in the ECM.

IMMUNOMODULATION Immunomodulation refers to the processes of induction of changes in the immune cells, either to secrete specific factors, to repress or activate them, or to induce their maturation. All B axis populations show a certain degree of immunomodulation, although B2/B3 show the greatest capacity.

Some of the most relevant immunomodulators are the members of the MHC, necessary for the activation of T cells. B2/B3 fibroblasts, and B1 to a lesser extent, express *HLA-B* and *HLA-F*, belonging to the MHC-I family; and HLA-DRB, associated with the MHC-II family. *HLA-B* and *HLA-F* present both autoantigens and exogenous antigens produced after infection to T $CD8^+$ cells (Neefjes et al., 2011); and *HLA-B* is constitutively expressed in all cells, whereas *HLA-F* expression is restricted to B cells and activated lymphocytes (Ishitani et al., 2003). Compared to the rest of the members of the MHC-I family, HLA-B loads the peptide and transports it to the membrane quicker, and shows a broader range of adaptability to peptides (Neefjes et al., 1988; Peh et al., 1998), which justifies its increased expression in immune fibroblasts. Additionally, HLA-F can act as a regulator by binding to activating and inhibitory receptors in NK and T cells (Lin et al., 2019a). On the other hand, *HLA-DRB1* is expressed mainly in APCs and captures exogenous antigens to process them and present their peptides to $CD4^+$ T cells (Neefjes et al., 2011). Therefore, fibroblasts may also act as APCs.

Additionally, these fibroblasts also express *CD74*, a molecule that serves two main functions. On the one hand, it is an intermediary element of the MHC-II antigen protein. A specific segment of CD74, the class II–associated Ii chain peptide (CLIP), binds to the MHC-II antigen binding site and prevents its premature binding to antigenic peptides (Rosenzweig, 2018). On the other hand, *CD74* can also be expressed as a surface receptor, where it forms a dimer with CD44 to activate NF-kB and related pathways and induce the expression of IL6, TNFα, TLR4 and the antiapoptotic protein Bcl-2 (Williams et al., 2022).

One key molecule for immune maturation is CD40. CD40 interacts with CD40L, expressed as a surface receptor in several immune types, such as DCs or T cells, inducing their maturation (Smith, 2005). For instance, activation in macrophages leads to the expression of TNF surface receptor expression, and activation in B cells leads to their differentiation into plasma cells (Kawabe et al., 1994). A protein similar to CD40 is TNFSF13B, which actually belongs to the same family, and is also expressed by B2/B3 (So et al., 2013). The binding of TNFRSF13B to a TRAF subfamily receptor induces DC activation and B cell maturation (Rickert et al., 2011).

A second family of proteins involved are the immunomodulatory interleukins IL15, IL33 and IL34–which were also chemotactic–; and the soluble form of IL11 receptor, IL11RA. IL11RA binds gp130, a transmembrane protein associated with IL6 receptors, and induces IL11 signalling, even on cells with no IL11 receptor (Lamertz et al., 2018; Lokau et al., 2016). IL11RA is involved in the activation of several downstream pathways, including MAPK, JAK-STAT, NF-kB or Akt (Balakrishnan et al., 2013). IL15 acts as an activator of adaptive responses, especially T, B and NK cells (Lodolce et al., 2002). In T cells, this process is also mediated by suppression of apoptosis, which enhances its function (Malamut et al., 2010). Regarding IL33, it is stored in the nucleus and released during injury, working as an alarmin (Haraldsen et al., 2009) that is activated by tryptases and chymases released by other immune cells, like mast cells (Eissmann et al., 2020). IL33 activates T, NK, DC and mast cells, as well as basophils, eosinophils and macrophages to produce other cytokines and chemokines (Bonilla et al., 2012; Moro et al., 2009; Pecaric-Petkovic et al., 2009; Price et al., 2010). Lastly, IL34 induces the differentiation of monocytes and macrophages through binding to CSF1R (Lin et al., 2008) and SDC1 (Segaliny et al., 2015).

A similar macrophage activator is CSF1, which also binds CSF1R and regulates macrophage differentiation. However, its range of action is systemic, compared to IL34, which is focused on the central nervous system and skin (Greter et al., 2012; Nakamichi et al., 2013). Other macrophage activators are SLCO2B1 and C3. Interestingly, C3 is secreted in blood vessels by adventitial fibroblasts in vesicles containing other components and induces macrophage reprogramming (Kumar et al., 2021).

Another set of markers indirectly related to immune processes are *CH25H* and *CYP7B1*. Both proteins are related in the production and metabolisation of oxysterols, a family of lipids related to cholesterol that is highly reactive and induces cytotoxic and pro-apoptotic responses by interfering with membrane lipids (Olkkonen et al., 2012). Additionally, oxysterol itself and metabolising enzymes indirectly are highly related to immune function. *CYP7B1* expression is induced by innate cells and is related to the regulation of immunoglobulins by B cells (Dulos et al., 2005) and, together with *CH25H* the activation of macrophages and DCs (Olkkonen et al., 2012). Also, 25-hydroxycholesterol, a product of CH25H, is involved in the induction of TNF and IL6, and dermal Tγδ17 cells require it for homing in that area (Frascoli et al., 2023).

Other markers related to the immune signalling mediated by lipids are *APOE* and *APOC1*. Aside from their *traditional* role in cholesterol and lipid transport and metabolism, these apolipoproteins show a relevant immune activity. For instance, Fuior et al., 2019 reported a significant effect of APOC1 in immune modulatory processes, and APOE is observed to suppress T cell proliferation, neutrophil activation, and regulate macrophage function (Zhang et al., 2010). Additionally, it is likely that the lipid transport function associated with the lipoproteins APOC1–present in VLDL and HDL–and APOE–present in VLDL and IDL–is necessary for the synthesis and metabolism of oxysterols by CH25H and CYP7B1.

Lastly, two additional markers show an immunomodulatory function: *IRF1* and *SOCS3*. On the one hand, *IRF1* TF is associated with the expression genes related to (1) IFN signalling (*IFNA/B*, *TNFSF10*, *ZBP1*), (2) regulation of cell cycle and proliferation (*TP53*, *CDKN1A*), (3) antibacterial response and angiogenesis (*NOS2*), (4) apoptosis (*CASP1/7/8*), (5) immune response (*IL12/15/17*, *PTGS2*), and (6) MHC expression (*B2M*, *PSME1*, *CIITA*) (Bowie et al., 2008; Dornan et al., 2004; Gao et al., 2009; Huang et al., 2009; Oshima et al., 2004; Park et al., 2007; Su et al., 2007). *SOCS3*, on the other hand, is expressed as a response to the binding of pro-inflammatory molecules–e.g. IL6, IL12, LPS, TNFα or IFNs–inhibiting the effector pathways activated by these molecules by binding to several cascade elements, like STATs or IL6R-bound JAK (Carow et al., 2014; Yin et al., 2015). Activation of *SOCS3* may serve a double purpose: first, it may reduce the overall immune response and act as a negative feedback regulator; and may also reduce the innate, acute response to favour a secondary, adaptive response.

TGF-β SIGNALLING Like B1, B2/B3 populations are related to the negative regulation of TGF-β signalling. There are 6 genes expressed by these populations involved in this pathway: *IL11RA*, *IL33*, *ACHE*, *PPP1R15A*, *SOCS3* and *HAS2*.

It has been observed that IL11 may induce IL33 expression in fibroblasts (Widjaja et al., 2022), and restrain them from switching to a myofibroblastic state (Gatti et al., 2021). However, IL11 by itself may show the potential to be either responsive to TGF-β signalling (Schafer et al., 2017) and also antifibrotic, at least in endothelial cells (Allanki et al., 2021). *ACHE*, acetylcholinesterase, has been shown to be indirectly related to TGF-β response inhibition (Stegemann et al., 2020).

Interestingly, regarding SOCS3, TGF-β induces the suppression of this gene in fibroblasts (Dees et al., 2020). Although the contrary phenomenon–SOCS3 inhibiting TGF-β signalling–has not been described, it is highly likely that it might be true. Lastly, it is observed that *HAS2* expression in myofibroblasts is increased, which also increases CD44 ECM levels and results in fibrosis (Li et al., 2011). However, in this scenario, *HAS2* expression is already linked to the myofibroblastic state, so we cannot directly assume that HAS2 is pro-fibrotic.

VASCULATURE AND ECM MODULATION Vasculature and ECM modulation are necessary processes during immune responses, and although a bit diffuse, the expression of certain genes by B2/B3 populations may be involved in some of these processes.

One of the genes implicated is *ACHE*, acetylcholinesterase. This enzyme *traditionally* hydrolyses the acetylcholine neurotransmitter in neuromuscular junctions and cholinergic synapses. However, acetylcholine is also necessary for vasodilation–it is involved in NO production–, and thus ACHE may modulate blood pressure (Amezcua et al., 1988). In fact, this process is used for the migration of ACHE$^+$ T cells (Fujii et al., 2017). Additionally, it is also highly likely that ACHE is involved in other functions due to the upregulation of acetylcholine-mediated response in several skin diseases like scleroderma and atopic dermatitis (Stegemann et al., 2020). Similarly, *IGFBP7*, a ligand of IGF receptors, is observed to diminish angiogenesis by diminishing the activity of COX2 and PGE2 secretion, which affects VEGF production (Tamura et al., 2009).

Regarding ECM modulation, 3 genes are putatively responsible for this activity: *COL6A5*, *CTSH* and *MGP*. *COL6A5* may be indirectly associated with ECM modulation since it attaches certain MMPs, including inflammatory MMP1 and MMP9 (Freise et al., 2009). Interestingly, this gene is expressed by B3 population as well as A2, which may indicate that it possesses a certain immunogenic degree.

CTSH, a cystein protease from the same family as *CTSK* expressed by A1/A4, is known for degrading lysosomal proteins. However, it also degrades the ECM for lymphocyte infiltration (Li et al., 2010). Lastly, MGP, which has already been mentioned to be involved in ECM and blood vessel decalcification, may be relevant for immune infiltration. Calcified ECM may be more complicated for immune cells to traverse, and therefore, its decalcification may facilitate this process.

12.2.3.3 *B4*

B4 is the last of the immune populations. It shows a clear, distinct, transcriptomic profile with markers such as *PPARG*, *FGF10*, *MYOC* or *ITM2A*. Interestingly, no immune population in mice was sufficiently matched to this population, probably indicating a unique function in humans. This population shows a diverse range of actions, including immunomodulatory actions, ECM and fibrosis response, lipid and compound metabolism, stress protection, vascular homeostasis, and, interestingly, it shows a relationship with HF.

IMMUNE FUNCTION Regarding immune function, B4 shows a mixed profile, with a predominance for immunomodulation and pro-regeneratory signals. One of its key markers, *FGF7*, for instance, is observed to participate in reparatory processes in gut inflammation, as well as the reduction of apoptosis, and neurite reparation (Chen et al., 2017; Marega et al., 2021); and the similar family marker *FGF10*, induces the chemotaxis of immune cells in lung by secretion of IL33 to activate repairing processes (Marega et al., 2021). However, in skin, besides its alarmin and chemotactic function described in B2/B3 fibroblasts, it may have a pro-inflammatory role by inducing IL-1 secretion by keratinocytes, and vice-versa (Russo et al., 2020). A similar function is observed for IGF1, which is observed to induce M2 polarisation in macrophages, related to anti-inflammatory and pro-regenerative signals (Yunna et al., 2020).

Regarding negative immunomodulation, several genes are involved. Two of the most relevant markers of B4 are *ITM2A* and *PPARG*. *ITM2A* is involved in contributing to T helper response (Kirchner et al., 1999; Tai et al., 2014), and is also involved in the induction of PD-L1, an immunosuppressive ligand (Zhang et al., 2021b; Zhou et al., 2019). On the other hand, *PPARG*, a widely studied adipose tissue TF, is known to suppress the expression of *TNF*, *IL1B*, *IL6*, *MMP9* innate response genes (Jiang et al., 1998), and Th1-related *IL12*, *CD80*, *CXCL10* and *RANTES* (Nencioni et al., 2002; Szatmari et al., 2004). It is also involved in the upregulation of lipid transport and metabolism genes such as CD36 (involved in the uptake of oxLDL), FABP4 or CD1d, involved in the presentation of lipids to T cells (Chawla et al., 2001; Szatmari et al., 2006; Tontonoz et al., 1998). Thus, its immune profile might be mixed, restricting acute and innate responses, and regulating more adaptative ones.

Other two markers involved in immune response downregulation are *SLPI* and *MGP*, mentioned in A4 and A1 populations. *SLPI* inhibits the maturation of IL-1β (Zakrzewicz et al., 2019), and reduces the expression of *IL6* in mouse adipocytes (Adapala et al., 2011). Similarly, MGP contains several Glu residues that can be carboxylated by vitamin K, attracting Ca^{2+} cations (Bashir et al., 2015; O'Shaughnessy et al., 2018) that confer them the ability to downregulate *TNF*, *IL1B* and NF-kB genes, as well as inhibition of Ca^{2+}-dependent inflammatory processes (Viegas et al., 2017). Lastly, the immune maturation molecule CD40 is not only expressed by B2/B3 fibroblasts, but also by B4 fibroblasts.

Besides immunomodulatory functions, B4 population is also observed to express certain chemoattractant molecules such as *CXCL12*, which attracts T cells in general (Bautista-Hernandez et al., 2017),

although it has higher attractive activity for T regs (Lin et al., 2009) and Langerhans cells (Ouwehand et al., 2008).

Lastly, regarding the complement system, B4 population expresses the components also expressed by B2/B3–C3, C6, C7–but it also actively expresses *CFH*, the inhibitor of the C3bBb convertase (Jozsi, 2017).

HF Interestingly, and contrary to B2/B3, B4 population expresses 10 genes involved in HF homeostasis, or expressed in this skin appendage. Some of these markers are *EFEMP1*, expressed in the bulge area (Takahashi et al., 2020); *WNT11* ligand is expressed in the dermal condensate at E14.5, and is expressed in adult HF DP and ORS (Lim et al., 2012); *HSPG2* is located in the BM adjacent to DP (Tsutsui et al., 2021); *IGF1* is expressed by DP cells (Panchaprateep et al., 2014); *FGF10* is located in DP and ORS (Zhang et al., 2018); and *MGST1* is a marker of mature sebocytes (Kobayashi et al., 2019).

Additionally, *IGF1* may be expressed in the HF regeneration context, since lower levels of IGF1 are correlated with androgenic alopecia (Panchaprateep et al., 2014), similarly to *FGF7/10*, which also induce new HF cycles by up-regulation of β-catenin (Greco et al., 2009; Zhang et al., 2018). Similarly to FGFs, *MYOC* also intervenes in Wnt signalling as it interacts with FZDs, SFRPs and WIF, inhibiting their function, and activating Wnt pathway (Kwon et al., 2009).

The last two markers are *PLA2G2A* and *PPARG*. *PLA2G2A* overexpression is related to adnexal hyperplasia, which results in alopecia (Grass et al., 1996), and *PPARG* is involved in HF morphogenesis and maturation, and its downregulation induces a delay in HF cycles (Islam et al., 2018).

ECM B4 population is actively involved in secreting factors that bind several ECM components. For instance, (1) HSPG2 perlecan binds heparan and chondroitin sulfate GAGs, laminins, collagens IV, V, VI, XI, elastin and FBLN2. It also binds, lipid with a LDLR-like domain, as well as Wnt morphogens (Hayes et al., 2022); (2) PODN podocan binds collagen I (Shimizu-Hirota et al., 2004); (3) FGF7/10 bind to HSPG (Marega et al., 2021), and may bind TNC (Jones et al., 2021), and (4) *MGP* binds elastic fibres.

Lastly, although it is not an ECM-component binder, *SLPI* is involved in ECM homeostasis. SLPI is an inhibitor of serine proteases–such as trypsin and cathepsin–, and MMP1/9 produced by immune cells (Nugteren et al., 2021; Zhang et al., 1997). Therefore, as done by A1/A4, it may regulate and avoid excess ECM degradation.

FIBROSIS/REGENERATION B4 population, similar to the rest of the immune fibroblasts, and in line with findings from previous markers from B4 and the rest of B populations, shows a pro-regenerative and antifibrotic response. *ADA*, apart from being necessary for immune signalling, is observed to regulate collagen production (Marucci et al., 2022) and may have an antifibrotic potential (Fernández et al., 2013). Similarly, *SLPI* suppresses TGF-β (Ashcroft et al., 2000); *IGF1* shows modulation of *ACTA* expression *in vitro* (Culley et al., 2021); and *GPX3* is involved in cardiac fibroblast regulation into a pro-regenerative fate under stress conditions (Li et al., 2022a).

Interestingly, this population also expresses *GDF10*, a TGF-β ligand from the BMP7 family. Therefore it is likely that this population has either a reduced TGF-β signalling or it is linked with non-canonical signalling.

PROTECTION AGAINST STRESS B4 is a highly active population, and due to this activity and the involvement with other immune populations, on last relevant function is derived from the expression of "protective" markers. Similar to B1, which expressed *SOD2* to reduce H_2O_2, B4 expresses *GPX3*, an extracellular enzyme that reduces H_2O_2, hydroxyperoxides and other oxidised species (Brigelius-Flohé, 2006; Maiorino et al., 1995), and its expression is upregulated by PPARG (Chung et al., 2009; Reddy et al., 2018). Similarly, B4 fibroblasts express *MGST1*, a metabolic enzyme involved in eicosanoid and glutathione metabolism. This factor, glutathione, is a highly active component against oxidative stress too (Pompella et al., 2003). Additionally, B4 expresses two markers with an associated chaperon function: *MYOC* and *ITM2A*. *MYOC* is a modulator of the actin cytoskeleton, but it also acts as a chaperone against ER stress (Anderssohn et al., 2011); and *ITM2A* contains a BRICHOS domain that is known to act probably as a chaperone in different scenarios (Hedlund et al., 2009). Lastly, *APOD* is observed to have a protective action against oxidative stress preventing lipid oxidation in different organisms (Ganfornina et al., 2008).

12.2.4 *D axis*

12.2.4.1 *D1*

D1 population is, together with D2 and E1, a small population that was previously unidentified but with very particular markers. Human-mouse comparisons revealed that D1 was transcriptomically similar to either y5 or v1 populations. However, at a closer examination, more genes from D1-y5 are selected markers based on their function. Some of these markers are *ABCA8*, *APOD*, *COL8A1*, *SOX9*, *TGFBI*, and *VIT*; of which *SOX9* is exclusively expressed by D1. There are other markers expressed exclusively by D1, such as the similar SOX gene, *SOX8*, *BAMBI*, *CDH19*, *FMO2*, or *ATP1A2*. Some of their functions will be mentioned later.

The most relevant functions of the D1 population, and D2, to a lesser extent, are related to the expression of nerve and blood vessel markers that are either expressed by or observed in these structures or whose function is associated with the activities performed by these cells.

NERVE FUNCTION The D1 population expresses several markers that are also markers of other cell populations. For instance, (1) *ABCA8* is highly expressed by oligodendrocytes and may play a role in sphingomyelin production (Kim et al., 2013b), (2) *SOX8* TF is involved in the development of astrocyte, oligodendrocyte and Schwann cell precursors (Hutton et al., 2009; Stolt et al., 2004; Takouda et al., 2021), and also plays a role in sphingomyelin production (Turnescu et al., 2017), (3) *CDH19* cadherin is expressed by Schwann cell precursors (Kim et al., 2020b), (4) *S100B* is also a marker of Schwann cells, although it is expressed in melanocytes, chondrocytes, adipocytes and immune cells (Donato et al., 2009); (5) *PIEZO2* is a mechanosensory Ca^{2+} channel involved in propriocepcion (Woo et al., 2015), located in Merkel cells and afferent terminals (Wu et al., 2017); and (6) *COL28A1*, although not being related with a clear neural function, it is almost-exclusively detected by terminally-differentiated Schwann cells and Merkel cells (Grimal et al., 2010).

Some of these and other D1 markers have additional relevant functions in neural environments. For instance, *ATP1A2* is a Na^+/K^+ channel that osmoregulates the cation gradients for proper neuronal excitability in nerves and muscles (Friedrich et al., 2016); *OGN* and *VIT*, usually involved in fibroblast proliferation, participate in neural development and neurite growth (Deckx et al., 2016; Whittaker

et al., 2002); and *SCN7A*, another Na^+ channel that acts jointly with *ATP1A2* in glial cells (Dolivo et al., 2021), acts as a Na^+ concentration regulator in homeostasis and TEWL-related injuries (Watanabe et al., 2000). In this case, coactivation of *ATP1A2* to compensate Na^+ influx leads to the hydrolysis of ATP and the production of lactate. This, together with an increase in endothelin ligands induced by *SCN7A*, leads to compensatory neural responses like an increase in water intake in the ECM or an increase in blood pressure (Dolivo et al., 2021).

VASCULATURE Jointly with neural functioning, and as we have observed with *SCN7A*, D1 population expresses more than 10 markers with a direct or indirect function related to vasculature.

Three interesting markers are *COL15A2* and *COL8A1/2*, which are shown to be expressed at the BM of microvessels or at proliferating vessels (Manon-Jensen et al., 2019a; Sutmuller et al., 1997). *EFNA1* and *NRP2*, also expressed by D2, act as ligands of receptors of several pathways involved in neural and vascular neogenesis (Hao et al., 2020; Islam et al., 2022). In fact, NRP2 is involved in vascular permeability and lymphangiogenesis as well (Favier et al., 2006; Harman et al., 2020; Yuan et al., 2002). In a similar fashion, *BAMBI* and *TGFBI*, related to TGF-β signalling, are also indirectly related to these responses (Guillot et al., 2012), although *TGFBI* specifically shows antiangiogenic potential (Son et al., 2013).

Other members that act by inducing the secretion of pro-angiogenic factors are *IGFBP7*, which secrets PGE2, which in turn induces angiogenesis (Tamura et al., 2009); *NFKB1* and *S100B*, which are involved in the secretion of iNOS (Donato et al., 2009; Liu et al., 2017); or *EPHX1* and *CYP4B1* which metabolise multiple substrates involved in angiogenic responses (Gautheron et al., 2020; Tang et al., 2010b). Lastly, both D1, D2 and E1 populations express the kallikrein *KLK1*, responsible for the production of bradykinin (Bellis et al., 2020), a vasoactive substance involved in blood vessel dilation by the production of NO and prostacyclin, as well as derived inflammatory processes (Pinheiro et al., 2022).

IMMUNE B axis is the principal immune axis. However, there are certain markers that are related to this family of responses specifically expressed by D1. Some markers, like *NFKB1* and *SOD2*, already described in B1, or *CCL2*, *CFH*, *RARRES* or *SOCS3*, also expressed by B axis population, showcase the putative function of D1 in immune roles; although due to the pleiotropism of some of these genes, their functions in this population could be different.

Specific to D1 are three additional markers: *C2orf40/ECRG4*, *CYP4B1* and *ATP1A2*. *ATP1A2*, aside from its function in nerve, is also observed to modulate LPS responses by NF-kB signalling (Leite et al., 2020); a similar activation mechanism of *C2orf40*, which also activates IFN-I mediated signalling (Moriguchi et al., 2016) and neutrophil recruitment (Dorschner et al., 2020). Lastly, *CYP4B1* is upregulated by pro-inflammatory cytokines (Smerdova et al., 2014), so downstream effects of this activation may also be immune-related.

TGF-β Regarding TGF-β signalling, D1 also expresses *BMP7* and *INHBA*, a combination postulated for C5 population to inhibit canonical signalling and activate the BMP-mediated alternative pathway. In addition, D1 expresses *GDF10*, which acts as *BMP7*.

In a similar fashion to HF populations, TGFB-β signalling in D1 is complex, with additional expression of markers such as *PPP1R15A*, which recruits Smad7 to inhibit TGF-β (Shi et al., 2004) and *BAMBI*, which is also inhibitory, by binding TGFBRI/TGFBRII complex (Huang et al., 2012a). On the other hand,

SCN7A expression linked to increased acidity in ECM–from sensing TEWL after injury–is related to latent TGF-β activation (La et al., 2016). In a similar fashion, *SOX9* TF may lead to fibrosis by induction of *ACTA2*, collagens and *LOXL2* (Gajjala et al., 2021; Scharf et al., 2019); and is related with hfSC niche reestablishment by TGF-β signalling (Berndt, 2014). This stemness state, although regulated by NOTCH in specific environments–e.g. gastric cancer–, and not TGF-β signalling, may be induced by *SPON2* (Badarinath et al., 2022).

WNT Wnt signalling, similar to C axis populations, is active in D axis populations. In D1, it is probable that inhibitory signalling is more dominant, with the expression of *DKK3*–sequesters LRP5/6 (Ahn et al., 2011)–, *SFRP4*–binds Wnt ligands and inhibits their action (Liang et al., 2019)–, *SHISA3*–impedes the translocation of Wnt receptors to the membrane (Furushima et al., 2007; Hedge et al., 2008)– or *NDRG2*–activates GSK-3β, which signals β-catenin degradation (Lee et al., 2022)–. Thus, the expression of *FZD2* non-canonical receptor (Gujral et al., 2014) may not be sufficient for this signalling.

COLLAGENS AND ECM REGULATION Interestingly, D1–and D2 to some extent–shows increased expression of 5 collagen genes: *COL8A1/2*, found in proliferating vessels (Sutmuller et al., 1997); microvessel-associated *COL15A1*; and nucleating *COL9A3* and *COL28A1*, found around Merkel and Schwann cells (Grimal et al., 2010). Another related ECM component related to collagen is *DCN*, not only expressed by A axis, but also by D1.

Additionally, D1 population expresses *CYP4B1*, which is related to *POSTN* expression (Zhao et al., 2013); or *VIT*–vitrin–, acting similar to COCH in cell adhesion and neural development (Whittaker et al., 2002). It also expresses several genes whose expression products bind or regulate multiple ECM components, such as *TGFBI*, which binds integrin $\alpha v\beta 3$, fibronectin, vitronectin, collagens, fibrinogen and VWF (Ruoslahti et al., 1987); *EGFR*, which binds AREG, TGF-α, or EREG (Harris, 2003; Normanno et al., 2006), and other ligands like TNC, DCN, or laminin-332 (Iyer et al., 2007; Santra et al., 2002; Schenk et al., 2003); and *ANGPTL7*, which is associated with a diminished expression of *FN1*, *COL1A1*, *COL4A1*, *COL5A1*, *MYOC* and *VCAN*, and an increased expression of *MMP1* (Comes et al., 2010), in line with a pro-inflammatory profile.

Although it is complex to establish a clear function of these genes in the context of the D1 population, it is possible that this type of fibroblast transforms "classical" ECM microenvironments into more specialised environments necessary for the correct development of HF, nerve or blood vessels.

VITAMIN A AND LIPID METABOLISM Regarding lipid and retinol metabolism, three related genes are expressed by D1: *ALDH1A3*, which transforms retinaldehyde into retinoic acid (Kedishvili, 2013); and two lipoprotein-related markers: *APOD* and *LDLR*. *APOD*, despite belonging to the apolipoprotein family, is also highly related to RBPs in function (Munussami et al., 2018); thus, it may be the agent transporting retinol into D1 fibroblasts. Lastly, *LDLR*, which binds VLDL and LDL in plasma (Go et al., 2012), may be related to lipid metabolism; or, considering the adipose tissue surrounding anagen HFs, may be an active component in the communication with hypodermis.

FMOS Lastly, the family of flavin mono-oxygenases (FMOs), is highly expressed by B4, D1 and D2 populations; and in the case of D1, is the only population expressing all of the members expressed in the samples from the analysis: *FMO1*, *FMO2* and *FMO3*. In general terms, each FMO metabolites, generally using FAD or NADPH as cofactors, a wide range of molecules including triethylamine and

other secondary and tertiary amines, thiols and other xenobiotics (Hisamuddin et al., 2007). Although FMOs activity has been linked with fibroblast activation (Yu et al., 2022) and cardiovascular disease (Schugar et al., 2015; Shih et al., 2019; Warrier et al., 2015), specific functions of these enzymes are yet unknown.

12.2.4.2 *D2*

D2 is the second population from the D axis. Although both populations differ in certain markers and possibly in some of their functions, their transcriptomic similarity is extremely high, and the most relevant genes are expressed in both populations. Regarding mouse-human comparisons, D2 shows a clear similarity to v1 population, with exclusive markers, such as *AQP3*, *CAV1/2*, *ITGA6*, *ITGB4*, *KRT19*, which are all exclusively expressed by D2. Additionally, D2 also expresses other exclusive markers, including *DACT1*, *ADAMTSL5*, *AQP3*, *GFRA2* and *NGFR*, many of which will be discussed throughout this section.

NERVE AND VASCULATURE Similar to D1, D2 population seems to be associated with nerve and blood vessel microenvironments. Regarding nerve markers, population D2 shows the expression of already mentioned D1 markers such as *ABCA8*, *S100B*, and *VIT*; but it also expresses specific markers. For instance, (1) *GFRA2*, a member of glial cell-derived neurotrophic factor (GDFN), is necessary for neurotransmitter release of several neuron types (Airaksinen et al., 2002); (2) *KRT19* and (3) *NGFR* are expressed in Merkel cells–as well as fibroblasts, mast cells and rete-ridge keratinocytes–(Botchkarev et al., 2006; Michel et al., 1996), and *NGFR* is necessary for Schwann cell cone formation (Cragnolini et al., 2008); and (4) *SLC22A3*, a channel involved in neurotransmitter transportation (Amphoux et al., 2006), which may be necessary for the regulation of osmolality in the neuron vicinity (Vialou et al., 2004).

Regarding vascular markers, D2 population also expresses D1 markers *NRP2* and *EFNA1*; as well as *CAVIN2*, which will be discused later, and *SEMA3C*, which binds NRP2 and shows a mixed angiogenic profile (Jiao et al., 2021; Karpus et al., 2019; Neufeld et al., 2008; Valiulytė et al., 2019).

IMMUNE D2 population expresses many immunogenic markers, but many of these have already been mentioned in previous populations: *CLU* (complement inhibitor, A1/A4), *CCL2* (B3), *SOCS3* (B3), *ITM2A* (B4) and *C2orf40* (D1). The only relevant new marker is *CCL13*, a pleiotropic chemokine that (1) binds CCR1/2/3/5 and induces chemotaxis of different immune cells, including Th and NK cells, mast cells, basophils, eosinophils and monocyte/macrophages (Mendez-Enriquez et al., 2013); (2) upregulates *TLR2/3/4/5* expression and induces DC maturation (Mendez-Enriquez et al., 2013); and (3) its expression is related to CD40 and MHC-II expression, involved in adaptive responses (Chiu et al., 2004).

Although D axis populations show some expression of regulatory immune modulators such as *SOCS3*, *CLU* and *ITM2A*, the expression of several chemokines and other markers indicate that their immune activity is not negligible.

HF AND ECM REDULATION D2 population also shows a wide range of markers associated with ECM and HF in homeostasis and other conditions.

Some markers that were mentioned to be expressed surrounding nerves, blood vessels and other structures have been shown to mark HF areas such as *KRT19*, which marks ORS in the bulge area (Michel et al., 1996); *SLC22A3*, located in HF, and SG–and may be associated with sebum production– (Takechi et al., 2021); *ADAMTSL5*, also located in ORS (Higgins et al., 2011); *GFRA2*, expressed in ORS, IRS, DP and CTS in anagen, although its CTS expression is maintained throughout the HF cycle (Adly et al., 2008); and *NGFR*, expressed in DP.

In fact, some markers, like *NGFR*, are also associated with the HF cycle or balding processes. For instance, *NGFR* is expressed in the anagen-to-catagen transition (Botchkarev et al., 2000; Enshell-Seijffers et al., 2010); *FGF7*, which signals the induction of new hair cycles (Geyfman et al., 2014; Greco et al., 2009); *CCL13*, whose immune function, if exacerbated, is associated with T cell accumulation in alopecia areata (Wang et al., 2021a); and *PTGDS*, which is highly expressed in androgenic alopecia (Garza et al., 2012).

Regarding ECM regulation, the D2 population expresses three different ADAMTSL members: *ADAMTSL3*, *ADAMTSL4*, *ADAMTSL5*. Their main function is the binding of fibrillin-1 (Sengle et al., 2012), and regulating its biogenesis (Gabriel et al., 2012), although each member has specific functions as well. Additionally, D2 also expresses *TGFBI* and *VIT*, which have been mentioned in D1.

WNT SIGNALLING Similar to D1, D2 population expresses several genes involved in these two pathways. Regarding Wnt signalling, apart from the already mentioned *NDRG2* and *SFRP4*, it expresses *DKK3*, *CAV1* and *DACT1*, all three represorst of Wnt signalling–canonical or overall–(Ahn et al., 2011; Esposito et al., 2021; Galbiati et al., 2000; Gao et al., 2022); and *DAAM1*, which activates PCP non-canonical pathway (Lai et al., 2009).

LIPID METABOLISM While the D1 population expressed genes related to cholesterol and general compound metabolism; markers expressed by D2 are slightly more related to eicosanoid metabolism. For instance, *EPHX1* metabolises the oxidation of the endocannabinoid 2-AG into arachidonic acid (Nithipatikom et al., 2014); *CYP4B1* CYP450 monooxygenase catalyses arachidonic acid oxidation; and *PTGDS* catalyses the conversion og PGH2 to PGD2 (Zhou et al., 2010).

CAVEOLINS Caveolins are a family of scaffolding proteins associated with caveolae formation. Caveolae are membrane invaginations of lipid rafts that endocytose and are generally involved in signal transduction (Anderson, 1998). D2 population specifically expresses three caveolae-related markers: *CAV1* and *CAV2* caveolins, and *CAVIN2*–Caveolae Associated Protein 2–. Generally, although *CAV1* is able to produce caveolae on its own (Scherer et al., 1997), caveolae are usually created by the interaction of these three proteins (Mora et al., 1999). In some scenarios *CAV1* and *CAV2* may have antagonistic functions (Almeida, 2017).

The functions of these genes are diverse and context-specific. For instance, *CAV1* is related to TGFBR1 sequestering from membrane rafts into caveolae to reduce TGF-β signalling (Hwangbo et al., 2015); may also sequester β-catetin to regulate Wnt signalling (Galbiati et al., 2000; Gao et al., 2022); may be related to immune responses by inducing T-cell proliferation (Ohnuma et al., 2007), favours neutrophil extravasation (Marmon et al., 2009) or ICAM1-mediated leucocyte adhesion (Bouzin et al., 2007). *CAV2* may be related to keratinocyte proliferation upon KGF action (Gassmann et al., 2000), may also act like *CAV1* in TGF-β signalling (Xie et al., 2011), or may be involved in lipid metabolism and

receptor trafficking (Sowa, 2011). Lastly, *CAVIN2* may control angiogenesis by regulating eNOS activity (Boopathy et al., 2017) and may also inhibit TNF-mediated NF-kB signalling (Annabi et al., 2017).

The immune profile associated with the expression of caveolin-related genes is in favour of a mature, adaptive response, in line with the immune markers presented in this section.

INTEGRINS Lastly, an interesting heterodimer expressed by D1 principally is the integrin $\alpha 6\beta 4$, formed by the expression of *ITGA6* and *ITGB4*. Similar to caveolae, the functions of this integrin dimer are diverse, but may be extremely relevant for this fibroblast population.

ITGA6 by itself is found to serve several functions. In mouse HF, ITGA6 is located at the BM surrounding DP and ORS (Tsutsui et al., 2021); and it was observed that HF ITGA6$^+$ cells were able to generate hair shaft, ORS and IRS *in vitro* and *in vivo* (Yang et al., 2014; Yang et al., 2020). This activity is not specific to HF, as ITGA6$^+$ populations tend to show stem potential across tissues (Yu et al., 2012). Additionally, it participates in Langerhans cell migration from the epidermis into the dermis and lymph nodes (Varlet et al., 1991) and may act as an ECM mechanosensor to induce myofibroblast activation under specific circumstances (Chen et al., 2016). This effect is mediated in keratinocytes by *EGFR* binding to integrin $\alpha 6\beta 4$. Under ECM stiffening, EGF bioavailability increases, activating EGFR and inducing migration and cell activation (Kleiser et al., 2020).

Regarding the heterodimer, it binds several ECM and BM members, including laminins 332 and 511, and colocalises with BM collagen IV, as well as other integrins, in DEJ and HF and SG BM (Blok et al., 2015; d'Ovidio et al., 2007). Additionally, integrin $\alpha 6\beta 4$ is tightly related to cell migration, both by playing a key role in the formation and stabilisation of hemidesmosomes–together with CD151 and COL17–, as well as its disassembly and translocation to lamellipodia (Walko et al., 2015; Wilhelmsen et al., 2006). Another relevant function is the maturation of the microvasculature by playing a suppressive role in angiogenesis (Hiran et al., 2003).

12.2.5 *E1*

Population E1 function, similar to D1 and D2, cannot be assigned based on belonging to any of the "big" axes. Additionally, there is no clear relationship between E1 and any mouse population. Looking at the combined PAGA graphs from Figure 45, we see that this population shares transcriptomic similarity with B4, D1 and C3 populations. Firstly, all three populations are directly or indirectly linked to HF, so it is also likely that this population will be too. Analysis of markers for this population reveals that, similarly to B4 or D axis, this population has a wide diversity of functions, including TGF-β and Wnt pathways, vasculature and neural regulation, proliferation, ECM, and immune signalling. Even more, due to the reduced, but diverse, number of markers,there is no function with a clear predominance, which reinforces the versatility of this population.

PROLIFERATION AND CELL MIGRATION One of the gene families relevant to E1 population is the insulin growth factor (IGF) signalling pathway, with the expression of *IGFBP2* and *IGF1*, which are associated with cell proliferation. *IGFBP2* may act by binding IGF1 or IGF2 (Shin et al., 2017). By binding to IGFs, it modulates their availability in the ECM. After proteolytic cleavage, IGFs may bind their receptors (Boughanem et al., 2021). Its functions are related to angiogenesis and cell proliferation in general (Boughanem et al., 2021). This is related to its implication in energy metabolism; since its

expression is increased after fasting, insulin suppresses it, and it is a target of PPARA (Shin et al., 2017). Similarly, IGF1 binds integrins $\alpha v\beta 3$ and $\alpha 6\beta 4$ to form a ternary complex with IGFR1 (Fujita et al., 2012; Saegusa et al., 2009) or with insulin receptor. In both cases, it strongly activates the Akt pathway, which stimulates growth and proliferation and inhibits apoptosis (Juin et al., 1999; Peruzzi et al., 1999). Interestingly, keratinocytes express IGF1R, while fibroblasts express IGF1; thus IGF1 in keratinocytes may work by paracrine modulation (Rudman et al., 1997).

Another proliferative marker is *SPON2*. It activates STAT3 signalling in cancer stem cells to reinforce their stemness state (Badarinath et al., 2022), and also upregulated NOTCH signalling in gastric cancer (Kang et al., 2020). This activation may be targeted to other cells, like macrophages, where SPON2 activation leads to its chemotaxis (Li et al., 2020a).

Lastly, the E1 population expresses two markers related to migration: *PTN* and *GMFG*. *PTN* binds to several receptors, including PTPRZ1, ALK, SDC3, ITGAvB3, and NRP1 to induce cell migration (Papadimitriou et al., 2022). GMFG is a protein commonly found in T and B cells, neurons and cancer cells; and is located in the filopodia, pseudopodia, or other prolongations used in the movement of the cell (Deretic et al., 2021; Lippert et al., 2012; Liu et al., 2022).

HF E1 population shows the expression of 4 markers linked to HF structure and homeostasis. Two of these markers, *IGF1* and *IGFBP2*, have already been mentioned before and are detected in DP and DS respectively (Hagner et al., 2020; Panchaprateep et al., 2014). IGF1 is presumed to be linked to HF renovation. In fact, lower levels of IGF1 and IGFBP2 are observed in androgenic alopecia patients (Panchaprateep et al., 2014). Similarly, *PGF*–placental growth factor–is also involved in maintaining HF growth (Hu et al., 2022; Yoon et al., 2014). Lastly, *LFNG*, a fucosyltransferase present in the Golgi apparatus that adds an N-acetylglucosamine to the fucose of the NOTCH1 receptor, reducing its activity (Barco Barrantes et al., 1999; Shimizu et al., 2001). NOTCH signalling is necessary for HF cells to maintain cell differentiation (Nowell et al., 2013). Therefore, NOTCH inhibition may be related to HF cycle, in line with some of these other markers.

IMMUNE Several previously described markers are also expressed by E1 fibroblasts, namely *MGP* and *ITM2A* from B4, and *ACKR3* from A1/A4. All these markers show a direct suppressive role in the immune system or in the clearance of immune products from the ECM. This population expresses two additional markers, *CMKLR1* and *TNFRSF21*. *CMKLR1* is the receptor for chemerin/RARRES2, predominantly expressed by B2 and B4 populations, and shown to have pro- and antiinflammatory functions; and TNFRSF21 is a proapoptotic TNF receptor (Dostert et al., 2019) that may decrease T cell and B cell proliferation (Liu et al., 2001; Schmidt et al., 2002), although it may also be necessary for their migration to the central nervous system (Schmidt et al., 2002). Therefore E1 population may show an immunomodulatory function, as transcriptomically similar populations like B4.

ECM E1 population expresses 4 ECM-related markers: the *MMP6* metalloproteinase, which can degrade collagen III, fibronectin and laminin (Cabral-Pacheco et al., 2020), and *TIMP3* to counteract MMPs functions. Additionally, it expresses two collagens, *COL15A1* and *COL26A1*: *COL15A1* has already been mentioned as a non-fibrillar collagen that is located at the BM of microvessels (Manon-Jensen et al., 2019a), nerves, adipocytes (Saarela et al., 1998), and HFs (Hagg et al., 1997); and *COL26A1*, specific of E1 population, is not sufficiently studied to have a clearly assigned function.

NEUROVASCULAR E1 population, similar to D1 and D2, expresses genes related to neural and vascular development. Regarding the neural-related markers *SCN7A* and *SLC22A3*, already commented on D-axis populations, are also expressed by the E1 population. This population specifically expresses *NTRK3*, a member of the neurotrophin receptor family, involved in different proliferation and differentiation pathways (Ruiz-Cordero et al., 2020). In adults, they are usually expressed in the peripheral and central nervous system and are thought to maintain the regular neuron balance (Hechtman, 2022). In skin, it is expressed in the placode morphogenesis, specifically in Merkel cell precursors (Jenkins et al., 2019).

Regarding vascular markers, most of them are specific to E1 population, and act via VEGF signalling. For instance, *PTN* acts as a competitor of VEGFA to bind VEGFR2, showing pro-angiogenic potential (Koutsioumpa et al., 2015); *PGF* binds VEGFR1 and may act synergistically with VEGF (Chang et al., 2012; Odorisio et al., 2006). Similarly, *PDGFD* is shown to increase the expression of VEGFA, as well as FGF1, INHBA and IL11 (Kim et al., 2015), thus it has a marked pleiotropic potential. Lastly, one of the many functions of IGFBP2 is also related to proliferation and angiogenesis (Boughanem et al., 2021).

WNT AND TGF-β SIGNALLING It comes as no surprise that, as with other functions shared with D1, D2, B4 and other populations, E1 expresses markers related with Wnt and TGF-β signalling.

Regarding TGF-β signalling, there is a mixture of activatory–*SLC22A3*, *EGR2*–and inhibitory/modulatory–*LTBP2*, *IGFBP2*–signalling. Moreover, both activatory and inhibitory cases are, in some cases circumstantial. For instance, *EGR2* is observed to be induced by TGF-β in SSc fibroblasts (Fang et al., 2011), and *SLC22A3* could be indirectly related to *TGFB1* increased expression (Vollmar et al., 2019). Thus we cannot ensure that activatory signalling is clear in this population. Regarding *IGFBP2*, it was observed to decrease αSMA expression via TGFB1 inhibition (Park et al., 2015), but due to its pleiotropism, this function may not be active in dermal fibroblasts.

12.3 CONCLUSION ON FIBROBLAST HETEROGENEITY AND ROLES IN SKIN

The results and conclusion derived from this thesis clearly show that fibroblasts are far from "ECM-producers" and instead show clearly how they interact with all the cell types within and surrounding the stroma, and how this translates into a large array of functions, many of them categorised into the interaction with immune cells, vasculature, nerves and related cell types.

A axis populations are clearly the main "ECM producers", with the expression of a large set of ECM components and their corresponding regulators. This expression is discriminated by their spatial location, with reticular and papillary fibroblasts expressing specific sets of genes related to adapted to the area.

For instance, A2 population, which could be termed as the "secretory papillary fibroblasts", secrete collagens, proteoglycans and other components bound to the DEJ and other structures surrounded by BMs–e.g., *COL4A2/4*, *COL6A1/2/3*, *COL13A1* *COL18A1*, *COMP*, *POSTN*–. In contraposition, A1/A3/A4 fibroblasts, which could be termed as "secretory reticular fibroblasts", secrete entirely different ECM components associated with providing structure and strength to the dermis–i.e. *COL1A2*, *COL3A1*, *COL12A1*, *PODN*, *DCN*–. Two interesting contraposed components are *TNXB*, expressed by A1/A3/A4, and *TNC*, by *A2*; which also have contraposed functions: *TNXB* is more associated with the restriction of cell motility, whereas *TNC* favours it. Moreover, A4, in addition to these genes, also expresses elas-

tic fibre-associated genes such as *ELN*, *FBN1* and *EMILIN2*; and thus could be coined as the "secretory elastic fibre fibroblast". Interestingly, similar to the *TNC*/*TNXB* binomium, two microfibril-associated markers, *MFAP5*/*MFAP2* are similarly expressed by reticular and papillary fibroblasts respectively.

Of note, several of the ECM-related markers, such as *MFAP2*, *COL7A1*, *COL6A1/2/3*, *COMP*, *TNC*, or *POSTN* are also expressed by HF-related populations from the C axis. This showcases that papillary fibroblasts are in close contact with HF structures, either by being located near the IFE, or other surrounding HF areas.

ECM production is related to ECM regulation, and in this matter, A2 differs from A1/A3/A4 too. Three main families of ECM-regulation components are necessary for proper ECM regulation: (1) pro-collagen cutting enzymes, (2) collagen and fibre maturation enzymes and (3) ECM-degradation enzymes. Both families of fibroblasts express all three families, but their components differ. *PCOLCE* and *PCOLCE2* pro-collagenases are expressed by reticular fibroblasts, in contraposition to *APCDD1* by papillary fibroblasts; *LOX*, *LOXL1* and *P4HA2* fibre maturation enzymes are expressed by reticular fibroblasts, whereas *LOXL2* is expressed by papillary fibroblasts; and, in order to degrade the ECM, reticular fibroblasts express *MMP2*, whereas papillary fibroblasts express *MMP2*.

Part of their ability to produce and regulate the ECM is linked to their ability to sense the stiffness and traverse the ECM. This ability is liked to several markers. A1 and A4 populations express *PIEZO2*, *DBN1* and *TPPP3*, the last two of them bound to F-actin and found in the terminal protrusions; A1/A3 and A2 populations express *SGCA/G* sarcoglycans; and A2 exclusively expresses *ANTXR2*, which interestingly binds papillary collagen VI.

Despite their relevance in ECM production and regulation, stating that A axis populations only produce ECM would be a mistake since they also take an active role in immune signalling, nerve and vascular activity, and the metabolism of several compounds. For instance, immune response regulation, which is highly relevant in fibroblasts, predominantly from B axis, is also important in A axis, primarily on reticular fibroblasts. There are several strategies to this immunomodulation, like (1) inhibition of pro-inflammatory cytokines–IL-1β, IL6–or other components–MMP1/9–by *SLPI*, *C1QTNF3* or *DPP4*; (2) complement inhibition by *CD55*, *CLU* and *DCN*; (3) clearance of proinflammatory chemokines from the ECM by truncation of chemokines (by DPP4) to only bind ACKRs expressed by these fibroblasts–e.g., ACKR3/4–to be scavenged and degraded; or by (4) degradation of immunogenic products like oxysterols by *HSD3B7*. The main reason for the immunomodulation might be to preserve the ECM integrity against the action of B axis fibroblasts or other pro-inflammatory or immune-related insults.

Regarding A3 population, most of the markers have been expressed by either A1 or A2. Therefore, it is possible that this population acts as a "bridge", not only transcriptomically but also in its physical location in the dermis, between the A2-populated papillary dermis, and A1-populated reticular dermis. Although few genes are markers of A3, the expression of *CES1* and *SLC47A2*, jointly with *MGST1*, may confer this population a potential to metabolise endogenous and exogenous substances that may be harmful in the ECM. Lastly, it is apparent that A4 population, despite being closely related to A1, serves a complementary function, with increased expression of vascular-related markers–*AGTR1*, *CD248*, *SEMA3C/E*, *NRP1*, *MGP*–. Including these and other relevant functions, the A4 population is likely to be a "specialised" A1 population that may be necessary for vascular structure.

Initial analyses of single-cell data revealed one immune-related fibroblast population (Solé-Boldo et al., 2020; Tabib et al., 2018; Vorstandlechner et al., 2020) which, compared to the traditional knowl-

edge of no fibroblasts exerting that specific function, it was considered as an advance. This analysis expands on this knowledge, revealing 4 fibroblast populations with a direct relationship with immune function exist, which can be simplified to 3 archetypal populations with clearly differentiated functions: B1, B2/B3 and B4.

B1 population is majorly associated with an innate and acute response, and could in fact be termed as the "acute immune response fibroblast", based on the expression of (1) immediate early genes, which activate the expression of necessary target genes; (2) acute phase cytokines and chemokines, such as IL6, most of which are necessary to attract innate cells such as neutrophils and monocytes, but also other APCs, and NK and T cells; and (3) ECM degradation proteins, mainly matrix metalloproteinases and members of the ADAM family.

B2 and B3 populations also participate in primary responses but are active in secondary and adaptive responses as well, and could be termed as "adaptive immune response fibroblast". This occurs either by (1) the expression of specialised immune cell chemoattractants, like *CCL19*, *CSF1*, *IL34*, *CXCL12*; (2) other components like adhesion molecules, that make the effect of chemoattractants feasible, or (3) by the expression of cytokines or other ligands, such as HLAs, *TNFRSF13* or *CD40*, that induce cell maturation.

Lastly, B4 may play a triple role. First, the expression of anti-fibrotic, pro-regulatory and protective factors, as well as anti-inflammatory and immunomodulatory signals, clearly shows that it is involved in the regulation of the immune response by several mechanisms, including the (1) direct response against damages to the ECM or other cell types, and also (2) by direct modulation of pro-inflammatory signals, inhibiting them. It is interesting to notice also that the inhibition is particularly apparent for the innate response, but not so for the adaptive one, likely because of the link between certain adaptive responses to pro-regenerative environments. Lastly, (3) it is highly likely to act not only on the ECM as a whole but also on HF, probably acting as the "immune population of HF", or the "immune regulator within HF". In any case, more analyses of cell type location must be performed to see whether the realm of action of this population is restricted to HF, or if it is extended to other areas in the dermis. Therefore, the B4 population could be termed as the "immunoregulatory fibroblast" or the "HF-associated immune fibroblast".

Interestingly, the hallmark function of B4 of protection against stress is also present in B1, possibly to protect itself and its surroundings from the increased expression of IEGs, increased production of ROS, and production of other cytotoxic components that may be necessary during the primary response.

An additional complementary function of B axis fibroblasts, and A axis to some extent, is the antifibrotic response by intervening in multiple parts of the TGF-β pathway like (1) reducing the expression of TGF-β–by *CXCL1*–, (2) inactivating the TGFBR–by *PPP1R15A*–or (3) other direct or indirect mechanisms that avoid the transformation of fibroblasts to a myofibroblastic state–by *IL11*–.

Regarding the C axis, the vast number of studies with specific markers shows that at least the two archetypal populations, C1 and C2, belong to the HF. Specifically, C1 can be coined as "dermal sheath" and C2 as "dermal papilla", based on the expression of established markers in single-cell literature, as well as other sources. Both populations are actively secreting and modifying ECM components, most of which are linkers of collagens, SLRPs and other components. In fact, components expressed by these populations, like *ACAN*, *F13A1*, *COL11A1*, or *COL24A1*, which are associated to cartilage-like or

calcified structures, shows that the ECM in HF, and specifically in DP, is different from the ECM found in the different dermal layers. Therefore, it is likely that the proliferation of these HF populations is highly dependent on this specific composition.

Unexpectedly, both populations are highly involved in HF cycle regulation, mainly based on the activation and repression of specific Wnt elements, which translated into a prolongation of anagen and shortening of telogen phases, although with some nuances. Although the specifics may be interesting for further studies, this analysis replicates the plethora of research performed on the relevance of Wnt pathway, among others, in the correct functioning of HF.

C3 and C5 populations are also associated with HF, and share transcriptomic similarities with C1, and to some extent with C2. C3 is clearly involved not only in the production of specialised ECM, as C1/C2, but also in its modulation and degradation, with markers like tenomodulin or kallikrein 4. Therefore, we could coin it as "HF ECM regulatory fibroblast", although the lack of clear markers may hinder determining its proper functions.

C5, in contrast to C3, shows a wider variety of functions, like Wnt and non-canonical TGF-β signalling, vascular and neural modulation and putative immune functions. It also expresses some striking markers, like KRT9, a palmoplantar keratin that is likely to be expressed in other skin structures, yet to be studied. Due to the lack of studies to compare the location of its markers, studying this population could be of relevance to bridge the single-cell knowledge to *in situ* biological applications. Regarding its putative functions, the most probable one could be involved with it being DP or DS in telogen phase. However, this is a big assumption without clear proofs and therefore further supporting evidence is necessary.

The three last interesting populations are D1, D2 and E1 which, despite being very similar and consisting of a relatively small proportion of fibroblasts, they display an incredibly diverse range of functions, being active in neural- and vascular-related processes, immune signalling, TGF-β and Wnt signalling. The high correspondence between mouse and human populations may indicate that these populations, despite not being mayor or associated with main axes, seem to be relevant for both organisms.

Despite their similarities, we do observe certain aspects that make these two populations different. For instance, D1 has an increased expression in neural-related genes (11 vs 8) and even more in vascular-related genes (11 vs 4); whereas D2 has a more apparent expression genes reported to be located in, or interacting with, HF (8). Additionally, each population has certain specific gene families that are more relevant to each of the populations, like collagens and FMO genes for D1 population, and lipid metabolism, caveolae-related genes, or ITGA4B6 complex-associated genes for D2.

This putative separation of functions may imply that both D axes populations seem to be relevant for skin appendage homeostasis, probably by interaction with vasculature, nerve endings–e.g. the lanceolate complex, other sensory terminals–, and even the APM (Martino et al., 2020). Due to the increased number of genes associated with HF, D2 may be more linked to HF functioning, or maybe a population directly associated with HF structures such as ORS or hair bulge; whereas the increased number of nerve- and vascular-associated genes may confer D1 a secondary role to these supporting structures around the HF.

Focusing on the nerve aspect of the D axis populations, specially of D1 but also of D2, is the increased number of markers linked to that structure may be a sign of not only association to nerve, but even to be part of nerve structures, like endoneural and epineural fibroblasts.

In classical literature on these fibroblasts, such as Richard et al., 2012, 2014, common fibroblast and/or perivascular markers are used, including *CD34*, *NG2*, *PDGFRB* or *NES*; which are not sufficiently specific to discriminate any fibroblast population from this analysis. A recent paper published by (Chen et al., 2021a) uses single-cell to determine the heterogeneity of peripheral nerve cells and differentiates epineural fibroblasts, expressing *Sfrp2*, *Dpt*, *Pcolce2*, *Adamts5*, *Sfrp4*, *Prrx1*, *Comp*, *Ly6c1*, from endoneural fibroblasts, which express *Sox9*, *Osr2*, *Wif1*, *Cdkn2a/b*, *Abca9*, *Plxdc1*, *Apod*. A more in depth analysis on DEGs of these populations reveals that endoneural fibroblast express a large set of markers that are also markers of D1 population, as well as D2: *Apod*, *Col15a1*, *Crispld2*, *Abca9*, *Angptl7*, *Matn2*, *Sox9* and *Gfra1*, among others. Interestingly, looking at the markers of perineural fibroblasts, some of them are A or C ECM or ECM modulatory markers–*Pi16*, *Bgn*, *Aspn*, *Gpx3*, or *Col14a1*–and others are more expressed in A4 population–*Clec3b*, *Pcolce2*, *Ccdc80*, *Sfrp4*, *Fbln1*, *Cd248*, *Gpx3*, *Sema3c*–.

Therefore, D1 fibroblasts–and A4 possibly–are likely to be these nerve fibroblast populations. As such, we could term D1 as "endoneural fibroblasts". On the other hand, the D2 population, despite its relevance of HF markers, we cannot ascertain a clear function. Thus, we coin the term "appendage multimodal fibroblast".

Lastly, E1 population due to the reduced number of relevant markers, and its diffuse set of functions, cannot be easily classified. The expression of IGF-related genes as well as other markers involved in proliferation, could be helpful to title this population as "proliferative multimodal fibroblast".

12.4 TABLES AND FIGURES

Table 13: **Comparison of Shin 2020 and Joost 2020 DP and DS populations** For each pair of clusters between Joost et al., 2020 and Shin et al., 2020, the top 50 genes were selected (based on Supplementary Material from each paper) and the Jaccard index was calculated. The overlapping markers for each compared pair are also shown.

Shin / Joost	DS1	DS2	tDP	aDP
CTS1	4.17 | Ly6a, Rnase4, Eln, Plpp3	11.11 | Mgp, Mfap5, Igfbp7, Igfbp4, Dpep1, Pmepa1, Tagln, Ntn1, Omd, Acta2	3.3 | S100a6, Lgals3, Aspn	1.01 | Corin
CTS2	12.36 | Col1a1, Mylk, Col5a2, Ramp1, Dcn, Sparc, Tnmd, Postn, Col6a1, Col1a2, Col5a1	7.53 | Col12a1, Col11a1, Igfbp2, Tagln, Cryab, Cdc42ep3, Acta2		1.01 | Fbn2
DP1		1.01 | Prss12	5.62 | Crabp1, Col23a1, Notum, Tppp3, Bmp4	6.38 | Rspo4, Sfrp1, Rspo3, Thsd7a, Mgst1, Emb
DP2			1.41 | Malat1	1.3 | Sostdc1
DP3			2.17 | Crabp1, Pappa2	12.36 | Inhba, Chodl, Sfrp1, Rspo3, Ptprz1, Rspo2, Ebf1, Sostdc1, Tmem176b, Mgst1, Steap1
DP4		1.45 | Prss12	3.23 | Col23a1, Bmp4	1.45 | Tmem176b

13

GENERAL DISCUSSION

Single-cell methods, specially scRNAseq, have opened a new approach to the study of cell and tissue heterogeneity. In this work, we have virtually recapitulated all dermal fibroblast heterogeneity into 15 human and 17 mouse populations, each of them classified with a set of robust markers, and in the case of human fibroblasts, with putative functions within the homeostatic skin. With these populations and their respective markers, we can extrapolate putative functions in other organs, tissues or even conditions, to shed some light on how gene pleiotropism shapes cell heterogeneity.

One of the hallmarks in fibroblast studies is the determination of pan-tissue fibroblasts, that is, which specific fibroblasts have a shared transcriptomic profile across all tissues, or determining which is that shared profile. Buechler et al., 2021 recently determined two main populations in their pan-organ study, present both in mouse and human organs: a *Pl16*$^{+}$ fibroblast, which is the main producer of the stromal ECM, and a *COL15A1*$^{+}$ fibroblast associated with BMs. In fact, collagen XV has been classically studied as one of the collagens secreted by fibroblasts to attach them to the BM, and thanks to their uncommonly large number of GAG binding sites, it links several CS and HS chains to keep a porous network underlying the BM (Bretaud et al., 2020).

Looking at the fibroblast populations described in the analysis, *Pl16*$^{+}$ / *Pi16*$^{+}$ and *COL15A1*$^{+}$ / *Col15a1*$^{+}$ populations exist in both organisms. In humans, *Pl16* is expressed by A1, A3 and A4 populations, the "canonical" reticular ECM producers; whereas *COL15A1* is expressed by B4, C1, D1 and E1 populations, which are mainly involved in structures that have BM, or which interact with structures that have them, such as blood vessels and nerves. In mice, *Pi16* is expressed by x1, y2, y3, z1 and z2 fibroblasts, which correlate with ECM-producing A1/A3/A4 human populations from different layers; and *Col15a1*, which is expressed by y2, y4, y5, w2 and x1 populations, some of them–y5, w2–linked to human *COL15A1*$^{+}$ populations.

Nerves also harbor several fibroblast types, divided into endoneural and perineural fibroblasts. These types, mentioned in Section 12.3 of the Discussion, and which have beeen studied using scRNAseq by (Chen et al., 2021b), replicates this Pi16/Col15a1 "universality": the perineurial fibroblasts are *Pi16*$^{+}$ and express a list of ECM-related markers such as *Dpt*, *Pcolce2*, *Fbln1*, *Fbln5*, *Col14a1*, *Bgn*, *Aspn*, *Lox* and *Loxl1*; whereas endoneurial fibroblasts are *Col15a1*$^{+}$ and express *Lama3*, *Angptl7*, *Pdpn*, components of the BM itself or which bind to the BM (Kaji et al., 2012; Pozzi et al., 2017).

Another scenario where fibroblasts are of relevance is in the context of highly prevalent and studied diseases, such as cancer and fibrotic illnesses.

Regarding cancer, cancer-associated fibroblasts (CAFs) are one of the key components of the tumour stroma. They are extensively studied as helpers in the development of a proper tumour microenvironment related to its malignancy, even before it is converted into a malignant tumour (Lockwood et al., 2003). Although research in CAFs shows a vast set of types, two main axes are widely described: *ACTA2*^{hi}*IL6*^{lo} myofibroblastic CAFs (myCAFs) which adopt a more pro-fibrotic profile, and are probably involved as a host response against tumour progression; and *ACTA2*^{lo}*IL6*^{hi} immune CAFs (iCAFs), which may be more tumour-promoting by the expression of specific chemokines related to the malignancy (Han et al., 2020).

Despite the heterogeneity of CAFs, different bibliographic records using single-cell methods have already been published to study the heterogeneity in different tumours and organisms. One recent paper published by (Elyada et al., 2019) performed a cross-species single-cell analysis on iCAFs and myCAFs in pancreatic cancer, revealing specific markers of each population. Interestingly, there is no clear association between human dermal fibroblast population markers and CAF markers. For instance, myCAFs are governed by markers expressed by C axis populations–*INHBA*– , *CALD1*, *BGN*, *ACTA2*, *POSTN*, *GRP*, *MMP11*, *EDNRA*–some of those, specifically *ACTA2* and *POSTN* are more upregulated in fibrotic processes. On the other hand, iCAF markers are expressed by B axis populations, as well as other axes, principally A axis. Traditional B markers like *CD74*, *APOE*, *SOD2*, *PNRC1*, *CXCL1* or *CXCL2* showcase the partial immune profile of iCAFs, although A axis markers like *FBLN2*, *CCDC80*, *FSTL1*, *ADH1B*, *CPE*, *HAS1*, *FBLN1*, *MFAP5*, *CRABP2* and *PPIC* may support the additional role of these CAFs maintaining the tumour microenvironment.

The second, and most relevant scenario studied in fibroblast-related diseases is fibrosis. Dermal fibrosis or skin inflammatory diseases related to fibrosis have been widely studied, and several single-cell datasets have already been published, with a variety of results. For instance, Deng et al., 2021 discovered one population in keloid expressing *POSTN*, *COL11A1* and *ACTA2*. Clearly, this population refers to DS in human samples, which is also described by Gao et al., 2021 to be increased in psoriasis.

On the other hand, other authors report the overexpression of certain markers that are not exclusive of one fibroblast population, or expansion of more than one population at the same time. For instance, Gur et al., 2022 observed an upregulation *COL1A1*, *COL3A1* and *COL5A1*, expressed by A and C axes, which goes in line with traditional studies where increased collagen expression is reported; but also C axis-predominant *BGN* and *POSTN*. *GPX3*, and *MGST1*, genes expressed to protect against stress, and markers or A4 and B4 populations in general, were also upregulated. Additionally, two *MYOC*$^{+}$–B4 marker–populations were found to be present in SSc samples, one of them specific to the disease. Therefore, in this scenario, not only are markers of C1 specific to fibrosis but markers of A1, A4 and B4 populations are also relevant.

Similar findings are reported by Vorstandlechner et al., 2021, who discovered an exclusive cluster in samples of scarring tissue that expresses *COL1A1*, *COL3A1*, *COL5A1/2* and *BGN*, as in Gur et al., 2022, but also *COL14A1*, *FN1* and *PCOLCE*, mainly expressed by A1/A3 populations, and *BGN* and *OGN*, expressed by C axis populations.

Lastly, Tabib et al., 2021 and Buechler et al., 2021 discovered enriched populations with specific markers, *PRSS23*–expressed by C2, A4 and E1–and *LRRC15*–expressed mainly in C axis–respectively. The *PRRS23*$^{+}$ population also expressed other relevant markers like *TNC*–expressed by A2, C3 and B3–, *ADAM12*–C1 and C3–, *SFRP4*–A4 and D–, *COL10A1*–A1–, and FNDC1–A2, A4 and C1–. Therefore, these two studies show that C and A4 populations may be highly involved in the development of fibrosis.

Therefore, looking at the heterogeneity of the markers and their respective populations, it is clear that there is not a clear fibroblast population that differentiates into myofibroblast and is responsible for the fibrotic process. Thus, we consider certain hypotheses that could explain the combined action of more than one fibroblast population.

Firstly, we recall that A axis populations express a vast array of markers such as *ADA*, *DPP4*, *CILP*, *CTHRC1*, *SOSTDC1* or *SLPI*, showing a marked antifibrotic potential. Therefore, it is likely that either A1/A4 are not myofibroblasts precursors *per se* or that, in order to acquire this myofibroblastic potential, the antifibrotic signature is inhibited. It is likely that, due to the abundance of axis C markers, axis C populations, and especially the C1 DS population, are the precursors of the myofibroblasts in the dermis. In this scenario, despite A1/A3/A4 populations being the main ECM producers, *COL1A1* and *COL3A1* expression may be induced as an activation step during myofibroblastic differentiation.

A second hypothesis is that although most of the papers studying dermal fibrosis focus on fibrosis as a whole, the underlying diseases are different. Therefore, it is possible that the probability of one population being pro-fibrotic can be context-dependent, and thus more studies are necessary to unravel disease-specific patterns. This hypothesis can be partially contested in cases where more than one pro-fibrotic population is expanded, like in Gur et al., 2022.

Lastly, it is possible that in contexts where more than one pro-fibrotic population arises, there are, in fact, different myofibroblasts, each one executing independent functions, as with dermal fibroblasts in homeostasis. This effect has already been studied in pro-inflammatory contexts. A cross-tissue study in four pro-inflammatory diseases–rheumatoid arthritis, inflammatory bowel disease, interstitial lung disease, and Sjögren's syndrome–using scRNAseq revealed two main fibroblast clusters expanded in all inflamed tissues (Korsunsky et al., 2022). The clusters were (1) *CXCL10*$^{+}$*CCL19*$^{+}$ immune-interacting fibroblasts, which could be similar to the B2/B3 population, and (2) *SPARC*$^{+}$*COL3A1*$^{+}$, which could be similar to A1/A3/A4 or C1 fibroblasts. The authors state that both populations serve specific roles, (1) being involved in the T cell, and (2) perivascular niches respectively. In a similar fashion, each fibroblast population contributing to the myofibroblastic populations could serve specific functions in that environment. For instance, C axis populations could be involved in the vasculature, or in promoting ECM deposition by A1/A4-derived myofibroblasts, and B4-derived myofibroblasts could be the regulators of the immune response, or the fibrotic process by itself.

The analysis of pro-fibrotic fibroblasts is one example showing the relevance of requiring previous knowledge of fibroblasts in homeostasis to make informed decisions about the fibroblasts affected or involved in disease. However, even when implementing diverse datasets to obtain that information, the analysis from this thesis suffers from certain limitations that have to be studied and future work that can be done to address some of these issues and to improve the quality of the analysis.

We see that the relevance to making informed decisions about the fibroblasts being affected by disease requires previous knowledge of fibroblasts in homeostasis. However, even when implementing diverse datasets to obtain that information, this analysis suffers from some limitations that have to

be studied, and future work that can be done to improve the quality of the analysis and the impact of the work in general.

Firstly, one big assumption of this study is that the selected cells during the scRNAseq analysis are indeed fibroblasts. All cells are selected based on a specific gene signature, including *COL1A1*, *PDGFRA*, *LUM* and *DCN*, which are already reported fibroblast markers in different scRNAseq studies. However, it has already been mentioned in section 2.2 of the Introduction that fibroblast definition exclusively on markers may limit the characterisation by excluding putative fibroblasts and including non-fibroblastic cells. It is true, however, that during the selection of major skin cell types–e.g. keratinocytes, endothelial cells, perivascular cells, etc.–, in which UMAP and clustering are used, all cells within the cluster assigned as "fibroblasts" tend to appear together instead of being scattered in different clusters, which might be an indicator of similarity and cohesion in major functions.

It could also be argued that certain fibroblast types, especially minor or "bridge" types, could be a result of doublets of other populations, or simply other cell types that have been co-isolated during the library preparation step. For instance, immune types B1 and B2 could be specific immune cell types being bound to fibroblasts. However, we believe that the core transcriptome of all populations being replicated in several datasets with several sample preparation protocols is strong a sign of these populations being their own and not a result of an artefact during library preparation.

Another limitation of this study is the lack of proper DS/DP characterisation in all hair cycle stages. The majority of HF cells analysed in human and mouse samples are captured during the anagen state, and thus cells in the catagen or telogen stages are shadowed by those in the anagen state. Therefore, it is likely that mouse and human fibroblast heterogeneity only reflects the partial heterogeneity of fibroblasts. Although it is true that Joost et al., 2020 provide samples in telogen state, the number of cells is not sufficiently high, at least compared to other datasets like Boothby et al., 2021; Shin et al., 2020; Shook et al., 2020 for the unbiased algorithm to recognise them as different entities. Therefore, further studies in human and mouse samples enriched for HFs in non-anagenic stages should be necessary to ensure full capture of HF fibroblast heterogeneity.

Fibroblast populations, and other key cell types in countless publications, are partially defined based on their presence in the UMAPs and other 2D DR methods, as shown in Figures 42 and 48. Although these methods are useful to have a first glance of the cell type heterogeneity and "structure" in a sample, they suffer from several distortions that may give the illusion of different cell types being similar because of their closeness in the UMAP, and vice-versa (Chari et al., 2021). In fact, we do observe that, depending on the dataset, certain populations like B4 and E1 appear to be extremely close, when in fact they are different axes.

To avoid falling into believing cluster similarity as a reality, we performed additional analyses, like PAGA, to ensure that population-population similarities were in hand with observations based on UMAPs, and also performed the population robustness analysis to check that obtained populations were not an artefact of the marker-to-population algorithm. In future iterations of the analysis, we will implement a similar robustness analysis in the population-to-marker algorithm to ensure that the whole pipeline is robust. Additionally, we could complement UMAP and PAGA with other DR methods like t-SNE, PHATE or FA2, and also by plotting different results of UMAP with other hyperparameters. However, due to the large number of datasets involved, the implementation of several methods may become impractical, and the result will probably remain unchanged.

Focusing on Chapter 12 where we discussed the putative fibroblast functions based on robust markers, there are several technical and biological limitations to this process. The main limitation is the bias in the process of marker selection. Due to the large sets of genes and their expression diversity across datasets, but also the lack of a proper metric to state the relevance of each marker in a fast way, the use of the A ~ B > C notation as a semiquantitative option is good to meet that need but is created based on a biased view on the UMAP and dot plot of the expression of each gene. This method ends up being inconsistent if multiple genes are assessed in a row, since the criteria to assign a population as more or less relevant to the expression of the gene is not completely accurate.

Secondly, several studies, especially the ones regarding cytokine signalling, were performed on cell lines, where experimental conditions can be variable, and false positives may occur (Viswanathan et al., 2021). Even in more "robust" cases studied in humans or in animal models, many of the conditions studied were not in skin or in homeostasis. For instance, markers associated with TGF-β signalling could be studied in cardiac myofibroblasts, or markers associated with vasculature could be studied in tumour angiogenesis. Therefore, the reported function of a marker can be exclusive of the system of study, and its extrapolation into dermal fibroblasts may result in the attribution of a false function.

Additionally, many of the selected markers are highly pleiotropic, and only some, or even none of their functions could be actually specific to dermal fibroblasts in homeostasis. Therefore, there is an implicit bias in the selection of markers and assignment of their individual functions to a collective function determined by several markers in each fibroblast population. In order to reduce this bias, and also the bias associated to the A ~ B > C notation, a specific function was assigned to a fibroblast if at least three or four markers with a similar function were reported.

This analysis can be also strengthened by performing validatory experiments of the retrieved population markers using immunofluorescence assays, similar to the ones performed for fibroblast axes in Figure 36. Considering that (1) most markers expressed by fibroblasts are also expressed by other non-fibroblastic populations, (2) many relevant fibroblast population-specific markers are not exclusive of a population and (3) protein expression does not correlate with RNA transcription levels, a specific approach to this characterisation should be performed. To address (1), at least one putative fibroblast marker, like PDGFRA, should colocalise with the rest of the markers. To perform the analysis, ideally, 4 channels should be used: one channel for a fibroblast marker, two channels for two markers that are expected to be coexpressed by the same population, and one marker of a similar population, to assess that there is no colocalisation of markers and, thus, they represent separate populations. Although arduous and idealised, this procedure would guarantee that the populations described robustly by scRNAseq are, in fact, real.

A second type of analysis that could be performed to gain more insight about the location of the fibroblasts is single-cell spatial sequencing. Between the two families of spatial sequencing, that is, (1) methods based on high-plex RNA imaging and (2) spatial barcoding, the most suitable ones would be the high-plex RNA imaging methods. The main drawback of spatial barcoding methods is their low resolution which, for fibroblasts specifically, has been observed to be insufficient to recognise different types (Thrane et al., 2023). Therefore, the best strategy to follow would be to use sets of markers specific to each fibroblast population, as well as markers of other structures–HF, SG, keratinocytes, endothelial and perivascular cells, nerves, etc.–to determine their identity and location.

Domcke et al., 2023 recently commented on the different forms in which cells can be classified. Within the 6 different forms, the most common two are the "wet" approach in which cell types are defined by their biological function and morphological features, and the "single", molecular approach, in which arbitrarily large transcriptomic differences between samples, or cells, are used to justify the presence of different cell types, in a similar fashion to the use of operational taxonomical units to differentiate microbial species. The main drawback of the "single" approach exposed by Domcke et al., 2023 is the methods' sensitivity and, more importantly, the inter-dataset heterogeneity. The advantage of the analysis developed during this thesis is that it overcomes this main drawback and, if it were validated through the previously proposed methods, it could describe the cell types in both the "wet" and the "single" approach, which thus imply a complete dermal fibroblast annotation.

Nonetheless, despite the limitation of this analysis and the foreseeable work, we believe that it already serves a big purpose: providing an unbiased atlas of fibroblast heterogeneity in homeostatic conditions. We believe this purpose to be paramount for the proper characterisation of diseased samples. Although in many datasets, diseased samples are collected and processed jointly with control samples, the intra-dataset variability, as well as any other confounding variables affecting the control dataset may provide biased results.

This effect can be overcome both with proper ground truth samples and by acknowledging the scalability issues from large-scale sequencing studies. As Lähnemann et al., 2020 mentions, and as we criticised in the reanalysis of Reynolds et al., 2021 described in Chapter 11, single-cell methods result in high-throughput results that are usually difficult to analyse. Whereas the throughput of experimental methodologies continues to scale, the scalability of computational frameworks does not follow the same trend, and the time and ability to perform an in-depth analysis required to understand such complex data become a bottleneck that is usually subjected to the fast paces of academic publishing.

Therefore, even if computational resources were on par with the amount of data generated, several other factors may lead to an over publication of false positives that then the scientific community builds subsequent research on, including overlooked bad data quality, lack of insight during the analysis and absence of understanding of the central assumptions under single-cell datasets and methods, many of them due to the lack of human and time resources to perform such analyses. In fact, the current scientific publication system, and in general, the biased competitivity of the academia leads to this false positive bias (Smaldino et al., 2016). Therefore, more effort should be allocated to performing robust analyses of existing datasets, instead of producing new large atlases without a prior research question.

In that sense, we believe the atlas of dermal fibroblasts to serve several purposes. First, it will cater to the researchers working on dermal fibroblasts to compare the populations obtained during their analysis with the ones already described from our study, either to ensure that fibroblast heterogeneity in control datasets is replicated and also to understand the differences in markers and populations of non-control samples. Second, and probably the most important one, this analysis already reveals one key aspect of the research about dermal fibroblasts, as well as other fibroblasts: the dogmatic concept of fibroblast as a cell type producing ECM in the surrounding stroma is extremely limited, and the fibroblast should be understood not only as a cell type, but a family of cells with extremely diverse functions revolved around maintaining the structure and general homeostasis of the stromal space and the cells and structures that lay within it.

Part VI

CONCLUSIONS

CONCLUSIONS

1. Human and mouse dermal fibroblasts have been characterised using a semi-supervised, robust algorithm. Human fibroblasts are divided into 15 populations grouped in 5 major axes (A-E), and mouse fibroblasts are divided into 17 populations grouped in 5 major axes (v-z).
2. Analysis of robust markers shows that fibroblasts from each human axis are separated by function: A axis is related to the production and regulation of ECM; B axis is related to immune surveillance, chemotaxis, maturation and regulation; C axis is related to specialised hair follicle populations; and D and E axes are involved in multiple supportive functions to other fibroblasts.
3. Comparison with previous literature shows that major mouse fibroblast axes (x-z) are separated not by function but by location in the skin layers: fibroblasts from x axis are located mainly above the hypodermal adipocyte layer, fibroblasts from y axis are located within the hypodermis, and fibroblasts from z axis are located below the panniculus carnosus. Fibroblasts from axes v and w are likely to be associated with dermal adnexa.
4. Human and mouse characterisation is highly robust and encompasses virtually all fibroblast heterogeneity within the dermal skin. Thus, with certain exceptions like fibroblasts from telogenic hair follicles, hypodermis or other adnexa, it is highly unlikely that new fibroblast populations will be discovered after the addition of newer datasets.
5. Mouse-human fibroblast marker comparison reveals vague similarities across axes, although detailed analysis using robust markers shows that only specific populations have an acceptable similarity, possibly due to the function vs location differentiation. These populations are mainly within ECM-producing and hair follicle-associated axes (A, C). Immune-related axis, B, shows almost no relevant similarities with mouse counterparts.
6. Individual datasets are highly heterogeneous and, thus, a combination of a large number of them is necessary to overcome the interdataset heterogeneity and obtain a robust characterisation of populations. Additionally, analysis of individual datasets alone may lead to a biased characterisation due to inherent dataset biases.

Part VII

APPENDIX

A

SUPPLEMENTARY TABLES

Table S1: **Computation time of a standard pipeline based on the number of cells.**

N cells	Filter	Log1p	Normal.	PCA	Harmony	bbknn	Neighbors	FS	UMAP	Leiden	DEGs	Total
913	1,7	0,0	0,0	4,5	6,6	25,1	0,2	6,6	2,7	0,0	7,4	**55,8**
939	1,4	0,0	0,0	2,7	6,7	4,5	0,2	6,9	2,7	0,0	5,4	**31,3**
933	0,9	0,0	0,0	1,3	9,2	2,1	0,2	6,8	2,6	0,1	6,9	**31,2**
1,959	2,2	0,1	0,0	6,5	72,9	9,0	0,4	9,8	4,5	0,1	8,2	**114,7**
1,965	2,0	0,1	0,0	5,8	94,5	15,7	0,6	19,3	10,1	0,3	24,1	**173,6**
1,963	1,6	0,1	0,0	5,8	213,0	16,4	0,7	26,2	16,6	0,4	26,8	**308,9**
4,978	3,3	0,5	0,2	14,4	131,2	45,9	1,1	13,5	10,1	0,3	13,6	**235,6**
4,985	1,1	0,1	0,0	6,4	103,4	14,8	1,1	13,4	12,4	0,7	18,1	**172,2**
4,992	1,2	0,1	0,0	9,5	378,7	22,5	1,3	27,0	22,5	0,9	24,8	**489,2**
9,998	2,6	0,6	0,2	20,8	741,5	85,2	27,3	20,0	19,7	2,4	35,9	**957,6**
9,990	2,2	0,6	0,2	26,6	659,4	84,8	1,7	29,0	24,4	2,0	34,3	**866,9**
9,987	2,6	0,6	0,2	22,8	499,8	91,6	1,7	36,2	14,3	0,9	22,6	**694,7**
19,998	2,6	1,0	0,4	17,0	509,8	87,0	3,5	27,3	8,6	2,4	19,0	**682,2**
19,995	2,0	0,4	0,2	13,6	318,0	68,7	3,5	26,8	8,4	2,2	18,8	**463,5**
19,991	2,0	0,4	0,2	13,3	450,0	65,2	3,4	26,7	8,3	2,2	19,9	**592,6**
50,000	3,6	1,1	0,4	26,3	339,7	173,4	8,1	49,8	23,0	7,5	39,4	**674,4**
50,000	3,7	1,1	0,4	28,1	464,5	171,8	8,1	50,8	23,4	11,1	39,3	**804,2**
50,000	3,7	1,1	0,4	25,2	514,6	171,9	8,0	49,9	23,1	8,1	41,8	**849,9**
100,000	6,4	2,2	0,8	47,8	516,7	354,5	15,3	86,9	54,5	25,2	68,2	**1182,4**
100,000	6,3	2,1	0,8	47,6	533,3	355,0	15,8	86,9	55,3	25,3	67,3	**1199,4**
100,000	6,4	2,1	0,8	47,7	549,8	354,4	17,3	86,6	55,4	46,9	72,4	**1243,5**
200,000	11,6	4,3	1,7	91,2	869,1	737,2	31,8	158,5	124,7	113,0	150,1	**2300,0**
200,000	11,7	4,3	1,7	91,0	892,4	735,6	32,7	159,0	127,6	115,1	151,8	**2329,7**
200,000	12,2	4,3	1,7	92,4	1067,6	737,9	32,2	156,2	128,1	90,3	150,8	**2480,4**
400,000	22,5	8,6	3,3	182,8	1634,1	1571,5	71,3	368,0	303,8	305,9	358,0	**4842,8**
400,000	23,0	8,6	3,3	180,9	1615,1	1572,0	71,1	367,6	307,5	377,5	367,1	**4907,2**
400,000	23,0	8,6	3,3	186,6	1621,5	1570,0	70,6	364,3	297,8	456,7	388,1	**5003,9**

Table S2: **Genes associated with ECM.** The list of genes is divided into the following categories: Collagens, Small Leucin Repeat Proteoglycans (SLRPs), elastic fibres and other elements.

Collagens		
COL1A2	A1 ~ A3 ~ A4 > C1 ~ C2 ~ C3	*Collagen I α chain 2*. Main fibrillar collagen in dermis. Composed of three helices expressed by COL1A1, COL1A2 and COL1A3 (Henriksen et al., 2019).
COL3A1	A1 ~ A3 ~ A4 ~ C1 ~ C3 > C2	*Collagen III α chain 1*. Fibrillar collagen abundant in the dermis, and usually found together with collagen I (Nielsen et al., 2019a).
COL4A2/4	A2 ~ B ~ C > D	*Collagen IV α chains 2/4*. Component of the lamina densa in the BM (Barbieri et al., 2014). It provides a network of fibrils to scaffold other elements like laminin 332.
COL5A1/2	C3 > C1 ~ A2 > A3 ~ A4	*Collagen V α chains 1/2*. Nucleating collagen essential for fibrillation of type I and III collagens (Leeming et al., 2019). It is predominantly found in the papillary dermis (Nauroy et al., 2017) and around HFs (Chanoki et al., 1988).
COL6A1/2/3	A2 > C1 ~ C3 > A3	*Collagen VI α chains 1/2/3*. The trimer is formed by combining A1/A2/A3 chains. It is found between the BM and the ECM and is a target of a vast array of ligands. In the DEJ, it assembles anchor cells to the ECM through interaction with integrins (Kaur et al., 2015).
COL6A5	A2 > B3	*Collagen VI α chain 5*. It may form a trimer with A1 and A2 chains. In the dermis, α5 chain is mainly located in the papillary area, intertwined with other collagen fibres although α1-3 can also be found around blood vessels (Sabatelli et al., 2011), in reticular dermis and hypodermis (Theocharidis et al., 2017). This collagen attaches KGF, PDGF and MMP1/2/3/8/9, among other elements (Freise et al., 2009).
COL7A1	C1 ~ A2 ~ C3	*Collagen VII α chain 1*. Component of the anchoring fibrils in the BM, which bind collagen I and III fibres (Barbieri et al., 2014). Its N-terminal domains also bind collagen IV and laminins 5/6 (Mortensen et al., 2019a). In mouse HF, it is located at the BM surrounding DP and ORS (Tsutsui et al., 2021).
COL8A1/2	D1 > A1 ~ D2	*Collagen VIII α chains 1/2*. Non-fibrillar collagen found in the cornea. It is found in proliferating vessels in other organs, forming a hexagonal lattice structure (Sutmuller et al., 1997).
COL9A3	D1 > D2	*Collagen IX α chain 3*. Nucleating collagen, predominantly found in the reticular dermis (Nauroy et al., 2017). It stabilises the collagen network by binding matrillin and proteoglycans, through which it controls the fibre diameter. It is commonly found in cartilaginous chondrocytes (He et al., 2019).
COL11A1	C1 > C3	*Collagen XI α chain 1*. Minor fibrillar collagen that regulates collagen I and II fibrillogenesis as nucleating point. It is usually found in tendons and cartilage (Luo et al., 2019; Sun et al., 2020). KO of COL11A1 in HF decreased adipogenic and chondrogenic differentiation capacity, while the upregulation of COL11A1 was associated with increased differentiation capacity and improved wound healing (Ahlers et al., 2022).
COL12A1	C1 > A1 ~ A4 ~ A3	*Collagen XII α chain 1*. FACIT collagen. It binds GAG chains and interacts with decorin, COMP, fibromodulin and tenascin. It stabilises the collagen I fibrils and prevents their crosslinking (Mortensen et al., 2019b) It is more abundant in the papillary dermis (Nauroy et al., 2017) but is also expressed in the reticular dermis (Marinkovich et al., 1992), and around HFs (Sasaki et al., 1996).

COL13A1	A2	*Collagen XIII α chain 1*. Non-fibrillar transmembrane collagen that binds to fibronectin, perlecan, vitronectin, type IV collagen and $\alpha 1\beta 1$ integrin (Siebuhr et al., 2019). In fibroblasts, it is bound to focal adhesions (Hägg et al., 2001) and collocalises with E-cadherin, suggesting a role within the adherens junction (Theocharidis et al., 2017). It is found at the DEJ and surrounding blood vessels and nerves (Peltonen et al., 1999). In mouse HF, it is almost exclusively located in the interface BM between the DP and the hair gem above the DP (Tsutsui et al., 2021).
COL14A1	A3 > A1 ~ A2	*Collagen XIV α chain 1*. FACIT collagen. It is more abundant in the reticular dermis (Nauroy et al., 2017).
COL15A1	E1 ~ C1 ~ B4 ~ D1	*Collagen XV α chain 1*. Non-fibrillar collagen that contains several noncollagenous interruptions, GAGs and chondroitin sulfate chains, that is located at the BM of microvessels (Manon-Jensen et al., 2019a), nerves, adipocytes (Saarela et al., 1998), and HFs (Hagg et al., 1997).
COL18A1	A2	*Collagen XVIII α chain 1*. Non-fibrillar collagen with GAGs and heparan sulfate chains (Halfter et al., 1998). It contains a C-terminal endostatin function, thereby inhibiting angiogenesis. It is usually found in vascular and epithelial BMs (Pehrsson et al., 2019). It is also abundant in the DEJ (Bonnet et al., 2017).
COL21A1	C1 > A2	*Collagen XXI α chain 1*. FACIT collagen associated with other fibrillar collagens, serving as a bridge. It may play a role in blood vessel assembly (Kehlet et al., 2019a).
COL23A1	A2	*Collagen XXIII α chain 1*. Transmembrane collagen present in several tissues. It is located in different epithelia, but its function is unknown (Kehlet et al., 2019b).
COL24A1	C2	*Collagen XXIV α chain 1*. Fibrillar collagen with two collagenous and three non-collagenous domains. It is highly expressed in bone, although it is not specific (Nielsen et al., 2019b). It has been observed in mouse skin during development (Koch et al., 2003).
COL26A1	E1	*Collagen XXVI α chain 1*. It contains several collagenous domains and three non-collagenous domains. It is mainly found in testes and ovaries (Manon-Jensen et al., 2019b).
COL28A1	D1	*Collagen XXVIII α chain 1*. It has been detected around terminally-differentiated Schwann cells and Merkel cells (Grimal et al., 2010).
SLRPs		
ASPN	C2 > C1 ~ C3 > C5	*Asporin*. It belongs to the SLRP family, although it is not a proteoglycan. It binds collagen (competing with DCN) and calcium, promoting ECM mineralisation (Kalamajski et al., 2009) In keloids, it prevented the 3D remodelling of the collagen mesh (Liu et al., 2021b).
BGN	C3 > A3 > C5	*Biglycan*. Belongs to the SLRP family. It contains two GAG chains, either chondroitin or dermatan sulfate (Roughley et al., 1989). They interact with different collagens (I, VI) via their protein or the GAGs (Johnstone et al., 1987; Schönherr et al., 1995).
DCN	A1 ~ A4 > A3 > B4 ~ D1	*Decorin*. It contains SLRPs and binds to collagen I fibrils and a large set of targets: fibronectin (Winnemöller et al., 1991), vitronectin to regulate collagenase expression (Huttenlocher et al., 1996).
FMOD	C2 > rest	*Fibromodulin*. Member of SLRP family. It has four keratan sulfate chains linked in the protein. It binds collagen I and inhibits collagen I and III fibrillogenesis in vitro.
PODN	A1 ~ B4 > rest	*Podocan*. Belongs to the SLRP family. Similar to other members, it increases hydration and viscoelasticity. It inhibits smooth muscle cell migration (Hutter et al., 2013). It also binds collagen I (Shimizu-Hirota et al., 2004).

POSTN	C3 > C1 > A2	*Periostin*. It contains several domains that bind to collagen I, fibronectin and tenascin C to facilitate ECM formation. Its been observed to localise at the DEJ, papillary dermis and around HFs (Yamaguchi, 2014).
OGN	C2 ~ D1 > A1 ~ A4 ~ C3 ~ C1	*Osteoglycin*. Belongs to the SLRP family. It is widely expressed in the skin and may be involved in fibroblast proliferation. In neural cells, it may induce neurite growth (Deckx et al., 2016).
PRG4	A4 > A1	*Proteoglycan 4*. It is classically studied as a lubricant molecule, which binds to hyaluronan and fibronectin, preventing attachment of cells to surfaces, especially in cartilage areas (Eguiluz et al., 2015; Rhee et al., 2005). Binding to CD44 internalizes it and targets the inflammasome, reducing IL1/8β expression. It also affects TLR-mediated cascades (e.g. NF-kB), showing antiinflammatory potential (Richendrfer et al., 2020). In oral fibroblasts, TGF-β induced PRG4 expression in vitro (Stähli et al., 2016). On the other hand, PRG4 reduced α-SMA expression in synoviocytes, showing antifibrotic potential (Qadri et al., 2020).
Elastic fibers		
ELN	A3 ~ C3 > A4 ~ A1	*Elastin*. It is one of the components of the elastic fibres, which completes after the binding of GAGs, fibrillin, heparan sulfate and othe ECM components (Gheduzzi et al., 2005). It is ~1000 times more flexible than collagen; and its degradation releases several DAMPs.
EMILIN2	A4 > A1	*Elastin microfibril interfacer 2*. They are bound to microfibrils in different organs by binding to fibrillin (Doliana et al., 2001; Schiavinato et al., 2016). It also binds to pro-apoptotic receptors, inducing cell death (Mongiat et al., 2010).
FBN1	A4 ~ A1	*Fibrillin-1*. Glicoprotein that serves as a structural component of calcium-binding microfibrils (Lee et al., 2004). These fibres can be associated with elastic fibers and elastin-independent networks (Jensen et al., 2016). It can also interact with BMPs or LTBPs to regulate TGF-β signalling, or other cell surface proteins and proteoglycans (Jensen et al., 2016). SSc patients show a decrease in FBN1 (Wipff et al., 2010). Philippeos et al., 2018 observe an increase of Fbn1 in $Dlk1^{+}Sca1^{+}$ pre-adipocytes.
MFAP2	C1 ~ C3 > A2 ~ C2	*Microfibrillar-associated protein 2*. Microfibril-associated glycoprotein that does not bind integrins (Gibson et al., 1999). It interacts with TGF-β and sequesters it into the microfibrils (Weinbaum et al., 2008).
MFAP5	A4 ~ A1 > C1 > A3 ~ D2	*Microfibrillar-associated protein 5*. Microfibril-associated glicoprotein that binds $\alpha 5\beta 3$ integrin (Gibson et al., 1999). It is a key component in the organisation of elastic fibres by interacting with fibrillin-1/2 (Penner et al., 2002).
Other proteins		
ACAN	C1 > C3	*Aggrecan*. Large proteoglycan with a hyaluronan core and chondroitin and keratan sulfate branch chains. It withstands high-pressure loads and is usually associated with cartilage (Roughley et al., 2014)
COCH	C2	*Cochlin*. ECM glycoprotein (Hynes et al., 2011). It is abundant in the cochlea and the eye and may play a role in cell adhesion and mechanosensation (Picciani et al., 2007; Robertson et al., 1997). No established function has been determined in the skin.
COMP	A2 > C3 > A1 ~ A3	*Cartilage oligomeric matrix protein*. It contains five glycoprotein subunits which bind different collagens (I, II, IX) (Farina et al., 2006). It is also found to bind XII and XIV collagens at the anchoring plaques, where it attaches mainly collagen I fibrils into the vicinity of the anchoring plaques to stabilise the DEJ structure (Agarwal et al., 2012).

DPT	T1 ~ C3 > C2 > B2 ~ B3	*Dermatopontin*. It is found to hold membranes in the intercellular junction. It binds to integrins and modulates the size and arrangement of collagen fibrils (Okamoto et al., 2006).
EFEMP1	B4 > A4 ~ B2	*EGF-containing fibulin-like extracellular matrix protein 1*. It contains several EGF-like repeats and a fibulin-type C-terminal domain (Sun et al., 1998). (Solé-Boldo et al., 2020) describe this marker within the pro-inflammatory population. It was also localised to the bulge area in HFs, according to Takahashi et al., 2020.
HSPG2	B4 ~ A1 > rest	*Perlecan*. The core protein binds three GAGs (heparan sulfate or chondroitin sulfate). It binds to a number of ECM and matrisome components such as laminin, collagen IV, V, VI, XI, elastin and FBLN2. It also binds lipids with an LDLR-like domain and Wnt morphogens. It promotes angiogenesis by binding to VEGF and VEGFR2 (Hayes et al., 2022). In mouse HF, it is located within a mesh-like deposition within the BM adjacent to the DP (Tsutsui et al., 2021).
MXRA5	C2 > C1 ~ E1 > C5	*Matrix Remodeling Associated 5*. MXRA5 is reported as a secreted glycoprotein, and it containes seven leucine-rich repeats and 12 immunoglobulin-like C2-type domains related to perlecan. TGF-β upregulated its expression; thus it may show a putative role in fibrosis (Poveda et al., 2016). It contributes to EMT in pancreatic cancer, and increases vimentin expression (Peng et al., 2023).
NPNT	C?	*Nephronectin*. Deposited by the HF BM bulge, also located in the DEJ, with putative integrin binding activity (Fujiwara et al., 2011).
PDPN	A1 ~ A3 ~ A4 > A2 > B1 ~B3	*Podoplanin*. Mucin-type transmembrane protein. It binds to CLEC2 in platelets, which increases aggregation, and in immune cells (NK cells, neutrophils and DCs), which increases motility (Kerrigan et al., 2009; Seymour et al., 2016; Sobanov et al., 2001). PDPN$^+$ cells have been found throughout the dermis, although with a higher presence in papillary dermis (Korosec et al., 2019). A subset of PDPN$^+$ cells form a reticular network near lymph nodes to facilitate leukocyte migration and antigen presentation via CCL19, CCL21 and IL7 (Fletcher et al., 2015).
SDC1	C3 > C2 ~ C5 > C1	*Syndecan-1*. Transmembrane heparan sulfate proteoglycan. It shows many interactions, including collagens, fibrin, tenascin, fibrillin, vitronectin; MMPs, ADAMTSs and other proteases; integrins, BMPs, Wnt morphogens, PDGF, VEGF, TGF-β, ApoB/E, cytokines, and complement proteins (Stepp et al., 2015). Thus, the number of functions is very diverse and includes neuron guidance, ECM organisation, wound healing, etc. (Stepp et al., 2015).
TNC	A2 ~ B3 ~ C3 > B2 ~ C2	*Tenascin C*. Glycoprotein with several domains that, after binding fibronectin, upregulate MMP expression (Tremble et al., 1994). It also binds integrins, perlecan, PDGF, TGF-β, and includes a large range of functions, mainly cell division and motility (Giblin et al., 2014). It favours Treg migration, proliferation and immunotolerance (Rüegg et al., 1989).
TNXB	A1 ~ A4 > A3	*Tenascin XB*. It interacts with type I, III, V, XII and XIV collagens, as well as decorin, integrins, etc. It has been observed to have opposite locations and functions to TNC. It modulates cell adhesion, may restrict keratinocyte and fibroblast movement, and may induce cell differentiation (Valcourt et al., 2015).
VIT	D1 ~ D2 > B4 ~ A2	*Vitrin*. It is similar to COCH and may be associated with cell adhesion and neural development (Whittaker et al., 2002).

Table S3: **Genes associated with ECM modulation.** The list of genes is divided into the following categories: MMPs and TIMPs, and other elements.

MMPs and TIMPs		
MMP1	B1	*Matrix metallopeptidase 1*. It can degrade collagens III > I, II, VIII, aggrecan and entactin (Cabral-Pacheco et al., 2020). It also activates MMP-2 (Kahari et al., 1997).
MMP2	A1 > A4 ~ A3 > A2 > C3	*Matrix metallopeptidase 2*. It can degrade collagens I, IV, V, VII, X, XI, fibronectin, elastin, laminin, and vitronectin. It also activates proMMP 9 and 13 (Cabral-Pacheco et al., 2020).
MMP3	B1	*Matrix metallopeptidase 3*. It can degrade collagens IV > V, IX, X, XI, aggrecan, vitronectin, fibronectin and laminin. It also activates proMMP 1, 8, 9 and 13 (Cabral-Pacheco et al., 2020; Kahari et al., 1997).
MMP11	C3 > C1 ~ A2	*Matrix metallopeptidase 11*. Acts like MMP3.
MMP16	E1 ~ C3 > C2 ~ C5	*Matrix metallopeptidase 16*. It can degrade collagen III mainly, fibronectin and laminin. It also activates proMMP 2 and 13 (Cabral-Pacheco et al., 2020).
MMP27	A1 ~ A3 ~ A4	*Matrix metallopeptidase 27*. It may not be secreted, and its functions are unknown (Cominelli et al., 2014).
TIMP1	B4 ~ E1 > B3 ~ A4 > rest	*Tissue inhibitor of metalloproteinases 1*. It inhibits MMPs 1, 2, 3, 7, 8, 9; as well as ADAM 10 (Cabral-Pacheco et al., 2020). It is shown to mediate angiogenesis through interaction with β1 integrin and CD63 (Jung et al., 2006; Toricelli et al., 2013). Acts as a key modulator of inflammatory processes (Schoeps et al., 2022). It increased photoprotection against UVB, by inhibiting TNFα and other pro-inflammatory processes (Yokose et al., 2012).
TIMP2	A1 ~ A4 ~ C2 > C3 ~ A2 ~ A3	*Tissue inhibitor of metalloproteinases 2*. It inhibits MMPs 1, 2, 3, 7, 8, 9, and 14; although MMP3 and 7 inhibition are weaker than TIMP1. It also binds ProMMP-2 and certain integrins (Cabral-Pacheco et al., 2020).
TIMP3	E1 > A1 ~ A4 ~ B4	*Tissue inhibitor of metalloproteinases 3*. It inhibits MMPs 1, 2, 3, 7, 8, and 9; as well as ADAM 10 and 17. It also binds ProMMP-2/9 and VEGFR2 (Cabral-Pacheco et al., 2020).
ADAM		
ADAMTS4	B1 > B3 > B2	*ADAM metallopeptidase With thrombospondin Type 1 Motif 4*. Degrades chondroitin sulfate proteoglycans such as ACAN or VCAN (Kelwick et al., 2015).
ADAMTS9	C1 ~ A2 > C2 ~ C3 ~ C5	*ADAM metallopeptidase With thrombospondin Type 1 Motif 9*. Degrades chondroitin sulfate proteoglycans such as ACAN or VCAN (Kelwick et al., 2015).
ADAMTS18	C1	*ADAM metallopeptidase with thrombospondin Type 1 Motif 18*. Located in the dermal papilla (Hagner et al., 2020). No clear function has been elucidated yet, but loss of function of this gene is related to bone, hematological, central nervous system and eye diseases. Suppression of this gene is also observed in several tumours; thus, it may act as a tumour supressor gene (Wei, 2014).
ADAMTSL1	A	*ADAMTS like 1*. It lacks MMP and disintegrin domains of ADAMTS proteins but contains other domains. Therefore, like all other ADAMTSLs, it lacks ADAMTS catalytic capability. It may play a role in the skin, but it is not known (Hirohata et al., 2002).
ADAMTSL3	D2 > B4	*ADAMTS like 3*. It binds fibrillin-1 fibers (Sengle et al., 2012).
ADAMTSL4	D2 > rest	*ADAMTS like 4*. It binds fibrillin-1 and accelarates its biogenesis (Gabriel et al., 2012).

ADAMTSL5	D2	*ADAMTS like 5*. Besides binding fibrillin-1, it also binds fibrillin-2. It has an NTR module that is present in SFRPs, C3, PCOLCEs and TIMPs (Bader et al., 2012). It is observed in ORS in HFs (Higgins et al., 2011). In melanocytes, it acts as an autoantigen in psoriasis (Bergen et al., 2020).
Other proteins		
AEBP1	A1 > A3 ~ C3 > A4	*AE binding protein 1*. Belongs to the carboxypeptidase A family. It regulates adipogenesis and smooth muscle differentiation and may be involved in wound healing and bad prognosis of certain cancers (Liu et al., 2018; Majdalawieh et al., 2010; Ohno et al., 1996).
ANGPTL7	D1	*Angiopoietin like 7*. It diminished the expression of FN1, COL1A1, COL4A1, COL5A1, MYOC and VCAN and increased the expression of MMP1 (Comes et al., 2010).
ANTXR1	A2 > rest	*ANTXR Cell Adhesion Molecule 1*. It is a type I transmembrane protein and a tumour-specific endothelial marker of myofibroblastic CAFs and other cancers (Huang et al., 2020; Kieffer et al., 2020). It can bind collagen VI (Nanda et al., 2004), and lack of it leads to the accumulation of collagens II and VI (Besschetnova et al., 2015). It also acts on the cytoskeleton by binding it to the ECM, and it is actively involved in wound healing (Gu et al., 2010).
CD44	A2 ~ A3 ~ B1 > A4 ~ B3 > rest	*Cluster of Differentiation 44*. - Also expressed on immune cells. It binds to hyaluronan chains and enhances T cell migration in vitro (Bertheim et al., 1994). It binds, apart from HA, fibronectin, proteoglycans, HB-GFs (like EGFR (D2)), and collagens I, IV and XIV (Bennett et al., 1995; Fujimoto et al., 2001; Ishii et al., 1993; Jalkanen et al., 1992; Kawashima et al., 2000). It is implicated in cell proliferation, migration, invasion and differentiation. It may maintain the levels of collagen I, N-cadherin and fibronectin in homeostasis (Tsuneki et al., 2015) and wounding, avoiding its accumulation (Govindaraju et al., 2019). It shows a high binding affinity to hyaluronic acid types, and most of its functions are mediated by their weight. For instance, low molecular weight hyaluronic acid mediates CD44-induced IL6, CXCL1 and CXCL2 expression in fibroblasts (Vistejnova et al., 2014). It acts as a negative regulator of TGF-B and PDGFRB (Porsch et al., 2014). It regulates αSMA (ACTA2) gene expression through several pathways, including p38 (Wang et al., 2019).
CHADL	C2	*Chondroadherin like*. It binds collagen *in vitro*, inhibits its fibrillogenesis, and KO increases collagen II and ACAN deposition (Tillgren et al., 2015).
CHST15	C2	*Carbohydrate sulfotransferase 15*. Transmembrane sulfotransferase that acts on specific chondroitin sulfate residues may influence the storage of certain proteases in mast cells and may affect osteoblast differentiation (Mizumoto et al., 2021). It also reduces the presence of chondroitin sulfate proteoglycans (Kai et al., 2017).
CTSK	A1 ~ C2 > A3 ~ A4 ~ C3	*Cathepsin K*. Decreases COL1A1 expression (Soundararajan et al., 2021), and catabolises elastin and collagen (Kondo et al., 2022). It also activates MMP-9 in certain environments (Christensen et al., 2015).
HAS2	B3 > A2 ~ B1	*Hyaluronan synthase 2*. It synthesises large hyaluronic acid chains necessary for tissue repair, and leukocyte homing (Sussmann et al., 2004), and protect against stressors (Wang et al., 2014). In myofibroblasts, HAS2 expression is increased, which increases CD44 ECM levels and results in fibrosis (Li et al., 2011).
KLK1	D1 ~ E1 > D2	*Kallikrein 1*. Activates kininogen-1 and MMP2, and MMP9. Kininogen-1 is activated into bradykinin, a strong vasodilator (Nauroy et al., 2020).

LOX	A1 > A3 ∼ A4 ∼ C1	*Lysyl oxidase*. Crosslinks collagen and elastin chains, which is necessary for stabilisation and maturation of ECM (Csiszar, 2001), although it is also related to ECM stiffness in ageing (Langton et al., 2012).
LOXL1	A1 ∼ A3 ∼ A4	*LOX like 1*. Acts like LOX, although it may have a more relevant function in elastin crosslinking (Liu et al., 2004).
LOXL2	C3 ∼ A2 > C1 ∼ C5	*LOX like 2*. Acts like LOX, but it also contains several SRCR domains that may be related to the innate immune system (Sarrias et al., 2004)
LOXL4	A	*LOX like 4*. Acts like LOXL2.
P3H2	C2	*Propyl 3-hydroxylase 2*. This enzyme catalyses the 3-hydroxylation of collagen proline residues. It has a high affinity to collagen IV, and lesser to collagen I (Tiainen et al., 2008).
P4HA2	A	*Propyl 4-hydroxylase subunit α 2*. This is a key enzyme in collagen synthesis by catalysing the 4-hydroxylation of proline residues (Wilhelm et al., 2023).
P4HA3	C3	*Propyl 4-hydroxylase subunit α 3*. Acts like P4HA3.
PCOLCE	A1 ∼ A3 ∼ A4 > rest	*Procollagen C-Endopeptidase Enhancer*.Binds to the pro-collagen form of collagen during its synthesis and may induce the cleavage of the pro-peptide by activating BMP1 (Takahara et al., 1994). Overexpression of PCOLCE is a general feature and marker of fibrosis (Lagoutte et al., 2021).
PCOLCE2	A4 > A1 ∼ A3	*Procollagen C-Endopeptidase Enhancer 2*. Acts like PCOLCE in the cleavage of pro-collagen I and II forms (Steiglitz et al., 2002). In adipocytes, it regulates HDL uptake (Xu et al., 2021a).
SPARC	C1 > C3 ∼ A3 ∼ A1 ∼ A4 > A2	*Secreted protein acidic and cystein rich*. Also known as osteonectin. Widely studied in bone, where it absorbs Ca^{2+} and binds it to collagen, PDGF and other ligands (Termine et al., 1981). In skin, it also promotes the production of collagens IV and VII and their accumulation in the DEJ (Nakamura et al., 2022).
TGFBI	D1 ∼ D2 > A2	*Transforming growth factor β induced*. This protein contains several domains and an RGD motif that binds integrin $\alpha v \beta 3$, fibronectin, vitronectin, collagens, fibrinogen and VWF (Ruoslahti et al., 1987). It has been shown to have antiangiogenic properties (Son et al., 2013). It has been shown to bind collagen XII (Runager et al., 2013).
WISP1	C3 ∼ A1	*WNT-inducible-signalling pathway protein 1*. This protein contains several domains, including IGFBP, von Willebrand type C repeats, thrombospondin type 1 repeat and cysteine knot motif. WISP1 has been observed to bind DCN and BGN (Desnoyers et al., 2001) and acts as a pro-mitogenic and pro-survival factor (Venkatachalam et al., 2009).
WISP2	A1 > A4 ∼ A3 > B4	*WNT-inducible-signalling pathway protein 2*. With a similar domain structure as WISP1, WISP2 inhibits the binding of fibrinogen to integrin receptors (Janjanam et al., 2021). It also is related to preadipocyte commitment and PPARG activation (Hammarstedt et al., 2013) and is upregulated in hypertrophic scars (Chaudet et al., 2020).

Table S4: **Genes associated with complement pathway.**

C3	B2 ~ B3 > B4 > A1	*Complement C3*. It is a member of classical and alternative pathways. It activates into C3a anaphylatoxin and C3b, which binds to C2a and C4b to activate C5. C5 splits into C5a and C5b. C5b, together with C6, C7, C8 and several C9 units, form the membrane attack complex (MAC) (Rutkowski et al., 2010). In synovial fibroblasts, C3 exposure within the cell primes them for future immune responses, which may lead to chronic inflammation (Afzali et al., 2021). Adventitial fibroblasts may release vesicles containing several components, including C3, which induces inflammatory reprogramming in macrophages (Kumar et al., 2021).
C6	B2 ~ B4 > B3	*Complement C6*. It is a member of the MAC, which causes cell lysis (Rutkowski et al., 2010).
C7	B2 ~ B4 > B1 ~ B3	*Complement C7*. It is a member of the MAC, which causes cell lysis (Rutkowski et al., 2010).
CD55/DAF	A4 ~ A1	*Decay-accelerating factor*. It recognises C4b and C3b fragments. Binding to C4 prevents the creation of C2aC4b (classical pathway C3 convertase). Binding to C3b prevents the creation of C3bBb (alternative pathway C3 convertase). Both convertases activate C3; thus, Cd55 reduces the formation of the MAC indirectly (Dho et al., 2018).
CFD	A1 ~ A4 ~ B4 > A3	*Complement factor D*. Cleaves CFB into Ba and Bb. Bb binds to C3b to create the C3bBb C3 convertase (alternate pathway) (Rutkowski et al., 2010). Thus, it favours the formation of the MAC.
CFH	B4 ~ D1 > A1 ~ A3 ~ B2	*Complement factor H*. It inhibits the C3bBb convertase, stopping the amplification loop (Jozsi, 2017). It also interacts with receptors like ITGAM, mediating the adhesion of human neutrophils and enhancing their activity (Losse et al., 2009).
CLU	A4 ~ A1 > B3 ~B2 ~D2 > rest	*Clusterin*. Inhibits MAC formation (Shinjyo et al., 2021)
DCN	A1 ~ A4 > A3 > B4 ~ D1	*Decorin*. Can bind C1q and inhibit the complement system (Krumdieck et al., 1992).

Table S5: **Genes associated with cytokine signalling].** The list of genes is divided into the following categories: CCL - CXCL - CX3C - CXCR cytokines, ACKRs, interleukins, TNFs, and other elements.

CCL - CXCL - CX3C - CXCR		
CCL2	B3 ~ B1 ~ D1 > D2	*C-C Motif Chemokine Ligand 2*. Powerful chemoattractant from bone-marrow-derived monocytes. (Vanbervliet et al., 2002), by binding to CCR2 or CCR4 (Craig et al., 2006). It is also required with CCL5 for epidermal-to-dermal migration of LCs after injury (Ouwehand et al., 2010). Hair plucking led to the release of CCL2 to recruit macrophages which help adjacent hair creation by activation of regenerative signals (Chen et al., 2015b). Among those signals are the activation of HFSCs by TNFα and by Notch signalling from Treg cells (Rahmani et al., 2020). IL6 induces CCL2 expression in fibroblasts and participates in their survival (Liu et al., 2007).

CCL13	D2 > D1	*C-C Motif Chemokine Ligand 13*. Binds to CCR1/2/3/5 and induces chemotaxis of different immune cells, including Th and NK cells, mast cells, basophils, eosinophils and monocyte/macrophages (Mendez-Enriquez et al., 2013). It plays a role in innate immunity by upregulating TLR2/3/4/5, and also in adaptive responses by inducing DC maturation and activation (Mendez-Enriquez et al., 2013). Its expression may be related to CD40 and MHC-II expression (Chiu et al., 2004). CCL13 overexpression in alopecia areata leads accumulation of T cells surrounding the HF (Wang et al., 2021a).
CCL19	B3 > B2	*C-C Motif Chemokine Ligand 19*.It is a chemoattractant of $CCR7^+$ cells, such as DCs, B cells, NK cells and several types of T cells (Laufer et al., 2019; Ohl et al., 2004; Reif et al., 2002; Robbiani et al., 2000) It induces migration, cytokine production and cell growth and differentiation by activation of several pathways, including PI3K, JAK-STAT or Ras/Raf/Erk (Yan et al., 2019) In APCs, CCR7 phosphorylation and internalisation leads to APC migration (Anderson et al., 2015; Tian et al., 2013) In certain tumours, stromal cells expressing CCL19 are responsible for T $CD8^+$ infiltration and antitumorigenic response (Cheng et al., 2018b).
CX3CL1	B3 > B2	*C-X3-C Motif Chemokine Ligand 1*. It is a chemoattractant of T cells, NK cells, monocytes and DCs via CX3CR1; expressed by neural, muscle and immune cells (Johnson et al., 2013; Limatola et al., 2014). It exists in an anchored form, which promotes leukocyte attachment in physiological conditions, and in a soluble form after cleavage by ADAM10/17, in inflammatory conditions (Thelen et al., 2016). IL1β and IFNγ costimulated CX3CL1 expression in lung fibroblasts in vitro and may have a role in pulmonary fibrosis (Isozaki et al., 2011). CX3CL1 expression is also observed in wound healing by promoting macrophage and fibroblast accumulation (Ishida et al., 2008); and in atopic dermatitis, by retaining T cells in the skin (Staumont-Salle et al., 2014).
CXCL1	B1 > B3	*C-X-C Motif Chemokine Ligand 1*. Binds to CXCR2 and acts different cell types–e.g. eosinophils, basophils, macrophages, immature DC and naïve T cells–having the most chemoattractant activity towards neutrophils (Bautista-Hernandez et al., 2017; Murphy, 2007). It also has a marked angiogenic potential, being implicated in tumorigenesis (melanoma cells selectively express CXCL1) (Murphy, 2007). Interestingly, TGF-β reduces its expression, although the inverse might not be true, since CXCR2 KO mice resulted in decreased TGF-β1 and collagen I expression (Zhang et al., 2020b) Its binding to GAGs–heparan, dermatan and chondroitin sulfate–is essential. CXCL1 is attached to GAGs and is liberated by MMPs, and GAGs seem to be necessary to fully activate CXCR2 (Wang et al., 2003). CXCL1 may polimerise in vivo. In fact, homodimerisation may be necessary for full CXCL1 activity (Sawant et al., 2020).
CXCL2	B1 > B3	*C-X-C Motif Chemokine Ligand 2*. Acts similarly to CXCL1 by binding to CXCR2 (Murphy, 2007). CXCL2 bound to ACKR1 may be necessary for neutrophils to perform diapedesis; CXCL1 is also necessary but not sufficient (Girbl et al., 2018). CXCL2 also binds to GAGs and exists in dimeric form (Sawant et al., 2020).
CXCL3	B1 > B3	*C-X-C Motif Chemokine Ligand 3*. Acts similarly to CXCL1 by binding to CXCR2 (Murphy, 2007).

CXCL12	B2 ~ B4 > B3	*C-X-C Motif Chemokine Ligand 12*. Attracts exogenous immune cells such as T cells (Bautista-Hernandez et al., 2017), and endogenous Langerhans cells from the epidermis into the dermis (Ouwehand et al., 2008). It binds PKM2, an enzyme-producing pyruvate at the end of the glycolysis pathway. Activation of CXCR4 or ACKR3 reduces its oligomeric form, enhancing glycolysis and pentose phosphate pathway (Luker et al., 2022) In keloid scars, overexpression of CXCL12 led to TGF-β-induced suppression of DPP4, and CXCR4-mediated immune infiltration (Chen et al., 2021b). Truncation by DPP4 leads to an "inactive" form that only binds to ACKR3 and may inactivate CXCR4 (Elmansi et al., 2022)
CXCR4	C5 > D	*C-X-C Motif Chemokine Receptor 4*. It binds CXCL12 and induces β-arrestin-mediated signalling (Lipfert et al., 2013). It may form a complex with ACKR3 since blockade of either receptor leads to a complete loss of CXCL12-dependent response (Lipfert et al., 2013) Its function is related to stem/progenitor cell migration (Chen et al., 2015b) It was expressed in epithelial cells and eosinophils infiltrating skin after burning. Blocking CXCR4 led to a reduction of eosinophil accumulation in the dermis (Avniel et al., 2006) In HF, it is expressed in DP and ORS cells. Overexpression of CXCL12 from dermal fibroblasts led to a delayed telogen-to-anagen transition (Zheng et al., 2022).
ACKR		
ACKR3	A1 ~ A4 ~ E1	*Atypical Chemokine Receptor 3*. It binds CXCL12 with higher affinity than CXCR4 (Meyrath et al., 2020). It "scavenges" CXCL12 by binding it to the receptor, internalising the complex, degrading CXCL12 and resurfacing ACKR3. This establishes a CXCL12 gradient used for cell type migration (Berahovich et al., 2013; Donà et al., 2013). It is involved in monocyte differentiation via JNK and p38 pathways (Ma et al., 2013) Its function is more related to stem/progenitor survival (Chen et al., 2015a).
ACKR4	A1 ~ A2 ~ A3	*Atypical Chemokine Receptor 4*. Acts as a scavenger receptor for CCL19 and CCL21 chemokines, creating a gradient favouring immune migration within lymph nodes (Bryce et al., 2016). In lymphatic vessels, it is expressed in afferent lymphatic collectors. T cells migrate by adhesion from $CCL21^{+}$ lymphatic capillaries into $CCL21^{dim}ACKR4^{+}$ collectors, where they lift and go with the lymph flow (Friess et al., 2022). In myocardial infarction, ACKR4 expression lead to IL6-mediated EMT (Zhang et al., 2021a).
Interleukins		
IL11RA	B2 ~ B3 > B4	*Interleukin 11 Receptor Subunit Alpha*. Soluble form of the receptor that binds IL11, which binds to gp130 (a transmembrane protein associated with IL6 receptor family) and induces IL11 signalling, even on cells with no IL11 receptor (Lamertz et al., 2018; Lokau et al., 2016). It is involved in the activation of several downstream pathways, including MAPK, JAK-STAT, NF-kB or Akt (Balakrishnan et al., 2013). May be involved in fibrosis since IL11 is identified as TGF-β responsive (Schafer et al., 2017); although in other models IL11 signalling led to an antifibrotic response by endothelial cells (Allanki et al., 2021). In rheumatoid arthritis, IL11 signalling led to fibroblast infiltration, as well as IL8 and VEGF expression, which induced endothelial migration and angiogenesis (Elshabrawy et al., 2018). Interestingly, IL6 is expressed in DP cells and IL11RA in the IRS and ORS; and balding may be related IL11RA/IL6-mediated anagen-to-catagen transition (Yu et al., 2007).
IL15RA	B ~ A1 ~ A3	*Interleukin 15 Receptor Subunit Alpha*. Mediates IL15-related signalling by binding IL15 and presenting it to the effector cell (Jakobisiak et al., 2011).

IL6	B1	*Interleukin 6*.Interleukin expressed at basal levels by many cells. Its expression is induced by LPS and other external agents that bind TLRs, as well as IL1, TNFα, TGFβ and other factors. The induction is mediated by STAT, NF-kB and other pathways. (Paquet et al., 1996; West, 2019). For instance, after epidermal injury, keratinocytes synthesis IL1α, which induces IL6 and IL8 production in fibroblasts (Boxman et al., 1996). It is one of the most readily inducible cytokines and activates the differentiation of B cells to plasma cells, peripheral T cells–Th2, Th17, Treg–, as well as ECM remodelling–the expression of collagen, GAGs–and fibrosis (Duncan et al., 1991; Paquet et al., 1996; West, 2019). It also induces the secretion of more IL6, as well as different chemokines (CCL2, CCL11) and adhesion molecules (ICAM-1, VCAM-1) (West, 2019). It is also associated with immediate response in acute infections (septic shock), inflammation and bacterial infections (Paquet et al., 1996). IL6 expression is necessary for cutaneous wound healing in mice (Gallucci et al., 2000).
IL15	B3 > B	*Interleukin 15*. Pleiotropic cytokine with roles in innate and adaptive responses, mainly activating B, T and NK cells (Lodolce et al., 2002). It also suppresses T cell apoptosis, thus enhancing its function (Malamut et al., 2010). It also induces the production of IL8 and CCL2 in monocytes to attract neutrophils and monocytes to the infection site (Badolato et al., 1997). IL15 is observed to be overexpressed in atopic dermatitis, mainly in fibroblasts, macrophages and DCs (Karlen et al., 2020).
IL32	B3 ~ B1 > B2 ~ A2	*Interleukin 32*. Pleiotropic cytokine with up to 9 isoforms (Gautam et al., 2021), and predominantly expressed intracellularly (Kim et al., 2005). Its role is not completely known, but it has been observed to switch between pro- and anti-inflammatory programs (Heinhuis et al., 2015); create an antimicrobial environment by upregulation of microbidical peptides (Montoya et al., 2014); or induce expression of pro-inflammatory IL1β, IL18 and TNFα cytokines (Alsaleh et al., 2010). Aberrant expression of IL32 has been linked to autoimmune diseases such as rheumatoid arthritis, IBS and type 1 diabetes (Albuquerque et al., 2021) and skin diseases like hidradenitis suppurativa (Thomi et al., 2017) or atopic dermatitis (Chang et al., 2022). It may be related to cholesterol metabolism in liver cells due to the correlative expression of ABCA1 and ABCG1 transporters, as well as LXRα nuclear receptor (Damen et al., 2018).
IL33	B2 ~ B3 > B4 > B1	*Interleukin 33*. Constitutively expressed by cells exposed to the environment (lung, skin, etc.) (Moussion et al., 2008), it is stored in the nucleus and released passively during injury, thus working as an "alarmin" that alerts the immune system after injury (Haraldsen et al., 2009). IL33 activates T, NK, DC and mast cells, as well as basophils, eosinophils and macrophages (Bonilla et al., 2012; Moro et al., 2009; Pecaric-Petkovic et al., 2009; Price et al., 2010). Activated mast cells release chymases and tryptases, which activate the rest of the immune cells directly or indirectly by cleaving IL33 into its active form (Eissmann et al., 2020). Pro-inflammatory IL11 may induce IL33 expression in fibroblasts (Widjaja et al., 2022), and restrain them from switching to a myofibroblastic state (Gatti et al., 2021).
IL34	B2 ~ B3	*Interleukin 34*. Promotes the growth differentiation of monocytes and macrophages through binding to CSF1R (Lin et al., 2008), and to a lesser extent, to syndecan 1 (Segaliny et al., 2015). In the skin, it may be related to LC homeostasis and self-renewal (Wang et al., 2015a), and protection from bacterial infections (Lin et al., 2019b).
TNFs		

CD70	A1 ~ A4	. It is usually found in activated T and B cells (Borst et al., 2005). In APCs, it binds CD27 in T and B cells and induces their activation and differentiation (Han et al., 2016). For instance, in the skin, LCs present CD70 to augment $CD8^+$ T cell presence in the epidermis (Polak et al., 2012). $CD70^+$ CAFs stimulated the migration and frequency of Tregs in vitro (Jacobs et al., 2018).
TNFRSF19	A2	*TNF Receptor Superfamily Member 19*. Target gene of Wnt/β-catenin pathway, it interacts with Lgr5 to inhibit Wnt signalling in intestinal stem cells (Fafilek et al., 2013) and may act via NF-kB signalling (Pispa et al., 2008). In mice, it marks the IFE and infundibulum cells in telogenic HF and hair bulb in anagen; in humans, it marks the basal epidermis (Kretzschmar et al., 2021). In cancer cells, it inhibits TGF-β by associating with TGFBR (Deng et al., 2018).
TNFRSF21	A2 ~ E1 > B2	*TNF Receptor Superfamily Member 21*. Proapoptotic TNF receptor (Dostert et al., 2019). It is implicated in axonal pruning via caspase 6 (Nikolaev et al., 2009). May decrease T cell and B cell proliferation (Liu et al., 2001; Schmidt et al., 2002), although it may also be necessary for their migration to the central nervous system (Schmidt et al., 2002).
TNFSF13B	B2 ~ B3	*TNF Superfamily Member 13b*. Binds to a TRAF subfamily receptor, which activates the non-canonical NF-kB pathway and belongs to the same family as CD40/CD40L (So et al., 2013). It is expressed by several immune cells, such as activated DCs, and it is critical for B-cell maturation (Rickert et al., 2011)
TNFSF14	B1	*TNF Superfamily Member 14*. Member of the lymphotoxin system, a network of LR pairs expressed by T cell activated during viral infection (Dostert et al., 2019). It is expressed by myeloid cells and immature DCs, where it is associated with DC maturation (Albarbar et al., 2015), and can be induced on activated T cells to regulate their proliferation and differentiation (Wang et al., 2001) TNFSF14 induced the expression of MMP9 and IL6 in synovial fibroblasts in rheumatoid arthritis (Pierer et al., 2007)
Other		
RARRES2	B2 ~ B4 > D2 > E1 ~ A1 ~ A3 > B1 ~ D1	*Retinoic Acid Receptor Responder 2*. Predominantly expressed in adipocyte and immune cells (Bozaoglu et al., 2007; Goralski et al., 2007), binding to CMKLR1 may induce adipogenesis and angiogenesis (Henau et al., 2016). Regarding the immunogenic potential, it shows pro- and anti-inflammatory properties depending on cleavage by proteases (Mattern et al., 2014) For instance, it was observed to enhance chemotaxis of immature DCs and monocytes but also reduce the recruitment of neutrophils, and macrophages info inflamed tissues (Cash et al., 2008; Luangsay et al., 2009) It may have a link with insulin resistance (Takahashi et al., 2008), high-fat diet-induced neuroinflammation (Yun et al., 2022), and food intake and metabolism in general (Tavolaro et al., 2015). However, its negative or positive effects on metabolic diseases may depend on its functions in acute versus chronic scenarios (Helfer et al., 2016), and most findings are controversial (Helfer et al., 2018). It may be involved in early psoriatic lesions (Albanesi et al., 2008), although it may be downregulated in advance stage psoriasis by IL17 and IL22 (Banas et al., 2015).
CMKLR1	E1 > A2	*Chemerin Chemokine-Like Receptor 1*. Receptor of adipokine chemerin/RARRES2 and omega-3 fatty acid derived molecule resolvin E1 (Arita et al., 2005; Wittamer et al., 2003).

CSF1	B3 > B > rest	*Colony Stimulating Factor 1*. Similar to IL34, binds CSF1R and regulates macrophage differentiation. However, its range of action is systemic, compared to IL34, which is focused on the central nervous system–maintenance of microglia–and the skin (Greter et al., 2012; Nakamichi et al., 2013). Produced by wound fibroblasts to fortify the immune response (Bautista-Hernandez et al., 2017). Its absence significantly delays wound closure (Li et al., 2016). During UV-induced damage, LC regeneration depended on CSF1-expressing infiltrated neutrophils (Wang et al., 2015b).
MDK	C3 > C1 > C2 ∼ C5	*Midkine*. Cytokine and growth factor that binds to proteoglycan and non-proteoglycan receptors–including α4/6 and several β integrins, LRP, NOTCH2 or N-syndecan– (Kurosawa et al., 2001; Muramatsu et al., 2000) and regulates cell proliferation, adhesion, survival, differentiation and inflammatory response (Kurosawa et al., 2001; Muramatsu et al., 2004; Stoica et al., 2002; Weckbach et al., 2014). It is also implicated in cancer progression (Filippou et al., 2019). In the immune response, it recruits neutrophils and macrophages (Weckbach et al., 2014) and then contributes to vessel regeneration by endothelial cell migration (Horiba et al., 2000) and angiogenesis through VEGFA expression induction (Hao et al., 2013). This latter effect is also induced in the hypoxic environment (Weckbach et al., 2012). In the skin, it may be involved in the synthesis stimulation of collagens I and III, TGFB1 and MMP2 (Yamada et al., 1997). Localised in the DP (Rendl et al., 2005) and assumed to be expressed in ORS (Woo et al., 2022) and sebaceous glands (Iwashita et al., 1999), it may induce autocrine signalling (Rezza et al., 2016), and has been reported to induce proliferation and migration of HF cells (Rendl et al., 2005).
SOCS3	B3 > B1 > B2 ∼ B4 > D2 ∼ D1 ∼ A2	*Suppressor Of Cytokine Signaling 3*. It is expressed as a response to the binding of pro-inflammatory molecules–e.g. IL6, IL12, LPS, TNFα or IFNs–inhibiting the effector pathways activated by these molecules (Carow et al., 2014; Yin et al., 2015). For instance, it binds JAK bound to IL6 receptor, or STATs, inhibiting its action (Yin et al., 2015). Despite acting as an immunomodulator, several pathologies have been found to show increased levels of SOCS3, such as IBD, rheumatoid arthritis, allergic diseases or atopic dermatitis; as a developed resistance to this inhibition (Horiuchi et al., 2006; Yin et al., 2015). Basal keratinocyte and HF SOCS3 expression may also be necessary for skin wound repair (Linke et al., 2010). Besides its immunomodulatory role, it also binds leptin and insulin receptors, and similar increases in SOCS3 have been associated with resistance and weight gain (Pedroso et al., 2018). TGF-β induces the suppression of SOCS3 to ensure its fibrotic potential in fibroblasts (Dees et al., 2020). Similarly, IL6 signalling may evade inhibition by activating ERK1/2 instead of STAT3 in dermal fibroblasts in vitro (Luckett-Chastain et al., 2012).
VCAM1	B3 > B2	*Vascular Cell Adhesion Molecule 1*. Usually expressed by endothelial cells after cytokine stimulation–i.e. IL1, TNFα, IL4, IL3–, which promotes the adhesion of lymphocytes, monocytes, eosinophils and basophils (Broide et al., 2014; Mantovani et al., 1998). Functions of VCAM1 in other contexts are understudied. In muscle, it is observed that VCAM1 is necessary for satellite cells to communicate with other satellite cells and immune cells and may affect myofibril growth (Choo et al., 2017). In post-infarction hearts, its expression is correlated with increased lymphangiogenesis (Iwamiya et al., 2020).

Table S6: **Genes associated with Wnt pathway.** The list of genes is divided into the following categories: Component of canonical signalling, component of non-canonical signalling, component of canonical and non-canonical signalling, positive modulator, negative modulator and ambivalent modulator.

Component of canonical signalling		
APCDD1	A2 > C2 ~ C5	*APC down-regulated 1*. Acts like PCOLCE but may also bind heparin and has a less dispersed expression. It also binds BMP1, another procollagen proteinase, enhancing its activity (Baicu et al., 2012).
FZD1	C2 > C3	*Frizzled 1*. Receptor of Wnt proteins in the canonical pathway, activated by WNT3A, WNT3 and WNT1; and probably not by WNT4, WNT5A, WNT5B, WNT6, WNT7A or WNT7B (Gazit et al., 1999).
WNT2	A4 > E1	*Wingless type 2*. Canonical Wnt ligand. It may promote ECM fibre deposition in fibroblasts and may contribute to keloids with an aberrant expression (Cai et al., 2017).
WNT10B	A4	*Wingless type 10B*. Canonical Wnt ligand. Like Wnt10a, it may promote DP proliferation and maintenance in vitro (Ouji et al., 2012). It is expressed in the dermal condensate at E14.5, and is necessary for HF generation and homeostasis (Chen et al., 2012). It also inhibits adipose tissue development (Longo et al., 2004).
Component of non-canonical signalling		
DAAM1	D2 > D1 > rest	*Disheveled-associated activator of morphogenesis 1*. It is a planar cell polarity Wnt pathway member by forming a complex with DVL, which activates cytoskeletal remodelling via activation of RhoA and Cdc42 (Lai et al., 2009).
DAAM2	C2 ~ A2 > C5	*Disheveled-associated activator of morphogenesis 2*. Potentiates Wnt signalling by creating Fzr/Axin/Dvl protein aggregates (Lee et al., 2012a, 2015). It is also a regulator of actin nucleation and elongation, filopodia formation and podocyte migration (Schneider et al., 2020).
FZD2	D1 > C1	*Frizzled 2*. Receptor of Wnt, with possible functions in the non-canonical pathway. It binds WNT5 and may regulate EMT and metastasis in certain contexts (Gujral et al., 2014).
PTK7	A2 ~ C2 ~ C3 > C5	*Tyrosine-protein kinase-like 7*. Ptk7 acts as a co-receptor in the non-canonical PCP Wnt pathway. It also binds LRPs and suppresses canonical Wnt signalling (Lhoumeau et al., 2011).
WNT5A	C5	*Wingless type 5A*. Non-canonical Wnt ligand. It is expressed in the dermal condensate at E14.5, and is necessary for HF generation and homeostasis (Chen et al., 2012). It is also expressed in keratinocytes and DP cells, contributing to their dedifferentiation, and in psoriatic epidermis, probably lowering the threshold to IFN response (Romanowska et al., 2009). In dermal fibroblasts, binding to Fzd3 promotes their adhesion (Kawasaki et al., 2007). It is expressed in DP and IRS cells (Lim et al., 2012).
WNT10A	C2	*Wingless type 10A*. Non-canonical Wnt ligand. It is expressed in the dermal condensate at E14.5, and is necessary for HF generation and homeostasis (Chen et al., 2012). Concretely, it is expressed in anagen DP and IRS and not in telogen (Reddy et al., 2001), (Lim et al., 2012). It also plays a key role in wound closure in mice (Wang et al., 2018a).
WNT11	A2 > B4	*Wingless type 11*. Non-canonical Wnt ligand. It is expressed in the dermal condensate at E14.5, and is necessary for HF generation and homeostasis (Chen et al., 2012). It is expressed in DS and ORS cells (Lim et al., 2012).

Component of canonical and non-canonical signalling		
CTHRC1	A1 ~ A4 > A3 > C3	*Collagen triple helix repeat containing 1*. Binds Wnt5A and ROR1/2 as well as FZRs, inducing canonical and noncanonical Wnt activation (Mei et al., 2020). Elevated levels also attenuate the TGF-β pathway by induction of proteasomal degradation of Smad2/3 complex (Myngbay et al., 2021).
FZD6	A1	*Frizzled 6*. Receptor of Wnt that may bind WNT4 (Lyons et al., 2004), shown to regulate mostly non-canonical pathways, but sometimes canonical pathways too (Corda et al., 2017).
FZD7	E1 > rest	*Frizzled 7*. Acts like PCOLCE, but may also bind heparin and has a less dispersed expression (Steiglitz et al., 2002). It also binds BMP1, another procollagen proteinase, enhancing its activity (Baicu et al., 2012).
Positive modulator		
LGR4	C1 > A	*Leucin-rich repeat-containing G-protein coupled receptor 4*. It is a regulator of the Wnt pathway. It binds RSPOs and forms a complex with ZNFR3, a Wnt repressor. LGR4/RSPO/ZNFR3 complex thus keeps ZNFR3 sequestered and allows Wnt pathway activation to occur (Ruffner et al., 2012).
LGR5	A1 > A4	*Leucin-rich repeat-containing G-protein coupled receptor 5*. Acts like LGR4.
MYOC	B4	*Myociclin*. It is a modulator of the actin cytoskeleton, expressed thoroughly across skin organs, and may act as a chaperon against cellular stressors (Anderssohn et al., 2011). Its function in the trabecular network is relevant since mutations in the gene lead to glaucoma (Borras, 2014). MYOC interacts with FZDs, SFRPs and WIF1, and may activate the Wnt pathway to modulate the actin cytoskeleton (Kwon et al., 2009).
RSPO1	A2 > A3 ~ B3	*R-spondin 1*. Activator of the canonical Wnt signalling pathway by acting as a ligand for LGR4-6 receptors and forming a complex with RNF43 and ZNRF3, which suppress Wnt signalling (Szenker-Ravi et al., 2018).
RSPO3	C5 > A2	*R-spondin 3*. Acts like RSPO1 as an indirect Wnt activator. Together with RSPO2, it is involved in HF dermal stem cell proliferation Hagner et al., 2020. Its expression is also upregulated in mouse DP (Tsutsui et al., 2021).
RSPO4	C2 > C3 ~ A2	*R-spondin 4*. Acts like RSPO1 as an indirect Wnt activator.
Negative modulator		
APCDD1L	A4	*APC down-regulated like 1*. It is predicted to be involved in the negative regulation of the Wnt pathway (Otsuki, 2005).
AXIN2	A2 > C2	*Axin 2*. It is a member of the APC|Axin|GSK3β complex that inhibits canonical Wnt through β-catenin inhibition. Concretely, β-catenin induces AXIN2 transcription, thus being a self-regulatory component of the pathway (Jho et al., 2002).
DACT1	D2	*Disheveled binding antagonist of β-catenin 1*. It represses canonical Wnt signalling by stabilising CTNNB1 (β-catenin gene), promoting its membrane localisation and inhibiting GSK3. DACT1 is induced by TGF-β (Esposito et al., 2021)
DKK1	A4 ~ A1	*Dickkopf 1*. Inhibits canonical Wnt signalling by sequestering LRP5/6 and Wnt and promoting its internalisation (Ahn et al., 2011).
DKK2	C2 > C3 ~ C5 ~ A2 > rest	*Dickkopf 2*. Acts like DKK1 as a Wnt repressor.
DKK3	C1 > A2 ~ D1 ~ D2 > E1	*Dickkopf 3*. Acts like DKK1 as a Wnt repressor.

NKD1	C5 > A2	*Naked cuticle 1*. It is a Wnt inhibitor which acts by either binding Dvl, thus impeding the necessary GSK3 inhibition for β-catenin translocation; or by binding to β-catenin and impeding its translocation to the nucleus (Angonin et al., 2013).
NKD2	A2 > B3 ~ C3	*Naked cuticle 2*. Acts similarly to NKD1 (Zhao et al., 2015).
NOTUM	C2?	*Notum, Palmotoleoyl-protein carboxylesterase*. NOTUM is a Wnt inhibitor. It hydrolyses the palmitoleoyl residue of the protein, which induces its multimerisation (Zhang et al., 2015).
SFRP1	E1 ~ C2 > C5 > C1 ~ C3	*Secreted frizzled related protein 1*. It binds several Wnt ligands and induces Wnt signalling suppression; and can also act as a suppressor by directly binding β-catenin (Baharudin et al., 2020).
SHISA3	D1 > D2	*Shisa Family Member 3*. Located at the ER, it inhibits Wnt and FGF pathways by hindering the transportation of the receptors to the surface (Furushima et al., 2007; Hedge et al., 2008). Its functions are yet to be studied, although it is observed to possibly suppress tumorigenesis (Chen et al., 2014). In HF, it may be expressed during telogen and anagen to support Shh signalling (Laron et al., 2018).
WIF1	A2 ~ A3 > A1 ~ C5	*WNT inhibitory factor 1*. Binds Wnt ligands and sequesters them, impeding their binding to the receptors (Gajos-Michniewicz et al., 2020).
Ambivalent modulator		
CAV1	D1 ~ D2 > rest	*Caveolin 1*. Encodes a scaffolding protein of the caveolae, small invaginations in the cell membrane involved in signal transduction, lipid metabolism (Vargas et al., 2002), as well as inducing T-cell proliferation, neutrophil extravasation and leucocyte adhesion (Bouzin et al., 2007; Marmon et al., 2009; Ohnuma et al., 2007). May regulate Wnt-mediated signalling by sequestering β-catenin into caveolae (Galbiati et al., 2000; Gao et al., 2022; Zhang et al., 2020a).
SFRP4	D1 ~ D2 > A4	*Secreted frizzled related protein 4*. Acts similarly to SFRP1 binding Wnt ligands. However, in some contexts like cancer, it is observed that it facilitates the association of β-catenin with TCF_4, which induces gene expression (Liang et al., 2019).

Table S7: **Genes associated with TGF-β pathway.** The list of genes is divided into the following categories: Positive modulator and negative/ambivalent modulator.

Positive modulator		
BMP7	C5 > D1	*Bone morphogenic protein 7*. Activates secondary TGF-β signalling by binding to its receptor and inducing Smad1/5 phosphorylation (Meng et al., 2016).
DPT	T1 ~ C3 > C2 > B2 ~ B3	*Dermatopontin*. Enhances TGFB1 activity via interaction with decorin (Okamoto et al., 1999).
GDF10	B4 > A1 ~ D1	*Growth differentiation factor 10*. It belongs to the BMP family–it was previously termed BMP3b–. Acts as BMP7.
Negative / ambivalent modulator		
ADAMTSL3	D2 > B4	*ADAMTS like 3*. It regulates collagen deposition and myofibroblast differentiation in heart by inhibition of TGF-β signalling (Rypdal et al., 2022).

ASPN	C2 > C1 ~ C3 > C5	*Asporin*. Inhibits TGF-β by binding to the BMP receptor (Tomoeda et al., 2008) as well as to TGFBR (Kizawa et al., 2005), although in other studies, TGF-β activation by several mechanisms are observed (Li et al., 2019).
BAMBI	D1	*BMP and activin membrane bound inhibitor*. Binds to TGF-β receptor I and impedes the formation of TGFBRI + RII active receptor complex (Huang et al., 2012a). It thus regulates angiogenesis and endothelial homeostasis, and gene KOs show aberrant vascular endothelium (Guillot et al., 2012).
CAV1	D1 ~ D2 > rest	*Caveolin 1*. Encodes a scaffolding protein of the caveolae, small invaginations in the cell membrane involved in signal transduction, lipid metabolism (Vargas et al., 2002), as well as inducing T-cell proliferation, neutrophil extravasation and leucocyte adhesion (Bouzin et al., 2007; Marmon et al., 2009; Ohnuma et al., 2007). It negatively regulates TGFβ signalling by sequestering TGBFR1 from membrane rafts into caveolae (Hwangbo et al., 2015).
CAV2	D2 > rest	*Caveolin 2*. It needs CAV1 presence to form caveolae and might stabilise it, although CAV1 alone can build them (Mora et al., 1999; Scherer et al., 1997). Although acting complementarily, CAV2 might antagonise CAV1 in certain processes such as angiogenesis or regulation of inflammatory processes (Almeida, 2017). Additionally, like CAV1, it acts as a negative regulator of TGF-β signalling (Xie et al., 2011).
CILP	A4 > A1	*Cartilage intermediate layer protein*. May inhibit Smad2/3 phosphorylation by binding to TGF-β, and also may inhibit the pathway by binding to TGFBR (Liu et al., 2020b)
CTSK	A1 ~ C2 > A3 ~ A4 ~ C3	*Cathepsin K*. Decreases TGFB2 expression (Soundararajan et al., 2021).
DCN	A1 ~ A4 > A3 > B4 ~ D1	*Decorin*. Inhibitor of TGF-β. Binds to TGF-β and prevents binding to its receptor (Ferdous et al., 2007). In specific conditions, it releases TGF-β, which leads to its activation (Bierbaum et al., 2017).
FMOD	C2 > rest	*Fibromodullin*. It regulates TGF-βeta activities by sequestering TGF-βeta into the extracellular matrix.(Zheng et al., 2014)
GDF15	A1	*Growth differentiation factor 15*. Ligand of the TGF-β family activates SMAD signalling and other TGF-β-related pathways such as ERK, AKT or PLC (Wang et al., 2021b). It is widely expressed and may be induced by endogenous or exogenous stress signals, showing protective activity in different organs.
HTRA1	C3 ~ C2 > C1 ~ A4 ~ A1 ~ A3	*HtrA serine peptidase 1*. Inhibits TGF-β pathway by cleavage of TGFB1 (Launay et al., 2008; Shiga et al., 2011).
INHBA	C5 > D1 ~ C1 ~ C3	*Inhibin subunit β A*. Inhibitor of TGF-β. Binds to TGF-β and prevents binding to its receptor (Ferdous et al., 2007). In specific conditions, it releases TGF-β, which leads to its activation (Bierbaum et al., 2017).
LTBP2	C3 > C1 ~ C2 ~ E1	*Late TGFβ binding protein 2*. Belongs to the LTBP family. It sequesters TGF-β1 dimer in the ECM. Upon cleavage of LTBP2, the dimer is released and activates the signalling cascade.
MFAP2	C1 ~ C3 > A2 ~ C2	*Microfibril associated protein 2*. Interacts with TGF-β and sequesters it into the microfibrils (Weinbaum et al., 2008).
NOG	A2 ~ C	*Noggin*. It inhibits several BMPs, including BMP7 (Blazquez-Medela et al., 2019).
PPP1R15A	B1 ~ B3 ~ D1 ~ D2	*Protein phosphatase 1 regulatory subunit 15A*. May inhibit TGF-β pathway by recruiting Smad7 and dephosphorylating TGF-βR (Shi et al., 2004). Its deletion induces lung fibrosis in mice (Monkley et al., 2021).

SLPI	A1 ~ A4 > B4	*Secretory Leukocyte Peptidase Inhibitor*. Inhibitor of serine proteases (trypsin, cathepsin, elastase, etc.) from neutrophil and other immune cells is highly expressed in mucous secretions (Nugteren et al., 2021) It may inhibit the production of MMP1 and MMP9 by monocytes (Zhang et al., 1997) and may inhibit the expression of IL1β and IL6 in leukocytes in different tissues (Zakrzewicz et al., 2019). In skin, it promotes wound healing via protease inhibition and suppression of TGF-β (Ashcroft et al., 2000)
SOSTDC1	A3	*Sclerostic domain containing 1*. Antagonist of BMP proteins (Faraahi et al., 2019) also induces reduced TGF-β1 responsiveness (Badshah et al., 2019).
TNC	A2 ~ B3 ~ C3 > B2 ~ C2	*Tenascin C*. Induced by TGF-β signalling (Jinnin et al., 2004), might regulate the availability of TGF-β1 dimer by binding to its latent form and liberating the active homodimer (Aubert et al., 2021).
TNXB	A1 ~ A4 > A3	*Tenascin XB*. It may bind LTBP and, after protease cleaving, LTBP bound to part of TNXB binds TGFBR, therefore modulating TGF-β response to ECM signals (Valcourt et al., 2015).

Table S8: **Genes associated with lipid metabolism.** The list of genes is divided into the following categories: Cholesterol and lipoproteins, eicosanoids and other lipids.

Cholesterol and lipoproteins		
ABCA8	D1 > B4 > C2 ~ D2	*ATP binding cassette subfamily A member 8*. ABC transporter that regulates cholesterol efflux and plays a role in sphingomyelin production in oligodendrocytes (Kim et al., 2013b).
APOC1	B4 ~ B2	*Apolipoprotein C1*. It is present in VLDL and HDL, depending on the fasting condition. It inhibits the binding of LDL and VLDL to their receptors. It inhibits CETP, which transforms HDL into VLDL. It has also been involved in immune modulatory processes (Fuior et al., 2019). In macrophages, APOC1 may induce M2 polarisation (Ren et al., 2022).
APOD	D1 > B4 | E1	*Apolipoprotein D*. It has low similarity to other apolipoproteins but has high homology to RBPs. Also, APOD activates LCAT, an enzyme in HDL that transports peripheral cholesterol to the liver (PMID: 2373967). It may show protective action against oxidative stress preventing lipid oxidation in different organisms (Ganfornina et al., 2008; Muffat et al., 2008), although in humans, it has been shown to accumulate in Alzheimer's disease, atherosclerosis and aged skin (Takaya et al., 2023).
APOE	B2 > B3 | B4 > B1	*Apolipoprotein E*. Apolipoprotein that binds to liver and peripheral cell receptors for catabolism of VLDL and IDL into higher density lipoproteins (Marcel et al., 1983). Different protein versions (E2/3/4) are related to risk and protective factors of immune responses and pathologies like Alzheimer's disease (Husain et al., 2021). In fibroblasts, APOE is upregulated in apoptosis and starvation-induced growth arrest (QUINN et al., 2004). In pancreatic cancer, it induced CXCL1 expression, promoting immune suppression (Kemp et al., 2021). It may also suppress T cell proliferation, and neutrophil activation, and regulate macrophage function (Zhang et al., 2010).
CES1	A3 > A1	*Carboxylesterase 1*. It metabolises many molecules (heroin, cocaine, etc.) into their active forms. In lipid metabolism, it converts monoacylglycerides to free fatty acids and glycerol; or transforms cholesterol ester to free cholesterol (Markey, 2010).

CH25H	B3 > B1 ~ B2	*Cholesterol 25-hydroxylase*. Transforms cholesterol into 25-hydroxycholesterol, which regulates cholesterol homeostasis, and is also involved in inflammatory and immune responses (induction of TNF and IL6), as well as showing antiviral properties (Zhao et al., 2020). In dermis, special T$\gamma\delta$17 cells require 25-hydroxycholesterol for homing in that area (Frascoli et al., 2023). 25-hydroxycholesterol inhibits the expression of pro-IL-1β under pro IL-1β signals such as IFN ot LPS, possibly acting as a negative feedback (Reboldi et al., 2014).
CYP4B1	A1 > B4 ~ D2	*Cytochrome P450 family 4 subfamily B member 1*. P450 family member that oxidises a large set of substrates, including fatty acids, arachidonic acid and cholesterol. Hypoxia, androgens, and other factors activate it (Röder et al., 2023). Retinoic acid induces synthesis of inflammatory eicosanoids in cornea (Ashkar et al., 2004). This synthesis may induce VEGF-derived angiogenesis (Seta et al., 2007).
CYP7B1	B3 ~ B2 > B1	*Cytochrome P450 family 7 subfamily B member 1*. P450 family member that hydroxylates oxysterols and steroids, including 25/27-OH-cholesterol (Stiles et al., 2009). Proinflammatory stimuli from innate immune cells (M0) induce CYP7B1 expression and may regulate immunoglobulin production from B cells (Dulos et al., 2005). Together with CH25H, it converts cholesterol to oxysterols, highly reactive molecules that induce cytotoxic and pro-apoptotic responses (by interfering with membrane lipids), as well as activation of macrophages and dendritic cells (Olkkonen et al., 2012) It may regulate the levels of cortisol and other glucocorticoids (Hennebert et al., 2007).
HSD3B7	A4 > A1	*Hydroxy-Delta-5-Steroid Dehydrogenase, 3 Beta- And Steroid Delta-Isomerase 7*. It is involved in the synthesis of bile acids and other hormonal steroids from cholesterol (Schwarz et al., 2000). It participates in degrading 7α,25-dihydroxycholesterol produced by CH25H and CYP7B1. This molecule binds EBI2, which controls B cell location and responses; thus HSD3B7 may modulate this effect (Yi et al., 2012).
LDLR	D1 > rest	*Low Density Lipoprotein Receptor*. It binds VLDL and LDL in plasma and transports it via endocytosis (Go et al., 2012). This process is also mediated by lipoprotein receptor-related protein (LRPs) (Go et al., 2012).
PCSK9	C2	*Proprotein Convertase Subtilisin/Kexin Type 9*. It binds LDLR, VLDLR and APOE receptors and mediates their degradation, which may increase LDL plasma levels (Ruscica et al., 2019), leading to vascular inflammation from oxidised LDLs, reduced interferon response, and reduced activation of MHC-I in T CD8+ cells. It also promotes lysosomal degradation of MHC-I (Seidah et al., 2021) and may induce pro-inflammatory responses in macrophages (Ricci et al., 2018). PCSK9 may be involved in keratinocyte hyper-proliferation in psoriasis-like inflammation (Luan et al., 2019).
Eicosanoids		
CES1	A3 > A1	*Carboxylesterase 1*. It blocks PGD2 glyceryl ester in monocytes/macrophages, enhancing its anti-inflammatory effects (Scheaffer et al., 2020).
DPEP1	C1	*Dipeptidase 1*. Catalises the conversion of several molecules and dipeptides, including leukotriene D4 into leukotriene E4 (Lee et al., 1983), or glutathione degradation byproduct cystidyl-bis-glycine. It also acts as a receptor for neutrophil recruitment (Choudhury et al., 2019). Described as a DS and DS cup marker by Ahlers et al., 2022; Tabib et al., 2021 together with COL11A1, it is highly upregulated as with other factors in myofibroblasts (Valenzi et al., 2019).

HPGD	A1 > rest	*15-Hydroxyprostaglandin Dehydrogenase*. Belongs to the alcohol dehydrogenase family and metabolises a wide range of prostaglandins, mainly to regulate their levels (Cho et al., 2006; Yan et al., 2004). For instance, it degrades PGE2, an IL-1 derived prostaglandin (produced by PTGS2), regulating its pro-inflammatory action (Arai et al., 2014). It also converts resolvins, which promote restoration of normal cellular function following the inflammation, to regulate ther action (Arita et al., 2006).
IGFBP7	B2 ~ D1 > B3 ~ B4 > B1 ~ E1	*Insulin Like Growth Factor Binding Protein 7*. It belongs to the IGFBP family and binds IGF-I/II with low affinity. It inhibits BRAF signalling and induces senescence and apoptosis (Wajapeyee et al., 2008). It may induce the formation of new HFs by activation of BMP and β-catenin signalling (Lee et al., 2012b). It diminished PTGS2 (COX2) activity and PGE2 secretion, which affects VEGF production and angiogenesis (Tamura et al., 2009).
MGST1	A1 ~ B4 > A3 ~ A4	*Microsomal Glutathione S-Transferase 1*. It is involved in the prostaglandin E2 (PGE2) synthesis pathway, induced by pro-inflammatory factors such as IL-1β (Jakobsson et al., 1999). It also shows some peroxidase activity, which may protect from oxidative stress (Morgenstern et al., 2011), and may explain the higher presence of MGST1 in certain stem cells in aged individuals (Prall et al., 2007; Tsitsipatis et al., 2022). Interestingly, Mgst1 is a marker of mouse SGs (Joost et al., 2016).
PLA2G2A	B4 ~ A4 > A1 ~ B1	*Phospholipase A2 Group IIA*. Member of the phospholipase family, it performs several functions. Related to lipid metabolism, it transforms phospholipids into arachidonic acid, the precursor of various signalling molecules, including prostaglandins (Sato et al., 2016). This isoform is secreted and may play a role in antibacterial protection or atherosclerosis (Murakami et al., 2011). Expression of PLA2G2A in transgenic mice produced adnexal hyperplasia and hyperkeratosis, resulting in alopecia, which may indicate some role of this enzyme in HF homeostasis (Grass et al., 1996). Other isoforms like 2D, 2E and 2F are widely involved in dermal homeostasis (Murakami et al., 2018)
PTGDS	D1 > B2 ~ B3 > A2 ~ C2 ~ D2	*Prostaglandin D2 Synthase*. Catalyses the conversion of PGH2 to PGD2, a prostaglandin involved in smooth muscle contraction/relaxation and a potent platelet aggregation inhibitor (Zhou et al., 2010). It is involved in various CNS functions and may have an anti-apoptotic role in oligodendrocytes (Zhou et al., 2010). Regarding HF function, PTGDS was observed to inhibit wound-induced HF neogenesis (Nelson et al., 2013) and was highly expressed in androgenic alopecia (Garza et al., 2012); by the union of PGD2 to Gpr44 receptor. PGD2 is also observed to facilitate mast-cell maturation (Taketomi et al., 2013).
PTGFR	C2 > rest	*Prostaglandin F Receptor*. Receptor of prostaglandin F2-α. In hair, it may be regulated by Wnt3 (Shin et al., 2010). It initiates the luteolysis of the corpus luteum and is also expressed in endometrium (Gao et al., 2018).
PTGIS	A4 > A1 ~ A2 ~ A3	*Prostaglandin I2 Synthase*. It transforms prostaglandin G2 into prostaglandin I2 (prostacyclin). It induces VEGF production by activating PPARD (Kamio et al., 2008; Wang et al., 2013). It reduced fibroblast contractibility by suppressing fibronectin (Kamio et al., 2007; Kohyama et al., 2002) and/or collagen I (Larsson-Callerfelt et al., 2013). PGI2 may also be involved in hair regeneration regulation since high levels were increased in bald scalps compared to hairy ones (Chovarda et al., 2021), and minoxidil is observed to inhibit prostacyclin (Messenger et al., 2004).
PTGS1	A2	*Prostaglandin-Endoperoxide Synthase 1*. Synthesises several prostaglandins and thromboxanes from arachidonic acid, which act in homeostatic processes, such as platelet function (TXA2) or stomach lining production (PGI2). It is widely expressed in most tissues (Heide et al., 2006).

PTGS2	B1	*Prostaglandin-Endoperoxide Synthase 2*. Induced by pro-inflammatory mediators such as IL1 or IL6, it transforms arachidonic acid into PGE2, which is involved in fever and pain signalling, among others (Heide et al., 2006). Due to its proangiogenic and antiapoptotic nature, its overexpression has been observed in certain tumours (Maturu et al., 2017). Increased levels are found in UV-damaged skin fibroblasts (Surowiak et al., 2014).
Other lipids		
CTSH	B3 > B2	*Cathepsin H*. Cysteine protease that degrades lysosomal proteins. It degrades the ECM together with MMPs, for lymphocyte infiltration (Li et al., 2010). Together with cholesteryl esterase may activate C5 protein (Bhakdi et al., 2004). In KO, filaggrin expression was reduced, and epidermal barrier was impaired. Macrophage and mast cell infiltration was observed (Naeem et al., 2017).
CYP1B1	D1 > C2	*Cytochrome P450 Family 1 Subfamily B Member 1*. Enzyme of the cytochrome P450 family. Pro-inflammatory cytokines upregulate it (Smerdova et al., 2014) and after UV-B exposure (Villard et al., 2002). It oxidises a wide range of substrates and induces several effects, such as angiogenesis under stress (Palenski et al., 2013; Tang et al., 2010b); or mediation of POSTN to regulate the trabecular meshwork in retina (Zhao et al., 2013).

Table S9: **Genes associated with Vitamin A metabolism.**

ALDH1A3	D1 > D2	*Aldehyde Dehydrogenase 1 Family Member A3*. It transforms retinaldehyde to retinoic acid, the active form of binding RXR-RAR TFs (Kedishvili, 2013).
CRABP1	C2 > C5	*Cellular Retinoic Acid Binding Protein 1*. It is similar to RBPs and transports retinoic acid into the nucleus. It competes with FABP5 in the binding of the molecule. While FABP5 promotes cell survival and proliferation, CRABP1 is more associated with apoptosis and growth arrest (Michalik et al., 2007).
CYP26B1	A2 > C5 ~ C2	*Cytochrome P450 Family 26 Subfamily B Member 1*. It degrades all-trans retinoic acid by hydroxylation (Isoherranen et al., 2019). This degradation reduces retinoic acid-mediated mast cell activation by fibroblasts (Kurashima et al., 2014). In HF, retinoic acid contributes to the refractory telogen phase, and CYP26B1 may take part in the modulation of this step (Hovland et al., 2020). It has also been observed that CYP26B1 retinoic acid level modulation is essential for HF morphogenesis (Okano et al., 2012).
RBP4	C1	*Retinol Binding Protein 4*. It carries retinol in the blood and hydrophilic environments into their target cells, where it binds the STRA6 transporter and is interiorised (Steinhoff et al., 2021).
RBP5	B2 ~ B3	*Retinol Binding Protein 5*. May act like RBP4.
TTR	A2	*Transthyretin*. It associates with RBPs to transport retinol into the cells (Steinhoff et al., 2021).

B

SUPPLEMENTARY FIGURES

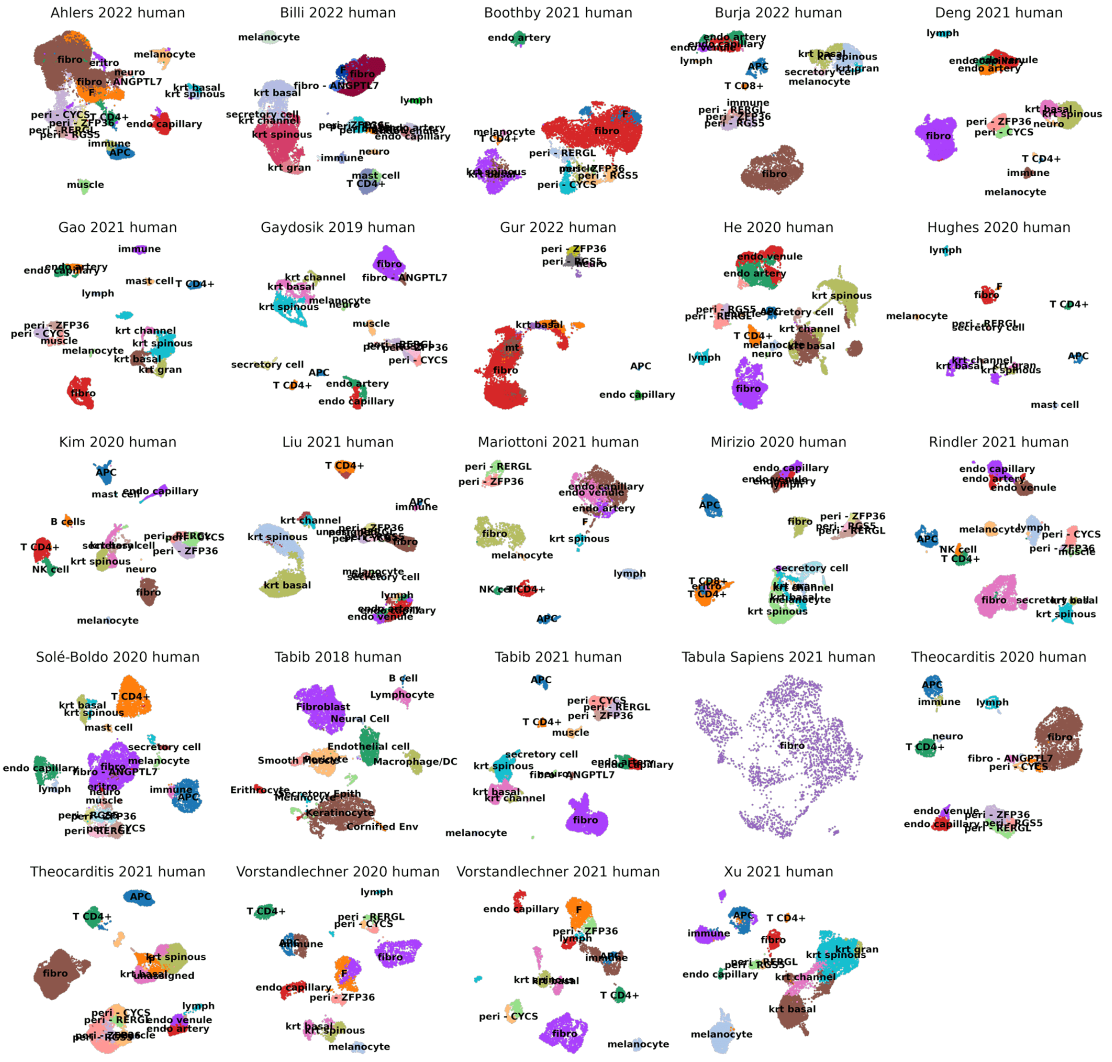


Figure S1: **UMAP plots of all human skin datasets from secondary analysis.** Colours of populations are not shared across datasets.

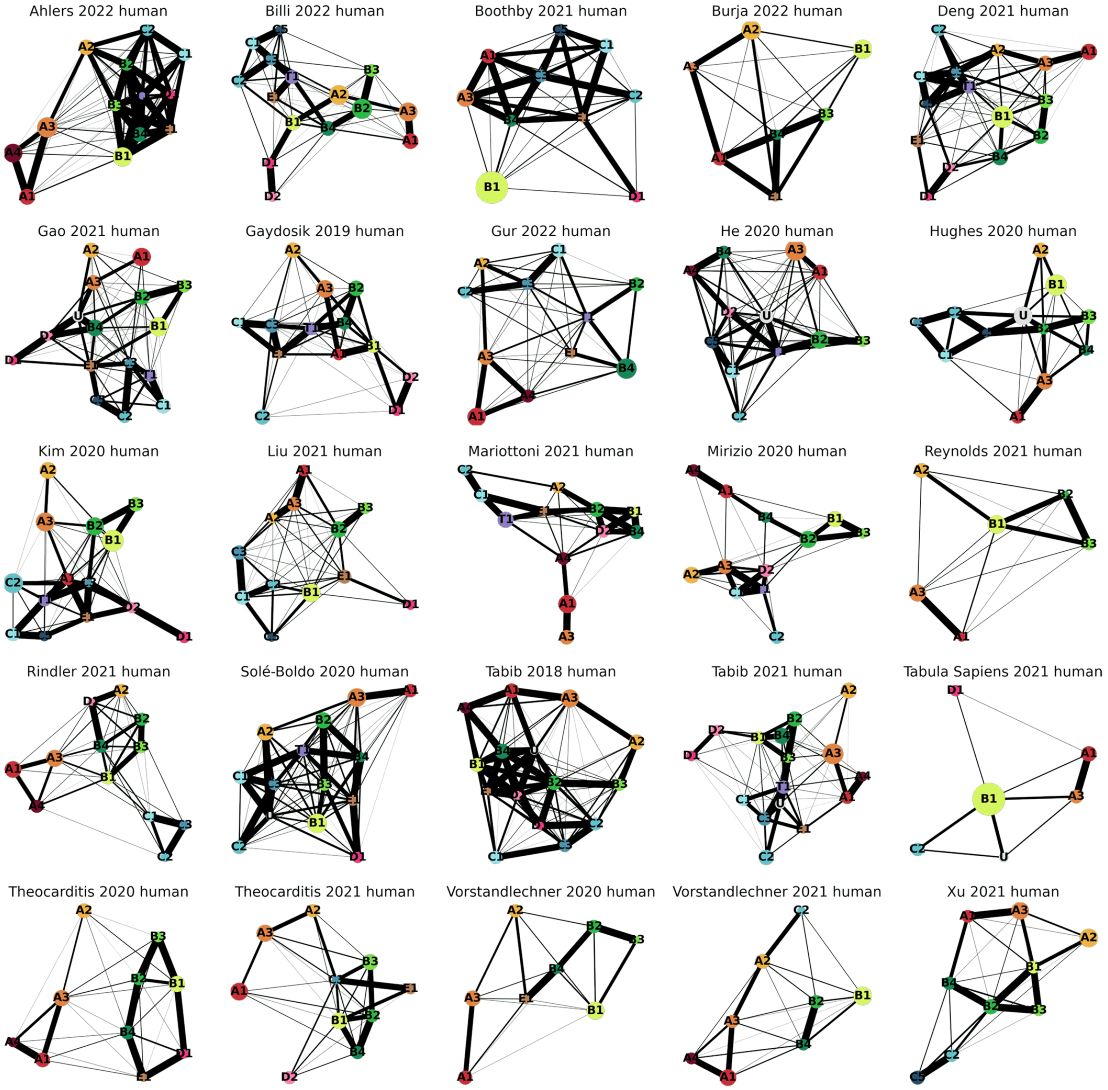


Figure S2: **Full PAGA graph of human dermal populations.** Colours of populations are shared across datasets. Thicker lines indicate a greater connectivity between nodes.

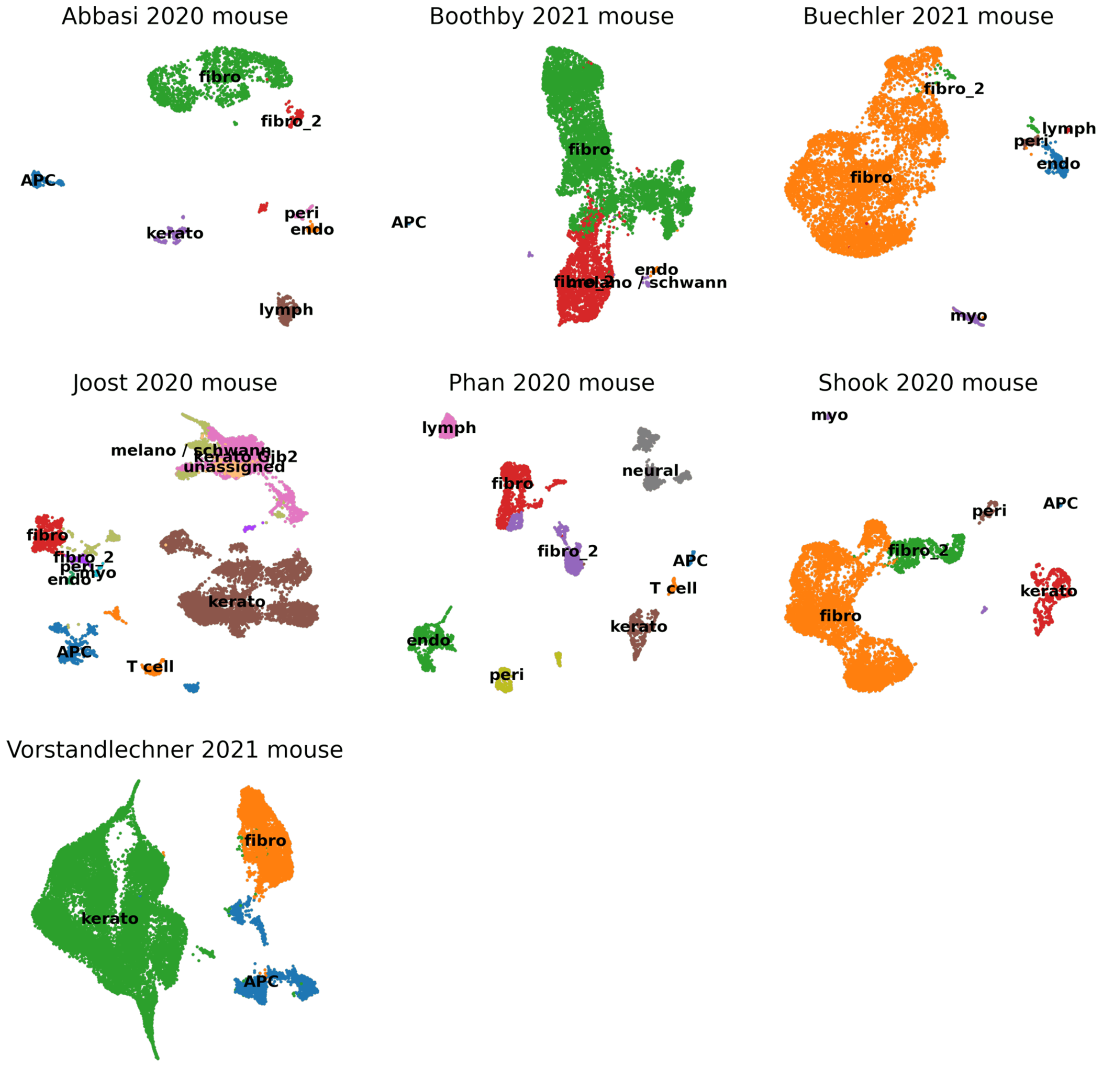


Figure S3: **UMAP plots of all mouse skin datasets from secondary analysis.** Colours of populations are not shared across datasets.

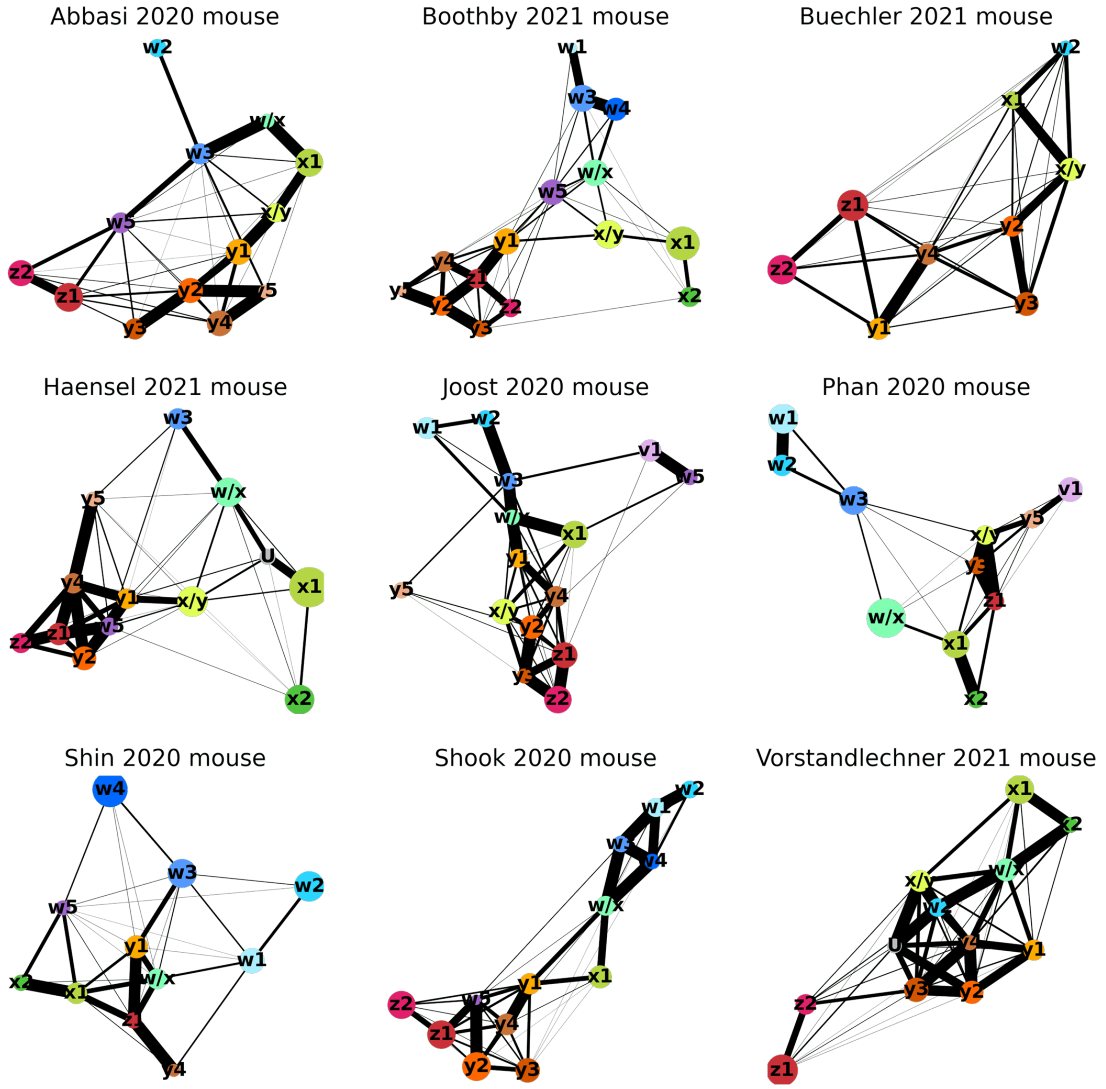


Figure S4: **Full PAGA graph of mouse dermal populations.** Colours of populations are shared across datasets. Thicker lines indicate a greater connectivity between nodes.

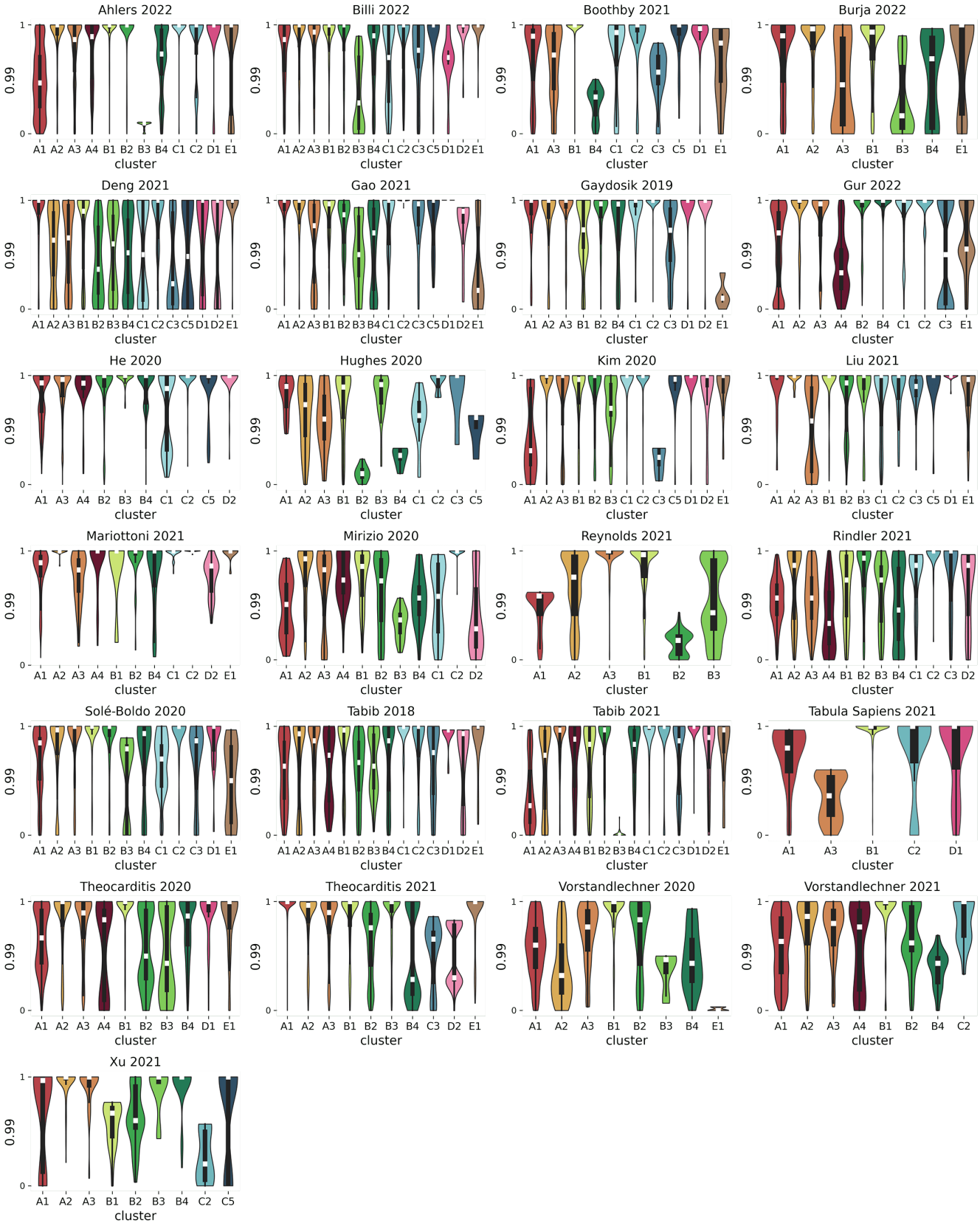


Figure S5: **Violinplot of robustness score in human datasets.** Each violin represents the distribution of robustness scores per dataset and population, across the representative cells. Boxes inside the violins represent the quartiles of the distribution, center bar represents the median, and whiskers extend to Q1 - 1.5 * IQR and Q3 + 1.5 * IQR.

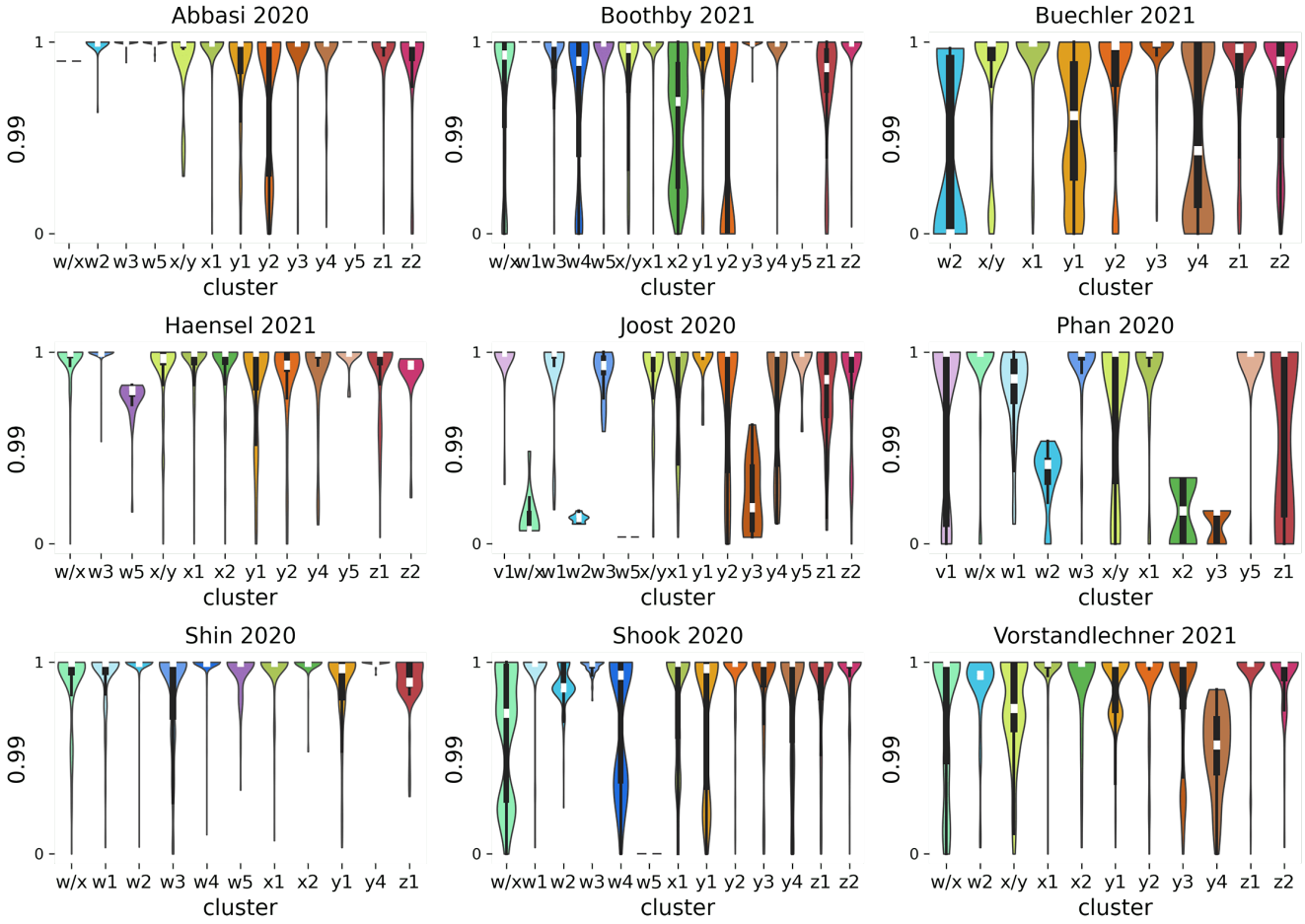


Figure S6: **Violinplot of robustness score in mouse datasets.** Each violin represents the distribution of robustness scores per dataset and population, across the representative cells. Boxes inside the violins represent the quartiles of the distribution, center bar represents the median, and whiskers extend to Q1 - 1.5 * IQR and Q3 + 1.5 * IQR.

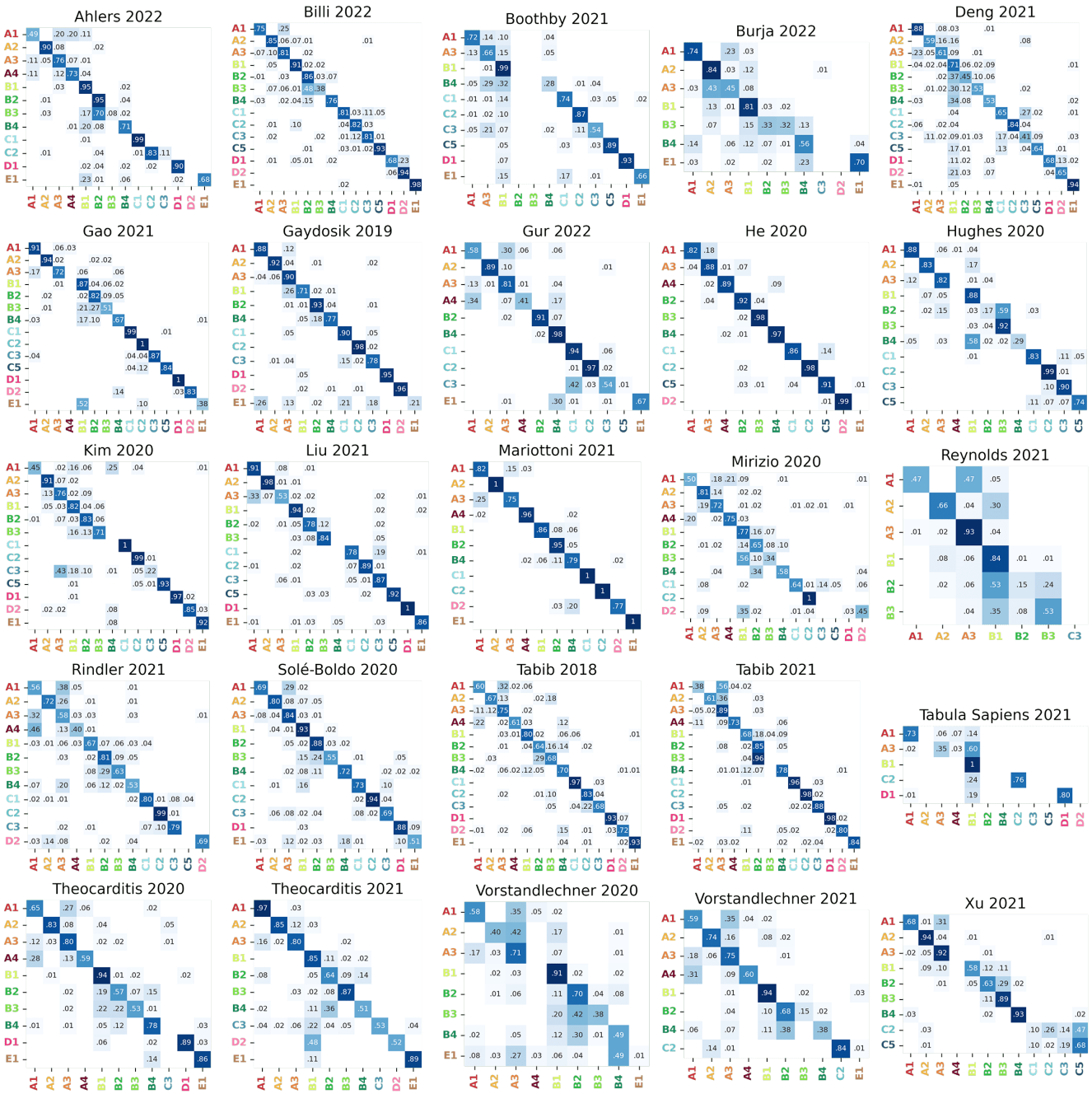


Figure S7: **Adjacency matrix of robust cluster assignation in human datasets.** Each row-column combination is the proportion of times the row cluster is assigned to the column cluster, considering all cells within the row clusters, and all the iterations.

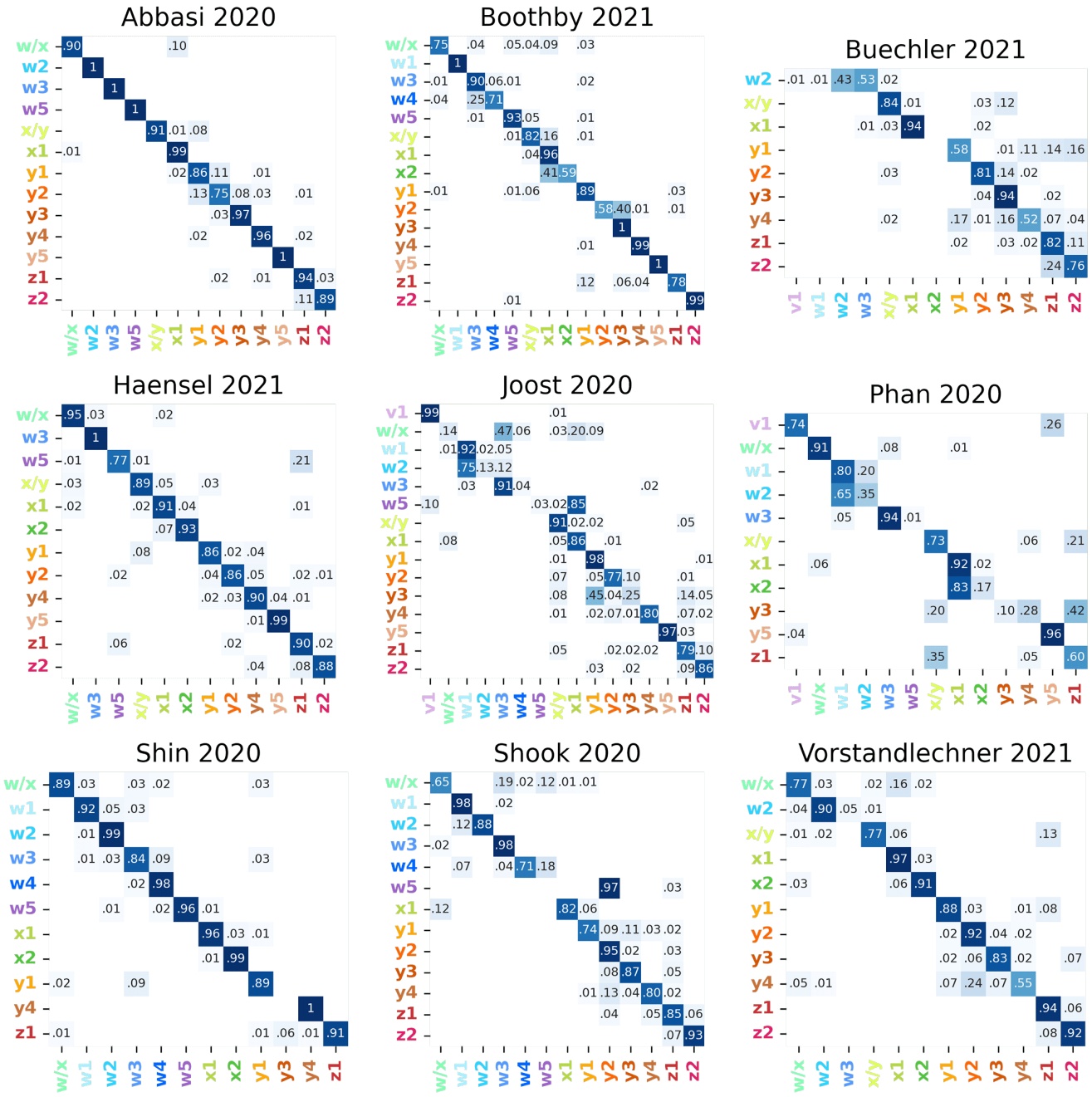


Figure S8: **Adjacency matrix of robust cluster assignation in mouse datasets.** Each row-column combination is the proportion of times the row cluster is assigned to the column cluster, considering all cells within the row clusters, and all the iterations.

BIBLIOGRAPHY

Abbasi, Sepideh et al. (2020). "Distinct regulatory programs control the latent regenerative potential of dermal fibroblasts during wound healing." In: *Cell Stem Cell* 27.3. DOI: 10.1016/j.stem.2020.07.008.

Abdo, Joseph M., Nikolai A. Sopko, and Stephen M. Milner (2020). "The applied anatomy of Human skin: A model for regeneration." In: *Wound Medicine* 28, p. 100179. DOI: 10.1016/j.wndm.2020.100179.

Abdullahi, A., S. Amini-Nik, and M. G. Jeschke (2014). "Animal models in burn research." In: *Cellular and Molecular Life Sciences* 71.17, 3241–3255. DOI: 10.1007/s00018-014-1612-5.

Adam, Mike, Andrew S. Potter, and S. Steven Potter (2017). "Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development." In: *Development* 144.19, 3625–3632. DOI: 10.1242/dev.151142.

Adapala, Venkata J, Kimberly K Buhman, and Kolapo M Ajuwon (2011). "Novel anti-inflammatory role of SLPI in adipose tissue and its regulation by high fat diet." In: *Journal of Inflammation* 8.1. DOI: 10.1186/1476-9255-8-5.

Adil, Asif et al. (2021). "Single-cell transcriptomics: Current methods and challenges in data acquisition and analysis." In: *Frontiers in Neuroscience* 15. DOI: 10.3389/fnins.2021.591122.

Adly, Mohamed A. et al. (2008). "Expression patterns of the glial cell line–derived neurotrophic factor, neurturin, their cognate receptors GFRα-1, GFRα-2, and a common signal transduction element c-Ret in the human skin hair follicles." In: *Journal of the American Academy of Dermatology* 58.2, pp. 238–250. DOI: 10.1016/j.jaad.2007.10.014.

Afzali, Behdad and Claudia Kemper (2021). "Fibroblast tissue priming—not so nice to C you!" In: *Immunity* 54.5, pp. 847–850. DOI: 10.1016/j.immuni.2021.04.010.

Agarwal, Pallavi et al. (2012). "Collagen XII and XIV, New Partners of Cartilage Oligomeric Matrix Protein in the Skin Extracellular Matrix Suprastructure." In: *Journal of Biological Chemistry* 287.27, pp. 22549–22559. DOI: 10.1074/jbc.m111.335935.

Aguilar-Ballester, María et al. (2020). "Impact of Cholesterol Metabolism in Immune Cell Function and Atherosclerosis." In: *Nutrients* 12.7, p. 2021. DOI: 10.3390/nu12072021.

Ahlers, Juliane M. et al. (2022). "Single-cell RNA profiling of human skin reveals age-related loss of dermal sheath cells and their contribution to a juvenile phenotype." In: *Frontiers in Genetics* 12. DOI: 10.3389/fgene.2021.797747.

Ahn, Victoria E. et al. (2011). "Structural Basis of Wnt Signaling Inhibition by Dickkopf Binding to LRP5/6." In: *Developmental Cell* 21.5, pp. 862–873. DOI: 10.1016/j.devcel.2011.09.003.

Airaksinen, Matti S. and Mart Saarma (2002). "The GDNF family: Signalling, biological functions and therapeutic value." In: *Nature Reviews Neuroscience* 3.5, pp. 383–394. DOI: 10.1038/nrn812.

Albanesi, Cristina et al. (2008). "Chemerin expression marks early psoriatic skin lesions and correlates with plasmacytoid dendritic cell recruitment." In: *Journal of Experimental Medicine* 206.1, pp. 249–258. DOI: 10.1084/jem.20080129.

Albarbar, Balid, Christopher Dunnill, and Nikolaos T. Georgopoulos (2015). "Regulation of cell fate by lymphotoxin (LT) receptor signalling: Functional differences and similarities of the LT system to other TNF superfamily (TNFSF) members." In: *Cytokine & Growth Factor Reviews* 26.6, pp. 659–671. DOI: 10.1016/j.cytogfr.2015.05.001.

Albuquerque, Rafael et al. (2021). "The role of Interleukin-32 in autoimmunity." In: *Scandinavian Journal of Immunology* 93.2. DOI: 10.1111/sji.13012.

Alkon, Natalia et al. (2022). "Single-cell analysis reveals innate lymphoid cell lineage infidelity in atopic dermatitis." In: *Journal of Allergy and Clinical Immunology* 149.2, 624–639. DOI: 10.1016/j.jaci.2021.07.025.

Allanki, Srinivas et al. (2021). "Interleukin-11 signaling promotes cellular reprogramming and limits fibrotic scarring during tissue regeneration." In: *Science Advances* 7.37. DOI: 10.1126/sciadv.abg6497.

Almeida, Cecilia Jacques Gonçalves de (2017). "Caveolin-1 and Caveolin-2 Can Be Antagonistic Partners in Inflammation and Beyond." In: *Frontiers in Immunology* 8. DOI: 10.3389/fimmu.2017.01530.

Almet, Axel A. et al. (2021). "The landscape of cell–cell communication through single-cell transcriptomics." In: *Current Opinion in Systems Biology* 26, 12–23. DOI: 10.1016/j.coisb.2021.03.007.

Alriquet, Marion et al. (2015). "Characterization of SLC transporters in human skin." In: *ADMET & DMPK* 3.1. DOI: 10.5599/admet.3.1.163.

Alsaleh, Ghada et al. (2010). "Innate immunity triggers IL-32 expression by fibroblast-like synoviocytes in rheumatoid arthritis." In: *Arthritis Research & Therapy* 12.4, R135. DOI: 10.1186/ar3073.

Amezcua, J.L. et al. (1988). "Acetylcholine induces vasodilatation in the rabbit isolated heart through the release of nitric oxide, the endogenous nitrovasodilator." In: *British Journal of Pharmacology* 95.3, pp. 830–834. DOI: 10.1111/j.1476-5381.1988.tb11711.x.

Amezquita, Robert A. et al. (2019). "Orchestrating single-cell analysis with bioconductor." In: *Nature Methods* 17.2, 137–145. DOI: 10.1038/s41592-019-0654-x.

Amodio, Matthew et al. (2017). "Exploring single-cell data with deep multitasking neural networks." In: DOI: 10.1101/237065.

Amphoux, Anne et al. (2006). "Differential pharmacological in vitro properties of organic cation transporters and regional distribution in rat brain." In: *Neuropharmacology* 50.8, pp. 941–952. DOI: 10.1016/j.neuropharm.2006.01.005.

Amsellem, Valerie et al. (2014). "ICAM-2 regulates vascular permeability and N-cadherin localization through ezrin-radixin-moesin (ERM) proteins and Rac-1 signalling." In: *Cell Communication and Signaling* 12.1. DOI: 10.1186/1478-811x-12-12.

Anan, T. et al. (2003). "Protease-activated receptor-1 (thrombin receptor) is expressed in mesenchymal portions of human hair follicle." In: *Journal of Investigative Dermatology* 121.4, 669–673. DOI: 10.1046/j.1523-1747.2003.12490.x.

Anderson, Caroline A, Roberto Solari, and James E Pease (2015). "Biased agonism at chemokine receptors: obstacles or opportunities for drug discovery?" In: *Journal of Leukocyte Biology* 99.6, pp. 901–909. DOI: 10.1189/jlb.2mr0815-392r.

Anderson, Richard G. W. (1998). "THE CAVEOLAE MEMBRANE SYSTEM." In: *Annual Review of Biochemistry* 67.1, pp. 199–225. DOI: 10.1146/annurev.biochem.67.1.199.

Anderssohn, Ann Marie et al. (2011). "Molecular Chaperone Function for Myocilin." In: *Investigative Opthalmology & Visual Science* 52.10, p. 7548. DOI: 10.1167/iovs.11-7723.

Andrews, Tallulah S and Martin Hemberg (2018). "M3Drop: dropout-based feature selection for scRNASeq." In: *Bioinformatics* 35.16, 2865–2867. DOI: 10.1093/bioinformatics/bty1044.

Andrews, Tallulah S. and Martin Hemberg (2019). "False signals induced by single-cell imputation." In: *F1000Research* 7, p. 1740. DOI: 10.12688/f1000research.16613.2.

Andrews, Tallulah (2020). *BRENNECKE_GETVARIABLEGENES: Identify highly variable genes in M3drop: Michaelis-menten modelling of dropouts in single-cell rnaseq*.

Angelidis, Ilias et al. (2019). "An atlas of the ageing lung mapped by single cell transcriptomics and deep tissue proteomics." In: *Nature Communications* 10.1. DOI: 10.1038/s41467-019-08831-9.

Angonin, Diane and Terence J. Van Raay (2013). "Nkd1 Functions as a Passive Antagonist of Wnt Signaling." In: *PLoS ONE* 8.8. Ed. by Alistair P. McGregor, e74666. DOI: 10.1371/journal.pone.0074666.

Annabi, Bayader, Alain Zgheib, and Borhane Annabi (2017). "Cavin-2 Functions as a Suppressive Regulator in TNF-induced Mesenchymal Stromal Cell Inflammation and Angiogenic Phenotypes." In: *International Journal of Stem Cells* 10.1, pp. 103–113. DOI: 10.15283/ijsc16032.

Anscombe, F. J. (1948). "The transformation of Poisson, binomial and negative-binomial data." In: *Biometrika* 35.3/4, p. 246. DOI: 10.2307/2332343.

Antonini, Dario et al. (2013). "An Intimate Relationship between Thyroid Hormone and Skin: Regulation of Gene Expression." In: *Frontiers in Endocrinology* 4. DOI: 10.3389/fendo.2013.00104.

Arai, Koji Y. et al. (2014). "Stimulatory effect of fibroblast-derived prostaglandin E2 on keratinocyte stratification in the skin equivalent." In: *Wound Repair and Regeneration* 22.6, pp. 701–711. DOI: 10.1111/wrr.12228.

Argelaguet, Ricard et al. (2021). "Computational principles and challenges in single-cell data integration." In: *Nature Biotechnology* 39.10, 1202–1215. DOI: 10.1038/s41587-021-00895-7.

Arita, Makoto et al. (2005). "Stereochemical assignment, antiinflammatory properties, and receptor for the omega-3 lipid mediator resolvin E1." In: *Journal of Experimental Medicine* 201.5, pp. 713–722. DOI: 10.1084/jem.20042031.

Arita, Makoto et al. (2006). "Metabolic Inactivation of Resolvin E1 and Stabilization of Its Anti-inflammatory Actions." In: *Journal of Biological Chemistry* 281.32, pp. 22847–22854. DOI: 10.1074/jbc.m603766200.

Armingol, Erick et al. (2020). "Deciphering cell–cell interactions and communication from gene expression." In: *Nature Reviews Genetics* 22.2, 71–88. DOI: 10.1038/s41576-020-00292-x.

Ascensión, Alex M. et al. (2021). "Human Dermal Fibroblast Subpopulations Are Conserved across Single-Cell RNA Sequencing Studies." In: *Journal of Investigative Dermatology* 141(7), 1735–1744. DOI: 10.1016/j.jid.2020.11.028.

Ascensión, Alex M. et al. (2022). "Triku: A feature selection method based on nearest neighbors for single-cell data." In: *GigaScience* 11. DOI: 10.1093/gigascience/giac017.

Ashburner, Michael et al. (2000). "Gene ontology: Tool for the unification of biology." In: *Nature Genetics* 25.1, 25–29. DOI: 10.1038/75556.

Ashcroft, Gillian S. et al. (2000). "Secretory leukocyte protease inhibitor mediates non-redundant functions necessary for normal wound healing." In: *Nature Medicine* 6.10, pp. 1147–1153. DOI: 10.1038/80489.

Ashkar, Silvia et al. (2004). "Retinoic Acid Induces Corneal Epithelial CYP4B1 Gene Expression and Stimulates the Synthesis of Inflammatory 12-Hydroxyeicosanoids." In: *Journal of Ocular Pharmacology and Therapeutics* 20.1, pp. 65–74. DOI: 10.1089/108076804772745473.

Aubert, Alexandre et al. (2021). "Latent TGF-β Activation Is a Hallmark of the Tenascin Family." In: *Frontiers in Immunology* 12. DOI: 10.3389/fimmu.2021.613438.

Avniel, Shani et al. (2006). "Involvement of the CXCL12/CXCR4 Pathway in the Recovery of Skin Following Burns." In: *Journal of Investigative Dermatology* 126.2, pp. 468–476. DOI: 10.1038/sj.jid.5700069.

Avsar, Orcun (2022). "Analysis of missense SNPs in the SLC47A1 and SLC47A2 genes affecting the pharmacokinetics of metformin: Computational approach." In: *Egyptian Journal of Medical Human Genetics* 23.1. DOI: 10.1186/s43042-022-00306-9.

Bacher, Rhonda and Christina Kendziorski (2016). "Design and computational analysis of single-cell RNA-sequencing experiments." In: *Genome Biology* 17.1. DOI: 10.1186/s13059-016-0927-y.

Badarinath, Krithika et al. (2022). "Snail maintains the stem/progenitor state of skin epithelial cells and carcinomas through the autocrine effect of matricellular protein Mindin." In: *Cell Reports* 40.12, p. 111390. DOI: 10.1016/j.celrep.2022.111390.

Bader, Hannah L. et al. (2012). "A disintegrin-like and metalloprotease domain containing thrombospondin type 1 motif-like 5 (ADAMTSL5) is a novel fibrillin-1-, fibrillin-2-, and heparin-binding member of the ADAMTS superfamily containing a netrin-like module." In: *Matrix Biology* 31.7-8, pp. 398–411. DOI: 10.1016/j.matbio.2012.09.003.

Badolato, R. et al. (1997). "Interleukin-15 (IL-15) induces IL-8 and monocyte chemotactic protein 1 production in human monocytes." In: *Blood* 90.7, pp. 2804–2809.

Badshah, I.I. et al. (2019). "Differential expression of secreted factors SOSTDC1 and ADAMTS8 cause profibrotic changes in linear morphoea fibroblasts." In: *British Journal of Dermatology* 180.5, pp. 1135–1149. DOI: 10.1111/bjd.17352.

Baharudin, Rashidah et al. (2020). "Epigenetics of SFRP1: The Dual Roles in Human Cancers." In: *Cancers* 12.2, p. 445. DOI: 10.3390/cancers12020445.

Bahrami, Shahram and Finn Drabløs (2016). "Gene regulation in the immediate-early response process." In: *Advances in Biological Regulation* 62, pp. 37–49. DOI: 10.1016/j.jbior.2016.05.001.

Baicu, Catalin F. et al. (2012). "Effects of the absence of procollagen C-endopeptidase enhancer-2 on myocardial collagen accumulation in chronic pressure overload." In: *American Journal of Physiology-Heart and Circulatory Physiology* 303.2, H234–H240. DOI: 10.1152/ajpheart.00227.2012.

Baines, K.J. and S.J. Renaud (2017). "Transcription Factors That Regulate Trophoblast Development and Function." In: *Progress in Molecular Biology and Translational Science*. Elsevier, pp. 39–88. DOI: 10.1016/bs.pmbts.2016.12.003.

Balakrishnan, Lavanya et al. (2013). "IL-11/IL11RA receptor mediated signaling: a web accessible knowledgebase." In: *Cell Communication & Adhesion* 20.3-4, pp. 81–86. DOI: 10.3109/15419061.2013.791683.

Banas, Magdalena et al. (2015). "The Expression and Regulation of Chemerin in the Epidermis." In: *PLOS ONE* 10.2. Ed. by Bernhard Ryffel, e0117830. DOI: 10.1371/journal.pone.0117830.

Baran-Gale, Jeanette, Tamir Chandra, and Kristina Kirschner (2017). "Experimental design for single-cell RNA sequencing." In: *Briefings in Functional Genomics* 17.4, 233–239. DOI: 10.1093/bfgp/elx035.

Barbieri, J.S., K. Wanat, and J. Seykora (2014). "Skin: Basic structure and function." In: *Pathobiology of Human Disease*, 1134–1144. DOI: 10.1016/b978-0-12-386456-7.03501-2.

Barco Barrantes, Ivén del et al. (1999). "Interaction between Notch signalling and Lunatic fringe during somite boundary formation in the mouse." In: *Current Biology* 9.9, pp. 470–480. DOI: 10.1016/s0960-9822(99)80212-7.

Barkas, Nikolas et al. (2019). "Joint analysis of heterogeneous single-cell RNA-seq dataset collections." In: *Nature Methods* 16.8, 695–698. DOI: 10.1038/s41592-019-0466-z.

Barnes, William G. et al. (2005). "β-Arrestin 1 and Gαq/11 Coordinately Activate RhoA and Stress Fiber Formation following Receptor Stimulation." In: *Journal of Biological Chemistry* 280.9, pp. 8041–8050. DOI: 10.1074/jbc.m412924200.

Baroni, Adone et al. (2012). "Structure and function of the epidermis related to barrier properties." In: *Clinics in Dermatology* 30.3, 257–262. DOI: 10.1016/j.clindermatol.2011.08.007.

Barrese, Vincenzo et al. (2020). "SMIT (Sodium-Myo-Inositol Transporter) 1 Regulates Arterial Contractility Through the Modulation of Vascular Kv7 Channels." In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 40.10, pp. 2468–2480. DOI: 10.1161/atvbaha.120.315096.

Barron, Martin and Jun Li (2016). "Identifying and removing the cell-cycle effect from single-cell RNA-sequencing data." In: *Scientific Reports* 6.1. DOI: 10.1038/srep33892.

Bartlett, M. S. (1947). "The use of transformations." In: *Biometrics* 3.1, p. 39. DOI: 10.2307/3001536.

Bashir, Ahmed et al. (2015). "Coronary Artery Calcium Assessment in CKD: Utility in Cardiovascular Disease Risk Assessment and Treatment?" In: *American Journal of Kidney Diseases* 65.6, pp. 937–948. DOI: 10.1053/j.ajkd.2015.01.012.

Bauer, Eugene A. and Jouni Uitto (1979). "Collagen in cutaneous diseases." In: *International Journal of Dermatology* 18.4, 251–270. DOI: 10.1111/j.1365-4362.1979.tb01926.x.

Bautista-Hernandez, Luis Antonio et al. (2017). "Fibroblasts: the unknown sentinels eliciting immune responses against microorganisms." In: *European Journal of Microbiology and Immunology* 7.3, pp. 151–157. DOI: 10.1556/1886.2017.00009.

Becht, Etienne et al. (2018). "Dimensionality reduction for visualizing single-cell data using UMAP." In: *Nature Biotechnology* 37.1, 38–44. DOI: 10.1038/nbt.4314.

Becker, Daniel et al. (2005). "TRPV4 exhibits a functional role in cell-volume regulation." In: *Journal of Cell Science* 118.11, 2435–2440. DOI: 10.1242/jcs.02372.

Bellis, Alessandro et al. (2020). "The Rationale of Neprilysin Inhibition in Prevention of Myocardial Ischemia-Reperfusion Injury during ST-Elevation Myocardial Infarction." In: *Cells* 9.9, p. 2134. DOI: 10 . 3390 / cells9092134.

Bellman, Richard Ernest (2015). *Dynamic Programming*. Dover.

Belotti, Dorina et al. (2011). "Targeting angiogenesis with compounds from the extracellular matrix." In: *The International Journal of Biochemistry & Cell Biology* 43.12, 1674–1685. DOI: 10.1016/j.biocel.2011.08.012.

Bendall, Sean C. et al. (2014). "Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development." In: *Cell* 157.3, 714–725. DOI: 10.1016/j.cell.2014.04.005.

Bennett, K L et al. (1995). "CD44 isoforms containing exon V3 are responsible for the presentation of heparin-binding growth factor." In: *Journal of Cell Biology* 128.4, pp. 687–698. DOI: 10.1083/jcb.128.4.687.

Berahovich, Robert D. et al. (2013). "Endothelial expression of CXCR7 and the regulation of systemic CXCL12 levels." In: *Immunology* 141.1, pp. 111–122. DOI: 10 . 1111 / imm . 12176.

Bergen, Lisa Lynn ten et al. (2020). "Current knowledge on autoantigens and autoantibodies in psoriasis." In: *Scandinavian Journal of Immunology* 92.4. DOI: 10.1111/sji.12945.

Bergsma, Derk J. et al. (1992). "Cloning and characterization of a human angiotensin II type 1 receptor." In: *Biochemical and Biophysical Research Communications* 183.3, pp. 989–995. DOI: 10.1016/s0006-291x(05)80288-8.

Berika, Mohamed and David Garrod (2014). "Desmosomal Adhesion In Vivo." In: *Cell Communication & Adhesion* 21.1, pp. 65–75. DOI: 10.3109/15419061.2013.876018.

Berndt, Jason D. (2014). "Sox Keep Hair Follicle Stem Cells Cozy." In: *Science Signaling* 7.314. DOI: 10 . 1126 / scisignal.2005209.

Bertheim, Ulf and Sten Hellstro¨m (1994). "The distribution of hyaluronan in human skin and mature, hypertrophic and keloid scars." In: *British Journal of Plastic Surgery* 47.7, pp. 483–489. DOI: 10.1016/0007-1226(94)90031-0.

Berthod, François et al. (2006). "Extracellular matrix deposition by fibroblasts is necessary to promote capillary-like tube formation in vitro." In: *Journal of Cellular Physiology* 207.2, 491–498. DOI: 10.1002/jcp.20584.

Besschetnova, Tatiana Y. et al. (2015). "Regulatory mechanisms of anthrax toxin receptor 1-dependent vascular and connective tissue homeostasis." In: *Matrix Biology* 42, pp. 56–73. DOI: 10.1016/j.matbio.2014.12.002.

Bhakdi, Sucharit et al. (2004). "Possible Protective Role for C-Reactive Protein in Atherogenesis." In: *Circulation* 109.15, pp. 1870–1876. DOI: 10.1161/01.cir.0000124228.08972.26.

Bhoopalam, Myan, Luis A. Garza, and Sashank K. Reddy (2020). "Wound Induced Hair Neogenesis – A Novel Paradigm for Studying Regeneration and Aging." In: *Frontiers in Cell and Developmental Biology* 8. DOI: 10.3389/fcell.2020.582346.

Bierbaum, S., V. Hintze, and D. Scharnweber (2017). "Artificial Extracellular Matrices to Functionalize Biomaterial Surfaces." In: *Comprehensive Biomaterials II*. Elsevier, pp. 147–178. DOI: 10.1016/b978-0-12-803581-8.10206-1.

Billi, Allison C. et al. (2022). "Nonlesional lupus skin contributes to inflammatory education of myeloid cells and primes for cutaneous inflammation." In: *Science Translational Medicine* 14.642. DOI: 10 . 1126 / scitranslmed . abn2263.

Blazquez-Medela, Ana M., Medet Jumabay, and Kristina I. Boström (2019). "Beyond the bone: Bone morphogenetic protein signaling in adipose tissue." In: *Obesity Reviews* 20.5, pp. 648–658. DOI: 10.1111/obr.12822.

Blok, J et al. (2015). "Increased Expression of Integrin $\alpha 6\beta 4$ in the Basement Membrane Zone Lining the Sebaceous Glands in Hidradenitis Suppurativa." In: *Acta Dermato Venereologica* 95.8, pp. 994–996. DOI: 10.2340/00015555-2186.

Blondel, Vincent D et al. (2008). "Fast unfolding of communities in large networks." In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10. DOI: 10.1088/1742-5468/2008/10/p10008.

Bodnár, Andrea et al. (2008). "A biophysical approach to IL-2 and IL-15 receptor function: Localization, conformation and interactions." In: *Immunology Letters* 116.2, 117–125. DOI: 10.1016/j.imlet.2007.12.014.

Bonilla, Weldy V. et al. (2012). "The Alarmin Interleukin-33 Drives Protective Antiviral CD8 T Cell Responses." In: *Science* 335.6071, pp. 984–989. DOI: 10 . 1126 / science . 1215418.

Bonnet, I. et al. (2017). "447 Collagen XVIII, a key interfacial component of the skin architecture." In: *Journal of Investigative Dermatology* 137.5, S77. DOI: 10.1016/j.jid.2017.02.466.

Booeshaghi, A Sina and Lior Pachter (2021). "Normalization of single-cell RNA-seq counts by log(x + 1) or log(1 + x)." In: *Bioinformatics* 37.15, 2223–2224. DOI: 10 . 1093 / bioinformatics/btab085.

Boon, Wah Chin et al. (2011). "Increasing cdna yields from single-cell quantities of mrna in standard laboratory reverse transcriptase reactions using acoustic mi-

crostreaming." In: *Journal of Visualized Experiments* 53. DOI: 10.3791/3144.

Boopathy, Gandhi T.K. et al. (2017). "Cavin-2 regulates the activity and stability of endothelial nitric-oxide synthase (eNOS) in angiogenesis." In: *Journal of Biological Chemistry* 292.43, pp. 17760–17776. DOI: 10.1074/jbc.m117.794743.

Boothby, Ian C. et al. (2021). "Early-life inflammation primes a T helper 2 cell–fibroblast niche in skin." In: *Nature* 599.7886, 667–672. DOI: 10.1038/s41586-021-04044-7.

Borras, Teresa (2014). "The Effects of Myocilin Expression on Functionally Relevant Trabecular Meshwork Genes: A Mini-Review." In: *Journal of Ocular Pharmacology and Therapeutics* 30.2-3, pp. 202–212. DOI: 10.1089/jop.2013.0218.

Borst, Jannie, Jenny Hendriks, and Yanling Xiao (2005). "CD27 and CD70 in T cell and B cell activation." In: *Current Opinion in Immunology* 17.3, pp. 275–281. DOI: 10.1016/j.coi.2005.04.004.

Botchkarev, Vladimir A. et al. (2000). "A role for p75 neurotrophin receptor in the control of apoptosis-driven hair follicle regression." In: *The FASEB Journal* 14.13, pp. 1931–1942. DOI: 10.1096/fj.99-0930com.

Botchkarev, Vladimir A. et al. (2006). "Neurotrophins in Skin Biology and Pathology." In: *Journal of Investigative Dermatology* 126.8, pp. 1719–1727. DOI: 10.1038/sj.jid.5700270.

Boughanem, Hatim et al. (2021). "Potential Role of Insulin Growth-Factor-Binding Protein 2 as Therapeutic Target for Obesity-Related Insulin Resistance." In: *International Journal of Molecular Sciences* 22.3, p. 1133. DOI: 10.3390/ijms22031133.

Bouzin, Caroline et al. (2007). "Effects of Vascular Endothelial Growth Factor on the Lymphocyte-Endothelium Interactions: Identification of Caveolin-1 and Nitric Oxide as Control Points of Endothelial Cell Anergy." In: *The Journal of Immunology* 178.3, pp. 1505–1511. DOI: 10.4049/jimmunol.178.3.1505.

Bowie, Michelle L., Catherine Ibarra, and Victoria L. Seewalt (2008). "IRF-1 Promotes Apoptosis in p53-damaged Basal-type Human Mammary Epithelial Cells: A Model for Early Basal-type Mammary Carcinogenesis." In: *Hormonal Carcinogenesis V*. Springer New York, pp. 367–374. DOI: 10.1007/978-0-387-69080-3_35.

Boxman, Ingeborg L. et al. (1996). "Role of fibroblasts in the regulation of proinflammatory interleukin IL-1, IL-6 and IL-8 levels induced by keratinocyte-derived IL-1." In: *Archives of Dermatological Research* 288.7, 391–398. DOI: 10.1007/bf02507108.

Bozaoglu, Kiymet et al. (2007). "Chemerin Is a Novel Adipokine Associated with Obesity and Metabolic Syndrome." In: *Endocrinology* 148.10, pp. 4687–4694. DOI: 10.1210/en.2007-0175.

Bray, Nicolas L et al. (2016). "Near-optimal probabilistic RNA-seq quantification." In: *Nature Biotechnology* 34.5, 525–527. DOI: 10.1038/nbt.3519.

Brennecke, Philip et al. (2013). "Accounting for technical noise in single-cell RNA-seq experiments." In: *Nature Methods* 10.11, 1093–1095. DOI: 10.1038/nmeth.2645.

Bretaud, Sandrine et al. (2020). "Collagen XV, a multifaceted multiplexin present across tissues and species." In: *Matrix Biology Plus* 6-7, p. 100023. DOI: 10.1016/j.mbplus.2020.100023.

Brigelius-Flohé, Regina (2006). "Glutathione peroxidases and redox-regulated transcription factors." In: *Biological Chemistry* 387.10/11. DOI: 10.1515/bc.2006.166.

Brink, Susanne C van den et al. (2017). "Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations." In: *Nature Methods* 14.10, 935–936. DOI: 10.1038/nmeth.4437.

Broide, David H. and P. Sriramarao (2014). "Cellular Adhesion in Inflammation." In: *Middleton's Allergy*. Elsevier, pp. 83–97. DOI: 10.1016/b978-0-323-08593-9.00007-3.

Bronaugh, Robert L., Raymond F. Stewart, and Elaine R. Congdon (1982). "Methods for in vitro percutaneous absorption studies II. animal models for human skin." In: *Toxicology and Applied Pharmacology* 62.3, 481–488. DOI: 10.1016/0041-008x(82)90149-1.

Brownlie, Rebecca J. and Rose Zamoyska (2013). "T cell receptor signalling networks: branched, diversified and bounded." In: *Nature Reviews Immunology* 13.4, pp. 257–269. DOI: 10.1038/nri3403.

Bryce, Steven A. et al. (2016). "ACKR4 on Stromal Cells Scavenges CCL19 To Enable CCR7-Dependent Trafficking of APCs from Inflamed Skin to Lymph Nodes." In: *The Journal of Immunology* 196.8, pp. 3341–3353. DOI: 10.4049/jimmunol.1501542.

Buechler, Matthew B. et al. (2021). "Cross-tissue organization of the fibroblast lineage." In: *Nature* 593.7860, 575–579. DOI: 10.1038/s41586-021-03549-5.

Bui, Triet M, Hannah L Wiesolek, and Ronen Sumagin (2020). "ICAM-1: A master regulator of cellular responses in inflammation, injury resolution, and tumorigenesis." In: *Journal of Leukocyte Biology* 108.3, pp. 787–799. DOI: 10.1002/jlb.2mr0220-549r.

Burja, Blaž et al. (2022). "An optimized tissue dissociation protocol for single-cell RNA sequencing analysis of fresh and cultured human skin biopsies." In: *Frontiers in Cell and Developmental Biology* 10. DOI: 10.3389/fcell.2022.872688.

Butkevich, Eugenia et al. (2015). "Drebrin-like protein DBN-1 is a sarcomere component that stabilizes actin filaments during muscle contraction." In: *Nature Communications* 6.1. DOI: 10.1038/ncomms8523.

Butler, Andrew et al. (2018). "Integrating single-cell transcriptomic data across different conditions, technolo-

gies, and species." In: *Nature Biotechnology* 36.5, 411–420. DOI: 10.1038/nbt.4096.

Bzdok, Danilo, Naomi Altman, and Martin Krzywinski (2018). "Statistics versus machine learning." In: *Nature Methods* 15.4, 233–234. DOI: 10.1038/nmeth.4642.

Böhm, Jan Niklas, Philipp Berens, and Dmitry Kobak (2022). "Attraction-Repulsion Spectrum in Neighbor Embeddings." In: *Journal of Machine Learning Research* 23.95, pp. 1–32.

Büttner, Maren et al. (2018). "A test metric for assessing single-cell RNA-seq batch correction." In: *Nature Methods* 16.1, 43–49. DOI: 10.1038/s41592-018-0254-1.

Cabral-Pacheco, Griselda A et al. (2020). "The Roles of Matrix Metalloproteinases and Their Inhibitors in Human Diseases." In: *International Journal of Molecular Sciences* 21.24, p. 9739. DOI: 10.3390/ijms21249739.

Cai, Yumei et al. (2017). "Downregulation of β-catenin blocks fibrosis via Wnt2 signaling in human keloid fibroblasts." In: *Tumor Biology* 39.6, p. 101042831770742. DOI: 10.1177/1010428317707423.

Cao, Junyue et al. (2017). "Comprehensive single-cell transcriptional profiling of a multicellular organism." In: *Science* 357.6352, 661–667. DOI: 10.1126/science.aam8940.

Cao, Junyue et al. (2019). "The single-cell transcriptional landscape of mammalian organogenesis." In: *Nature* 566.7745, 496–502. DOI: 10.1038/s41586-019-0969-x.

Capolupo, Laura et al. (2022). "Sphingolipids control dermal fibroblast heterogeneity." In: *Science* 376.6590. DOI: 10.1126/science.abh1623.

Carithers, Latarsha J. and Helen M. Moore (2015). "The genotype-tissue expression (gtex) project." In: *Biopreservation and Biobanking* 13.5, 307–308. DOI: 10.1089/bio.2015.29031.hmm.

Carmon, Kendra S. et al. (2011). "R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/β-catenin signaling." In: *Proceedings of the National Academy of Sciences* 108.28, 11452–11457. DOI: 10.1073/pnas.1106083108.

Carow, Berit and Martin E. Rottenberg (2014). "SOCS3, a Major Regulator of Infection and Inflammation." In: *Frontiers in Immunology* 5. DOI: 10.3389/fimmu.2014.00058.

Cash, Jenna L. et al. (2008). "Synthetic chemerin-derived peptides suppress inflammation through ChemR23." In: *Journal of Experimental Medicine* 205.4, pp. 767–775. DOI: 10.1084/jem.20071601.

Caterina, Michael J. et al. (1997). "The capsaicin receptor: A heat-activated ion channel in the pain pathway." In: *Nature* 389.6653, 816–824. DOI: 10.1038/39807.

Chang, Howard Y. et al. (2002). "Diversity, topographic differentiation, and positional memory in human fibroblasts." In: *Proceedings of the National Academy of Sciences* 99.20, 12877–12882. DOI: 10.1073/pnas.162488599.

Chang, Jing et al. (2022). "IL-32 promotes the occurrence of atopic dermatitis by activating the JAK1/microRNA-155 axis." In: *Journal of Translational Medicine* 20.1. DOI: 10.1186/s12967-022-03375-x.

Chang, Susan M. and Derek R. Johnson (2012). "Biologic therapy for malignant glioma." In: *Brain Tumors*. Elsevier, pp. 102–113. DOI: 10.1016/b978-0-443-06967-3.00006-5.

Chanoki, M. et al. (1988). "Immunohistochemical localization of type V collagen in normal human skin." In: *Archives of Dermatological Research* 280.3, pp. 145–151. DOI: 10.1007/bf00456844.

Chari, Tara and Lior Pachter (2021). "The Specious Art of Single-Cell Genomics." In: *bioRxiv*. DOI: 10.1101/2021.08.25.457696.

Chaudet, Kristine M. et al. (2020). "Wnt Signaling Pathway Proteins in Scar, Hypertrophic Scar, and Keloid: Evidence for a Continuum?" In: *The American Journal of Dermatopathology* 42.11, pp. 842–847. DOI: 10.1097/dad.0000000000001661.

Chawla, Ajay et al. (2001). "A PPARγ-LXR-ABCA1 Pathway in Macrophages Is Involved in Cholesterol Efflux and Atherogenesis." In: *Molecular Cell* 7.1, pp. 161–171. DOI: 10.1016/s1097-2765(01)00164-2.

Chen, Bing et al. (2021a). "Single Cell Transcriptome Data Analysis Defines the Heterogeneity of Peripheral Nerve Cells in Homeostasis and Regeneration." In: *Frontiers in Cellular Neuroscience* 15. DOI: 10.3389/fncel.2021.624826.

Chen, Chih-Chiang et al. (2015a). "Organ-Level Quorum Sensing Directs Regeneration in Hair Stem Cell Populations." In: *Cell* 161.2, pp. 277–290. DOI: 10.1016/j.cell.2015.02.016.

Chen, Chun-Chieh et al. (2014). "Shisa3 Is Associated with Prolonged Survival through Promoting β-Catenin Degradation in Lung Cancer." In: *American Journal of Respiratory and Critical Care Medicine* 190.4, pp. 433–444. DOI: 10.1164/rccm.201312-2256oc.

Chen, Demeng et al. (2012). "Dermal β-catenin activity in response to epidermal Wnt ligands is required for fibroblast proliferation and hair follicle initiation." In: *Development* 139.8, pp. 1522–1533. DOI: 10.1242/dev.076463.

Chen, Guobao et al. (2018a). "SCA-1+ cardiac fibroblasts promote development of heart failure." In: *European Journal of Immunology* 48.9, 1522–1538. DOI: 10.1002/eji.201847583.

Chen, Huaping et al. (2016). "Mechanosensing by the α6-integrin confers an invasive fibroblast phenotype and mediates lung fibrosis." In: *Nature Communications* 7.1. DOI: 10.1038/ncomms12564.

Chen, Jian et al. (2017). "Neuron and microglia/macrophage-derived FGF10 activate neuronal FGFR2/PI3K/Akt signaling and inhibit microglia/macrophages TLR4/NF-κB-dependent neuroinflammation to improve functional recovery after spinal cord injury." In: *Cell Death & Disease* 8.10, e3090–e3090. DOI: 10.1038/cddis.2017.490.

Chen, Qiang et al. (2015b). "CXCR7 Mediates Neural Progenitor Cells Migration to CXCL12 Independent of CXCR4." In: *Stem Cells* 33.8, pp. 2574–2585. DOI: 10.1002/stem.2022.

Chen, Wenan et al. (2018b). "UMI-count modeling and differential expression analysis FOR single-cell RNA sequencing." In: *Genome Biology* 19.1. DOI: 10.1186/s13059-018-1438-9.

Chen, ZongAn et al. (2021b). "Dysregulation of DPP4-CXCL12 Balance by TGF-β1/SMAD Pathway Promotes $CXCR4^+$ Inflammatory Cell Infiltration in Keloid Scars." In: *Journal of Inflammation Research* Volume 14, pp. 4169–4180. DOI: 10.2147/jir.s326385.

Cheng, Chun-Chun et al. (2018a). "Hair follicle epidermal stem cells define a niche for tactile sensation." In: *eLife* 7. DOI: 10.7554/elife.38883.

Cheng, Hung-Chi, Mossaad Abdel-Ghany, and Bendicht U. Pauli (2003). "A Novel Consensus Motif in Fibronectin Mediates Dipeptidyl Peptidase IV Adhesion and Metastasis." In: *Journal of Biological Chemistry* 278.27, pp. 24600–24607. DOI: 10.1074/jbc.m303424200.

Cheng, Hung-Wei et al. (2018b). "CCL19-producing fibroblastic stromal cells restrain lung carcinoma growth by promoting local antitumor T-cell responses." In: *Journal of Allergy and Clinical Immunology* 142.4, 1257–1271.e4. DOI: 10.1016/j.jaci.2017.12.998.

Cheng, Jeffrey B. et al. (2018c). "Transcriptional programming of normal and inflamed human epidermis at single-cell resolution." In: *Cell Reports* 25.4, 871–883. DOI: 10.1016/j.celrep.2018.09.006.

Chiu, Bo-Chin et al. (2004). "Impaired Lung Dendritic Cell Activation in CCR2 Knockout Mice." In: *The American Journal of Pathology* 165.4, pp. 1199–1209. DOI: 10.1016/s0002-9440(10)63380-9.

Cho, Hoon et al. (2006). "Role of glutamine 148 of human 15-hydroxyprostaglandin dehydrogenase in catalytic oxidation of prostaglandin E2." In: *Bioorganic & Medicinal Chemistry* 14.19, pp. 6486–6491. DOI: 10.1016/j.bmc.2006.06.030.

Chong, Sy Giin et al. (2019). "Fibrocytes and fibroblasts—where are we now." In: *The International Journal of Biochemistry & Cell Biology* 116, p. 105595. DOI: 10.1016/j.biocel.2019.105595.

Choo, Hyo-Jung et al. (2017). "A tale of two niches: differential functions for VCAM-1 in satellite cells under basal and injured conditions." In: *American Journal of Physiology-Cell Physiology* 313.4, pp. C392–C404. DOI: 10.1152/ajpcell.00119.2017.

Choudhury, Saurav Roy et al. (2019). "Dipeptidase-1 Is an Adhesion Receptor for Neutrophil Recruitment in Lungs and Liver." In: *Cell* 178.5, 1205–1221.e17. DOI: 10.1016/j.cell.2019.07.017.

Chovarda, Eleni et al. (2021). "The role of prostaglandins in androgenetic alopecia." In: *International Journal of Dermatology* 60.6, pp. 730–735. DOI: 10.1111/ijd.15378.

Chovatiya, Gopal et al. (2021). "High-resolution single-cell transcriptomics reveals heterogeneity of self-renewing hair follicle stem cells." In: *Experimental Dermatology* 30.4, pp. 457–471. DOI: 10.1111/exd.14262.

Choy, Meng S. et al. (2015). "Structural and Functional Analysis of the GADD34:PP1 eIF2α Phosphatase." In: *Cell Reports* 11.12, pp. 1885–1891. DOI: 10.1016/j.celrep.2015.05.043.

Christensen, Jon and V Prasad Shastri (2015). "Matrix-metalloproteinase-9 is cleaved and activated by Cathepsin K." In: *BMC Research Notes* 8.1. DOI: 10.1186/s13104-015-1284-8.

Chung, Sung Soo et al. (2009). "Glutathione Peroxidase 3 Mediates the Antioxidant Effect of Peroxisome Proliferator-Activated Receptor γ in Human Skeletal Muscle Cells." In: *Molecular and Cellular Biology* 29.1, pp. 20–30. DOI: 10.1128/mcb.00544-08.

Cibelli, Jose et al. (2013). "Strategies for improving animal models for Regenerative Medicine." In: *Cell Stem Cell* 12.3, 271–274. DOI: 10.1016/j.stem.2013.01.004.

Coifman, R. R. et al. (2005). "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion Maps." In: *Proceedings of the National Academy of Sciences* 102.21, 7426–7431. DOI: 10.1073/pnas.0500334102.

Cole, Michael B. et al. (2019). "Performance assessment and selection of normalization procedures for single-cell RNA-seq." In: *Cell Systems* 8.4. DOI: 10.1016/j.cels.2019.03.010.

Collins, Charlotte A., Kai Kretzschmar, and Fiona M. Watt (2011). "Reprogramming adult dermis to a neonatal state through epidermal activation of β-catenin." In: *Development* 138.23, 5189–5199. DOI: 10.1242/dev.064592.

Collins, Charlotte A. and Fiona M. Watt (2008). "Dynamic regulation of retinoic acid-binding proteins in developing, adult and neoplastic skin reveals roles for β-catenin and notch signalling." In: *Developmental Biology* 324.1, 55–67. DOI: 10.1016/j.ydbio.2008.08.034.

Comes, Nuria, LaKisha K. Buie, and Teresa Borras (2010). "Evidence for a role of angiopoietin-like 7 (ANGPTL7) in extracellular matrix formation of the human trabecular meshwork: implications for glaucoma." In: *Genes to Cells* 16.2, pp. 243–259. DOI: 10.1111/j.1365-2443.2010.01483.x.

Cominelli, Antoine et al. (2014). "A Unique C-terminal Domain Allows Retention of Matrix Metalloproteinase-27 in the Endoplasmic Reticulum." In: *Traffic* 15.4, pp. 401–417. DOI: 10.1111/tra.12149.

Cong, Yongzheng et al. (2021). "Ultrasensitive single-cell proteomics workflow identifies 1000 protein groups per mammalian cell." In: *Chemical Science* 12.3, 1001–1006. DOI: 10.1039/d0sc03636f.

Conroy, Kylie P., Laura J. Kitto, and Neil C. Henderson (2016). "αv integrins: Key regulators of tissue fibrosis." In: *Cell*

and Tissue Research 365.3, 511–519. DOI: 10.1007/s00441-016-2407-9.

Corda, G and A Sala (2017). "Non-canonical WNT/PCP signalling in cancer: Fzd6 takes centre stage." In: *Oncogenesis* 6.7, e364–e364. DOI: 10.1038/oncsis.2017.69.

Correa-Gallegos, Donovan and Yuval Rinkevich (2021). "Cutting into wound repair." In: *The FEBS Journal*. DOI: 10.1111/febs.16078.

Costin, Gertrude-E. and Vincent J. Hearing (2007). "Human skin pigmentation: Melanocytes modulate skin color in response to stress." In: *The FASEB Journal* 21.4, 976–994. DOI: 10.1096/fj.06-6649rev.

Cragnolini, Andrea B. and Wilma J. Friedman (2008). "The function of p75NTR in glia." In: *Trends in Neurosciences* 31.2, pp. 99–104. DOI: 10.1016/j.tins.2007.11.005.

Craig, Matt J. and Robert D. Loberg (2006). "CCL2 (Monocyte Chemoattractant Protein-1) in cancer bone metastases." In: *Cancer and Metastasis Reviews* 25.4, pp. 611–619. DOI: 10.1007/s10555-006-9027-x.

Crick, Francis (1970). "Central Dogma of Molecular Biology." In: *Nature* 227.5258, 561–563. DOI: 10.1038/227561a0.

Csiszar, Katalin (2001). "Lysyl oxidases: A novel multifunctional amine oxidase family." In: *Progress in Nucleic Acid Research and Molecular Biology*, 1–32. DOI: 10.1016/s0079-6603(01)70012-8.

Ctortecka, Claudia et al. (2021). "An automated workflow for multiplexed single-cell proteomics sample preparation at unprecedented sensitivity." In: *bioRxiv*. DOI: 10.1101/2021.04.14.439828.

Culley, Oliver J. et al. (2021). "Differential Expression of Insulin-Like Growth Factor 1 and Wnt Family Member 4 Correlates With Functional Heterogeneity of Human Dermal Fibroblasts." In: *Frontiers in Cell and Developmental Biology* 9. DOI: 10.3389/fcell.2021.628039.

Dai, Gucan et al. (2016). "Osmoregulatory inositol transporter SMIT1 modulates electrical activity by adjusting PI(4,5)P 2 levels." In: *Proceedings of the National Academy of Sciences* 113.23. DOI: 10.1073/pnas.1606348113.

Damen, Michelle S.M.A. et al. (2018). "Interleukin-32 upregulates the expression of ABCA1 and ABCG1 resulting in reduced intracellular lipid concentrations in primary human hepatocytes." In: *Atherosclerosis* 271, pp. 193–202. DOI: 10.1016/j.atherosclerosis.2018.02.027.

Darby, Ian A. and Tim D. Hewitson (2007). "Fibroblast differentiation in wound healing and fibrosis." In: *International Review of Cytology*, 143–179. DOI: 10.1016/s0074-7696(07)57004-x.

DasGupta, R. and E. Fuchs (1999). "Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation." In: *Development* 126.20, 4557–4568. DOI: 10.1242/dev.126.20.4557.

Davoodi, Jamshid et al. (2007). "The Simpson–Golabi–Behmel syndrome causative Glypican-3, binds to and inhibits the dipeptidyl peptidase activity of CD26." In: *PROTEOMICS* 7.13, pp. 2300–2310. DOI: 10.1002/pmic.200600654.

Deckx, Sophie, Stephane Heymans, and Anna-Pia Papageorgiou (2016). "The diverse functions of osteoglycin: a deceitful dwarf, or a master regulator of disease?" In: *The FASEB Journal* 30.8, pp. 2651–2661. DOI: 10.1096/fj.201500096r.

Dees, Clara et al. (2020). "TGF-β–induced epigenetic deregulation of SOCS3 facilitates STAT3 signaling to promote fibrosis." In: *Journal of Clinical Investigation* 130.5, pp. 2347–2363. DOI: 10.1172/jci122462.

Demaria, Marco et al. (2014). "An essential role for senescent cells in optimal wound healing through secretion of PDGF-AA." In: *Developmental Cell* 31.6, 722–733. DOI: 10.1016/j.devcel.2014.11.012.

Dempsey, Timothy M. et al. (2019). "Clinical effectiveness of antifibrotic medications for idiopathic pulmonary fibrosis." In: *American Journal of Respiratory and Critical Care Medicine* 200.2, 168–174. DOI: 10.1164/rccm.201902-0456oc.

Denda, Mitsuhiro et al. (2007). "Epidermal keratinocytes as the forefront of the sensory system." In: *Experimental Dermatology* 16.3, 157–161. DOI: 10.1111/j.1600-0625.2006.00529.x.

Deng, Cheng-Cheng et al. (2021). "Single-cell RNA-seq reveals fibroblast heterogeneity and increased mesenchymal fibroblasts in human fibrotic skin diseases." In: *Nature Communications* 12.1. DOI: 10.1038/s41467-021-24110-y.

Deng, Chengcheng et al. (2018). "TNFRSF19 Inhibits TGFβ Signaling through Interaction with TGFβ Receptor Type I to Promote Tumorigenesis." In: *Cancer Research* 78.13, pp. 3469–3483. DOI: 10.1158/0008-5472.can-17-3205.

Deng, Dong et al. (2015). "Molecular basis of ligand recognition and transport by glucose transporters." In: *Nature* 526.7573, pp. 391–396. DOI: 10.1038/nature14655.

Denisenko, Elena et al. (2020). "Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows." In: *Genome Biology* 21.1. DOI: 10.1186/s13059-020-02048-6.

Dennis, Edward A. and Paul C. Norris (2015). "Eicosanoid storm in infection and inflammation." In: *Nature Reviews Immunology* 15.8, pp. 511–523. DOI: 10.1038/nri3859.

Deretic, Nikola et al. (2021). "The Actin-Disassembly Protein Glia Maturation Factor γ Enhances Actin Remodeling and B Cell Antigen Receptor Signaling at the Immune Synapse." In: *Frontiers in Cell and Developmental Biology* 9. DOI: 10.3389/fcell.2021.647063.

Desnoyers, Luc, David Arnott, and Diane Pennica (2001). "WISP-1 Binds to Decorin and Biglycan." In: *Journal of Biological Chemistry* 276.50, pp. 47599–47607. DOI: 10.1074/jbc.m108339200.

Dho, So Hee, Jae Cheong Lim, and Lark Kyun Kim (2018). "Beyond the Role of CD55 as a Complement Component." In: *Immune Network* 18.1. DOI: 10.4110/in.2018.18.e11.

Dhouailly, D. and P. Sengel (1975). "Propriétés phanérogènes des cellules dermiques de peau glabre d'Oiseau ou de Mammifère." In: *Comptes rendus de l'Académie des Sciences*. 14th ser. 281, 1007–1010.

Dijk, David van et al. (2018). "Recovering gene interactions from single-cell data using data diffusion." In: *Cell* 174.3. DOI: 10.1016/j.cell.2018.05.061.

Ding, Jiarui et al. (2020). "Systematic comparison of single-cell and single-nucleus RNA-sequencing methods." In: *Nature Biotechnology* 38, pp. 737–746. DOI: 10.1038/s41587-020-0465-8.

Dobin, Alexander et al. (2012). "Star: Ultrafast universal RNA-seq aligner." In: *Bioinformatics* 29.1, 15–21. DOI: 10.1093/bioinformatics/bts635.

Docheva, Denitsa et al. (2005). "Tenomodulin Is Necessary for Tenocyte Proliferation and Tendon Maturation." In: *Molecular and Cellular Biology* 25.2, pp. 699–705. DOI: 10.1128/mcb.25.2.699-705.2005.

Doeing, Diana C, Jessica L Borowicz, and Elahé T Crockett (2003). "Gender dimorphism in differential peripheral blood leukocyte counts in mice using cardiac, tail, foot, and saphenous vein puncture methods." In: *BMC Clinical Pathology* 3.1. DOI: 10.1186/1472-6890-3-3.

Dolgalev, Igor and Anastasia N. Tikhonova (2021). "Connecting the dots: Resolving the bone marrow niche heterogeneity." In: *Frontiers in Cell and Developmental Biology* 9. DOI: 10.3389/fcell.2021.622519.

Doliana, Roberto et al. (2001). "Isolation and Characterization of EMILIN-2, a New Component of the Growing EMILINs Family and a Member of the EMI Domain-containing Superfamily." In: *Journal of Biological Chemistry* 276.15, pp. 12003–12011. DOI: 10.1074/jbc.m011591200.

Dolivo, David et al. (2021). "The Nax (SCN7A) channel: an atypical regulator of tissue homeostasis and disease." In: *Cellular and Molecular Life Sciences* 78.14, pp. 5469–5488. DOI: 10.1007/s00018-021-03854-2.

Domcke, Silvia and Jay Shendure (2023). "A reference cell tree will serve science better than a reference cell atlas." In: *Cell* 186.6, pp. 1103–1114. DOI: 10.1016/j.cell.2023.02.016.

Domínguez Conde, C. et al. (2022). "Cross-tissue immune cell analysis reveals tissue-specific features in humans." In: *Science* 376.6594. DOI: 10.1126/science.abl5197.

Donati, Giacomo et al. (2014). "Epidermal WNT/ β-catenin signaling regulates adipocyte differentiation via secretion of adipogenic factors." In: *Proceedings of the National Academy of Sciences* 111.15. DOI: 10.1073/pnas.1312880111.

Donato, Rosario et al. (2009). "S100B's double life: Intracellular regulator and extracellular signal." In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1793.6, pp. 1008–1022. DOI: 10.1016/j.bbamcr.2008.11.009.

Dong, Wei, Charikar Moses, and Kai Li (2011). "Efficient K-nearest neighbor graph construction for generic similarity measures." In: *Proceedings of the 20th international conference on World wide web*. DOI: 10.1145/1963405.1963487.

Donà, Erika et al. (2013). "Directional tissue migration through a self-generated chemokine gradient." In: *Nature* 503.7475, pp. 285–289. DOI: 10.1038/nature12635.

Dornan, David et al. (2004). "Interferon Regulatory Factor 1 Binding to p300 Stimulates DNA-Dependent Acetylation of p53." In: *Molecular and Cellular Biology* 24.22, pp. 10083–10098. DOI: 10.1128/mcb.24.22.10083-10098.2004.

Dorschner, Robert A. et al. (2020). "ECRG4 regulates neutrophil recruitment and CD44 expression during the inflammatory response to injury." In: *Science Advances* 6.11. DOI: 10.1126/sciadv.aay0518.

Dostert, Catherine et al. (2019). "The TNF Family of Ligands and Receptors: Communication Modules in the Immune System and Beyond." In: *Physiological Reviews* 99.1, pp. 115–160. DOI: 10.1152/physrev.00045.2017.

Driskell, Ryan R. et al. (2009). "Sox2-positive dermal papilla cells specify hair follicle type in mammalian epidermis." In: *Development* 136.16, 2815–2823. DOI: 10.1242/dev.038620.

Driskell, Ryan R. et al. (2011). "Hair follicle dermal papilla cells at a glance." In: *Journal of Cell Science* 124.8, 1179–1182. DOI: 10.1242/jcs.082446.

Driskell, Ryan R. et al. (2012). "Clonal growth of dermal papilla cells in hydrogels reveals intrinsic differences between sox2-positive and -negative cells in vitro and in vivo." In: *Journal of Investigative Dermatology* 132.4, 1084–1093. DOI: 10.1038/jid.2011.428.

Driskell, Ryan R. et al. (2013). "Distinct fibroblast lineages determine dermal architecture in skin development and Repair." In: *Nature* 504.7479, 277–281. DOI: 10.1038/nature12783.

Driskell, Ryan R. et al. (2014). "Defining dermal adipose tissue." In: *Experimental Dermatology* 23.9, 629–631. DOI: 10.1111/exd.12450.

Dulbecco, R et al. (1983). "Functional changes of intermediate filaments in fibroblastic cells revealed by a monoclonal antibody." In: *Proceedings of the National Academy of Sciences* 80.7, 1915–1918. DOI: 10.1073/pnas.80.7.1915.

Dulos, John et al. (2005). "CYP7B expression and activity in fibroblast-like synoviocytes from patients with rheumatoid arthritis: Regulation by proinflammatory cytokines." In: *Arthritis & Rheumatism* 52.3, pp. 770–778. DOI: 10.1002/art.20950.

Duncan, Matthew R and Brian Berman (1991). "Stimulation of Collagen and Glycosaminoglycan Production in Cul-

tured Human Adult Dermal Fibroblasts by Recombinant Human Interleukin 6." In: *Journal of Investigative Dermatology* 97.4, pp. 686–692. DOI: 10.1111/1523-1747.ep12483971.

Duval, Mathias Marie (1879). *Atlas d'Embryologiee*. Masson, G.

Eberwine, J et al. (1992). "Analysis of gene expression in single live neurons." In: *Proceedings of the National Academy of Sciences* 89.7, 3010–3014. DOI: 10.1073/pnas.89.7.3010.

Efremova, Mirjana et al. (2020). "CellPhoneDB: Inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes." In: *Nature Protocols* 15.4, 1484–1506. DOI: 10.1038/s41596-020-0292-x.

Eguiluz, Roberto C. Andresen et al. (2015). "Fibronectin mediates enhanced wear protection of lubricin during shear." In: *Biomacromolecules* 16.9, pp. 2884–2894. DOI: 10.1021/acs.biomac.5b00810.

Eissmann, Moritz F., Michael Buchert, and Matthias Ernst (2020). "IL33 and Mast Cells—The Key Regulators of Immune Responses in Gastrointestinal Cancers?" In: *Frontiers in Immunology* 11. DOI: 10.3389/fimmu.2020.01389.

Eldar, Avigdor and Michael B. Elowitz (2010). "Functional roles for noise in genetic circuits." In: *Nature* 467.7312, 167–173. DOI: 10.1038/nature09326.

Elmansi, Ahmed M. et al. (2022). "DPP4-Truncated CXCL12 Alters CXCR4/ACKR3 Signaling, Osteogenic Cell Differentiation, Migration, and Senescence." In: *ACS Pharmacology & Translational Science* 6.1, pp. 22–39. DOI: 10.1021/acsptsci.2c00040.

Elmentaite, Rasa et al. (2020). "Single-cell sequencing of developing human gut reveals transcriptional links to childhood crohn's disease." In: *Developmental Cell* 55.6. DOI: 10.1016/j.devcel.2020.11.010.

Elmentaite, Rasa et al. (2022). "Single-cell atlases: Shared and tissue-specific cell types across human organs." In: *Nature Reviews Genetics* 23.7, 395–410. DOI: 10.1038/s41576-022-00449-w.

Elshabrawy, Hatem A. et al. (2018). "IL-11 facilitates a novel connection between RA joint fibroblasts and endothelial cells." In: *Angiogenesis* 21.2, pp. 215–228. DOI: 10.1007/s10456-017-9589-y.

Elyada, Ela et al. (2019). "Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts." In: *Cancer Discovery* 9.8, pp. 1102–1123. DOI: 10.1158/2159-8290.cd-19-0094.

Emrich, Scott J. et al. (2006). "Gene discovery and annotation using LCM-454 transcriptome sequencing." In: *Genome Research* 17.1, 69–73. DOI: 10.1101/gr.5145806.

Eng, Chee-Huat Linus et al. (2019). "Transcriptome-scale super-resolved imaging in tissues by RNA seqfish+." In: *Nature* 568.7751, 235–239. DOI: 10.1038/s41586-019-1049-y.

Enshell-Seijffers, David et al. (2010). "β-catenin Activity in the Dermal Papilla Regulates Morphogenesis and Regeneration of Hair." In: *Developmental Cell* 18.4, pp. 633–642. DOI: 10.1016/j.devcel.2010.01.016.

Eraslan, Gökcen et al. (2018). "Single cell RNA-seq denoising using a deep count autoencoder." In: DOI: 10.1101/300681.

– (2019). "Single-cell RNA-seq denoising using a deep count autoencoder." In: *Nature Communications* 10.1. DOI: 10.1038/s41467-018-07931-2.

Esposito, Mark et al. (2021). "TGF-β-induced DACT1 biomolecular condensates repress Wnt signalling to promote bone metastasis." In: *Nature Cell Biology* 23.3, pp. 257–267. DOI: 10.1038/s41556-021-00641-w.

Everts, Helen B. et al. (2004). "Hair Cycle-Specific Immunolocalization of Retinoic Acid Synthesizing Enzymes Aldh1a2 and Aldh1a3 Indicate Complex Regulation." In: *Journal of Investigative Dermatology* 123.2, pp. 258–263. DOI: 10.1111/j.0022-202x.2004.23223.x.

Everts, Helen B. et al. (2007). "Immunolocalization of Enzymes, Binding Proteins, and Receptors Sufficient for Retinoic Acid Synthesis and Signaling During the Hair Cycle." In: *Journal of Investigative Dermatology* 127.7, pp. 1593–1604. DOI: 10.1038/sj.jid.5700753.

Fafilek, Bohumil et al. (2013). "Troy, a Tumor Necrosis Factor Receptor Family Member, Interacts With Lgr5 to Inhibit Wnt Signaling in Intestinal Stem Cells." In: *Gastroenterology* 144.2, pp. 381–391. DOI: 10.1053/j.gastro.2012.10.048.

Fang, Feng et al. (2011). "The Early Growth Response Gene Egr2 (Alias Krox20) Is a Novel Transcriptional Target of Transforming Growth Factor-β that Is Up-Regulated in Systemic Sclerosis and Mediates Profibrotic Responses." In: *The American Journal of Pathology* 178.5, pp. 2077–2090. DOI: 10.1016/j.ajpath.2011.01.035.

Fang, Robert C. and Thomas A. Mustoe (2008). "Animal models of wound healing: Uility in Transgenic Mice." In: *Journal of Biomaterials Science, Polymer Edition* 19.8, 989–1005. DOI: 10.1163/156856208784909327.

Fang, Wei Bin et al. (2015). "TGF-β Negatively Regulates CXCL1 Chemokine Expression in Mammary Fibroblasts through Enhancement of Smad2/3 and Suppression of HGF/c-Met Signaling Mechanisms." In: *PLOS ONE* 10.8. Ed. by Yi Li, e0135063. DOI: 10.1371/journal.pone.0135063.

Fang, Zhuoqing et al. (2021). *zqfang/GSEApy: gseapy-v0.10.3*. Version v0.10.3. DOI: 10.5281/zenodo.4553090.

Faraahi, Z. et al. (2019). "Sostdc1: A soluble BMP and Wnt antagonist that is induced by the interaction between myeloma cells and osteoblast lineage cells." In: *Bone* 122, pp. 82–92. DOI: 10.1016/j.bone.2019.02.012.

Farage, Miranda A., Kenneth W. Miller, and Howard I. Maibach (2010). "Degenerative changes in aging skin." In:

Textbook of Aging Skin, 25–35. DOI: 10.1007/978-3-540-89656-2_4.

Farina, Giuseppina et al. (2006). "Cartilage oligomeric matrix protein is overexpressed by scleroderma dermal fibroblasts." In: *Matrix Biology* 25.4, pp. 213–222. DOI: 10.1016/j.matbio.2006.01.007.

Favier, Benoit et al. (2006). "Neuropilin-2 interacts with VEGFR-2 and VEGFR-3 and promotes human endothelial cell survival and migration." In: *Blood* 108.4, pp. 1243–1250. DOI: 10.1182/blood-2005-11-4447.

Ferdous, Zannatul et al. (2007). "Decorin-transforming Growth Factor-β Interaction Regulates Matrix Organization and Mechanical Characteristics of Three-dimensional Collagen Matrices." In: *Journal of Biological Chemistry* 282.49, pp. 35887–35898. DOI: 10.1074/jbc.m705180200.

Fernández, Patricia et al. (2013). "Extracellular Generation of Adenosine by the Ectonucleotidases CD39 and CD73 Promotes Dermal Fibrosis." In: *The American Journal of Pathology* 183.6, pp. 1740–1746. DOI: 10.1016/j.ajpath.2013.08.024.

Festa, Eric et al. (2011). "Adipocyte lineage cells contribute to the skin stem cell niche to drive hair cycling." In: *Cell* 146.5, 761–771. DOI: 10.1016/j.cell.2011.07.019.

Filippou, Panagiota S., George S. Karagiannis, and Anastasia Constantinidou (2019). "Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target." In: *Oncogene* 39.10, pp. 2040–2054. DOI: 10.1038/s41388-019-1124-8.

Filippou, Panagiota S. et al. (2020). "Kallikrein-related peptidases protein expression in lymphoid tissues suggests potential implications in immune response." In: *Clinical Biochemistry* 77, pp. 41–47. DOI: 10.1016/j.clinbiochem.2019.12.015.

Finak, Greg et al. (2015). "MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data." In: *Genome Biology* 16.1. DOI: 10.1186/s13059-015-0844-5.

Finnegan, Alex et al. (2019). "Single-cell transcriptomics reveals spatial and temporal turnover of keratinocyte differentiation regulators." In: *Frontiers in Genetics* 10. DOI: 10.3389/fgene.2019.00775.

Fletcher, Anne L., Sophie E. Acton, and Konstantin Knoblich (2015). "Lymph node fibroblastic reticular cells in health and disease." In: *Nature Reviews Immunology* 15.6, pp. 350–361. DOI: 10.1038/nri3846.

Fontaine, J., C. Le Lièvre, and N.M. Le Douarin (1977). "What is the developmental fate of the neural crest cells which migrate into the pancreas in the avian embryo?" In: *General and Comparative Endocrinology* 33.3, 394–404. DOI: 10.1016/0016-6480(77)90055-7.

Frascoli, Michela et al. (2023). "Skin $\gamma\delta$ T cell inflammatory responses are hardwired in the thymus by oxysterol sensing via GPR183 and calibrated by dietary cholesterol." In: *Immunity* 56.3, 562–575.e6. DOI: 10.1016/j.immuni.2023.01.025.

Freeman, Murray F. and John W. Tukey (1950). "Transformations related to the angular and the square root." In: *The Annals of Mathematical Statistics* 21.4, 607–611. DOI: 10.1214/aoms/1177729756.

Freise, Christian et al. (2009). "The alpha 2 chain of collagen type VI sequesters latent proforms of matrix-metalloproteinases and modulates their activation and activity." In: *Matrix Biology* 28.8, pp. 480–489. DOI: 10.1016/j.matbio.2009.08.001.

Friedman, Scott L. et al. (2013). "Therapy for fibrotic diseases: Nearing the starting line." In: *Science Translational Medicine* 5.167. DOI: 10.1126/scitranslmed.3004700.

Friedrich, Thomas, Neslihan N. Tavraz, and Cornelia Junghans (2016). "ATP1A2 Mutations in Migraine: Seeing through the Facets of an Ion Pump onto the Neurobiology of Disease." In: *Frontiers in Physiology* 7. DOI: 10.3389/fphys.2016.00239.

Friess, Mona C. et al. (2022). "Mechanosensitive ACKR4 scavenges CCR7 chemokines to facilitate T cell de-adhesion and passive transport by flow in inflamed afferent lymphatics." In: *Cell Reports* 38.5, p. 110334. DOI: 10.1016/j.celrep.2022.110334.

Frumkin, Dan et al. (2008). "Amplification of multiple genomic loci from single cells isolated by laser microdissection of tissues." In: *BMC Biotechnology* 8.1. DOI: 10.1186/1472-6750-8-17.

Fu, Dun Jack et al. (2014). "Keratin 9 Is Required for the Structural Integrity and Terminal Differentiation of the Palmoplantar Epidermis." In: *Journal of Investigative Dermatology* 134.3, pp. 754–763. DOI: 10.1038/jid.2013.356.

Fuior, Elena V. and Anca V. Gafencu (2019). "Apolipoprotein C1: Its Pleiotropic Effects in Lipid Metabolism and Beyond." In: *International Journal of Molecular Sciences* 20.23, p. 5939. DOI: 10.3390/ijms20235939.

Fujii, Takeshi et al. (2017). "Expression and Function of the Cholinergic System in Immune Cells." In: *Frontiers in Immunology* 8. DOI: 10.3389/fimmu.2017.01085.

Fujimoto, Takashi et al. (2001). "CD44 binds a chondroitin sulfate proteoglycan, aggrecan." In: *International Immunology* 13.3, pp. 359–366. DOI: 10.1093/intimm/13.3.359.

Fujita, Masaaki et al. (2012). "Cross-talk between Integrin $\alpha6\beta4$ and Insulin-like Growth Factor-1 Receptor (IGF1R) through Direct $\alpha6\beta4$ Binding to IGF1 and Subsequent $\alpha6\beta4$-IGF1-IGF1R Ternary Complex Formation in Anchorage-independent Conditions." In: *Journal of Biological Chemistry* 287.15, pp. 12491–12500. DOI: 10.1074/jbc.m111.304170.

Fujiwara, Hironobu et al. (2011). "The Basement Membrane of Hair Follicle Stem Cells Is a Muscle Cell Niche." In: *Cell* 144.4, pp. 577–589. DOI: 10.1016/j.cell.2011.01.014.

Fujiwara, Sayaka et al. (2014). "SOX10 Transactivates S100B to Suppress Schwann Cell Proliferation and to Promote Myelination." In: *PLoS ONE* 9.12. Ed. by Petri Kursula, e115400. DOI: 10.1371/journal.pone.0115400.

Furushima, Kenryo et al. (2007). "Mouse homologues of Shisa antagonistic to Wnt and Fgf signalings." In: *Developmental Biology* 306.2, pp. 480–492. DOI: 10.1016/j.ydbio.2007.03.028.

Gabbiani, G., G. B. Ryan, and G. Majno (1971). "Presence of modified fibroblasts in granulation tissue and their possible role in wound contraction." In: *Experientia* 27.5, 549–550. DOI: 10.1007/bf02147594.

Gabriel, Luis A. R. et al. (2012). "ADAMTSL4, a Secreted Glycoprotein Widely Distributed in the Eye, Binds Fibrillin-1 Microfibrils and Accelerates Microfibril Biogenesis." In: *Investigative Opthalmology & Visual Science* 53.1, p. 461. DOI: 10.1167/iovs.10-5955.

Gajjala, Prathibha R. et al. (2021). "Dysregulated overexpression of Sox9 induces fibroblast activation in pulmonary fibrosis." In: *JCI Insight* 6.20. DOI: 10.1172/jci.insight.152503.

Gajos-Michniewicz, Anna and Malgorzata Czyz (2020). "WNT Signaling in Melanoma." In: *International Journal of Molecular Sciences* 21.14, p. 4852. DOI: 10.3390/ijms21144852.

Galbiati, Ferruccio et al. (2000). "Caveolin-1 Expression Inhibits Wnt/β-Catenin/Lef-1 Signaling by Recruiting β-Catenin to Caveolae Membrane Domains." In: *Journal of Biological Chemistry* 275.30, pp. 23368–23377. DOI: 10.1074/jbc.m002020200.

Gallucci, Randle M. et al. (2000). "Impaired cutaneous wound healing in interleukin-6-deficient and immunosuppressed mice." In: *The FASEB Journal* 14.15, pp. 2525–2531. DOI: 10.1096/fj.00-0073com.

Ganfornina, Maria D. et al. (2008). "Apolipoprotein D is involved in the mechanisms regulating protection from oxidative stress." In: *Aging Cell* 7.4, 506–515. DOI: 10.1111/j.1474-9726.2008.00395.x.

Ganier, Clarisse et al. (2022). "Fibroblast heterogeneity in healthy and wounded skin." In: *Cold Spring Harbor Perspectives in Biology* 14.6. DOI: 10.1101/cshperspect.a041238.

Gao, J et al. (2009). "IRF-1 transcriptionally upregulates PUMA, which mediates the mitochondrial apoptotic pathway in IRF-1-induced apoptosis in cancer cells." In: *Cell Death & Differentiation* 17.4, pp. 699–709. DOI: 10.1038/cdd.2009.156.

Gao, Jie et al. (2022). "Up-regulation of caveolin 1 mediated by chitosan activates Wnt/β-catenin pathway in chronic refractory wound diabetic rat model." In: *Bioengineered* 13.1, pp. 1388–1398. DOI: 10.1080/21655979.2021.2017625.

Gao, Long et al. (2018). "PTGFR activation promotes the expression of PTGS-2 and growth factors via activation of the PKC signaling pathway in bovine endometrial epithelial cells." In: *Animal Reproduction Science* 199, pp. 30–39. DOI: 10.1016/j.anireprosci.2018.10.003.

Gao, Yuge et al. (2021). "Single cell transcriptional zonation of human psoriasis skin identifies an alternative immunoregulatory axis conducted by skin resident cells." In: *Cell Death and Disease* 12, p. 450. DOI: 10.1038/s41419-021-03724-6.

Garcia-Alonso, Luz et al. (2021). "Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro." In: *Nature Genetics* 53.12, 1698–1711. DOI: 10.1038/s41588-021-00972-2.

Garcia, Roman J. et al. (2008). "Endothelin 3 Induces Skin Pigmentation in a Keratin-Driven Inducible Mouse Model." In: *Journal of Investigative Dermatology* 128.1, pp. 131–142. DOI: 10.1038/sj.jid.5700948.

Gargiulo, Simona et al. (2016). "The role of oxysterols in vascular ageing." In: *The Journal of Physiology* 594.8, pp. 2095–2113. DOI: 10.1113/jp271168.

Garza, Luis A. et al. (2012). "Prostaglandin D2 Inhibits Hair Growth and Is Elevated in Bald Scalp of Men with Androgenetic Alopecia." In: *Science Translational Medicine* 4.126. DOI: 10.1126/scitranslmed.3003122.

Gassmann, Marcus G. and Sabine Werner (2000). "Caveolin-1 and -2 Expression Is Differentially Regulated in Cultured Keratinocytes and within the Regenerating Epidermis of Cutaneous Wounds." In: *Experimental Cell Research* 258.1, pp. 23–32. DOI: 10.1006/excr.2000.4904.

Gatti, Francesca et al. (2021). "Nuclear IL-33 restrains the early conversion of fibroblasts to an extracellular matrix-secreting phenotype." In: *Scientific Reports* 11.1. DOI: 10.1038/s41598-020-80509-5.

Gautam, Anuradha and Bhaswati Pandit (2021). "IL32: The multifaceted and unconventional cytokine." In: *Human Immunology* 82.9, pp. 659–667. DOI: 10.1016/j.humimm.2021.05.002.

Gautheron, Jérémie and Isabelle Jéru (2020). "The Multifaceted Role of Epoxide Hydrolases in Human Health and Disease." In: *International Journal of Molecular Sciences* 22.1, p. 13. DOI: 10.3390/ijms22010013.

Gay, Denise et al. (2020). "Phagocytosis of WNT inhibitor SFRP4 by late wound macrophages drives chronic WNT activity for fibrotic skin healing." In: *Science Advances* 6.12. DOI: 10.1126/sciadv.aay3704.

Gaydosik, Alyxzandria M. et al. (2019). "Single-cell lymphocyte heterogeneity in advanced cutaneous T-cell lymphoma skin tumors." In: *Clinical Cancer Research* 25.14, 4443–4454. DOI: 10.1158/1078-0432.ccr-19-0148.

Gazit, Arnona et al. (1999). "Human frizzled 1 interacts with transforming Wnts to transduce a TCF dependent tran-

scriptional response." In: *Oncogene* 18.44, pp. 5959–5966. DOI: 10.1038/sj.onc.1202985.

Gerber, Peter Arne et al. (2014). "The top skin-associated genes: A comparative analysis of human and mouse skin transcriptomes." In: *Biological Chemistry* 395.6, 577–591. DOI: 10.1515/hsz-2013-0279.

Gest, H. (2004). "The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the royal society." In: *Notes and Records of the Royal Society of London* 58.2, 187–201. DOI: 10.1098/rsnr.2004.0055.

Geyfman, Mikhail et al. (2014). "Resting no more: Re-defining telogen, the maintenance stage of the hair growth cycle." In: *Biological Reviews* 90.4, 1179–1196. DOI: 10.1111/brv.12151.

Ghaffari, Abdi et al. (2006). "Fibroblast extracellular matrix gene expression in response to keratinocyte-releasable stratifin." In: *Journal of Cellular Biochemistry* 98.2, 383–393. DOI: 10.1002/jcb.20782.

Gharaee-Kermani, Mehrnaz, Elizabeth M. Denholm, and Sem H. Phan (1996). "Costimulation of fibroblast collagen and transforming growth factor β1 gene expression by monocyte chemoattractant protein-1 via specific receptors." In: *Journal of Biological Chemistry* 271.30, 17779–17784. DOI: 10.1074/jbc.271.30.17779.

Gheduzzi, Dealba et al. (2005). "Heparan sulphate interacts with tropoelastin, with some tropoelastin peptides and is present in human dermis elastic fibers." In: *Matrix Biology* 24.1, pp. 15–25. DOI: 10.1016/j.matbio.2004.12.001.

Giblin, Sean P and Kim S Midwood (2014). "Tenascin-C: Form versus function." In: *Cell Adhesion & Migration* 9.1-2, pp. 48–82. DOI: 10.4161/19336918.2014.987587.

Gibson, Mark A., David I. Leavesley, and Leonie K. Ashman (1999). "Microfibril-associated Glycoprotein-2 Specifically Interacts with a Range of Bovine and Human Cell Types via αVβ3 Integrin." In: *Journal of Biological Chemistry* 274.19, pp. 13060–13065. DOI: 10.1074/jbc.274.19.13060.

Gilchrest, Barbara A. (1983). "In vitro assessment of keratinocyte aging." In: *Journal of Investigative Dermatology* 81.1. DOI: 10.1111/1523-1747.ep12541084.

Giovine, Francesco S di and Gordon W Duff (1990). "Interleukin 1: the first interleukin." In: *Immunology Today* 11, pp. 13–20. DOI: 10.1016/0167-5699(90)90005-t.

Girbl, Tamara et al. (2018). "Distinct Compartmentalization of the Chemokines CXCL1 and CXCL2 and the Atypical Receptor ACKR1 Determine Discrete Stages of Neutrophil Diapedesis." In: *Immunity* 49.6, 1062–1076.e6. DOI: 10.1016/j.immuni.2018.09.018.

Glatte, Patrick et al. (2019). "Architecture of the cutaneous autonomic nervous system." In: *Frontiers in Neurology* 10. DOI: 10.3389/fneur.2019.00970.

Glim, Judith E. et al. (2013). "Platelet derived growth factor-cc secreted by M2 macrophages induces alpha-smooth muscle actin expression by dermal and gingival fibroblasts." In: *Immunobiology* 218.6, 924–929. DOI: 10.1016/j.imbio.2012.10.004.

Go, Gwang-woong and Arya Mani (2012). "Low-Density Lipoprotein Receptor (LDLR) Family Orchestrates Cholesterol Homeostasis." In: *Yale Journal of Biology and Medicine* 85, 19–28.

Goldsmith, Lowell A. et al. (2012). *Fitzpatrick's dermatology in General Medicine (8th ed.)* McGraw-Hill.

Gong, Wuming et al. (2018). "Drimpute: Imputing dropout events in single cell RNA sequencing data." In: *BMC Bioinformatics* 19.1. DOI: 10.1186/s12859-018-2226-y.

Goralski, Kerry B. et al. (2007). "Chemerin, a Novel Adipokine That Regulates Adipogenesis and Adipocyte Metabolism." In: *Journal of Biological Chemistry* 282.38, pp. 28175–28188. DOI: 10.1074/jbc.m700793200.

Gordon, Christopher T. et al. (2015). "Mutations in the Endothelin Receptor Type A Cause Mandibulofacial Dysostosis with Alopecia." In: *The American Journal of Human Genetics* 96.4, pp. 519–531. DOI: 10.1016/j.ajhg.2015.01.015.

Gosline, J. et al. (2002). "Elastic proteins: Biological roles and mechanical properties." In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 357.1418, 121–132. DOI: 10.1098/rstb.2001.1022.

Govindaraju, Priya et al. (2019). "CD44-dependent inflammation, fibrogenesis, and collagenolysis regulates extracellular matrix remodeling and tensile strength during cutaneous wound healing." In: *Matrix Biology* 75-76, pp. 314–330. DOI: 10.1016/j.matbio.2018.06.004.

Grass, D S et al. (1996). "Expression of human group II PLA2 in transgenic mice results in epidermal hyperplasia in the absence of inflammatory infiltrate." In: *Journal of Clinical Investigation* 97.10, pp. 2233–2241. DOI: 10.1172/jci118664.

Greco, Valentina et al. (2009). "A Two-Step Mechanism for Stem Cell Activation during Hair Regeneration." In: *Cell Stem Cell* 4.2, pp. 155–169. DOI: 10.1016/j.stem.2008.12.009.

Green, Christopher Daniel et al. (2018). "A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-seq." In: *Developmental Cell* 46.5. DOI: 10.1016/j.devcel.2018.07.025.

Greene, Catherine M. et al. (2004). "Secretory Leucoprotease Inhibitor Impairs Toll-Like Receptor 2- and 4-Mediated Responses in Monocytic Cells." In: *Infection and Immunity* 72.6, pp. 3684–3687. DOI: 10.1128/iai.72.6.3684-3687.2004.

Greter, Melanie et al. (2012). "Stroma-Derived Interleukin-34 Controls the Development and Maintenance of Langerhans Cells and the Maintenance of Microglia." In: *Immunity* 37.6, pp. 1050–1060. DOI: 10.1016/j.immuni.2012.11.001.

Griffin, Michelle F. et al. (2021). "The role of Wnt signaling in skin fibrosis." In: *Medicinal Research Reviews* 42.1, 615–628. DOI: 10.1002/med.21853.

Griffiths, Jonathan A. et al. (2018). "Detection and removal of barcode swapping in single-cell RNA-seq data." In: *Nature Communications* 9.1. DOI: 10.1038/s41467-018-05083-x.

Grimal, Sophie et al. (2010). "Collagen XXVIII is a distinctive component of the peripheral nervous system nodes of ranvier and surrounds nonmyelinating glial cells." In: *Glia* 58.16, pp. 1977–1987. DOI: 10.1002/glia.21066.

Gron, Birgitte et al. (2002). "Oral fibroblasts produce more HGF and KGF than skin fibroblasts in response to coculture with keratinocytes." In: *APMIS* 110.12, 892–898. DOI: 10.1034/j.1600-0463.2002.1101208.x.

Grün, Dominic and Alexander van Oudenaarden (2015). "Design and analysis of single-cell sequencing experiments." In: *Cell* 163.4, 799–810. DOI: 10.1016/j.cell.2015.10.039.

Grün, Dominic (2019). "Revealing dynamics of gene expression variability in cell state space." In: *Nature Methods* 17.1, 45–49. DOI: 10.1038/s41592-019-0632-3.

Gu, Jingsheng, Victor Faundez, and Erica Werner (2010). "Endosomal recycling regulates anthrax toxin receptor 1/tumor endothelial marker 8-dependent cell spreading." In: *Experimental Cell Research* 316.12, pp. 1946–1957. DOI: 10.1016/j.yexcr.2010.03.026.

Guerrero-Juarez, Christian F. and Maksim V. Plikus (2018). "Emerging nonmetabolic functions of Skin Fat." In: *Nature Reviews Endocrinology* 14.3, 163–173. DOI: 10.1038/nrendo.2017.162.

Guerrero-Juarez, Christian F. et al. (2019). "Single-cell analysis reveals fibroblast heterogeneity and myeloid-derived adipocyte progenitors in murine skin wounds." In: *Nature Communications* 10.1. DOI: 10.1038/s41467-018-08247-x.

Guillot, Nicolas et al. (2012). "BAMBI Regulates Angiogenesis and Endothelial Homeostasis through Modulation of Alternative TGFβ Signaling." In: *PLoS ONE* 7.6. Ed. by Christos Chatziantoniou, e39406. DOI: 10.1371/journal.pone.0039406.

Gujral, Taranjit S. et al. (2014). "A Noncanonical Frizzled2 Pathway Regulates Epithelial-Mesenchymal Transition and Metastasis." In: *Cell* 159.4, pp. 844–856. DOI: 10.1016/j.cell.2014.10.032.

Guo, Bei et al. (2020). "New Insights Into Implications of CTRP3 in Obesity, Metabolic Dysfunction, and Cardiovascular Diseases: Potential of Therapeutic Interventions." In: *Frontiers in Physiology* 11. DOI: 10.3389/fphys.2020.570270.

Guo, Guoji et al. (2010). "Resolution of Cell Fate Decisions revealed by single-cell gene expression analysis from zygote to blastocyst." In: *Developmental Cell* 18.4, 675–685. DOI: 10.1016/j.devcel.2010.02.012.

Guo, Hongyu and Jun Li (2021). "ScSorter: Assigning cells to known cell types according to Marker genes." In: *Genome Biology* 22.1. DOI: 10.1186/s13059-021-02281-7.

Gur, Chamutal et al. (2022). "LGR5 expressing skin fibroblasts define a major cellular hub perturbed in scleroderma." In: *Cell* 185.8. DOI: 10.1016/j.cell.2022.03.011.

Gurtner, Geoffrey C. et al. (2008). "Wound repair and regeneration." In: *Nature* 453.7193, 314–321. DOI: 10.1038/nature07039.

Haensel, Daniel et al. (2020). "Defining epidermal basal cell states during skin homeostasis and wound healing using single-cell transcriptomics." In: *Cell Reports* 30.11. DOI: 10.1016/j.celrep.2020.02.091.

Hafemeister, Christoph and Rahul Satija (2019). "Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression." In: *Genome Biology* 20.1. DOI: 10.1186/s13059-019-1874-1.

Hagg, Paivi M. et al. (1997). "Cloning of Mouse Type XV Collagen Sequences and Mapping of the Corresponding Gene to 4B1-3." In: *Genomics* 45.1, pp. 31–41. DOI: 10.1006/geno.1997.4884.

Haghverdi, Laleh et al. (2016). "Diffusion pseudotime robustly reconstructs lineage branching." In: *Nature Methods* 13.10, 845–848. DOI: 10.1038/nmeth.3971.

Haghverdi, Laleh et al. (2018). "Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors." In: *Nature Biotechnology* 36.5, 421–427. DOI: 10.1038/nbt.4091.

Hagner, Andrew et al. (2020). "Transcriptional Profiling of the Adult Hair Follicle Mesenchyme Reveals R-spondin as a Novel Regulator of Dermal Progenitor Function." In: *iScience* 23.4, p. 101019. DOI: 10.1016/j.isci.2020.101019.

Halai, Krishma et al. (2013). "ICAM-2 facilitates luminal neutrophil-endothelial cell interactions in vivo." In: *Journal of Cell Science*. DOI: 10.1242/jcs.137463.

Halfter, Willi et al. (1998). "Collagen XVIII Is a Basement Membrane Heparan Sulfate Proteoglycan." In: *Journal of Biological Chemistry* 273.39, pp. 25404–25412. DOI: 10.1074/jbc.273.39.25404.

Hammarstedt, Ann et al. (2013). "WISP2 regulates preadipocyte commitment and PPARγ activation by BMP4." In: *Proceedings of the National Academy of Sciences* 110.7, pp. 2563–2568. DOI: 10.1073/pnas.1211255110.

Han, Bobby Kwanghoon, Nancy J. Olsen, and Andrea Bottaro (2016). "The CD27–CD70 pathway and pathogenesis of autoimmune disease." In: *Seminars in Arthritis and Rheumatism* 45.4, pp. 496–501. DOI: 10.1016/j.semarthrit.2015.08.001.

Han, Chencheng, Tongyan Liu, and Rong Yin (2020). "Biomarkers for cancer-associated fibroblasts." In: *Biomarker Research* 8.1. DOI: 10.1186/s40364-020-00245-w.

Hao, Huifang et al. (2013). "Inhibition of the Growth Factor MDK/Midkine by a Novel Small Molecule Compound to Treat Non-Small Cell Lung Cancer." In: *PLoS ONE* 8.8. Ed. by Masaru Katoh, e71093. DOI: 10.1371/journal.pone.0071093.

Hao, Yongping and Guang Li (2020). "Role of EFNA1 in tumorigenesis and prospects for cancer therapy." In: *Biomedicine and Pharmacotherapy* 130, p. 110567. DOI: 10.1016/j.biopha.2020.110567.

Haque, Ashraful et al. (2017). "A practical guide to single-cell RNA-sequencing for Biomedical Research and Clinical Applications." In: *Genome Medicine* 9.1. DOI: 10.1186/s13073-017-0467-4.

Haraldsen, Guttorm et al. (2009). "Interleukin-33 – cytokine of dual function or novel alarmin?" In: *Trends in Immunology* 30.5, pp. 227–233. DOI: 10.1016/j.it.2009.03.003.

Harman, Jennifer L. et al. (2020). "Emerging Roles for Neuropilin-2 in Cardiovascular Disease." In: *International Journal of Molecular Sciences* 21.14, p. 5154. DOI: 10.3390/ijms21145154.

Harris, R (2003). "EGF receptor ligands." In: *Experimental Cell Research* 284.1, pp. 2–13. DOI: 10.1016/s0014-4827(02)00105-2.

Hatamochi, A., K. Fujiwara, and H. Ueki (1984). "Effects of histamine on collagen synthesis by cultured fibroblasts derived from Guinea Pig Skin." In: *Archives of Dermatological Research* 277.1, 60–64. DOI: 10.1007/bf00406482.

Hayashi, Tetsutaro et al. (2010). "Single-cell gene profiling of planarian stem cells using fluorescent activated cell sorting and its "index sorting" function for Stem Cell Research." In: *Development, Growth & Differentiation* 52.1, 131–144. DOI: 10.1111/j.1440-169x.2009.01157.x.

Haydont, Valerie et al. (2020). "Fibroblasts from the human skin dermo-hypodermal junction are distinct from dermal papillary and reticular fibroblasts and from mesenchymal stem cells and exhibit a specific molecular profile related to extracellular matrix organization and modeling." In: *Cells* 9.2, p. 368. DOI: 10.3390/cells9020368.

Haydont, Valérie et al. (2019). "Genome-wide profiling of adult human papillary and reticular fibroblasts identifies Acan, Col Xi α1, and PSG1 as general biomarkers of dermis ageing, and kank4 as an exemplary effector of papillary fibroblast ageing, related to contractility." In: *Mechanisms of Ageing and Development* 177, 157–181. DOI: 10.1016/j.mad.2018.06.003.

Hayes, Anthony J. et al. (2022). "Perlecan, A Multi-Functional, Cell-Instructive, Matrix-Stabilizing Proteoglycan With Roles in Tissue Development Has Relevance to Connective Tissue Repair and Regeneration." In: *Frontiers in Cell and Developmental Biology* 10. DOI: 10.3389/fcell.2022.856261.

Hayflick, L. (1965). "The limited in vitro lifetime of human diploid cell strains." In: *Experimental Cell Research* 37.3, 614–636. DOI: 10.1016/0014-4827(65)90211-9.

Hazell, Georgina G. J. et al. (2016). "PI16 is a shear stress and inflammation-regulated inhibitor of MMP2." In: *Scientific Reports* 6.1. DOI: 10.1038/srep39553.

He, Helen et al. (2020). "Single-cell transcriptome analysis of human skin identifies novel fibroblast subpopulation and enrichment of immune subsets in atopic dermatitis." In: *Journal of Allergy and Clinical Immunology* 145.6, 1615–1628. DOI: 10.1016/j.jaci.2020.01.042.

He, Y., S. Sardar, and M.A. Karsdal (2019). "Type IX collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 83–89. DOI: 10.1016/b978-0-12-817068-7.00009-4.

He, Yu-Yao et al. (2022). "The Fra-1: Novel role in regulating extensive immune cell states and affecting inflammatory diseases." In: *Frontiers in Immunology* 13. DOI: 10.3389/fimmu.2022.954744.

Hechtman, Jaclyn F. (2022). "NTRK insights: best practices for pathologists." In: *Modern Pathology* 35.3, pp. 298–305. DOI: 10.1038/s41379-021-00913-8.

Hedge, Thomas A. and Ivor Mason (2008). "Expression of Shisa2, a modulator of both Wnt and Fgf signaling, in the chick embryo." In: *The International Journal of Developmental Biology* 52.1, pp. 81–85. DOI: 10.1387/ijdb.072355th.

Hedlund, Joel, Jan Johansson, and Bengt Persson (2009). "BRICHOS - a superfamily of multidomain proteins with diverse functions." In: *BMC Research Notes* 2.1, p. 180. DOI: 10.1186/1756-0500-2-180.

Heide, Huub J. L. van der et al. (2006). "Rofecoxib inhibits heterotopic ossification after total hip arthroplasty." In: *Archives of Orthopaedic and Trauma Surgery* 127.7, pp. 557–561. DOI: 10.1007/s00402-006-0243-1.

Heimberg, Graham et al. (2016). "Low dimensionality in gene expression data enables the accurate extraction of transcriptional programs from shallow sequencing." In: *Cell Systems* 2.4, 239–250. DOI: 10.1016/j.cels.2016.04.001.

Heinhuis, Bas et al. (2015). "Alternatively spliced isoforms of IL-32 differentially influence cell death pathways in cancer cell lines." In: *Carcinogenesis* 37.2, pp. 197–205. DOI: 10.1093/carcin/bgv172.

Heitman, Nicholas et al. (2020). "Dermal sheath contraction powers stem cell niche relocation during hair cycle regression." In: *Science* 367.6474, 161–166. DOI: 10.1126/science.aax9131.

Helfer, Gisela and Qing-Feng Wu (2018). "Chemerin: a multifaceted adipokine involved in metabolic disorders." In: *Journal of Endocrinology* 238.2, R79–R94. DOI: 10.1530/joe-18-0174.

Helfer, Gisela et al. (2016). "A neuroendocrine role for chemerin in hypothalamic remodelling and photoperi-

odic control of energy balance." In: *Scientific Reports* 6.1. DOI: 10.1038/srep26830.

Henau, Olivier De et al. (2016). "Signaling Properties of Chemerin Receptors CMKLR1, GPR1 and CCRL2." In: *PLOS ONE* 11.10. Ed. by Arun Shukla, e0164179. DOI: 10.1371/journal.pone.0164179.

Henderson, Neil C et al. (2013). "Targeting of αv Integrin identifies a core molecular pathway that regulates fibrosis in several organs." In: *Nature Medicine* 19.12, 1617–1624. DOI: 10.1038/nm.3282.

Hennebert, Olivier et al. (2007). "Dehydroepiandrosterone 7α-hydroxylation in human tissues: Possible interference with type 1 11β-hydroxysteroid dehydrogenase-mediated processes." In: *The Journal of Steroid Biochemistry and Molecular Biology* 104.3-5, pp. 326–333. DOI: 10.1016/j.jsbmb.2007.03.026.

Henriksen, K. and M.A. Karsdal (2019). "Type I collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 1–12. DOI: 10.1016/b978-0-12-817068-7.00001-x.

Heruth, Daniel P et al. (2012). "RNA-seq analysis of synovial fibroblasts brings new insights into rheumatoid arthritis." In: *Cell & Bioscience* 2.1, p. 43. DOI: 10.1186/2045-3701-2-43.

Hesse, Matthias et al. (2001). "Differential regulation of nitric oxide synthase-2 and arginase-1 by type 1/type 2 cytokines in vivo: Granulomatous pathology is shaped by the pattern of l-arginine metabolism." In: *The Journal of Immunology* 167.11, 6533–6544. DOI: 10.4049/jimmunol.167.11.6533.

Heumos, Lukas et al. (2023). "Best practices for single-cell analysis across modalities." In: *Nature Reviews Genetics*. DOI: 10.1038/s41576-023-00586-w.

Hicks, Stephanie C et al. (2015). "Missing data and technical variability in single-cell RNA- sequencing experiments." In: DOI: 10.1101/025528.

Hie, Brian, Bryan Bryson, and Bonnie Berger (2019). "Efficient integration of heterogeneous single-cell transcriptomes using Scanorama." In: *Nature Biotechnology* 37.6, 685–691. DOI: 10.1038/s41587-019-0113-3.

Hie, Brian et al. (2020). "Computational methods for single-cell RNA sequencing." In: *Annual Review of Biomedical Data Science* 3.1, 339–364. DOI: 10.1146/annurev-biodatasci-012220-100601.

Higgins, C.A. et al. (2017). "Multifaceted role of hair follicle dermal cells in bioengineered skins." In: *British Journal of Dermatology* 176.5, 1259–1269. DOI: 10.1111/bjd.15087.

Higgins, Claire A., Gillian E. Westgate, and Colin A.B. Jahoda (2011). "Modulation in Proteolytic Activity Is Identified as a Hallmark of Exogen by Transcriptional Profiling of Hair Follicles." In: *Journal of Investigative Dermatology* 131.12, pp. 2349–2357. DOI: 10.1038/jid.2011.227.

Hilario, Jona D. et al. (2009). "Semaphorin 5A is a bifunctional axon guidance cue for axial motoneurons in vivo." In: *Developmental Biology* 326.1, pp. 190–200. DOI: 10.1016/j.ydbio.2008.11.007.

Hinz, Boris et al. (2007). "The myofibroblast - one function, multiple origins." In: *The American Journal of Pathology* 170.6, 1807–1816. DOI: 10.2353/ajpath.2007.070112.

Hinz, Boris et al. (2012). "Recent developments in myofibroblast biology." In: *The American Journal of Pathology* 180.4, 1340–1355. DOI: 10.1016/j.ajpath.2012.02.004.

Hiran, Tejindervir S. et al. (2003). "Endothelial expression of the α6β4 integrin is negatively regulated during angiogenesis." In: *Journal of Cell Science* 116.18, pp. 3771–3781. DOI: 10.1242/jcs.00681.

Hirohata, Satoshi et al. (2002). "Punctin, a Novel ADAMTS-like Molecule, ADAMTSL-1, in Extracellular Matrix." In: *Journal of Biological Chemistry* 277.14, pp. 12182–12189. DOI: 10.1074/jbc.m109665200.

Hisamuddin, Irfan M and Vincent W Yang (2007). "Genetic polymorphisms of human flavin-containing monooxygenase 3: implications for drug metabolism and clinical perspectives." In: *Pharmacogenomics* 8.6, pp. 635–643. DOI: 10.2217/14622416.8.6.635.

Hogaboam, Cory M. et al. (1998). "Novel roles for chemokines and fibroblasts in interstitial fibrosis." In: *Kidney International* 54.6, 2152–2159. DOI: 10.1046/j.1523-1755.1998.00176.x.

Horiba, Mitsuru et al. (2000). "Neointima formation in a restenosis model is suppressed in midkine-deficient mice." In: *Journal of Clinical Investigation* 105.4, pp. 489–495. DOI: 10.1172/jci7208.

Horiuchi, Y., S-J. Bae, and I. Katayama (2006). "Overexpression of the suppressor of cytokine signalling 3 (SOCS3) in severe atopic dermatitis." In: *Clinical and Experimental Dermatology* 31.1, pp. 100–104. DOI: 10.1111/j.1365-2230.2005.01979.x.

Horst, Britt ter et al. (2018). "Advances in keratinocyte delivery in burn wound care." In: *Advanced Drug Delivery Reviews* 123, 18–32. DOI: 10.1016/j.addr.2017.06.012.

Hou, Wenpin et al. (2020). "A systematic evaluation of single-cell RNA-sequencing imputation methods." In: *bioRxiv*. DOI: 10.1101/2020.01.29.925974.

Houzelstein, D. et al. (2000). "The expression of the homeobox gene MSX1 reveals two populations of dermal progenitor cells originating from the somites." In: *Development* 127.10, 2155–2164. DOI: 10.1242/dev.127.10.2155.

Hovland, Damla et al. (2020). "Vitamin A Metabolism During Refractory Telogen." In: *Current Developments in Nutrition* 4, p. 4140109. DOI: 10.1093/cdn/nzaa041_013.

Hsu, Lauren L. and Aedín C. Culhane (2023). "Correspondence analysis for Dimension Reduction, batch integration, and visualization of single-cell RNA-seq data." In: *Scientific Reports* 13.1. DOI: 10.1038/s41598-022-26434-1.

Hu, Liuying et al. (2022). "Vitamins and their derivatives synergistically promote hair shaft elongation ex vivo via PlGF/VEGFR-1 signalling activation." In: *Journal of Dermatological Science* 108.1, pp. 2–11. DOI: 10.1016/j.jdermsci.2022.09.003.

Hu, Qing-Mei et al. (2017). "Induction of retinal-dependent calcium influx in human melanocytes by UVA or UVB radiation contributes to the stimulation of melanosome transfer." In: *Cell Proliferation* 50.6. DOI: 10.1111/cpr.12372.

Huang, Fei and Ye-Guang Chen (2012a). "Regulation of TGF-β receptor activity." In: *Cell & Bioscience* 2.1, p. 9. DOI: 10.1186/2045-3701-2-9.

Huang, Mo et al. (2018). "Saver: Gene expression recovery for single-cell RNA sequencing." In: *Nature Methods* 15.7, 539–542. DOI: 10.1038/s41592-018-0033-z.

Huang, Xiangwei et al. (2012b). "Matrix stiffness–induced myofibroblast differentiation is mediated by intrinsic mechanotransduction." In: *American Journal of Respiratory Cell and Molecular Biology* 47.3, 340–348. DOI: 10.1165/rcmb.2012-0050oc.

Huang, Xiaodong, Jie Zhang, and Yongbin Zheng (2020). "ANTXR1 Is a Prognostic Biomarker and Correlates With Stromal and Immune Cell Infiltration in Gastric Cancer." In: *Frontiers in Molecular Biosciences* 7. DOI: 10.3389/fmolb.2020.598221.

Huang, Yunlong et al. (2009). "Type I Interferons and Interferon Regulatory Factors Regulate TNF-Related Apoptosis-Inducing Ligand (TRAIL) in HIV-1-Infected Macrophages." In: *PLoS ONE* 4.4. Ed. by Derya Unutmaz, e5397. DOI: 10.1371/journal.pone.0005397.

Hughes, G. (1968). "On the mean accuracy of statistical pattern recognizers." In: *IEEE Transactions on Information Theory* 14.1, 55–63. DOI: 10.1109/tit.1968.1054102.

Hughes, Travis K. et al. (2020). "Second-strand synthesis-based massively parallel scrna-seq reveals cellular states and molecular features of human inflammatory skin pathologies." In: *Immunity* 53.4. DOI: 10.1016/j.immuni.2020.09.015.

Husain, Mohammed Amir, Benoit Laurent, and Melanie Plourde (2021). "APOE and Alzheimer's Disease: From Lipid Transport to Physiopathology and Therapeutics." In: *Frontiers in Neuroscience* 15. DOI: 10.3389/fnins.2021.630502.

Huttenlocher, A et al. (1996). "Decorin regulates collagenase gene expression in fibroblasts adhering to vitronectin." In: *Matrix Biology* 15.4, pp. 239–250. DOI: 10.1016/s0945-053x(96)90115-8.

Hutter, Randolph et al. (2013). "Novel Small Leucine-Rich Repeat Protein Podocan Is a Negative Regulator of Migration and Proliferation of Smooth Muscle Cells, Modulates Neointima Formation, and Is Expressed in Human Atheroma." In: *Circulation* 128.22, pp. 2351–2363. DOI: 10.1161/circulationaha.113.004634.

Hutton, S.R. and L.H. Pevny (2009). "Sox Gene Expression." In: *Encyclopedia of Neuroscience*. Elsevier, pp. 145–151. DOI: 10.1016/b978-008045046-9.01079-2.

Hwangbo, C et al. (2015). "Syntenin regulates TGF-β1-induced Smad activation and the epithelial-to-mesenchymal transition by inhibiting caveolin-mediated TGF-β type I receptor internalization." In: *Oncogene* 35.3, pp. 389–401. DOI: 10.1038/onc.2015.100.

Hynes, R. O. and A. Naba (2011). "Overview of the Matrisome–An Inventory of Extracellular Matrix Constituents and Functions." In: *Cold Spring Harbor Perspectives in Biology* 4.1, a004903–a004903. DOI: 10.1101/cshperspect.a004903.

Hägg, Pasi et al. (2001). "Type XIII collagen: a novel cell adhesion component present in a range of cell–matrix adhesions and in the intercalated discs between cardiac muscle cells." In: *Matrix Biology* 19.8, pp. 727–742. DOI: 10.1016/s0945-053x(00)00119-0.

Ibrahim, Mohamed M et al. (2015). "Myofibroblasts contribute to but are not necessary for wound contraction." In: *Laboratory Investigation* 95.12, pp. 1429–1438. DOI: 10.1038/labinvest.2015.116.

Ilicic, Tomislav et al. (2016). "Classification of low quality cells from single-cell RNA-seq data." In: *Genome Biology* 17.1. DOI: 10.1186/s13059-016-0888-1.

Illumina (2021). *Welcome to immense Discovery Power*.

Imoto, Yusuke et al. (2022). "Resolution of the curse of dimensionality in single-cell RNA sequencing data analysis." In: *Life Science Alliance* 5.12. DOI: 10.26508/lsa.202201591.

Inoue, Yu et al. (2012). "Bimodal effect of retinoic acid on melanocyte differentiation identified by time-dependent analysis." In: *Pigment Cell & Melanoma Research* 25.3, pp. 299–311. DOI: 10.1111/j.1755-148x.2012.00988.x.

Iribar, Haizea, Usue Etxaniz, and Ander Izeta (2018). "Diversity of adult stem cell niches in the dermal compartment of skin." In: *Reference Module in Biomedical Sciences*. DOI: 10.1016/b978-0-12-801238-3.65470-3.

Ishida, Yuko, Ji-Liang Gao, and Philip M. Murphy (2008). "Chemokine Receptor CX3CR1 Mediates Skin Wound Healing by Promoting Macrophage and Fibroblast Accumulation and Function." In: *The Journal of Immunology* 180.1, pp. 569–579. DOI: 10.4049/jimmunol.180.1.569.

Ishii, S. et al. (1993). "CD44 participates in the adhesion of human colorectal carcinoma cells to laminin and type IV collagen." In: *Surgical Oncology* 2.4, pp. 255–264. DOI: 10.1016/0960-7404(93)90015-q.

Ishitani, Akiko et al. (2003). "Protein Expression and Peptide Binding Suggest Unique and Interacting Functional Roles for HLA-E, F, and G in Maternal-Placental Immune Recognition." In: *The Journal of Immunology* 171.3, pp. 1376–1384. DOI: 10.4049/jimmunol.171.3.1376.

Islam, Nasif and Luis A. Garza (2018). "Adipose and Hair Function: An aPPARent Connection." In: *Journal of Investigative Dermatology* 138.3, pp. 480–482. DOI: 10.1016/j.jid.2017.10.019.

Islam, Ridwan et al. (2022). "Role of Neuropilin-2-mediated signaling axis in cancer progression and therapy resistance." In: *Cancer and Metastasis Reviews* 41.3, pp. 771–787. DOI: 10.1007/s10555-022-10048-0.

Islam, Saiful et al. (2011). "Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq." In: *Genome Research* 21.7, 1160–1167. DOI: 10.1101/gr.110882.110.

Islam, Saiful et al. (2014). "Quantitative single-cell RNA-seq with unique molecular identifiers." In: *Nature Methods* 11.2, 163–166. DOI: 10.1038/nmeth.2772.

Isoherranen, Nina and Guo Zhong (2019). "Biochemical and physiological importance of the CYP26 retinoic acid hydroxylases." In: *Pharmacology & Therapeutics* 204, p. 107400. DOI: 10.1016/j.pharmthera.2019.107400.

Isozaki, Takeo et al. (2011). "Synergistic induction of CX3CL1 by interleukin-1β and interferon-γ in human lung fibroblasts: involvement of signal transducer and activator of transcription 1 signaling pathways." In: *Translational Research* 157.2, pp. 64–70. DOI: 10.1016/j.trsl.2010.11.007.

Ito, Yuriko et al. (2007). "Isolation of murine hair-inducing cells using the cell surface marker prominin-1/CD133." In: *Journal of Investigative Dermatology* 127.5, 1052–1060. DOI: 10.1038/sj.jid.5700665.

Iwamiya, Takahiro et al. (2020). "Human cardiac fibroblasts expressing VCAM1 improve heart function in postinfarct heart failure rat models by stimulating lymphangiogenesis." In: *PLOS ONE* 15.9. Ed. by Vincenzo Lionetti, e0237810. DOI: 10.1371/journal.pone.0237810.

Iwashita, N et al. (1999). "Expression of midkine in normal and burn sites of rat skin." In: *Burns* 25.2, pp. 119–124. DOI: 10.1016/s0305-4179(98)00120-x.

Iwayama, Tomoaki et al. (2015). "PDGFRα signaling drives adipose tissue fibrosis by targeting progenitor cell plasticity." In: *Genes & Development* 29.11, 1106–1119. DOI: 10.1101/gad.260554.115.

Iyer, Anand Krishnan V. et al. (2007). "Tenascin cytotactin epidermal growth factor-like repeat binds epidermal growth factor receptor with low affinity." In: *Journal of Cellular Physiology* 211.3, pp. 748–758. DOI: 10.1002/jcp.20986.

Jacob, Tina et al. (2022). "Molecular and spatial design of early skin development." In: *bioRxiv*. DOI: 10.1101/2022.12.28.522081.

Jacobs, Julie et al. (2018). "Unveiling a CD70-positive subset of cancer-associated fibroblasts marked by pro-migratory activity and thriving regulatory T cell accumulation." In: *OncoImmunology* 7.7, e1440167. DOI: 10.1080/2162402x.2018.1440167.

Jacomy, Mathieu et al. (2014). "FORCEATLAS2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software." In: *PLoS ONE* 9.6. DOI: 10.1371/journal.pone.0098679.

Jaitin, D. A. et al. (2014). "Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types." In: *Science* 343.6172, 776–779. DOI: 10.1126/science.1247651.

Jakobisiak, Marek, Jakub Golab, and Witold Lasek (2011). "Interleukin 15 as a promising candidate for tumor immunotherapy." In: *Cytokine & Growth Factor Reviews* 22.2, pp. 99–108. DOI: 10.1016/j.cytogfr.2011.04.001.

Jakobsson, Per-Johan et al. (1999). "Identification of human prostaglandin E synthase: A microsomal, glutathione-dependent, inducible enzyme, constituting a potential novel drug target." In: *Proceedings of the National Academy of Sciences* 96.13, pp. 7220–7225. DOI: 10.1073/pnas.96.13.7220.

Jalkanen, S and M Jalkanen (1992). "Lymphocyte CD44 binds the COOH-terminal heparin-binding domain of fibronectin." In: *Journal of Cell Biology* 116.3, pp. 817–825. DOI: 10.1083/jcb.116.3.817.

Janjanam, Jagadeesh et al. (2021). "Matricellular Protein WISP2 Is an Endogenous Inhibitor of Collagen Linearization and Cancer Metastasis." In: *Cancer Research* 81.22, pp. 5666–5677. DOI: 10.1158/0008-5472.can-20-3982.

Janson, David G. et al. (2012). "Different gene expression patterns in human papillary and reticular fibroblasts." In: *Journal of Investigative Dermatology* 132.11, 2565–2572. DOI: 10.1038/jid.2012.192.

Jenkins, Blair A et al. (2019). "The cellular basis of mechanosensory Merkel-cell innervation during development." In: *eLife* 8. Ed. by Jeremy Nathans and Marianne E Bronner, e42633. ISSN: 2050-084X. DOI: 10.7554/eLife.42633.

Jensen, Sacha A. and Penny A. Handford (2016). "New insights into the structure, assembly and biological roles of 10–12 nm connective tissue microfibrils from fibrillin-1 studies." In: *Biochemical Journal* 473.7, pp. 827–838. DOI: 10.1042/bj20151108.

Jerber, Julie et al. (2021). "Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation." In: *Nature Genetics* 53.3, 304–312. DOI: 10.1038/s41588-021-00801-6.

Jho, Eek hoon et al. (2002). "Wnt/β-Catenin/Tcf Signaling Induces the Transcription of Axin2, a Negative Regulator of the Signaling Pathway." In: *Molecular and Cellular Biology* 22.4, pp. 1172–1183. DOI: 10.1128/mcb.22.4.1172-1183.2002.

Jiang, Chengyu, Adrian T. Ting, and Brian Seed (1998). "PPAR-γ agonists inhibit production of monocyte inflammatory cytokines." In: *Nature* 391.6662, pp. 82–86. DOI: 10.1038/34184.

Jiang, Dongsheng and Yuval Rinkevich (2018a). "Defining skin fibroblastic cell types beyond CD90." In: *Frontiers in Cell and Developmental Biology* 6. DOI: 10.3389/fcell.2018.00133.

Jiang, Dongsheng and Karin Scharffetter-Kochanek (2020). "Mesenchymal stem cells adaptively respond to environmental cues thereby improving granulation tissue formation and wound healing." In: *Frontiers in Cell and Developmental Biology* 8. DOI: 10.3389/fcell.2020.00697.

Jiang, Dongsheng et al. (2018b). "Two succeeding fibroblastic lineages drive dermal development and the transition from regeneration to scarring." In: *Nature Cell Biology* 20.4, 422–431. DOI: 10.1038/s41556-018-0073-8.

Jiang, Ruochen et al. (2022). "Statistics or biology: The Zero-inflation controversy about scRNA-Seq Data." In: *Genome Biology* 23.1. DOI: 10.1186/s13059-022-02601-5.

Jiao, Bo et al. (2021). "Class-3 semaphorins: Potent multifunctional modulators for angiogenesis-associated diseases." In: *Biomedicine & Pharmacotherapy* 137, p. 111329. DOI: 10.1016/j.biopha.2021.111329.

Jin, Suoqin et al. (2021). "Inference and analysis of cell-cell communication using CellChat." In: *Nature Communications* 12.1. DOI: 10.1038/s41467-021-21246-9.

Jinnin, Masatoshi et al. (2004). "Tenascin-C upregulation by transforming growth factor-β in human dermal fibroblasts involves Smad3, Sp1, and Ets1." In: *Oncogene* 23.9, pp. 1656–1667. DOI: 10.1038/sj.onc.1207064.

Johnson, Eric M., William Kath, and Madhav Mani (2022). "EMBEDR: Distinguishing signal from noise in single-cell omics data." In: *Patterns* 3.3, p. 100443. DOI: 10.1016/j.patter.2022.100443.

Johnson, Louise A. and David G. Jackson (2013). "The chemokine CX3CL1 promotes trafficking of dendritic cells through inflamed lymphatics." In: *Journal of Cell Science*. DOI: 10.1242/jcs.135343.

Johnson, P. L., E. O. Butcher, and G. Bevelander (1945). "The distribution of alkaline phosphatase in the cyclic growth of the rat hair follicle." In: *The Anatomical Record* 93.4, 355–361. DOI: 10.1002/ar.1090930406.

Johnson, W. Evan, Cheng Li, and Ariel Rabinovic (2006). "Adjusting batch effects in microarray expression data using empirical Bayes methods." In: *Biostatistics* 8.1, 118–127. DOI: 10.1093/biostatistics/kxj037.

Johnson, William B. and Joram Lindenstrauss (1984). "Extensions of Lipschitz mappings into a Hilbert space." In: *Conference on Modern Analysis and Probability*, 189–206. DOI: 10.1090/conm/026/737400.

Johnstone, R M et al. (1987). "Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes)." In: *Journal of Biological Chemistry* 262.19, pp. 9412–9420. DOI: 10.1016/s0021-9258(18)48095-7.

Jolliffe, Ian T. and Jorge Cadima (2016). "Principal component analysis: A review and recent developments." In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065, p. 20150202. DOI: 10.1098/rsta.2015.0202.

Jones, Matthew R., Lei Chong, and Saverio Bellusci (2021). "Fgf10/Fgfr2b Signaling Orchestrates the Symphony of Molecular, Cellular, and Physical Processes Required for Harmonious Airway Branching Morphogenesis." In: *Frontiers in Cell and Developmental Biology* 8. DOI: 10.3389/fcell.2020.620667.

Jones, Robert C. et al. (2022). "The tabula sapiens: A multiple-organ, single-cell transcriptomic atlas of humans." In: *Science* 376.6594. DOI: 10.1126/science.abl4896.

Joost, Simon et al. (2016). "Single-cell transcriptomics reveals that differentiation and spatial signatures shape epidermal and hair follicle heterogeneity." In: *Cell Systems* 3.3. DOI: 10.1016/j.cels.2016.08.010.

Joost, Simon et al. (2020). "The molecular anatomy of mouse skin during hair growth and rest." In: *Cell Stem Cell* 26.3. DOI: 10.1016/j.stem.2020.01.012.

Jou, W. MIN et al. (1972). "Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein." In: *Nature* 237.5350, 82–88. DOI: 10.1038/237082a0.

Jozsi, Mihaly (2017). "Factor H Family Proteins in Complement Evasion of Microorganisms." In: *Frontiers in Immunology* 8. DOI: 10.3389/fimmu.2017.00571.

Juhlin, L. (1997). "Hyaluronan in skin." In: *Journal of Internal Medicine* 242.1, 61–66. DOI: 10.1046/j.1365-2796.1997.00175.x.

Juin, P. et al. (1999). "c-Myc-induced sensitization to apoptosis is mediated through cytochrome c release." In: *Genes & Development* 13.11, pp. 1367–1381. DOI: 10.1101/gad.13.11.1367.

Jung, Ki-Kyung et al. (2006). "Identification of CD63 as a tissue inhibitor of metalloproteinase-1 interacting cell surface protein." In: *The EMBO Journal* 25.17, pp. 3934–3942. DOI: 10.1038/sj.emboj.7601281.

Junker, Johan P.E. et al. (2009). "Adipogenic, chondrogenic and osteogenic differentiation of clonally derived human dermal fibroblasts." In: *Cells Tissues Organs* 191.2, 105–118. DOI: 10.1159/000232157.

Kadler, Karl E, Adele Hill, and Elizabeth G Canty-Laird (2008). "Collagen fibrillogenesis: Fibronectin, integrins, and minor collagens as organizers and nucleators." In: *Current Opinion in Cell Biology* 20.5, 495–501. DOI: 10.1016/j.ceb.2008.06.008.

Kahari, Veli-Matti and Ulpu Saarialho-Kere (1997). "Matrix metalloproteinases in skin." In: *Experimental Dermatology* 6.5, pp. 199–213. DOI: 10.1111/j.1600-0625.1997.tb00164.x.

Kai, Yoshiro et al. (2017). "Silencing of Carbohydrate Sulfotransferase 15 Hinders Murine Pulmonary Fibrosis Development." In: *Molecular Therapy - Nucleic Acids* 6, pp. 163–172. DOI: 10.1016/j.omtn.2016.12.008.

Kaji, Chiaki et al. (2012). "The expression of podoplanin and classic cadherins in the mouse brain." In: *Journal of Anatomy* 220.5, pp. 435–446. DOI: 10.1111/j.1469-7580.2012.01484.x.

Kalamajski, Sebastian et al. (2009). "Asporin competes with decorin for collagen binding, binds calcium and promotes osteoblast collagen mineralization." In: *Biochemical Journal* 423.1, pp. 53–59. DOI: 10.1042/bj20090542.

Kamio, Koichiro et al. (2007). "Prostacyclin Analogs Inhibit Fibroblast Contraction of Collagen Gels through the cAMP-PKA Pathway." In: *American Journal of Respiratory Cell and Molecular Biology* 37.1, pp. 113–120. DOI: 10.1165/rcmb.2007-0009oc.

Kamio, Koichiro et al. (2008). "Prostacyclin analogs stimulate VEGF production from human lung fibroblasts in culture." In: *American Journal of Physiology-Lung Cellular and Molecular Physiology* 294.6, pp. L1226–L1232. DOI: 10.1152/ajplung.00129.2007.

Kang, Hyeon-Gu et al. (2020). "SPON2 Is Upregulated through Notch Signaling Pathway and Promotes Tumor Progression in Gastric Cancer." In: *Cancers* 12.6, p. 1439. DOI: 10.3390/cancers12061439.

Karlen, Helène et al. (2020). "IL-15 Expression Pattern in Atopic Dermatitis." In: *International Archives of Allergy and Immunology* 181.6, pp. 417–421. DOI: 10.1159/000508515.

Karpus, Olga N. et al. (2019). "Colonic CD90$^+$ Crypt Fibroblasts Secrete Semaphorins to Support Epithelial Growth." In: *Cell Reports* 26.13, 3698–3708.e5. DOI: 10.1016/j.celrep.2019.02.101.

Katoh, Masaru (2009). "FGFR2 Abnormalities Underlie a Spectrum of Bone, Skin, and Cancer Pathologies." In: *Journal of Investigative Dermatology* 129.8, pp. 1861–1867. DOI: 10.1038/jid.2009.97.

Kaur, Jasvir and Dieter P. Reinhardt (2015). "Extracellular Matrix (ECM) Molecules." In: *Stem Cell Biology and Tissue Engineering in Dental Sciences*. Elsevier, pp. 25–45. DOI: 10.1016/b978-0-12-397157-9.00003-5.

Kawabe, Tsutomu et al. (1994). "The immune responses in CD40-deficient mice: Impaired immunoglobulin class switching and germinal center formation." In: *Immunity* 1.3, pp. 167–178. DOI: 10.1016/1074-7613(94)90095-7.

Kawasaki, Aya et al. (2007). "Wnt5a promotes adhesion of human dermal fibroblasts by triggering a phosphatidylinositol-3 kinase/Akt signal." In: *Cellular Signalling* 19.12, pp. 2498–2506. DOI: 10.1016/j.cellsig.2007.07.023.

Kawashima, Hiroto et al. (2000). "Binding of a Large Chondroitin Sulfate/Dermatan Sulfate Proteoglycan, Versican, to L-selectin, P-selectin, and CD44." In: *Journal of Biological Chemistry* 275.45, pp. 35448–35456. DOI: 10.1074/jbc.m003387200.

Kazanskaya, Olga et al. (2004). "R-spondin2 is a secreted activator of Wnt/ β-catenin signaling and is required for xenopus myogenesis." In: *Developmental Cell* 7.4, 525–534. DOI: 10.1016/j.devcel.2004.07.019.

Keays, Kathryne Melissa et al. (2005). "Laser capture microdissection and single-cell RT-PCR without RNA purification." In: *Journal of Immunological Methods* 302.1-2, 90–98. DOI: 10.1016/j.jim.2005.04.018.

Kedishvili, Natalia Y. (2013). "Enzymology of retinoic acid biosynthesis and degradation." In: *Journal of Lipid Research* 54.7, pp. 1744–1760. DOI: 10.1194/jlr.r037028.

Kehlet, S.N., H. Jessen, and M.A. Karsdal (2019a). "Type XXI collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 171–174. DOI: 10.1016/b978-0-12-817068-7.00021-5.

– (2019b). "Type XXIII collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 181–185. DOI: 10.1016/b978-0-12-817068-7.00023-9.

Kelwick, Richard et al. (2015). "The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) family." In: *Genome Biology* 16.1. DOI: 10.1186/s13059-015-0676-3.

Kemp, Samantha B. et al. (2021). "Apolipoprotein E Promotes Immune Suppression in Pancreatic Cancer through NF-κB–Mediated Production of CXCL1." In: *Cancer Research* 81.16, pp. 4305–4318. DOI: 10.1158/0008-5472.can-20-3929.

Kerrigan, Ann M. et al. (2009). "CLEC-2 Is a Phagocytic Activation Receptor Expressed on Murine Peripheral Blood Neutrophils." In: *The Journal of Immunology* 182.7, pp. 4150–4157. DOI: 10.4049/jimmunol.0802808.

Kharchenko, Peter V. (2021). "The triumphs and limitations of computational methods for scrna-seq." In: *Nature Methods* 18.7, 723–732. DOI: 10.1038/s41592-021-01171-x.

Kharchenko, Peter V, Lev Silberstein, and David T Scadden (2014). "Bayesian approach to single-cell differential expression analysis." In: *Nature Methods* 11.7, 740–742. DOI: 10.1038/nmeth.2967.

Kieffer, Yann et al. (2020). "Single-Cell Analysis Reveals Fibroblast Clusters Linked to Immunotherapy Resistance in Cancer." In: *Cancer Discovery* 10.9, pp. 1330–1351. DOI: 10.1158/2159-8290.cd-19-1384.

Kikuchi, Akira et al. (2011). "New insights into the mechanism of Wnt Signaling Pathway Activation." In: *International Review of Cell and Molecular Biology*, 21–71. DOI: 10.1016/b978-0-12-386035-4.00002-1.

Kim, Dongwon et al. (2019). "Noncoding dsRNA induces retinoic acid synthesis to stimulate hair follicle regeneration via TLR3." In: *Nature Communications* 10.1. DOI: 10.1038/s41467-019-10811-y.

Kim, Doyoung et al. (2020a). "Targeted therapy guided by single-cell transcriptomic analysis in drug-induced hypersensitivity syndrome: A case report." In: *Nature Medicine* 26.2, 236–243. DOI: 10.1038/s41591-019-0733-7.

Kim, Han-Seop et al. (2020b). "Directly induced human Schwann cell precursors as a valuable source of Schwann cells." In: *Stem Cell Research & Therapy* 11.1. DOI: 10.1186/s13287-020-01772-x.

Kim, Hangeun et al. (2013a). "Hepatitis C Virus Suppresses C9 Complement Synthesis and Impairs Membrane Attack Complex Function." In: *Journal of Virology* 87.10, pp. 5858–5867. DOI: 10.1128/jvi.00174-13.

Kim, Injune et al. (2000). "Angiopoietin-1 regulates endothelial cell survival through the phosphatidylinositol 3'-kinase/AKT signal transduction pathway." In: *Circulation Research* 86.1, 24–29. DOI: 10.1161/01.res.86.1.24.

Kim, Jaehwan et al. (2021). "Single-cell transcriptomics applied to emigrating cells from psoriasis elucidate pathogenic vs. regulatory immune cell subsets." In: *Journal of Allergy and Clinical Immunology*. DOI: 10.1016/j.jaci.2021.04.021.

Kim, Ji Hye et al. (2015). "Functional Regulation of Adipose-Derived Stem Cells by PDGF-D." In: *Stem Cells* 33.2, pp. 542–556. DOI: 10.1002/stem.1865.

Kim, Kyung-Ah et al. (2008). "R-Spondin family members regulate the Wnt pathway by a common mechanism." In: *Molecular Biology of the Cell* 19.6, 2588–2596. DOI: 10.1091/mbc.e08-02-0187.

Kim, Soo-Hyun et al. (2005). "Interleukin-32." In: *Immunity* 22.1, pp. 131–142. DOI: 10.1016/j.immuni.2004.12.003.

Kim, Woojin Scott et al. (2013b). "ABCA8 stimulates sphingomyelin production in oligodendrocytes." In: *Biochemical Journal* 452.3, pp. 401–410. DOI: 10.1042/bj20121764.

Kinchen, James et al. (2018). "Structural remodeling of the human colonic mesenchyme in inflammatory bowel disease." In: *Cell* 175.2. DOI: 10.1016/j.cell.2018.08.067.

Kirchner, Jacqueline and Michael J. Bevan (1999). "Itm2a Is Induced during Thymocyte Selection and T Cell Activation and Causes Downregulation of Cd8 When Overexpressed in $Cd4^{+}Cd8^{+}$ Double Positive Thymocytes." In: *Journal of Experimental Medicine* 190.2, pp. 217–228. DOI: 10.1084/jem.190.2.217.

Kiselev, Vladimir Yu, Tallulah S. Andrews, and Martin Hemberg (2019). "Challenges in unsupervised clustering of single-cell RNA-seq data." In: *Nature Reviews Genetics* 20.5, 273–282. DOI: 10.1038/s41576-018-0088-9.

Kiselev, Vladimir Yu, Andrew Yiu, and Martin Hemberg (2018). "scmap: projection of single-cell RNA-seq data across data sets." In: *Nature Methods* 15.5, 359–362. DOI: 10.1038/nmeth.4644.

Kiselev, Vladimir Yu et al. (2017). "SC3: Consensus clustering of single-cell RNA-seq data." In: *Nature Methods* 14.5, 483–486. DOI: 10.1038/nmeth.4236.

Kivioja, Teemu et al. (2012). "Counting absolute numbers of molecules using unique molecular identifiers." In: *Nature Methods* 9.1, 72–74. DOI: 10.1038/nmeth.1778.

Kizawa, Hideki et al. (2005). "An aspartic acid repeat polymorphism in asporin inhibits chondrogenesis and increases susceptibility to osteoarthritis." In: *Nature Genetics* 37.2, pp. 138–144. DOI: 10.1038/ng1496.

Klein, Allon M. et al. (2015). "Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells." In: *Cell* 161.5, 1187–1201. DOI: 10.1016/j.cell.2015.04.044.

Kleiser, Svenja and Alexander Nyström (2020). "Interplay between Cell-Surface Receptors and Extracellular Matrix in Skin." In: *Biomolecules* 10.8, p. 1170. DOI: 10.3390/biom10081170.

Klopfenstein, D. V. et al. (2018). "GOATOOLS: A python library for gene ontology analyses." In: *Scientific Reports* 8.1. DOI: 10.1038/s41598-018-28948-z.

Kobak, Dmitry and Philipp Berens (2019). "The art of using t-SNE for single-cell transcriptomics." In: *Nature Communications* 10.1. DOI: 10.1038/s41467-019-13056-x.

Kobak, Dmitry and George C. Linderman (2021). "Initialization is critical for preserving global data structure in both T-Sne and Umap." In: *Nature Biotechnology* 39.2, 156–157. DOI: 10.1038/s41587-020-00809-z.

Kobayashi, Tetsuro et al. (2019). "Homeostatic Control of Sebaceous Glands by Innate Lymphoid Cells Regulates Commensal Bacteria Equilibrium." In: *Cell* 176.5, 982–997.e16. DOI: 10.1016/j.cell.2018.12.031.

Koch, Manuel et al. (2003). "Collagen XXIV, a Vertebrate Fibrillar Collagen with Structural Features of Invertebrate Collagens." In: *Journal of Biological Chemistry* 278.44, pp. 43236–43244. DOI: 10.1074/jbc.m302112200.

Kohyama, Tadashi et al. (2002). "Prostacyclin analogs inhibit fibroblast migration." In: *American Journal of Physiology-Lung Cellular and Molecular Physiology* 283.2, pp. L428–L432. DOI: 10.1152/ajplung.00432.2001.

Kolodziejczyk, Aleksandra A. et al. (2015). "The technology and biology of single-cell RNA sequencing." In: *Molecular Cell* 58.4, 610–620. DOI: 10.1016/j.molcel.2015.04.005.

Komatsu, Nahoko et al. (2003). "Expression and Localization of Tissue Kallikrein mRNAs in Human Epidermis and Appendages." In: *Journal of Investigative Dermatology* 121.3, pp. 542–549. DOI: 10.1046/j.1523-1747.2003.12363.x.

Kondo, Shinya et al. (2022). "Elastin microfibril interface-located protein 1 and its catabolic enzyme, cathepsin K, regulate the age-related structure of elastic fibers in the skin." In: *Journal of Cosmetic Dermatology* 21.10, pp. 4796–4804. DOI: 10.1111/jocd.14789.

Korkmaz, Kemal S. et al. (2001). "Distinctly Different Gene Structure ofKLK4/KLK-L1/Prostase/ARM1 Compared with Other Members of the Kallikrein Family: Intracellular Localization, Alternative cDNA Forms, and Regulation by Multiple Hormones." In: *DNA and Cell Biology* 20.7, pp. 435–445. DOI: 10.1089/104454901750361497.

Korosec, Ana et al. (2019). "Lineage identity and location within the dermis determine the function of papillary and reticular fibroblasts in human skin." In: *Journal of Investigative Dermatology* 139.2, 342–351. DOI: 10.1016/j.jid.2018.07.033.

Korsunsky, Ilya et al. (2019). "Fast, sensitive and accurate integration of single-cell data with Harmony." In: *Nature Methods* 16.12, 1289–1296. DOI: 10.1038/s41592-019-0619-0.

Korsunsky, Ilya et al. (2021). "Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast phenotypes in four chronic inflammatory diseases." In: *bioRxiv*. DOI: 10.1101/2021.01.11.426253.

Korsunsky, Ilya et al. (2022). "Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast phenotypes in four chronic inflammatory diseases." In: *Clinical and Translational Medicine* 3.7, 481–518.e14. DOI: 10.1016/j.medj.2022.05.002.

Koutsioumpa, Marina et al. (2015). "Receptor protein tyrosine phosphatase beta/zeta is a functional binding partner for vascular endothelial growth factor." In: *Molecular Cancer* 14.1. DOI: 10.1186/s12943-015-0287-3.

Kowalczyk, Monika S. et al. (2015). "Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells." In: *Genome Research* 25.12, 1860–1872. DOI: 10.1101/gr.192237.115.

Kraft, Vanessa A. N. et al. (2019). "GTP Cyclohydrolase 1/Tetrahydrobiopterin Counteract Ferroptosis through Lipid Remodeling." In: *ACS Central Science* 6.1, pp. 41–53. DOI: 10.1021/acscentsci.9b01063.

Krausgruber, Thomas et al. (2020). "Structural cells are key regulators of organ-specific immune responses." In: *Nature* 583.7815, 296–302. DOI: 10.1038/s41586-020-2424-4.

Kretzschmar, Kai et al. (2021). "Troy/Tnfrsf19 marks epidermal cells that govern interfollicular epidermal renewal and cornification." In: *Stem Cell Reports* 16.9, pp. 2379–2394. DOI: 10.1016/j.stemcr.2021.07.007.

Kristensen, J.H. and M.A. Karsdal (2016). "Elastin." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 197–201. DOI: 10.1016/b978-0-12-809847-9.00030-1.

Kroll, J. and J. Waltenberger (2000). "Regulation der endothelfunktion und der angiogenese durch den vaskulären endothelialen wachstumsfaktor-A (VEGF-A)." In: *Zeitschrift für Kardiologie* 89.3, 206–218. DOI: 10.1007/s003920050472.

Krumdieck, R et al. (1992). "The proteoglycan decorin binds C1q and inhibits the activity of the C1 complex." In: *The Journal of Immunology* 149.11, pp. 3695–3701. DOI: 10.4049/jimmunol.149.11.3695.

Kuleshov, Maxim V. et al. (2016). "ENRICHR: A comprehensive gene set enrichment analysis web server 2016 update." In: *Nucleic Acids Research* 44.W1. DOI: 10.1093/nar/gkw377.

Kumar, Suhas (2015). *Fundamental Limits to Moore's Law*. DOI: 10.48550/ARXIV.1511.05956.

Kumar, Sushil et al. (2021). "Complement-containing small extracellular vesicles from adventitial fibroblasts induce proinflammatory and metabolic reprogramming in macrophages." In: *JCI Insight* 6.21. DOI: 10.1172/jci.insight.148382.

Kurashima, Yosuke et al. (2014). "The Enzyme Cyp26b1 Mediates Inhibition of Mast Cell Activation by Fibroblasts to Maintain Skin-Barrier Homeostasis." In: *Immunity* 40.4, pp. 530–541. DOI: 10.1016/j.immuni.2014.01.014.

Kurosawa, Nobuyuki et al. (2001). In: *Glycoconjugate Journal* 18.6, pp. 499–507. DOI: 10.1023/a:1016042303253.

Kwon, Heung-Sun et al. (2009). "Myocilin Is a Modulator of Wnt Signaling." In: *Molecular and Cellular Biology* 29.8, pp. 2139–2154. DOI: 10.1128/mcb.01274-08.

L. Lun, Aaron T., Karsten Bach, and John C. Marioni (2016). "Pooling across cells to normalize single-cell RNA sequencing data with many zero counts." In: *Genome Biology* 17.1. DOI: 10.1186/s13059-016-0947-7.

La Manno, Gioele et al. (2018). "RNA velocity of single cells." In: *Nature* 560.7719, 494–498. DOI: 10.1038/s41586-018-0414-6.

La, Jennifer et al. (2016). "Downregulation of TGF-β Receptor-2 Expression and Signaling through Inhibition of Na/K-ATPase." In: *PLOS ONE* 11.12. Ed. by Donald Gullberg, e0168363. DOI: 10.1371/journal.pone.0168363.

Laberge, Remi-Martin et al. (2015). "MTOR regulates the pro-tumorigenic senescence-associated secretory phenotype by promoting IL1A translation." In: *Nature Cell Biology* 17.8, pp. 1049–1061. DOI: 10.1038/ncb3195.

Lafzi, Atefeh et al. (2018). "Tutorial: Guidelines for the experimental design of single-cell RNA sequencing studies." In: *Nature Protocols* 13.12, 2742–2757. DOI: 10.1038/s41596-018-0073-y.

Lagoutte, Priscillia et al. (2021). "Procollagen C-proteinase enhancer-1 (PCPE-1), a potential biomarker and therapeutic target for fibrosis." In: *Matrix Biology Plus* 11, p. 100062. DOI: 10.1016/j.mbplus.2021.100062.

Lähnemann, David et al. (2020). "Eleven grand challenges in single-cell data science." In: *Genome Biology* 21.1. DOI: 10.1186/s13059-020-1926-6.

Lai, Shih-Lei, Andy J Chien, and Randall T Moon (2009). "Wnt/Fz signaling and the cytoskeleton: potential roles in tumorigenesis." In: *Cell Research* 19.5, pp. 532–545. DOI: 10.1038/cr.2009.41.

Lamertz, Larissa et al. (2018). "Soluble gp130 prevents interleukin-6 and interleukin-11 cluster signaling but not intracellular autocrine responses." In: *Science Signaling* 11.550. DOI: 10.1126/scisignal.aar7388.

Langton, Abigail K. et al. (2012). "Cross-linking of structural proteins in ageing skin: an in situ assay for the detection of amine oxidase activity." In: *Biogerontology* 14.1, pp. 89–97. DOI: 10.1007/s10522-012-9394-3.

Larigot, Lucie et al. (2022). "Aryl Hydrocarbon Receptor and Its Diverse Ligands and Functions: An Exposome Receptor." In: *Annual Review of Pharmacology and Toxicology*

62.1, pp. 383–404. DOI: 10.1146/annurev-pharmtox-052220-115707.

Laron, Efrat Avigad, Emil Aamar, and David Enshell-Seijffers (2018). "The Mesenchymal Niche of the Hair Follicle Induces Regeneration by Releasing Primed Progenitors from Inhibitory Effects of Quiescent Stem Cells." In: *Cell Reports* 24.4, 909–921.e3. DOI: 10.1016/j.celrep.2018.06.084.

Larsson-Callerfelt, Anna-Karin et al. (2013). "Defective alterations in the collagen network to prostacyclin in COPD lung fibroblasts." In: *Respiratory Research* 14.1, p. 21. DOI: 10.1186/1465-9921-14-21.

Lau, Wim de et al. (2011). "LGR5 homologues associate with Wnt receptors and mediate R-spondin signalling." In: *Nature* 476.7360, 293–297. DOI: 10.1038/nature10337.

Laufer, Julia M. et al. (2019). "CCR7 Is Recruited to the Immunological Synapse, Acts as Co-stimulatory Molecule and Drives LFA-1 Clustering for Efficient T Cell Adhesion Through ZAP70." In: *Frontiers in Immunology* 9. DOI: 10.3389/fimmu.2018.03115.

Launay, S et al. (2008). "HtrA1-dependent proteolysis of TGF-β controls both neuronal maturation and developmental survival." In: *Cell Death & Differentiation* 15.9, pp. 1408–1416. DOI: 10.1038/cdd.2008.82.

Lause, Jan, Philipp Berens, and Dmitry Kobak (2021). "Analytic Pearson residuals for normalization of single-cell RNA-seq umi data." In: *Genome Biology* 22.1. DOI: 10.1186/s13059-021-02451-7.

LeBleu, Valerie S. and Eric G. Neilson (2020). "Origin and functional heterogeneity of fibroblasts." In: *The FASEB Journal* 34.3, 3519–3536. DOI: 10.1096/fj.201903188r.

Lee, C. W. et al. (1983). "Conversion of leukotriene D4 to leukotriene E4 by a dipeptidase released from the specific granule of human polymorphonuclear leucocytes." In: *Immunology* 48.1, pp. 27–35.

Lee, Hangnoh et al. (2016). "External RNA controls consortium beta version update." In: *Journal of Genomics* 4, 19–22. DOI: 10.7150/jgen.16082.

Lee, Hyun Kyoung and Benjamin Deneen (2012a). "Daam2 Is Required for Dorsal Patterning via Modulation of Canonical Wnt Signaling in the Developing Spinal Cord." In: *Developmental Cell* 22.1, pp. 183–196. DOI: 10.1016/j.devcel.2011.10.025.

Lee, Hyun Kyoung et al. (2015). "Daam2-PIP5K Is a Regulatory Pathway for Wnt Signaling and Therapeutic Target for Remyelination in the CNS." In: *Neuron* 85.6, pp. 1227–1243. DOI: 10.1016/j.neuron.2015.02.024.

Lee, Ki Won, Seyeon Lim, and Kwang Dong Kim (2022). "The Function of N-Myc Downstream-Regulated Gene 2 (NDRG2) as a Negative Regulator in Tumor Cell Metastasis." In: *International Journal of Molecular Sciences* 23.16, p. 9365. DOI: 10.3390/ijms23169365.

Lee, Mi Hye et al. (2012b). "Conditioned media obtained from human outer root sheath follicular keratinocyte culture activates signalling pathways that contribute to maintenance of hair-inducing capacity and increases trichogenicity of cultured dermal cells." In: *Experimental Dermatology* 21.10, pp. 793–795. DOI: 10.1111/j.1600-0625.2012.01570.x.

Lee, Stephen S.J et al. (2004). "Structure of the Integrin Binding Fragment from Fibrillin-1 Gives New Insights into Microfibril Organization." In: *Structure* 12.4, pp. 717–729. DOI: 10.1016/j.str.2004.02.023.

Leeming, D.J. and M.A. Karsdal (2019). "Type V collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 51–57. DOI: 10.1016/b978-0-12-817068-7.00005-7.

Leite, J. A. et al. (2020). "The α2 Na^+/K^+-ATPase isoform mediates LPS-induced neuroinflammation." In: *Scientific Reports* 10.1. DOI: 10.1038/s41598-020-71027-5.

Lendahl, Urban, Lars Muhl, and Christer Betsholtz (2022). "Identification, discrimination and heterogeneity of fibroblasts." In: *Nature Communications* 13.1. DOI: 10.1038/s41467-022-30633-9.

Lethias, Claire et al. (2006). "A model of tenascin-X integration within the collagenous network." In: *FEBS Letters* 580.26, pp. 6281–6285. DOI: 10.1016/j.febslet.2006.10.037.

Lewin, Benjamin et al. (2006). *Cells*. 1st ed. Jones & Bartlett Learning.

Lewthwaite, A.J. et al. (2015). "Novel GCH1 variant in Dopa-responsive dystonia and Parkinson's disease." In: *Parkinsonism & Related Disorders* 21.4, pp. 394–397. DOI: 10.1016/j.parkreldis.2015.01.004.

Leyva-Castillo, Juan Manuel et al. (2022). "Single-cell transcriptome profile of mouse skin undergoing antigen-driven allergic inflammation recapitulates findings in atopic dermatitis skin lesions." In: *Journal of Allergy and Clinical Immunology*. DOI: 10.1016/j.jaci.2022.03.002.

Lhoumeau, Anne-Catherine et al. (2011). "PTK7: A cell polarity receptor with multiple facets." In: *Cell Cycle* 10.8, pp. 1233–1236. DOI: 10.4161/cc.10.8.15368.

Li, Guoxing et al. (2022a). "Gpx3 and Egr1 Are Involved in Regulating the Differentiation Fate of Cardiac Fibroblasts under Pressure Overload." In: *Oxidative Medicine and Cellular Longevity* 2022. Ed. by Tao Li, pp. 1–21. DOI: 10.1155/2022/3235250.

Li, Hengcun et al. (2019). "Cytoplasmic Asporin promotes cell migration by regulating TGF-β/Smad2/3 pathway and indicates a poor prognosis in colorectal cancer." In: *Cell Death & Disease* 10.2. DOI: 10.1038/s41419-019-1376-9.

Li, Huirong et al. (2017). "Epilation induces hair and skin pigmentation through an EDN3/EDNRB-dependent regenerative response of melanocyte stem cells." In: *Scientific Reports* 7.1. DOI: 10.1038/s41598-017-07683-x.

Li, Jingyi Jessica, Peter J Bickel, and Mark D Biggin (2014). "System wide analyses have underestimated protein abundances and the importance of transcription in mammals." In: *PeerJ* 2. DOI: 10.7717/peerj.270.

Li, Nan et al. (2020a). "Transcriptional Activation of Matricellular Protein Spondin2 (SPON2) by BRG1 in Vascular Endothelial Cells Promotes Macrophage Chemotaxis." In: *Frontiers in Cell and Developmental Biology* 8. DOI: 10.3389/fcell.2020.00794.

Li, Qingyang et al. (2021). "Single-cell transcriptome profiling reveals vascular endothelial cell heterogeneity in Human skin." In: *Theranostics* 11.13, 6461–6476. DOI: 10.7150/thno.54917.

Li, Rongbo et al. (2018a). "PDGFRA marks a cellular lineage with distinct contributions to myofibroblasts in lung maturation and injury response." In: *eLife* 7. DOI: 10.7554/elife.36865.

Li, Wei Vivian and Jingyi Jessica Li (2018b). "An accurate and robust imputation method scImpute for single-cell RNA-seq data." In: *Nature Communications* 9.1. DOI: 10.1038/s41467-018-03405-7.

Li, Xiaodong et al. (2010). "Increased Expression of Cathepsins and Obesity-Induced Proinflammatory Cytokines in Lacrimal Glands of Male NOD Mouse." In: *Investigative Opthalmology & Visual Science* 51.10, p. 5019. DOI: 10.1167/iovs.09-4523.

Li, Yuejuan et al. (2011). "Severe lung fibrosis requires an invasive fibroblast phenotype regulated by hyaluronan and CD44." In: *Journal of Experimental Medicine* 208.7, pp. 1459–1471. DOI: 10.1084/jem.20102510.

Li, Yunyuan, Reza Baradar Jalili, and Aziz Ghahary (2016). "Accelerating skin wound healing by M-CSF through generating SSEA-1 and -3 stem cells in the injured sites." In: *Scientific Reports* 6.1. DOI: 10.1038/srep28979.

Li, Yuxia et al. (2022b). "Loss of Acta2 in cardiac fibroblasts does not prevent the myofibroblast differentiation or affect the cardiac repair after myocardial infarction." In: *Journal of Molecular and Cellular Cardiology* 171, pp. 117–132. DOI: 10.1016/j.yjmcc.2022.08.003.

Li, Zhenzhen et al. (2020b). "Extracellular S100A4 as a key player in fibrotic diseases." In: *Journal of Cellular and Molecular Medicine* 24.11, 5973–5983. DOI: 10.1111/jcmm.15259.

Liang, Chi-Jung et al. (2019). "SFRPs Are Biphasic Modulators of Wnt-Signaling-Elicited Cancer Stem Cell Properties beyond Extracellular Control." In: *Cell Reports* 28.6, 1511–1525.e5. DOI: 10.1016/j.celrep.2019.07.023.

Lichtenberger, Beate M., Maria Mastrogiannaki, and Fiona M. Watt (2016). "Epidermal β-catenin activation remodels the dermis via paracrine signalling to distinct fibroblast lineages." In: *Nature Communications* 7.1. DOI: 10.1038/ncomms10537.

Lim, Chae Ho et al. (2018). "Hedgehog stimulates hair follicle neogenesis by creating inductive dermis during murine skin wound healing." In: *Nature Communications* 9.1. DOI: 10.1038/s41467-018-07142-9.

Lim, X. and R. Nusse (2012). "Wnt Signaling in Skin Development, Homeostasis, and Disease." In: *Cold Spring Harbor Perspectives in Biology* 5.2, a008029–a008029. DOI: 10.1101/cshperspect.a008029.

Lim, Xinhong et al. (2016). "Axin2 marks quiescent hair follicle bulge stem cells that are maintained by autocrine Wnt/β-catenin signaling." In: *Proceedings of the National Academy of Sciences* 113.11. DOI: 10.1073/pnas.1601599113.

Lima, Joana Esteves de et al. (2020). "BMP signalling directs a fibroblast-to-myoblast conversion at the connective tissue/muscle interface to pattern limb muscles." In: DOI: 10.1101/2020.07.20.211342.

Limatola, Cristina and Richard M. Ransohoff (2014). "Modulating neurotoxicity through CX3CL1/CX3CR1 signaling." In: *Frontiers in Cellular Neuroscience* 8. DOI: 10.3389/fncel.2014.00229.

Lin, Aifen and Wei-Hua Yan (2019a). "The Emerging Roles of Human Leukocyte Antigen-F in Immune Modulation and Viral Infection." In: *Frontiers in Immunology* 10. DOI: 10.3389/fimmu.2019.00964.

Lin, Dasheng et al. (2017). "Tenomodulin is essential for prevention of adipocyte accumulation and fibrovascular scar formation during early tendon healing." In: *Cell Death & Disease* 8.10, e3116–e3116. DOI: 10.1038/cddis.2017.510.

Lin, Haishan et al. (2008). "Discovery of a Cytokine and Its Receptor by Functional Screening of the Extracellular Proteome." In: *Science* 320.5877, pp. 807–811. DOI: 10.1126/science.1154370.

Lin, Li et al. (2020). "Normalizing single-cell RNA sequencing data with internal spike-in-like genes." In: *NAR Genomics and Bioinformatics* 2.3. DOI: 10.1093/nargab/lqaa059.

Lin, WeiYu et al. (2019b). "Function of CSF1 and IL34 in Macrophage Homeostasis, Inflammation, and Cancer." In: *Frontiers in Immunology* 10. DOI: 10.3389/fimmu.2019.02019.

Lin, Yi et al. (2009). "CXCL12 enhances exogenous $CD4^+CD25^+$ T cell migration and prevents embryo loss in non-obese diabetic mice." In: *Fertility and Sterility* 91.6, pp. 2687–2696. DOI: 10.1016/j.fertnstert.2008.01.109.

Lin, Yuanbin et al. (2022). "Single-cell RNA-seq of UVB-radiated skin reveals landscape of photoaging-related inflammation and protection by Vitamin D." In: *Gene* 831, p. 146563. DOI: 10.1016/j.gene.2022.146563.

Lindeboom, Rik G.H., Aviv Regev, and Sarah A. Teichmann (2021). "Towards a human cell atlas: Taking Notes from the past." In: *Trends in Genetics* 37.7, 625–630. DOI: 10.1016/j.tig.2021.03.007.

Linderman, George C. et al. (2019). "Fast interpolation-based T-Sne for improved visualization of single-cell RNA-seq data." In: *Nature Methods* 16.3, 243–245. DOI: 10.1038/s41592-018-0308-4.

Linke, Andreas et al. (2010). "The Suppressor of Cytokine Signaling (SOCS)-3 Determines Keratinocyte Proliferative and Migratory Potential during Skin Repair." In: *Journal of Investigative Dermatology* 130.3, pp. 876–885. DOI: 10.1038/jid.2009.344.

Lipfert, J. et al. (2013). "CXCR4 and CXCR7 form a functional receptor unit for SDF-1/CXCL12 in primary rodent microglia." In: *Neuropathology and Applied Neurobiology* 39.6, pp. 667–680. DOI: 10.1111/nan.12015.

Lippert, Dustin ND and John A Wilkins (2012). "Glia maturation factor gamma regulates the migration and adherence of human T lymphocytes." In: *BMC Immunology* 13.1. DOI: 10.1186/1471-2172-13-21.

Lister, Ryan et al. (2008). "Highly integrated single-base resolution maps of the epigenome in Arabidopsis." In: *Cell* 133.3, 523–536. DOI: 10.1016/j.cell.2008.03.029.

Liu, Haikun et al. (2019). "Endothelins (EDN1, EDN2, EDN3) and their receptors (EDNRA, EDNRB, EDNRB2) in chickens: Functional analysis and tissue distribution." In: *General and Comparative Endocrinology* 283, p. 113231. DOI: 10.1016/j.ygcen.2019.113231.

Liu, Hao and Shuai Huang (2020a). "Role of oxysterol-binding protein-related proteins in malignant human tumours." In: *World Journal of Clinical Cases* 8.1, pp. 1–10. DOI: 10.12998/wjcc.v8.i1.1.

Liu, Hua, Bo Chen, and Brenda Lilly (2008). "Fibroblasts potentiate blood vessel formation partially through secreted factor TIMP-1." In: *Angiogenesis* 11.3, 223–234. DOI: 10.1007/s10456-008-9102-8.

Liu, Jared et al. (2021a). "Single-cell RNA sequencing of psoriatic skin identifies pathogenic TC17 cell subsets and reveals distinctions between CD8+ T cells in autoimmunity and cancer." In: *Journal of Allergy and Clinical Immunology* 147.6, 2370–2380. DOI: 10.1016/j.jaci.2020.11.028.

Liu, Jinqi et al. (2001). "Enhanced $CD4^+$ T Cell Proliferation and Th2 Cytokine Production in DR6-Deficient Mice." In: *Immunity* 15.1, pp. 23–34. DOI: 10.1016/s1074-7613(01)00162-5.

Liu, Jun yan et al. (2018). "AEBP1 promotes epithelial-mesenchymal transition of gastric cancer cells by activating the NF-κB pathway and predicts poor outcome of the patients." In: *Scientific Reports* 8.1. DOI: 10.1038/s41598-018-29878-6.

Liu, Junhui et al. (2022). "Expression and Prognostic Role of Glia Maturation Factor-γ in Gliomas." In: *Frontiers in Molecular Neuroscience* 15. DOI: 10.3389/fnmol.2022.906762.

Liu, Libangxi et al. (2020b). "Cartilage intermediate layer protein affects the progression of intervertebral disc degeneration by regulating the extracellular microenvironment (Review)." In: *International Journal of Molecular Medicine* 47.2, pp. 475–484. DOI: 10.3892/ijmm.2020.4832.

Liu, Longwei et al. (2021b). "Asporin inhibits collagen matrix-mediated intercellular mechanocommunications between fibroblasts during keloid progression." In: *The FASEB Journal* 35.7. DOI: 10.1096/fj.202100111r.

Liu, Ting et al. (2017). "NF-κB signaling in inflammation." In: *Signal Transduction and Targeted Therapy* 2.1. DOI: 10.1038/sigtrans.2017.23.

Liu, Xiangde et al. (2007). "The CC Chemokine Ligand 2 (CCL2) Mediates Fibroblast Survival through IL-6." In: *American Journal of Respiratory Cell and Molecular Biology* 37.1, pp. 121–128. DOI: 10.1165/rcmb.2005-0253oc.

Liu, Xiaoqing et al. (2004). "Elastic fiber homeostasis requires lysyl oxidase–like 1 protein." In: *Nature Genetics* 36.2, pp. 178–182. DOI: 10.1038/ng1297.

Liu, Yansheng, Andreas Beyer, and Ruedi Aebersold (2016). "On the dependency of cellular protein levels on mrna abundance." In: *Cell* 165.3, 535–550. DOI: 10.1016/j.cell.2016.03.014.

Lockwood, David S R et al. (2003). "Tumor progression in hepatocellular carcinoma: Relationship with tumor stroma and parenchymal disease." In: *Journal of Gastroenterology and Hepatology* 18.6, pp. 666–672. DOI: 10.1046/j.1440-1746.2003.03018.x.

Lodolce, James P. et al. (2002). "Regulation of lymphoid homeostasis by interleukin-15." In: *Cytokine & Growth Factor Reviews* 13.6, pp. 429–439. DOI: 10.1016/s1359-6101(02)00029-1.

Lodyga, Monika and Boris Hinz (2020). "TGF- β1 – a truly transforming growth factor in fibrosis and immunity." In: *Seminars in Cell & Developmental Biology* 101, 123–139. DOI: 10.1016/j.semcdb.2019.12.010.

Lodyga, Monika et al. (2019). "Cadherin-11–mediated adhesion of macrophages to myofibroblasts establishes a profibrotic niche of active TGF- β." In: *Science Signaling* 12.564. DOI: 10.1126/scisignal.aao3469.

Lokau, Juliane et al. (2016). "Proteolytic Cleavage Governs Interleukin-11 Trans-signaling." In: *Cell Reports* 14.7, pp. 1761–1773. DOI: 10.1016/j.celrep.2016.01.053.

Longo, Kenneth A. et al. (2004). "Wnt10b Inhibits Development of White and Brown Adipose Tissues." In: *Journal of Biological Chemistry* 279.34, pp. 35503–35509. DOI: 10.1074/jbc.m402937200.

Longo, Sophia K. et al. (2021). "Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics." In: *Nature Reviews Genetics* 22.10, 627–644. DOI: 10.1038/s41576-021-00370-8.

Lopez, Romain et al. (2018). "Deep generative modeling for single-cell transcriptomics." In: *Nature Methods* 15.12, 1053–1058. DOI: 10.1038/s41592-018-0229-2.

Losse, Josephine, Peter F. Zipfel, and Mihaly Jozsi (2009). "Factor H and Factor H-Related Protein 1 Bind to Human

Neutrophils via Complement Receptor 3, Mediate Attachment to Candida albicans, and Enhance Neutrophil Antimicrobial Activity." In: *The Journal of Immunology* 184.2, pp. 912–921. DOI: 10.4049/jimmunol.0901702.

Luan, Chao et al. (2019). "Potentiation of Psoriasis-Like Inflammation by PCSK9." In: *Journal of Investigative Dermatology* 139.4, pp. 859–867. DOI: 10.1016/j.jid.2018.07.046.

Luangsay, Souphalone et al. (2009). "Mouse ChemR23 Is Expressed in Dendritic Cell Subsets and Macrophages, and Mediates an Anti-Inflammatory Activity of Chemerin in a Lung Disease Model." In: *The Journal of Immunology* 183.10, pp. 6489–6499. DOI: 10.4049/jimmunol.0901037.

Lubbers, R et al. (2017). "Production of complement components by cells of the immune system." In: *Clinical and Experimental Immunology* 188.2, pp. 183–194. DOI: 10.1111/cei.12952.

Luckett-Chastain, Lerin R. et al. (2012). "SOCS3 Modulates Interleukin-6R Signaling Preference in Dermal Fibroblasts." In: *Journal of Interferon & Cytokine Research* 32.5, pp. 207–215. DOI: 10.1089/jir.2011.0086.

Luecken, Malte D and Fabian J Theis (2019). "Current best practices in single-cell RNA-seq analysis: a tutorial." In: *Molecular Systems Biology* 15.6. DOI: 10.15252/msb.20188746.

Luecken, Malte D. et al. (2021). "Benchmarking atlas-level data integration in single-cell genomics." In: *Nature Methods* 19.1, 41–50. DOI: 10.1038/s41592-021-01336-8.

Luker, Kathryn E. and Gary D. Luker (2022). "The CXCL12/CXCR4/ACKR3 Signaling Axis Regulates PKM2 and Glycolysis." In: *Cells* 11.11, p. 1775. DOI: 10.3390/cells11111775.

Lun, Aaron T. et al. (2019). "EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data." In: *Genome Biology* 20.1. DOI: 10.1186/s13059-019-1662-y.

Lun, Aaron (2018). "Overcoming systematic errors caused by log-transformation of normalized single-cell RNA sequencing data." In: *bioRxiv*. DOI: 10.1101/404962. eprint: https://www.biorxiv.org/content/early/2018/08/31/404962.full.pdf.

Luo, Y.Y. et al. (2019). "Type XI collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 99–106. DOI: 10.1016/b978-0-12-817068-7.00011-2.

Lynch, Magnus D. and Fiona M. Watt (2018). "Fibroblast heterogeneity: Implications for human disease." In: *Journal of Clinical Investigation* 128.1, 26–35. DOI: 10.1172/jci93555.

Lyons, Jon P et al. (2004). "Wnt-4 activates the canonical β-catenin-mediated Wnt pathway and binds Frizzled-6 CRD: functional implications of Wnt/β-catenin activity in kidney epithelial cells." In: *Experimental Cell Research* 298.2, pp. 369–387. DOI: 10.1016/j.yexcr.2004.04.036.

Ma, Keran, Lynsie A.M. Thomason, and JoAnne McLaurin (2012). "scyllo-Inositol, Preclinical, and Clinical Data for Alzheimer's Disease." In: *Current State of Alzheimer's Disease Research and Therapeutics*. Elsevier, pp. 177–212. DOI: 10.1016/b978-0-12-394816-8.00006-4.

Ma, Sai et al. (2020). "Chromatin potential identified by shared single-cell profiling of RNA and Chromatin." In: *Cell* 183.4. DOI: 10.1016/j.cell.2020.09.056.

Ma, Wanshu et al. (2013). "Induction of C-X-C Chemokine Receptor Type 7 (CXCR7) Switches Stromal Cell-derived Factor-1 (SDF-1) Signaling and Phagocytic Activity in Macrophages Linked to Atherosclerosis." In: *Journal of Biological Chemistry* 288.22, pp. 15481–15494. DOI: 10.1074/jbc.m112.445510.

Maaten, Laurens van der (2014). "Accelerating t-SNE using Tree-Based Algorithms." In: *Journal of Machine Learning Research* 15.93, pp. 3221–3245.

Maaten, Laurens van der and Geoffrey Hinton (2008a). "Visualizing Data using t-SNE." In: *Journal of Machine Learning Research* 9.86, pp. 2579–2605.

– (2008b). "Visualizing Data using t-SNE." In: *Journal of Machine Learning Research* 9.86, pp. 2579–2605.

MacQueen, James B. (2019). "Some Methods for classification and Analysis of Multivariate Observations." In: *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability* 1.1.

Macosko, Evan Z. et al. (2015). "Highly parallel genome-wide expression profiling of individual cells using Nanoliter droplets." In: *Cell* 161.5, 1202–1214. DOI: 10.1016/j.cell.2015.05.002.

Maiorino, Fursini M. et al. (1995). "[5] Diversity of glutathione peroxidases." In: *Methods in Enzymology*. Elsevier, pp. 38–53. DOI: 10.1016/0076-6879(95)52007-4.

Majdalawieh, Amin and Hyo-Sung Ro (2010). "PPARγ1 and LXRα face a new regulator of macrophage cholesterol homeostasis and inflammatory responsiveness, AEBP1." In: *Nuclear Receptor Signaling* 8.1, nrs.08004. DOI: 10.1621/nrs.08004.

Majno, G. et al. (1971). "Contraction of granulation tissue in vitro: Similarity to smooth muscle." In: *Science* 173.3996, 548–550. DOI: 10.1126/science.173.3996.548.

Malamut, Georgia et al. (2010). "IL-15 triggers an antiapoptotic pathway in human intraepithelial lymphocytes that is a potential new target in celiac disease–associated inflammation and lymphomagenesis." In: *Journal of Clinical Investigation* 120.6, pp. 2131–2143. DOI: 10.1172/jci41344.

Manon-Jensen, T., A. Arvanitidis, and M.A. Karsdal (2019a). "Type XV collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 127–131. DOI: 10.1016/b978-0-12-817068-7.00015-x.

Manon-Jensen, T., N.G. Kjeld, and M.A. Karsdal (2019b). "Type XXVI collagen." In: *Biochemistry of Collagens, Laminins*

and Elastin. Elsevier, pp. 197–200. DOI: 10.1016/b978-0-12-817068-7.00026-4.

Mantovani, Alberto and Elisabetta Dejana (1998). "Endothelium." In: *Encyclopedia of Immunology*. Elsevier, pp. 802–806. DOI: 10.1006/rwei.1999.0212.

Marangoni, Roberta Goncalves et al. (2015). "Myofibroblasts in murine cutaneous fibrosis originate from adiponectin-positive intradermal progenitors." In: *Arthritis & Rheumatology* 67.4, 1062–1073. DOI: 10.1002/art.38990.

Marcel, Yves L., Camilla Vezina, and Ross W. Milne (1983). "Cholesteryl ester and apolipoprotein E transfer between human high density lipoproteins and chylomicrons." In: *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 750.2, pp. 411–417. DOI: 10.1016/0005-2760(83)90047-4.

Marchese, Cinzia et al. (2001). "Fibroblast growth factor 10 induces proliferation and differentiation of human primary cultured keratinocytes." In: *Journal of Investigative Dermatology* 116.4, 623–628. DOI: 10.1046/j.0022-202x.2001.01280.x.

Marega, Manuela, Chengshui Chen, and Saverio Bellusci (2021). "Cross-Talk Between Inflammation and Fibroblast Growth Factor 10 During Organogenesis and Pathogenesis: Lessons Learnt From the Lung and Other Organs." In: *Frontiers in Cell and Developmental Biology* 9. DOI: 10.3389/fcell.2021.656883.

Marinkovich, M P et al. (1992). "The dermal-epidermal junction of human skin contains a novel laminin variant." In: *Journal of Cell Biology* 119.3, 695–703. DOI: 10.1083/jcb.119.3.695.

Mariottoni, Paula et al. (2021). "Single-cell RNA sequencing reveals cellular and transcriptional changes associated with M1 macrophage polarization in Hidradenitis Suppurativa." In: *Frontiers in Medicine* 8. DOI: 10.3389/fmed.2021.665873.

Markey, G. M. (2010). "Carboxylesterase 1 (Ces1): from monocyte marker to major player." In: *Journal of Clinical Pathology* 64.2, pp. 107–109. DOI: 10.1136/jcp.2010.084657.

Marks, James G. and Jeffrey J. Miller (2019). "Structure and function of the skin." In: *Lookingbill and Marks' Principles of Dermatology*, 2–10. DOI: 10.1016/b978-0-323-43040-1.00002-6.

Marmon, Shana et al. (2009). "Caveolin-1 Expression Determines the Route of Neutrophil Extravasation through Skin Microvasculature." In: *The American Journal of Pathology* 174.2, pp. 684–692. DOI: 10.2353/ajpath.2009.080091.

Martin, T. A., K. Harding, and W. G. Jiang (2001). "Matrix-bound fibroblasts regulate angiogenesis by modulation of VE-cadherin." In: *European Journal of Clinical Investigation* 31.11, 931–938. DOI: 10.1046/j.1365-2362.2001.00914.x.

Martinez, Fernando O., Laura Helming, and Siamon Gordon (2009). "Alternative activation of macrophages: An immunologic functional perspective." In: *Annual Review of Immunology* 27.1, 451–483. DOI: 10.1146/annurev.immunol.021908.132532.

Martinez, Fernando Oneissi (2008). "Macrophage activation and polarization." In: *Frontiers in Bioscience* 13.13, p. 453. DOI: 10.2741/2692.

Martino, Pieter A., Nicholas Heitman, and Michael Rendl (2020). "The dermal sheath: An emerging component of the hair follicle stem cell niche." In: *Experimental Dermatology* 30.4, pp. 512–521. DOI: 10.1111/exd.14204.

Martinot, V. et al. (1994). "Comparative study of split thickness skin grafts taken from the scalp and thigh in children." In: *Burns* 20.2, 146–150. DOI: 10.1016/s0305-4179(06)80012-4.

Marucci, Gabriella et al. (2022). "The possible role of the nucleoside adenosine in countering skin aging: A review." In: *BioFactors* 48.5, pp. 1027–1035. DOI: 10.1002/biof.1881.

Mastrogiannaki, Maria et al. (2016). "β-catenin stabilization in skin fibroblasts causes fibrotic lesions by preventing adipocyte differentiation of the reticular dermis." In: *Journal of Investigative Dermatology* 136.6, 1130–1142. DOI: 10.1016/j.jid.2016.01.036.

Matsumoto, Hirotaka and Hisanori Kiryu (2016). "Scoup: A probabilistic model based on the ornstein–uhlenbeck process to analyze single-cell expression data during differentiation." In: *BMC Bioinformatics* 17.1. DOI: 10.1186/s12859-016-1109-3.

Mattern, Andreas, Tristan Zellmann, and Annette G. Beck-Sickinger (2014). "Processing, signaling, and physiological function of chemerin." In: *IUBMB Life* 66.1, pp. 19–26. DOI: 10.1002/iub.1242.

Maturu, Paramahamsa et al. (2017). "Role of Cyclooxygenase-2 Pathway in Creating an Immunosuppressive Microenvironment and in Initiation and Progression of Wilms' Tumor." In: *Neoplasia* 19.3, pp. 237–249. DOI: 10.1016/j.neo.2016.07.009.

Maximow, Alexander (1905). "Über die zellformen des lockeren bindegewebes." In: *Archiv für Mikroskopische Anatomie* 67.1, 680–757. DOI: 10.1007/bf02979969.

Mayer, Christian et al. (2018). "Developmental diversification of cortical inhibitory interneurons." In: *Nature* 555.7697, 457–462. DOI: 10.1038/nature25999.

Mazzarello, Paolo (1999). "A unifying concept: The history of cell theory." In: *Nature Cell Biology* 1.1. DOI: 10.1038/8964.

Mazzoni, Jenna et al. (2017). "The Wnt Inhibitor Apcdd1 Coordinates Vascular Remodeling and Barrier Maturation of Retinal Blood Vessels." In: *Neuron* 96.5, 1055–1069.e6. DOI: 10.1016/j.neuron.2017.10.025.

McAndrews, Kathleen M et al. (2022). "Dermal αSMA+ myofibroblasts orchestrate skin wound repair via β1 integrin

and independent of type I collagen production." In: *The EMBO Journal* 41.7. DOI: 10.15252/embj.2021109470.

McCarthy, Davis J. et al. (2017). "Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R." In: *Bioinformatics*. DOI: 10.1093/bioinformatics/btw777.

McGee, Heather M. et al. (2013). "IL-22 promotes fibroblast-mediated wound repair in the skin." In: *Journal of Investigative Dermatology* 133.5, 1321–1329. DOI: 10.1038/jid.2012.463.

McGinnis, Christopher S., Lyndsay M. Murrow, and Zev J. Gartner (2019). "DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors." In: *Cell Systems* 8.4. DOI: 10.1016/j.cels.2019.03.003.

McGrath, J. A., R. A. Eady, and F. M. Pope (2004). "Anatomy and Organization of Human Skin." In: *Rook's Textbook of Dermatology*, 45–128. DOI: 10.1002/9780470750520.ch3.

McInnes, Leland, John Healy, and James Melville (2018). "Umap: Uniform manifold approximation and projection for dimension reduction." In: *arXiv preprint arXiv:1802.03426*.

McLarren, Keith W. et al. (2011). "Ship-deficient mice develop spontaneous intestinal inflammation and arginase-dependent fibrosis." In: *The American Journal of Pathology* 179.1, 180–188. DOI: 10.1016/j.ajpath.2011.03.018.

Mei, Dan et al. (2020). "The Role of CTHRC1 in Regulation of Multiple Signaling and Tumor Progression and Metastasis." In: *Mediators of Inflammation* 2020, pp. 1–13. DOI: 10.1155/2020/9578701.

Melnik, Bodo and Gerd Schmitz (2008). "FGFR2 signaling and the pathogenesis of acne." In: *Journal der Deutschen Dermatologischen Gesellschaft* 6.9, pp. 721–728. DOI: 10.1111/j.1610-0387.2008.06822.x.

Melsted, Páll, Vasilis Ntranos, and Lior Pachter (2019). "The Barcode, UMI, set format and BUStools." In: *Bioinformatics* 35.21, 4472–4473. DOI: 10.1093/bioinformatics/btz279.

Mendez-Enriquez, E. and E. A. Garcia-Zepeda (2013). "The multiple faces of CCL13 in immunity and inflammation." In: *Inflammopharmacology* 21.6, pp. 397–406. DOI: 10.1007/s10787-013-0177-5.

Meng, Xiao-ming, David J. Nikolic-Paterson, and Hui Yao Lan (2016). "TGF- β: The master regulator of fibrosis." In: *Nature Reviews Nephrology* 12.6, 325–338. DOI: 10.1038/nrneph.2016.48.

Mereu, Elisabetta et al. (2020). "Benchmarking single-cell RNA-sequencing protocols for cell atlas projects." In: *Nature Biotechnology* 38, pp. 747–755. DOI: 10.1038/s41587-020-0469-4.

Merline, Rosetta, Roland M. Schaefer, and Liliana Schaefer (2009). "The matricellular functions of small leucine-rich proteoglycans (SLRPs)." In: *Journal of Cell Communication and Signaling* 3.3-4, pp. 323–335. DOI: 10.1007/s12079-009-0066-2.

Mesler, Arlee L. et al. (2017). "Hair follicle terminal differentiation is orchestrated by distinct early and late matrix progenitors." In: *Cell Reports* 19.4, 809–821. DOI: 10.1016/j.celrep.2017.03.077.

Messenger, A.G. and J. Rundegren (2004). "Minoxidil: mechanisms of action on hair growth." In: *British Journal of Dermatology* 150.2, pp. 186–194. DOI: 10.1111/j.1365-2133.2004.05785.x.

Meyrath, Max et al. (2020). "The atypical chemokine receptor ACKR3/CXCR7 is a broad-spectrum scavenger for opioid peptides." In: *Nature Communications* 11.1. DOI: 10.1038/s41467-020-16664-0.

Michalik, Liliane and Walter Wahli (2007). "Guiding Ligands to Nuclear Receptors." In: *Cell* 129.4, pp. 649–651. DOI: 10.1016/j.cell.2007.05.001.

Michel, M. et al. (1996). "Keratin 19 as a biochemical marker of skin stem cells in vivo and in vitro: keratin 19 expressing cells are differentially localized in function of anatomic sites, and their number varies with donor age and culture stage." In: *Journal of Cell Science* 109.5, pp. 1017–1028. DOI: 10.1242/jcs.109.5.1017.

Mimitou, Eleni P. et al. (2019). "Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells." In: *Nature Methods* 16.5, 409–412. DOI: 10.1038/s41592-019-0392-0.

Mirizio, Emily et al. (2020). "Single-cell transcriptome conservation in a comparative analysis of fresh and cryopreserved human skin tissue: Pilot in localized scleroderma." In: *Arthritis Research & Therapy* 22.1. DOI: 10.1186/s13075-020-02343-4.

Mistriotis, Panagiotis and Stelios T. Andreadis (2013). "Hair Follicle: A Novel Source of Multipotent Stem Cells for Tissue Engineering and Regenerative Medicine." In: *Tissue Engineering Part B: Reviews* 19.4, pp. 265–278. DOI: 10.1089/ten.teb.2012.0422.

Mizumoto, Shuji and Shuhei Yamada (2021). "An Overview of in vivo Functions of Chondroitin Sulfate and Dermatan Sulfate Revealed by Their Deficient Mice." In: *Frontiers in Cell and Developmental Biology* 9. DOI: 10.3389/fcell.2021.764781.

Moffitt, Jeffrey R. et al. (2018). "Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region." In: *Science* 362.6416. DOI: 10.1126/science.aau5324.

Mohyeldin, Ahmed, Tomás Garzón-Muvdi, and Alfredo Quiñones-Hinojosa (2010). "Oxygen in Stem Cell Biology: A Critical Component of the Stem Cell Niche." In: *Cell Stem Cell* 7.2, 150–161. DOI: 10.1016/j.stem.2010.07.007.

Mok, Ka-Wai et al. (2019). "Dermal condensate niche fate specification occurs prior to formation and is placode

progenitor dependent." In: *Developmental Cell* 48.1. DOI: 10.1016/j.devcel.2018.11.034.

Mongiat, Maurizio et al. (2010). "The Extracellular Matrix Glycoprotein Elastin Microfibril Interface Located Protein 2: A Dual Role in the Tumor Microenvironment." In: *Neoplasia* 12.4, 294–IN1. DOI: 10.1593/neo.91930.

Monkley, Susan et al. (2021). "Sensitization of the UPR by loss of PPP1R15A promotes fibrosis and senescence in IPF." In: *Scientific Reports* 11.1. DOI: 10.1038/s41598-021-00769-7.

Montagna, William (1967). "Comparative anatomy and physiology of the skin." In: *Archives of Dermatology* 96.4, p. 357. DOI: 10.1001/archderm.1967.01610040007003.

Montoya, Dennis et al. (2014). "IL-32 is a molecular marker of a host defense network in human tuberculosis." In: *Science Translational Medicine* 6.250. DOI: 10.1126/scitranslmed.3009546.

Moon, Kevin R. et al. (2019). "Visualizing structure and transitions in high-dimensional biological data." In: *Nature Biotechnology* 37.12, 1482–1492. DOI: 10.1038/s41587-019-0336-3.

Moore, G.E. (1998). "Cramming more components onto integrated circuits." In: *Proceedings of the IEEE* 86.1, 82–85. DOI: 10.1109/jproc.1998.658762.

Mora, Rosalia et al. (1999). "Caveolin-2 Localizes to the Golgi Complex but Redistributes to Plasma Membrane, Caveolae, and Rafts when Co-expressed with Caveolin-1." In: *Journal of Biological Chemistry* 274.36, pp. 25708–25717. DOI: 10.1074/jbc.274.36.25708.

Morgenstern, Ralf, Jie Zhang, and Katarina Johansson (2011). "Microsomal glutathione transferase 1: mechanism and functional roles." In: *Drug Metabolism Reviews* 43.2, pp. 300–306. DOI: 10.3109/03602532.2011.558511.

Moriguchi, Tetsuo et al. (2016). "Ecrg4 contributes to the anti-glioma immunosurveillance through type-I interferon signaling." In: *OncoImmunology* 5.12, e1242547. DOI: 10.1080/2162402x.2016.1242547.

Morisaki, Naoko, Atsushi Ohuchi, and Shigeru Moriwaki (2013). "The Role of Neprilysin in Regulating the Hair Cycle." In: *PLoS ONE* 8.2. Ed. by Stefan Strack, e55947. DOI: 10.1371/journal.pone.0055947.

Moro, Kazuyo et al. (2009). "Innate production of TH2 cytokines by adipose tissue-associated c-Kit$^+$Sca-1$^+$ lymphoid cells." In: *Nature* 463.7280, pp. 540–544. DOI: 10.1038/nature08636.

Mortazavi, Ali et al. (2008). "Mapping and quantifying mammalian transcriptomes by RNA-seq." In: *Nature Methods* 5.7, 621–628. DOI: 10.1038/nmeth.1226.

Mortensen, J.H. and M.A. Karsdal (2019a). "Type VII collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 69–74. DOI: 10.1016/b978-0-12-817068-7.00007-0.

Mortensen, J.H., T. Manon-Jensen, and M.A. Karsdal (2019b). "Type XII collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 107–113. DOI: 10.1016/b978-0-12-817068-7.00012-4.

Moses, Lambda and Lior Pachter (2022). "Museum of Spatial Transcriptomics." In: *Nature Methods* 19.5, 534–546. DOI: 10.1038/s41592-022-01409-2.

Mostafa, Wedad Z. and Rehab A. Hegazy (2015). "Vitamin D and the skin: Focus on a complex relationship: A Review." In: *Journal of Advanced Research* 6.6, 793–804. DOI: 10.1016/j.jare.2014.01.011.

Moussion, Christine, Nathalie Ortega, and Jean-Philippe Girard (2008). "The IL-1-Like Cytokine IL-33 Is Constitutively Expressed in the Nucleus of Endothelial Cells and Epithelial Cells In Vivo: A Novel 'Alarmin'?" In: *PLoS ONE* 3.10. Ed. by Derya Unutmaz, e3331. DOI: 10.1371/journal.pone.0003331.

Muffat, Julien, David W. Walker, and Seymour Benzer (2008). "Human ApoD, an apolipoprotein up-regulated in neurodegenerative diseases, extends lifespan and increases stress resistance in Drosophila." In: *Proceedings of the National Academy of Sciences* 105.19, pp. 7088–7093. DOI: 10.1073/pnas.0800896105.

Muhl, Lars et al. (2020). "Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination." In: *Nature Communications* 11.1. DOI: 10.1038/s41467-020-17740-1.

Munkres, James (1957). "Algorithms for the Assignment and Transportation Problems." In: *Journal of the Society of Industrial800 and Applied Mathematics* 5.1.

Munsky, Brian, Gregor Neuert, and Alexander van Oudenaarden (2012). "Using gene expression noise to understand gene regulation." In: *Science* 336.6078, 183–187. DOI: 10.1126/science.1216379.

Munussami, Ganapathiraman et al. (2018). "In Silico Study on Retinoid-binding Modes in Human RBP and ApoD Lipocalins." In: *Biotechnology and Bioprocess Engineering* 23.2, pp. 158–167. DOI: 10.1007/s12257-018-0032-z.

Murakami, M. et al. (2011). "Secreted phospholipase A2 revisited." In: *Journal of Biochemistry* 150.3, pp. 233–255. DOI: 10.1093/jb/mvr088.

Murakami, Makoto, Kei Yamamoto, and Yoshitaka Taketomi (2018). "Phospholipase A2 in skin biology: new insights from gene-manipulated mice and lipidomics." In: *Inflammation and Regeneration* 38.1. DOI: 10.1186/s41232-018-0089-2.

Muramatsu, Hisako et al. (2000). "LDL Receptor-Related Protein as a Component of the Midkine Receptor." In: *Biochemical and Biophysical Research Communications* 270.3, pp. 936–941. DOI: 10.1006/bbrc.2000.2549.

Muramatsu, Hisako et al. (2004). "$\alpha 4\beta 1$- and $\alpha 6\beta 1$-integrins are functional receptors for midkine, a heparin-binding growth factor." In: *Journal of Cell Science* 117.22, pp. 5405–5415. DOI: 10.1242/jcs.01423.

Murphy, Philip M. (2007). "CXC Chemokine Receptors." In: *xPharm: The Comprehensive Pharmacology Reference*. Elsevier, pp. 1–3. DOI: 10.1016/b978-008055232-3.60163-1.

Muszbek, László et al. (2011). "Factor XIII: A Coagulation Factor With Multiple Plasmatic and Cellular Functions." In: *Physiological Reviews* 91.3, pp. 931–972. DOI: 10.1152/physrev.00016.2010.

Myngbay, Askhat et al. (2021). "The Role of Collagen Triple Helix Repeat-Containing 1 Protein (CTHRC1) in Rheumatoid Arthritis." In: *International Journal of Molecular Sciences* 22.5, p. 2426. DOI: 10.3390/ijms22052426.

Müller-Röver, Sven et al. (2000). "Intercellular Adhesion Molecule-1 and Hair Follicle Regression." In: *Journal of Histochemistry & Cytochemistry* 48.4, pp. 557–568. DOI: 10.1177/002215540004800413.

Müller-Röver, Sven et al. (2001). "A comprehensive guide for the accurate classification of murine hair follicles in distinct hair cycle stages." In: *Journal of Investigative Dermatology* 117.1, 3–15. DOI: 10.1046/j.0022-202x.2001.01377.x.

Na, Chan Hyun et al. (2019). "Integrated Transcriptomic and Proteomic Analysis of Human Eccrine Sweat Glands Identifies Missing and Novel Proteins." In: *Molecular & Cellular Proteomics* 18.7, pp. 1382–1395. DOI: 10.1074/mcp.ra118.001101.

Naeem, Aishath S. et al. (2017). "A mechanistic target of rapamycin complex 1/2 (mTORC1)/V-Akt murine thymoma viral oncogene homolog 1 (AKT1)/cathepsin H axis controls filaggrin expression and processing in skin, a novel mechanism for skin barrier disruption in patients with atopic dermatitis." In: *Journal of Allergy and Clinical Immunology* 139.4, pp. 1228–1241. DOI: 10.1016/j.jaci.2016.09.052.

Nakamichi, Yuko, Nobuyuki Udagawa, and Naoyuki Takahashi (2013). "IL-34 and CSF-1: similarities and differences." In: *Journal of Bone and Mineral Metabolism* 31.5, pp. 486–495. DOI: 10.1007/s00774-013-0476-3.

Nakamura, Tomomi et al. (2022). "SPARC promotes production of type IV and VII collagen and their skin basement membrane accumulation." In: *Journal of Dermatological Science* 107.2, pp. 109–112. DOI: 10.1016/j.jdermsci.2022.07.007.

Naldaiz-Gastesi, Neia et al. (2018). "The panniculus carnosus muscle: An evolutionary enigma at the intersection of distinct research fields." In: *Journal of Anatomy* 233.3, 275–288. DOI: 10.1111/joa.12840.

Nanda, Akash et al. (2004). "TEM8 Interacts with the Cleaved C5 Domain of Collagen $\alpha 3$(VI)." In: *Cancer Research* 64.3, pp. 817–820. DOI: 10.1158/0008-5472.can-03-2408.

Natarajan, Kedar Nath et al. (2019). "Comparative analysis of sequencing technologies for single-cell transcriptomics." In: *Genome Biology* 20.1. DOI: 10.1186/s13059-019-1676-5.

Nauroy, Pauline and Alexander NystrÃ¶m (2020). "Kallikreins: Essential epidermal messengers for regulation of the skin microenvironment during homeostasis, repair and disease." In: *Matrix Biology Plus* 6-7, p. 100019. DOI: 10.1016/j.mbplus.2019.100019.

Nauroy, Pauline et al. (2017). "Human dermal fibroblast subpopulations display distinct gene signatures related to cell behaviors and matrisome." In: *Journal of Investigative Dermatology* 137.8, 1787–1789. DOI: 10.1016/j.jid.2017.03.028.

Naylor, Elizabeth C., Rachel E.B. Watson, and Michael J. Sherratt (2011). "Molecular aspects of skin ageing." In: *Maturitas* 69.3, 249–256. DOI: 10.1016/j.maturitas.2011.04.011.

Neefjes, Jacques J. and Hidde L. Ploegh (1988). "Allele and locus-specific differences in cell surface expression and the association of HLA class I heavy chain with $\beta 2$-microglobulin: differential effects of inhibition of glycosylation on class I subunit association." In: *European Journal of Immunology* 18.5, pp. 801–810. DOI: 10.1002/eji.1830180522.

Neefjes, Jacques et al. (2011). "Towards a systems understanding of MHC class I and MHC class II antigen presentation." In: *Nature Reviews Immunology* 11.12, pp. 823–836. DOI: 10.1038/nri3084.

Nelson, Amanda M. et al. (2013). "Prostaglandin D2 Inhibits Wound-Induced Hair Follicle Neogenesis through the Receptor, Gpr44." In: *Journal of Investigative Dermatology* 133.4, pp. 881–889. DOI: 10.1038/jid.2012.398.

Nencioni, Alessio et al. (2002). "Dendritic Cell Immunogenicity Is Regulated by Peroxisome Proliferator-Activated Receptor γ." In: *The Journal of Immunology* 169.3, pp. 1228–1235. DOI: 10.4049/jimmunol.169.3.1228.

Neufeld, Gera and Ofra Kessler (2008). "The semaphorins: versatile regulators of tumour progression and tumour angiogenesis." In: *Nature Reviews Cancer* 8.8, pp. 632–645. DOI: 10.1038/nrc2404.

Niehrs, Christof (2012). "The complex world of Wnt receptor signalling." In: *Nature Reviews Molecular Cell Biology* 13.12, 767–779. DOI: 10.1038/nrm3470.

Nielsen, M.J. et al. (2019a). "Type III collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 23–36. DOI: 10.1016/b978-0-12-817068-7.00003-3.

Nielsen, S. Holm and M.A. Karsdal (2019b). "Type XXIV collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 187–190. DOI: 10.1016/b978-0-12-817068-7.00024-0.

Nikolaev, Anatoly et al. (2009). "APP binds DR6 to trigger axon pruning and neuron death via distinct caspases." In: *Nature* 457.7232, pp. 981–989. DOI: 10.1038/nature07767.

Nithipatikom, Kasem et al. (2014). "A novel activity of microsomal epoxide hydrolase: metabolism of the endocannabinoid 2-arachidonoylglycerol." In: *Journal of Lipid*

Research 55.10, pp. 2093–2102. DOI: 10.1194/jlr.m051284.

Noguchi, Satoru et al. (1995). "Mutations in the Dystrophin-Associated Protein γ-Sarcoglycan in Chromosome 13 Muscular Dystrophy." In: *Science* 270.5237, pp. 819–822. DOI: 10.1126/science.270.5237.819.

Nolano, Maria et al. (2012). "Cutaneous innervation of the human face as assessed by skin biopsy." In: *Journal of Anatomy* 222.2, 161–169. DOI: 10.1111/joa.12001.

Normanno, Nicola et al. (2006). "Epidermal growth factor receptor (EGFR) signaling in cancer." In: *Gene* 366.1, pp. 2–16. DOI: 10.1016/j.gene.2005.10.018.

Nowell, C. and F. Radtke (2013). "Cutaneous Notch Signaling in Health and Disease." In: *Cold Spring Harbor Perspectives in Medicine* 3.12, a017772–a017772. DOI: 10.1101/cshperspect.a017772.

Nugteren, Sandrine and Janneke N. Samsom (2021). "Secretory Leukocyte Protease Inhibitor (SLPI) in mucosal tissues: Protects against inflammation, but promotes cancer." In: *Cytokine & Growth Factor Reviews* 59, pp. 22–35. DOI: 10.1016/j.cytogfr.2021.01.005.

Nusse, Roel and Harold E. Varmus (1982). "Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome." In: *Cell* 31.1, 99–109. DOI: 10.1016/0092-8674(82)90409-3.

O'Shaughnessy, Michelle M et al. (2018). "Cause of kidney disease and cardiovascular events in a national cohort of US patients with end-stage renal disease on dialysis: a retrospective analysis." In: *European Heart Journal* 40.11, pp. 887–898. DOI: 10.1093/eurheartj/ehy422.

Obiezu, Chistina V. et al. (2006). "Human kallikrein 4: enzymatic activity, inhibition, and degradation of extracellular matrix proteins." In: *Biological Chemistry* 387.6, pp. 749–759. DOI: 10.1515/bc.2006.094.

Odorisio, Teresa et al. (2006). "The placenta growth factor in skin angiogenesis." In: *Journal of Dermatological Science* 41.1, pp. 11–19. DOI: 10.1016/j.jdermsci.2005.08.008.

Ohkawara, Bisei, Andrei Glinka, and Christof Niehrs (2011). "RSPO3 binds syndecan 4 and induces WNT/PCP signaling via clathrin-mediated endocytosis to promote morphogenesis." In: *Developmental Cell* 20.3, 303–314. DOI: 10.1016/j.devcel.2011.01.006.

Ohl, Lars et al. (2004). "CCR7 Governs Skin Dendritic Cell Migration under Inflammatory and Steady-State Conditions." In: *Immunity* 21.2, pp. 279–288. DOI: 10.1016/j.immuni.2004.06.014.

Ohno, Ikko et al. (1996). "A cDNA Cloning of Human AEBP1 from Primary Cultured Osteoblasts and Its Expression in a Differentiating Osteoblastic Cell Line." In: *Biochemical and Biophysical Research Communications* 228.2, pp. 411–414. DOI: 10.1006/bbrc.1996.1675.

Ohnuma, Kei et al. (2007). "Caveolin-1 Triggers T-cell Activation via CD26 in Association with CARMA1." In: *Journal of Biological Chemistry* 282.13, pp. 10117–10131. DOI: 10.1074/jbc.m609157200.

Okamoto, Osamu and Sakuhei Fujiwara (2006). "Dermatopontin, a Novel Player in the Biology of the Extracellular Matrix." In: *Connective Tissue Research* 47.4, pp. 177–189. DOI: 10.1080/03008200600846564.

Okamoto, Osamu et al. (1999). "Dermatopontin interacts with transforming growth factor β and enhances its biological activity." In: *Biochemical Journal* 337.3, 537–541. DOI: 10.1042/bj3370537.

Okano, Junko et al. (2012). "Cutaneous retinoic acid levels determine hair follicle development and downgrowth." In: *Journal of Biological Chemistry* 287.47, 39304–39315. DOI: 10.1074/jbc.m112.397273.

Olkkonen, Vesa M., Olivier Beaslas, and Eija NissilÃ¤ (2012). "Oxysterols and Their Cellular Effectors." In: *Biomolecules* 2.1, pp. 76–103. DOI: 10.3390/biom2010076.

Orlandini, Maurizio and Salvatore Oliviero (2001). "In fibroblasts VEGF-D expression is induced by cell-cell contact mediated by cadherin-11." In: *Journal of Biological Chemistry* 276.9, 6576–6581. DOI: 10.1074/jbc.m009573200.

Oshima, Shigeru et al. (2004). "Interferon Regulatory Factor 1 (IRF-1) and IRF-2 Distinctively Up-Regulate Gene Expression and Production of Interleukin-7 in Human Intestinal Epithelial Cells." In: *Molecular and Cellular Biology* 24.14, pp. 6298–6310. DOI: 10.1128/mcb.24.14.6298-6310.2004.

Osorio, Daniel et al. (2019). "Single-cell expression variability implies cell function." In: *Cells* 9.1, p. 14. DOI: 10.3390/cells9010014.

Ospelt, Caroline et al. (2010). "Inhibition of fibroblast activation protein and dipeptidylpeptidase 4 increases cartilage invasion by rheumatoid arthritis synovial fibroblasts." In: *Arthritis & Rheumatism* 62.5, pp. 1224–1235. DOI: 10.1002/art.27395.

Otsuki, T. (2005). "Signal Sequence and Keyword Trap in silico for Selection of Full-Length Human cDNAs Encoding Secretion or Membrane Proteins from Oligo-Capped cDNA Libraries." In: *DNA Research* 12.2, pp. 117–126. DOI: 10.1093/dnares/12.2.117.

Ouji, Yukiteru, Shigeaki Ishizaka, and Masahide Yoshikawa (2012). "Dermal Papilla Cells Serially Cultured with Wnt-10b Sustain their Hair Follicle Induction Activity after Transplantation into Nude Mice." In: *Cell Transplantation* 21.10, pp. 2313–2324. DOI: 10.3727/096368912x636867.

Ouwehand, Krista et al. (2008). "CXCL12 is essential for migration of activated Langerhans cells from epidermis to dermis." In: *European Journal of Immunology* 38.11, pp. 3050–3059. DOI: 10.1002/eji.200838384.

Ouwehand, Krista et al. (2010). "Epidermis-to-dermis migration of immature Langerhans cells upon topical irritant exposure is dependent on CCL2 and CCL5." In: *European Journal of Immunology* 40.7, pp. 2026–2034. DOI: 10.1002/eji.200940150.

O'Flanagan, Ciara H. et al. (2019). "Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses." In: *Genome Biology* 20.1. DOI: 10.1186/s13059-019-1830-0.

Pai, Chen-Hsueh et al. (2020). "Targeting fibroblast CD248 attenuates CCL17-expressing macrophages and tissue fibrosis." In: *Scientific Reports* 10.1. DOI: 10.1038/s41598-020-73194-x.

Palenski, Tammy L. et al. (2013). "Cyp1B1 expression promotes angiogenesis by suppressing NF-κB activity." In: *American Journal of Physiology-Cell Physiology* 305.11, pp. C1170–C1184. DOI: 10.1152/ajpcell.00139.2013.

Panchaprateep, Ratchathorn and Pravit Asawanonda (2014). "Insulin-like growth factor-1: roles in androgenetic alopecia." In: *Experimental Dermatology* 23.3, pp. 216–218. DOI: 10.1111/exd.12339.

Pangas, Stephanie A and Teresa K Woodruff (2000). "Activin signal transduction pathways." In: *Trends in Endocrinology & Metabolism* 11.8, 309–314. DOI: 10.1016/s1043-2760(00)00294-0.

Panteleyev, Andrei A. et al. (1997). "Keratin 17 Gene Expression during the Murine Hair Cycle." In: *Journal of Investigative Dermatology* 108.3, pp. 324–329. DOI: 10.1111/1523-1747.ep12286476.

Papadimitriou, Evangelia et al. (2022). "On the role of pleiotrophin and its receptors in development and angiogenesis." In: *The International Journal of Developmental Biology* 66.1-2-3, pp. 115–124. DOI: 10.1387/ijdb.210122ep.

Paquet, Philippe and Gerald E. Pierard (1996). "Interleukin-6 and the Skin." In: *International Archives of Allergy and Immunology* 109.4, pp. 308–317. DOI: 10.1159/000237257.

Parekh, Swati et al. (2018). "Zumis - a fast and flexible pipeline to process RNA sequencing data with UMIS." In: *GigaScience* 7.6. DOI: 10.1093/gigascience/giy059.

Park, Junsoo et al. (2007). "Elevated level of SUMOylated IRF-1 in tumor cells interferes with IRF-1-mediated apoptosis." In: *Proceedings of the National Academy of Sciences* 104.43, pp. 17028–17033. DOI: 10.1073/pnas.0609852104.

Park, Soo Hyun, Kyoung Woo Kim, and Jae Chan Kim (2015). "The Role of Insulin-Like Growth Factor Binding Protein 2 (IGFBP2) in the Regulation of Corneal Fibroblast Differentiation." In: *Investigative Opthalmology & Visual Science* 56.12, p. 7293. DOI: 10.1167/iovs.15-16616.

Parthasarathy, Sampath et al. (2009). "Oxidized Low-Density Lipoprotein." In: *Methods in Molecular Biology*. Humana Press, pp. 403–417. DOI: 10.1007/978-1-60327-029-8_24.

Paszek, Pawel (2007). "Modeling stochasticity in gene regulation: Characterization in the terms of the underlying distribution function." In: *Bulletin of Mathematical Biology* 69.5, 1567–1601. DOI: 10.1007/s11538-006-9176-7.

Paus, Ralf et al. (2003). "The hair follicle and immune privilege." In: *Journal of Investigative Dermatology Symposium Proceedings* 8.2, 188–194. DOI: 10.1046/j.1087-0024.2003.00807.x.

Pecaric-Petkovic, Tatjana et al. (2009). "Human basophils and eosinophils are the direct target leukocytes of the novel IL-1 family member IL-33." In: *Blood* 113.7, pp. 1526–1534. DOI: 10.1182/blood-2008-05-157818.

Pedroso, João A.B., Angela M. Ramos-Lobo, and Jose Donato (2018). "SOCS3 as a future target to treat metabolic disorders." In: *Hormones* 18.2, pp. 127–136. DOI: 10.1007/s42000-018-0078-5.

Peh, Chen Au et al. (1998). "HLA-B27–Restricted Antigen Presentation in the Absence of Tapasin Reveals Polymorphism in Mechanisms of HLA Class I Peptide Loading." In: *Immunity* 8.5, pp. 531–542. DOI: 10.1016/s1074-7613(00)80558-0.

Pehrsson, M., C.L. Bager, and M.A. Karsdal (2019). "Type XVIII collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 149–162. DOI: 10.1016/b978-0-12-817068-7.00018-5.

Peltonen, Sirkku et al. (1999). "A Novel Component of Epidermal Cell–Matrix and Cell–Cell Contacts: Transmembrane Protein Type XIII Collagen." In: *Journal of Investigative Dermatology* 113.4, pp. 635–642. DOI: 10.1046/j.1523-1747.1999.00736.x.

Peng, Shi qing et al. (2023). "Identification of matrix-remodeling associated 5 as a possible molecular oncotarget of pancreatic cancer." In: *Cell Death & Disease* 14.2. DOI: 10.1038/s41419-023-05684-5.

Penner, Akiyo S. et al. (2002). "Microfibril-associated Glycoprotein-2 Interacts with Fibrillin-1 and Fibrillin-2 Suggesting a Role for MAGP-2 in Elastic Fiber Assembly." In: *Journal of Biological Chemistry* 277.38, pp. 35044–35049. DOI: 10.1074/jbc.m206363200.

Perrine, Shane (2007). "Endothelin 3." In: *xPharm: The Comprehensive Pharmacology Reference*. Elsevier, pp. 1–2. DOI: 10.1016/b978-008055232-3.61684-8.

Peruzzi, Francesca et al. (1999). "Multiple Signaling Pathways of the Insulin-Like Growth Factor 1 Receptor in Protection from Apoptosis." In: *Molecular and Cellular Biology* 19.10, pp. 7203–7215. DOI: 10.1128/mcb.19.10.7203.

Peterson, Vanessa M et al. (2017). "Multiplexed quantification of proteins and transcripts in single cells." In: *Nature Biotechnology* 35.10, 936–939. DOI: 10.1038/nbt.3973.

Petrova, Tatiana V. and Gou Young Koh (2017). "Organ-specific lymphatic vasculature: From development to pathophysiology." In: *Journal of Experimental Medicine* 215.1, 35–49. DOI: 10.1084/jem.20171868.

Phan, Quan M et al. (2020). "LEF1 expression in fibroblasts maintains developmental potential in adult skin to regenerate wounds." In: *eLife* 9. DOI: 10.7554/elife.60066.

Philippeos, Christina et al. (2018). "Spatial and single-cell transcriptional profiling identifies functionally distinct

human dermal fibroblast subpopulations." In: *Journal of Investigative Dermatology* 138.4, 811–825. DOI: 10.1016/j.jid.2018.01.016.

Pias, Erin K. et al. (2003). "Differential Effects of Superoxide Dismutase Isoform Expression on Hydroperoxide-induced Apoptosis in PC-12 Cells." In: *Journal of Biological Chemistry* 278.15, pp. 13294–13301. DOI: 10.1074/jbc.m208670200.

Picciani, Renata et al. (2007). "Cochlin in the eye: Functional implications." In: *Progress in Retinal and Eye Research* 26.5, pp. 453–469. DOI: 10.1016/j.preteyeres.2007.06.002.

Picelli, Simone et al. (2013). "Smart-seq2 for sensitive full-length transcriptome profiling in single cells." In: *Nature Methods* 10.11, 1096–1098. DOI: 10.1038/nmeth.2639.

Pierer, M. et al. (2007). "The TNF superfamily member LIGHT contributes to survival and activation of synovial fibroblasts in rheumatoid arthritis." In: *Rheumatology* 46.7, pp. 1063–1070. DOI: 10.1093/rheumatology/kem063.

Pierson, Emma and Christopher Yau (2015). "Zifa: Dimensionality Reduction for zero-inflated single-cell gene expression analysis." In: *Genome Biology* 16.1. DOI: 10.1186/s13059-015-0805-z.

Pinheiro, Alessandro S., Sadiq Silbak, and Alvin H. Schmaier (2022). "Bradykinin – An elusive peptide in measuring and understanding." In: *Research and Practice in Thrombosis and Haemostasis* 6.2, e12673. DOI: 10.1002/rth2.12673.

Pispa, Johanna et al. (2008). "Edar and Troy signalling pathways act redundantly to regulate initiation of hair follicle development." In: *Human Molecular Genetics* 17.21, pp. 3380–3391. DOI: 10.1093/hmg/ddn232.

Pizzagalli, Mattia D., Ariel Bensimon, and Giulio Superti-Furga (2020). "A guide to plasma membrane solute carrier proteins." In: *The FEBS Journal* 288.9, pp. 2784–2835. DOI: 10.1111/febs.15531.

Plikus, Maksim V. et al. (2017). "Regeneration of fat cells from myofibroblasts during wound healing." In: *Science* 355.6326, 748–752. DOI: 10.1126/science.aai8792.

Plikus, Maksim V. et al. (2021). "Fibroblasts: Origins, definitions, and functions in health and disease." In: *Cell* 184.15, 3852–3872. DOI: 10.1016/j.cell.2021.06.024.

Ploeger, Diana TA et al. (2013). "Cell plasticity in wound healing: Paracrine factors of M1/ M2 polarized macrophages influence the phenotypical state of dermal fibroblasts." In: *Cell Communication and Signaling* 11.1, p. 29. DOI: 10.1186/1478-811x-11-29.

Polak, Marta E. et al. (2012). "CD70–CD27 Interaction Augments $CD8^+$ T-Cell Activation by Human Epidermal Langerhans Cells." In: *Journal of Investigative Dermatology* 132.6, pp. 1636–1644. DOI: 10.1038/jid.2012.26.

Polanski, Krzysztof et al. (2020). "BBKNN: Fast Batch Alignment of Single Cell Transcriptomes." In: *Bioinformatics* 36(3), 964–965. DOI: 10.1093/bioinformatics/btz625.

Pollen, Alex A et al. (2014). "Low-coverage single-cell mrna sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex." In: *Nature Biotechnology* 32.10, 1053–1058. DOI: 10.1038/nbt.2967.

Pompella, Alfonso et al. (2003). "The changing faces of glutathione, a cellular protagonist." In: *Biochemical Pharmacology* 66.8, pp. 1499–1503. DOI: 10.1016/s0006-2952(03)00504-5.

Porsch, Helena et al. (2014). "Platelet-derived Growth Factor β-Receptor, Transforming Growth Factor β Type I Receptor, and CD44 Protein Modulate Each Other's Signaling and Stability." In: *Journal of Biological Chemistry* 289.28, pp. 19747–19757. DOI: 10.1074/jbc.m114.547273.

Porter, Keith R. (1964). "Cell fine structure and biosynthesis of intercellular macromolecules." In: *Biophysical Journal* 4.1, 167–196. DOI: 10.1016/s0006-3495(64)86936-8.

Potter, Andrew S. and S. Steven Potter (2019). "Dissociation of tissues for single-cell analysis." In: *Methods in Molecular Biology*, 55–62. DOI: 10.1007/978-1-4939-9021-4_5.

Poveda, Jonay et al. (2016). "MXRA5 is a TGF-β1-regulated human protein with anti-inflammatory and anti-fibrotic properties." In: *Journal of Cellular and Molecular Medicine* 21.1, pp. 154–164. DOI: 10.1111/jcmm.12953.

Pozzi, Ambra, Peter D. Yurchenco, and Renato V. Iozzo (2017). "The nature and biology of basement membranes." In: *Matrix Biology* 57-58, pp. 1–11. DOI: 10.1016/j.matbio.2016.12.009.

Prall, Wolf C. et al. (2007). "Age-related transcription levels of KU70, MGST1 and BIK in $CD34^+$ hematopoietic stem and progenitor cells." In: *Mechanisms of Ageing and Development* 128.9, pp. 503–510. DOI: 10.1016/j.mad.2007.06.008.

Price, April E. et al. (2010). "Systemically dispersed innate IL-13–expressing cells in type 2 immunity." In: *Proceedings of the National Academy of Sciences* 107.25, pp. 11489–11494. DOI: 10.1073/pnas.1003988107.

Purcell, James W. et al. (2018). "LRRC15 is a novel mesenchymal protein and stromal target for antibody–drug conjugates." In: *Cancer Research* 78.14, 4059–4072. DOI: 10.1158/0008-5472.can-18-0327.

Purohit, Abhilasha et al. (2014). "Semaphorin 5A mediated cellular navigation: Connecting nervous system and cancer." In: *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1846.2, pp. 485–493. DOI: 10.1016/j.bbcan.2014.09.006.

QUINN, Carmel M. et al. (2004). "Induction of fibroblast apolipoprotein E expression during apoptosis, starvation-induced growth arrest and mitosis." In: *Biochemical Journal* 378.3, pp. 753–761. DOI: 10.1042/bj20031352.

Qadri, Marwa et al. (2020). "Proteoglycan-4 regulates fibroblast to myofibroblast transition and expression of

fibrotic genes in the synovium." In: *Arthritis Research & Therapy* 22.1. DOI: 10.1186/s13075-020-02207-x.

Qin, Zhaoping, Gary J. Fisher, and Taihao Quan (2013). "Cysteine-rich protein 61 (CCN1) domain-specific stimulation of matrix metalloproteinase-1 expression through $\alpha v\,\beta 3$ integrin in human skin fibroblasts." In: *Journal of Biological Chemistry* 288.17, 12386–12394. DOI: 10.1074/jbc.m112.424358.

Qiu, Peng (2020). "Embracing the dropouts in single-cell RNA-seq analysis." In: *Nature Communications* 11.1169. DOI: 10.1038/s41467-020-14976-9.

Qiu, Xiaojie et al. (2017). "Reversed graph embedding resolves complex single-cell trajectories." In: *Nature Methods* 14.10, 979–982. DOI: 10.1038/nmeth.4402.

Quan, Taihao and Gary J. Fisher (2015). "Role of age-associated alterations of the dermal extracellular matrix microenvironment in human skin aging: A mini-review." In: *Gerontology* 61.5, 427–434. DOI: 10.1159/000371708.

Quenouille, Maurice H. (1956). "Notes on bias in estimation." In: *Biometrika* 43.3-4, 353–360. DOI: 10.1093/biomet/43.3-4.353.

Quinn, Thomas P. et al. (2017). "Propr: An R-package for identifying proportionally abundant features using compositional data analysis." In: *Scientific Reports* 7.1. DOI: 10.1038/s41598-017-16520-0.

Qwarnström, E E et al. (1991). "Interleukin 1 beta induces rapid phosphorylation and redistribution of Talin: A possible mechanism for modulation of fibroblast focal adhesion." In: *Proceedings of the National Academy of Sciences* 88.4, 1232–1236. DOI: 10.1073/pnas.88.4.1232.

Rahmani, Waleed, Sarthak Sinha, and Jeff Biernaskie (2020). "Immune modulation of hair follicle regeneration." In: *npj Regenerative Medicine* 5.1. DOI: 10.1038/s41536-020-0095-2.

Rahmani, Waleed et al. (2014). "Hair follicle dermal stem cells regenerate the dermal sheath, repopulate the dermal papilla, and modulate hair type." In: *Developmental Cell* 31.5, 543–558. DOI: 10.1016/j.devcel.2014.10.022.

Raj, Arjun et al. (2006). "Stochastic mrna synthesis in mammalian cells." In: *PLoS Biology* 4.10. DOI: 10.1371/journal.pbio.0040309.

Ramsköld, Daniel et al. (2012). "Full-length mrna-seq from single-cell levels of RNA and individual circulating tumor cells." In: *Nature Biotechnology* 30.8, 777–782. DOI: 10.1038/nbt.2282.

Ramón y Cajal, Santiago (1896). *Manual de Anatomía Patológica general, seguida de un Resumen de Microscopia Aplicada á La histología Y Bacteriología patológicas*. Nicolás Moya.

Ran, Di et al. (2020). "ScDoc: Correcting drop-out events in single-cell RNA-seq data." In: *Bioinformatics* 36.15, 4233–4239. DOI: 10.1093/bioinformatics/btaa283.

Rao, V. Srinivasa et al. (2014). "Protein-protein interaction detection: Methods and analysis." In: *International Journal of Proteomics* 2014, 1–12. DOI: 10.1155/2014/147648.

Raser, Jonathan M. and Erin K. O'Shea (2004). "Control of stochasticity in eukaryotic gene expression." In: *Science* 304.5678, 1811–1814. DOI: 10.1126/science.1098641.

Reboldi, Andrea et al. (2014). "25-Hydroxycholesterol suppresses interleukin-1–driven inflammation downstream of type I interferon." In: *Science* 345.6197, pp. 679–684. DOI: 10.1126/science.1254790.

Reddy, Aravind T. et al. (2018). "Role of GPx3 in PPARγ-induced protection against COPD-associated oxidative stress." In: *Free Radical Biology and Medicine* 126, pp. 350–357. DOI: 10.1016/j.freeradbiomed.2018.08.014.

Reddy, Seshamma et al. (2001). "Characterization of Wnt gene expression in developing and postnatal hair follicles and identification of Wnt5a as a target of Sonic hedgehog in hair follicle morphogenesis." In: *Mechanisms of Development* 107.1-2, pp. 69–82. DOI: 10.1016/s0925-4773(01)00452-x.

Reed, Charles C. and Renato V. Iozzo (2002). "The role of decorin in collagen fibrillogenesis and skin homeostasis." In: *Glycoconjugate Journal* 19.4/5, 249–255. DOI: 10.1023/a:1025383913444.

Regn, Michael et al. (2016). "Peptidase inhibitor 16 is a membrane-tethered regulator of chemerin processing in the myocardium." In: *Journal of Molecular and Cellular Cardiology* 99, pp. 57–64. DOI: 10.1016/j.yjmcc.2016.08.010.

Reif, Karin et al. (2002). "Balanced responsiveness to chemoattractants from adjacent zones determines B-cell position." In: *Nature* 416.6876, pp. 94–99. DOI: 10.1038/416094a.

Ren, Liwen et al. (2022). "Systematic pan-cancer analysis identifies APOC1 as an immunological biomarker which regulates macrophage polarization and promotes tumor metastasis." In: *Pharmacological Research* 183, p. 106376. DOI: 10.1016/j.phrs.2022.106376.

Ren, Xianwen et al. (2021). "Covid-19 immune features revealed by a large-scale single-cell transcriptome Atlas." In: *Cell* 184.7. DOI: 10.1016/j.cell.2021.01.053.

Rendl, Michael, Lisa Lewis, and Elaine Fuchs (2005). "Molecular dissection of mesenchymal–epithelial interactions in the hair follicle." In: *PLoS Biology* 3.11. DOI: 10.1371/journal.pbio.0030331.

Reynolds, Gary et al. (2021). "Developmental cell programs are co-opted in inflammatory skin disease." In: *Science* 371.6527. DOI: 10.1126/science.aba6500.

Rezza, Amélie et al. (2016). "Signaling networks among stem cell precursors, transit-amplifying progenitors, and their niche in developing hair follicles." In: *Cell Reports* 14.12, 3001–3018. DOI: 10.1016/j.celrep.2016.02.078.

Rezzonico, Roger, Danielle Burger, and Jean-Michel Dayer (1998). "Direct contact between T lymphocytes and human dermal fibroblasts or synoviocytes down-regulates types I and III collagen production via cell-associated cytokines." In: *Journal of Biological Chemistry* 273.30, 18720–18728. DOI: 10.1074/jbc.273.30.18720.

Rhee, David K. et al. (2005). "The secreted glycoprotein lubricin protects cartilage surfaces and inhibits synovial cell overgrowth." In: *Journal of Clinical Investigation* 115.3, pp. 622–631. DOI: 10.1172/jci200522263.

Ricci, Chiara et al. (2018). "PCSK9 induces a pro-inflammatory response in macrophages." In: *Scientific Reports* 8.1. DOI: 10.1038/s41598-018-20425-x.

Richard, Laurence et al. (2012). "Endoneurial Fibroblast-Like Cells." In: *Journal of Neuropathology & Experimental Neurology* 71.11, pp. 938–947. DOI: 10.1097/nen.0b013e318270a941.

Richard, Laurence et al. (2014). "Characterization of Endoneurial Fibroblast-like Cells from Human and Rat Peripheral Nerves." In: *Journal of Histochemistry & Cytochemistry* 62.6, pp. 424–435. DOI: 10.1369/0022155414530994.

Richendrfer, Holly and Gregory D. Jay (2020). "Lubricin as a Therapeutic and Potential Biomarker in Sepsis." In: *Critical Care Clinics* 36.1, pp. 55–67. DOI: 10.1016/j.ccc.2019.08.005.

Rickert, Robert C., Julia Jellusova, and Ana V. Miletic (2011). "Signaling by the tumor necrosis factor receptor superfamily in B-cell biology and disease." In: *Immunological Reviews* 244.1, pp. 115–133. DOI: 10.1111/j.1600-065x.2011.01067.x.

Rindler, Katharina et al. (2021). "Single-cell RNA sequencing reveals markers of disease progression in primary cutaneous T-cell lymphoma." In: *Molecular Cancer* 20.1. DOI: 10.1186/s12943-021-01419-2.

Rinkevich, Yuval et al. (2015). "Identification and isolation of a dermal lineage with intrinsic fibrogenic potential." In: *Science* 348.6232. DOI: 10.1126/science.aaa2151.

Rinn, John L et al. (2006). "Anatomic demarcation by positional variation in fibroblast gene expression programs." In: *PLoS Genetics* 2.7. DOI: 10.1371/journal.pgen.0020119.

Rinn, John L. et al. (2008). "A dermal HOX transcriptional program regulates site-specific epidermal fate." In: *Genes & Development* 22.3, pp. 303–307. DOI: 10.1101/gad.1610508.

Rivera-Gonzalez, Guillermo C. et al. (2016). "Skin adipocyte stem cell self-renewal is regulated by a PDGFA/AKT-signaling axis." In: *Cell Stem Cell* 19.6, 738–751. DOI: 10.1016/j.stem.2016.09.002.

Robbiani, Davide F et al. (2000). "The Leukotriene C_4 Transporter MRP1 Regulates CCL19 (MIP-3β, ELC)–Dependent Mobilization of Dendritic Cells to Lymph Nodes." In: *Cell* 103.5, pp. 757–768. DOI: 10.1016/s0092-8674(00)00179-3.

Robertson, Nahid G. et al. (1997). "Mapping and Characterization of a Novel Cochlear Gene in Human and in Mouse: A Positional Candidate Gene for a Deafness Disorder, DFNA9." In: *Genomics* 46.3, pp. 345–354. DOI: 10.1006/geno.1997.5067.

Roeder, Robert G. and William J. Rutter (1969). "Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms." In: *Nature* 224.5216, 234–237. DOI: 10.1038/224234a0.

Rognoni, Emanuel et al. (2016). "Inhibition of β-catenin signalling in dermal fibroblasts enhances hair follicle regeneration during wound healing." In: *Development*. DOI: 10.1242/dev.131797.

Rognoni, Emanuel et al. (2018). "Fibroblast state switching orchestrates dermal maturation and wound healing." In: *Molecular Systems Biology* 14.8. DOI: 10.15252/msb.20178174.

Rojahn, Thomas B. et al. (2020). "Single-cell transcriptomics combined with interstitial fluid proteomics defines cell-type-specific immune regulation in atopic dermatitis." In: *Journal of Allergy and Clinical Immunology* 146(5), pp. 1056–1069. DOI: 10.1016/j.jaci.2020.03.041.

Romanowska, Malgorzata et al. (2009). "Wnt5a Exhibits Layer-Specific Expression in Adult Skin, Is Upregulated in Psoriasis, and Synergizes with Type 1 Interferon." In: *PLoS ONE* 4.4. Ed. by Elizabeth Didier, e5354. DOI: 10.1371/journal.pone.0005354.

Rosenberg, Alexander B. et al. (2018). "Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding." In: *Science* 360.6385, 176–182. DOI: 10.1126/science.aam8999.

Rosenquist, Thomas A. and Gail R. Martin (1996). "Fibroblast growth factor signalling in the hair growth cycle: Expression of the fibroblast growth factor receptor and ligand genes in the murine hair follicle." In: *Developmental Dynamics* 205.4, pp. 379–386. DOI: 10.1002/(sici)1097-0177(199604)205:4<379::aid-aja2>3.0.co;2-f.

Rosenzweig, Sergio D. (2018). "Old vaccines, new diseases: when BCG meets SPPL2a." In: *Nature Immunology* 19.9, pp. 906–907. DOI: 10.1038/s41590-018-0193-0.

Roughley, P J and R J White (1989). "Dermatan sulphate proteoglycans of human articular cartilage. The properties of dermatan sulphate proteoglycans I and II." In: *Biochemical Journal* 262.3, pp. 823–827. DOI: 10.1042/bj2620823.

Roughley, Peter J and John S Mort (2014). "The role of aggrecan in normal and osteoarthritic cartilage." In: *Journal of Experimental Orthopaedics* 1.1. DOI: 10.1186/s40634-014-0008-7.

Rousseeuw, Peter J. (1987). "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis." In:

Journal of Computational and Applied Mathematics 20, 53–65. DOI: 10.1016/0377-0427(87)90125-7.

Ruano, Juan et al. (2016). "Molecular and cellular profiling of scalp psoriasis reveals differences and similarities compared to skin psoriasis." In: *PLOS ONE* 11.2. DOI: 10.1371/journal.pone.0148450.

Rubin, Mark A. et al. (2004). "Overexpression, Amplification, and Androgen Regulation of TPD52 in Prostate Cancer." In: *Cancer Research* 64.11, pp. 3814–3822. DOI: 10.1158/0008-5472.can-03-3881.

Rudman, Sarah M. et al. (1997). "The Role of IGF-I in Human Skin and its Appendages: Morphogen as Well as Mitogen?" In: *Journal of Investigative Dermatology* 109.6, pp. 770–777. DOI: 10.1111/1523-1747.ep12340934.

Ruffner, Heinz et al. (2012). "R-spondin potentiates Wnt/β-catenin signaling through orphan receptors LGR4 and LGR5." In: *PLoS ONE* 7.7. DOI: 10.1371/journal.pone.0040976.

Ruiz-Cordero, Roberto and Dianna L. Ng (2020). "Neurotrophic receptor tyrosine kinase (NTRK) fusions and their role in cancer." In: *Cancer Cytopathology* 128.11, pp. 775–779. DOI: 10.1002/cncy.22350.

Runager, Kasper et al. (2013). "The Insoluble TGFBIp Fraction of the Cornea Is Covalently Linked via a Disulfide Bond to Type XII Collagen." In: *Biochemistry* 52.16, pp. 2821–2827. DOI: 10.1021/bi400212m.

Ruoslahti, Erkki and Michael D. Pierschbacher (1987). "New Perspectives in Cell Adhesion: RGD and Integrins." In: *Science* 238.4826, pp. 491–497. DOI: 10.1126/science.2821619.

Ruscica, Massimiliano et al. (2019). "PCSK9 inhibition and inflammation: A narrative review." In: *Atherosclerosis* 288, pp. 146–155. DOI: 10.1016/j.atherosclerosis.2019.07.015.

Russo, Barbara, Nicolò C. Brembilla, and Carlo Chizzolini (2020). "Interplay Between Keratinocytes and Fibroblasts: A Systematic Review Providing a New Angle for Understanding Skin Fibrotic Disorders." In: *Frontiers in Immunology* 11. DOI: 10.3389/fimmu.2020.00648.

Rutkowski, Martin J. et al. (2010). "The complement cascade as a mediator of tissue growth and regeneration." In: *Inflammation Research* 59.11, pp. 897–905. DOI: 10.1007/s00011-010-0220-6.

Rypdal, Karoline B. et al. (2022). "ADAMTSL3 knock-out mice develop cardiac dysfunction and dilatation with increased TGFβ signalling after pressure overload." In: *Communications Biology* 5.1. DOI: 10.1038/s42003-022-04361-1.

Röder, Annika et al. (2023). "Spotlight on CYP4B1." In: *International Journal of Molecular Sciences* 24.3, p. 2038. DOI: 10.3390/ijms24032038.

Rüegg, C R, R Chiquet-Ehrismann, and S S Alkan (1989). "Tenascin, an extracellular matrix protein, exerts immunomodulatory activities." In: *Proceedings of the National Academy of Sciences* 86.19, pp. 7437–7441. DOI: 10.1073/pnas.86.19.7437.

Saalbach, Anja, Uwe F. Haustein, and Ulf Anderegg (2000). "A ligand of human thy-1 is localized on polymorphonuclear leukocytes and monocytes and mediates the binding to activated thy-1-positive microvascular endothelial cells and fibroblasts." In: *Journal of Investigative Dermatology* 115.5, 882–888. DOI: 10.1046/j.1523-1747.2000.00104.x.

Saarela, Janna et al. (1998). "Complete primary structure of two variant forms of human type XVIII collagen and tissue-specific differences in the expression of the corresponding transcripts." In: *Matrix Biology* 16.6, pp. 319–328. DOI: 10.1016/s0945-053x(98)90003-8.

Sabatelli, Patrizia et al. (2011). "Expression of the Collagen VI α5 and α6 Chains in Normal Human Skin and in Skin of Patients with Collagen VI-Related Myopathies." In: *Journal of Investigative Dermatology* 131.1, pp. 99–107. DOI: 10.1038/jid.2010.284.

Saegusa, Jun et al. (2009). "The Direct Binding of Insulin-like Growth Factor-1 (IGF-1) to Integrin αvβ3 Is Involved in IGF-1 Signaling." In: *Journal of Biological Chemistry* 284.36, pp. 24106–24114. DOI: 10.1074/jbc.m109.013201.

Saelens, Wouter et al. (2019). "A comparison of single-cell trajectory inference methods." In: *Nature Biotechnology* 37.5, 547–554. DOI: 10.1038/s41587-019-0071-9.

Saeys, Yvan, Iñaki Inza, and Pedro Larrañaga (2007). "A review of feature selection techniques in bioinformatics." In: *Bioinformatics* 23.19, pp. 2507–2517. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btm344. eprint: https://academic.oup.com/bioinformatics/article-pdf/23/19/2507/452322/btm344.pdf.

Safran, Marilyn et al. (2021). "The GeneCards Suite." In: *Practical Guide to Life Science Databases*. Ed. by Imad Abugessaisa and Takeya Kasukawa. Singapore: Springer Nature Singapore, pp. 27–56. ISBN: 978-981-16-5812-9. DOI: 10.1007/978-981-16-5812-9_2.

Saikia, Paramananda et al. (2018). "Basement membranes in the cornea and other organs that commonly develop fibrosis." In: *Cell and Tissue Research* 374.3, 439–453. DOI: 10.1007/s00441-018-2934-7.

Saito, Momomi, Maho Hamasaki, and Masabumi Shibuya (2003). "Induction of Tube Formation by angiopoietin-1 in Endothelial Cell/fibroblast co-culture is dependent on endogenous VEGF." In: *Cancer Science* 94.9, 782–790. DOI: 10.1111/j.1349-7006.2003.tb01519.x.

Saliba, Antoine-Emmanuel et al. (2014). "Single-cell RNA-seq: Advances and future challenges." In: *Nucleic Acids Research* 42.14, 8845–8860. DOI: 10.1093/nar/gku555.

Saluzzo, Simona et al. (2021). "Delayed antiretroviral therapy in HIV-infected individuals leads to irreversible depletion of skin- and mucosa-resident memory T cells." In: *Immunity* 54.12. DOI: 10.1016/j.immuni.2021.10.021.

Salzer, Marion Claudia et al. (2018). "Identity noise and adipogenic traits characterize dermal fibroblast aging." In: *Cell* 175.6. DOI: 10.1016/j.cell.2018.10.012.

Santos, Celio XC et al. (2016). "Targeted redox inhibition of protein phosphatase 1 by Nox4 regulates eIF2α-mediated stress signaling." In: *The EMBO Journal* 35.3, pp. 319–334. DOI: 10.15252/embj.201592394.

Santra, Manoranjan, Charles C. Reed, and Renato V. Iozzo (2002). "Decorin Binds to a Narrow Region of the Epidermal Growth Factor (EGF) Receptor, Partially Overlapping but Distinct from the EGF-binding Epitope." In: *Journal of Biological Chemistry* 277.38, pp. 35671–35681. DOI: 10.1074/jbc.m205317200.

Sarrias, Maria Rosa et al. (2004). "The Scavenger Receptor Cysteine-Rich (SRCR) Domain: An Ancient and Highly Conserved Protein Module of the Innate Immune System." In: *Critical Reviews in Immunology* 24.1, pp. 1–38. DOI: 10.1615/critrevimmunol.v24.i1.10.

Sasaki, M. and J. Enami (1996). "Structure and Expression of a Murine Homologue of sky Receptor Tyrosine Kinase Gene." In: *Journal of Biochemistry* 120.2, pp. 264–270. DOI: 10.1093/oxfordjournals.jbchem.a021408.

Satija, Rahul et al. (2015). "Spatial reconstruction of single-cell gene expression data." In: *Nature Biotechnology* 33.5, 495–502. DOI: 10.1038/nbt.3192.

Sato, Akira et al. (2009). "WNT5A regulates distinct signalling pathways by binding to FRIZZLED2." In: *The EMBO Journal* 29.1, 41–54. DOI: 10.1038/emboj.2009.322.

Sato, Hiroyasu, Yoshitaka Taketomi, and Makoto Murakami (2016). "Metabolic regulation by secreted phospholipase A2." In: *Inflammation and Regeneration* 36.1. DOI: 10.1186/s41232-016-0012-7.

Sawant, Kirti V et al. (2020). "Neutrophil recruitment by chemokines Cxcl1/KC and Cxcl2/MIP2: Role of Cxcr2 activation and glycosaminoglycan interactions." In: *Journal of Leukocyte Biology* 109.4, pp. 777–791. DOI: 10.1002/jlb.3a0820-207r.

Schafer, Sebastian et al. (2017). "IL-11 is a crucial determinant of cardiovascular fibrosis." In: *Nature* 552.7683, pp. 110–115. DOI: 10.1038/nature24676.

Scharf, Gesine M. et al. (2019). "Inactivation of Sox9 in fibroblasts reduces cardiac fibrosis and inflammation." In: *JCI Insight* 4.15. DOI: 10.1172/jci.insight.126721.

Scheaffer, Hannah L. et al. (2020). "Inactivation of CES1 Blocks Prostaglandin D2 Glyceryl Ester Catabolism in Monocytes/Macrophages and Enhances Its Anti-inflammatory Effects, Whereas the Pro-inflammatory Effects of Prostaglandin E2 Glyceryl Ester Are Attenuated." In: *ACS Omega* 5.45, pp. 29177–29188. DOI: 10.1021/acsomega.0c03961.

Schena, Mark et al. (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." In: *Science* 270.5235, 467–470. DOI: 10.1126/science.270.5235.467.

Schenk, Susann et al. (2003). "Binding to EGF receptor of a laminin-5 EGF-like fragment liberated during MMP-dependent mammary gland involution." In: *Journal of Cell Biology* 161.1, pp. 197–209. DOI: 10.1083/jcb.200208145.

Scherer, Philipp E. et al. (1997). "Cell-type and Tissue-specific Expression of Caveolin-2." In: *Journal of Biological Chemistry* 272.46, pp. 29337–29346. DOI: 10.1074/jbc.272.46.29337.

Schiavinato, Alvise et al. (2016). "Targeting of EMILIN-1 and EMILIN-2 to Fibrillin Microfibrils Facilitates their Incorporation into the Extracellular Matrix." In: *Journal of Investigative Dermatology* 136.6, pp. 1150–1160. DOI: 10.1016/j.jid.2016.02.021.

Schlake, Thomas (2007). "Determination of hair structure and shape." In: *Seminars in Cell & Developmental Biology* 18.2, 267–273. DOI: 10.1016/j.semcdb.2007.01.005.

Schmid, Andreas et al. (2021). "Anti-Inflammatory Effects of C1q/Tumor Necrosis Factor-Related Protein 3 (CTRP3) in Endothelial Cells." In: *Cells* 10.8, p. 2146. DOI: 10.3390/cells10082146.

Schmidt, Clint S. et al. (2002). "Enhanced B Cell Expansion, Survival, and Humoral Responses by Targeting Death Receptor 6." In: *Journal of Experimental Medicine* 197.1, pp. 51–62. DOI: 10.1084/jem.20020617.

Schneider, Marlon R., Ruth Schmidt-Ullrich, and Ralf Paus (2009). "The hair follicle as a dynamic Miniorgan." In: *Current Biology* 19.3. DOI: 10.1016/j.cub.2008.12.005.

Schneider, Ronen et al. (2020). "DAAM2 Variants Cause Nephrotic Syndrome via Actin Dysregulation." In: *The American Journal of Human Genetics* 107.6, pp. 1113–1128. DOI: 10.1016/j.ajhg.2020.11.008.

Schoeps, Benjamin, Julian Frädrich, and Achim Krüger (2022). "Cut loose TIMP-1: an emerging cytokine in inflammation." In: *Trends in Cell Biology*. DOI: 10.1016/j.tcb.2022.08.005.

Scholz, Claus J. et al. (2018). "FASTGenomics: An analytical ecosystem for single-cell RNA sequencing data." In: DOI: 10.1101/272476.

Schugar, Rebecca C. and J. Mark Brown (2015). "Emerging roles of flavin monooxygenase 3 in cholesterol metabolism and atherosclerosis." In: *Current Opinion in Lipidology* 26.5, pp. 426–431. DOI: 10.1097/mol.0000000000000215.

Schuster, Ronen et al. (2022). "The role of myofibroblasts in physiological and pathological tissue repair." In: *Cold Spring Harbor Perspectives in Biology* 15.1. DOI: 10.1101/cshperspect.a041231.

Schwanhäusser, Björn et al. (2011). "Global quantification of mammalian gene expression control." In: *Nature* 473.7347, 337–342. DOI: 10.1038/nature10098.

Schwann, Theodor (1839). *Mikroskopische Untersuchungen über die uebereinstimmung in der Struktur und dem wachsthum der thiere und pflanzen*. Sander.

Schwarz, Margrit et al. (2000). "The bile acid synthetic gene 3β-hydroxy-Δ5-C27-steroid oxidoreductase is mutated in progressive intrahepatic cholestasis." In: *Journal of Clinical Investigation* 106.9, pp. 1175–1184. DOI: 10.1172/jci10902.

Schäbitz, A. et al. (2022). "Spatial transcriptomics landscape of lesions from non-communicable inflammatory skin diseases." In: *Nature Communications* 13.1. DOI: 10.1038/s41467-022-35319-w.

Schönherr, Elke et al. (1995). "Interaction of Biglycan with Type I Collagen." In: *Journal of Biological Chemistry* 270.6, pp. 2776–2783. DOI: 10.1074/jbc.270.6.2776.

Schütze, Sebastian, Ian J. Orozco, and Thomas J. Jentsch (2016). "KCNQ Potassium Channels Modulate Sensitivity of Skin Down-hair (D-hair) Mechanoreceptors." In: *Journal of Biological Chemistry* 291.11, pp. 5566–5575. DOI: 10.1074/jbc.m115.681098.

Seatter, Michael J. et al. (1998). "QLS Motif in Transmembrane Helix VII of the Glucose Transporter Family Interacts with the C-1 Position of d-Glucose and Is Involved in Substrate Selection at the Exofacial Binding Site." In: *Biochemistry* 37.5, pp. 1322–1326. DOI: 10.1021/bi972322u.

Segaliny, Aude I. et al. (2015). "Syndecan-1 regulates the biological activities of interleukin-34." In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1853.5, pp. 1010–1021. DOI: 10.1016/j.bbamcr.2015.01.023.

Segalla, Lydia, Salvatore Chirumbolo, and Andrea Sbarbati (2021). "Dermal white adipose tissue: Much more than a metabolic, lipid-storage organ?" In: *Tissue and Cell* 71, p. 101583. DOI: 10.1016/j.tice.2021.101583.

Seidah, Nabil G and Annik Prat (2021). "The Multifaceted Biology of PCSK9." In: *Endocrine Reviews* 43.3, pp. 558–582. DOI: 10.1210/endrev/bnab035.

Sena, Johnny A. et al. (2018). "Unique molecular identifiers reveal a novel sequencing artefact with implications for RNA-seq based gene expression analysis." In: *Scientific Reports* 8.1. DOI: 10.1038/s41598-018-31064-7.

Sengle, Gerhard et al. (2012). "Microenvironmental Regulation by Fibrillin-1." In: *PLoS Genetics* 8.1. Ed. by Marshall S. Horwitz, e1002425. DOI: 10.1371/journal.pgen.1002425.

Sennett, Rachel and Michael Rendl (2012). "Mesenchymal–epithelial interactions during hair follicle morphogenesis and cycling." In: *Seminars in Cell & Developmental Biology* 23.8, 917–927. DOI: 10.1016/j.semcdb.2012.08.011.

Sennett, Rachel et al. (2015). "An integrated transcriptome atlas of embryonic hair follicle progenitors, their niche, and the developing skin." In: *Developmental Cell* 34.5, 577–591. DOI: 10.1016/j.devcel.2015.06.023.

Serrano-Falcón, C., M.A. Fernández-Pugnaire, and S. Serrano-Ortega (2013). "Hair and Scalp Evaluation: The Trichogram." In: *Actas Dermo-Sifiliográficas (English Edition)* 104.10, pp. 867–876. DOI: 10.1016/j.adengl.2013.03.009.

Seta, Francesca et al. (2007). "Inhibition of VEGF expression and corneal neovascularization by siRNA targeting cytochrome P450 4B1." In: *Prostaglandins & Other Lipid Mediators* 84.3-4, pp. 116–127. DOI: 10.1016/j.prostaglandins.2007.05.001.

Setty, Manu et al. (2016). "Wishbone identifies bifurcating developmental trajectories from single-cell data." In: *Nature Biotechnology* 34.6, 637–645. DOI: 10.1038/nbt.3569.

Seykora, John T. (2010). "Grabbing amphiregulin by the tail to better understand keratinocyte growth." In: *Journal of Investigative Dermatology* 130.8, 1966–1968. DOI: 10.1038/jid.2010.199.

Seymour, Christopher W. et al. (2016). "Assessment of Clinical Criteria for Sepsis." In: *JAMA* 315.8, p. 762. DOI: 10.1001/jama.2016.0288.

Shao, Xin et al. (2020). "CellTalkDB: A manually curated database of ligand–receptor interactions in humans and mice." In: *Briefings in Bioinformatics* 22.4. DOI: 10.1093/bib/bbaa269.

Sharp, Lester W. (1984). *Introduction to cytology*. McGraw-Hill.

Shaw, Tanya J and Emanuel Rognoni (2020). "Dissecting fibroblast heterogeneity in health and fibrotic disease." In: *Current Rheumatology Reports* 22.8. DOI: 10.1007/s11926-020-00903-w.

Shetty, Ankitha et al. (2022). "A systematic comparison of FOSL1, FOSL2 and BATF-mediated transcriptional regulation during early human Th17 differentiation." In: *Nucleic Acids Research* 50.9, pp. 4938–4958. DOI: 10.1093/nar/gkac256.

Shi, Weibin et al. (2004). "GADD34–PP1c recruited by Smad7 dephosphorylates TGFβ type I receptor." In: *Journal of Cell Biology* 164.2, pp. 291–300. DOI: 10.1083/jcb.200307151.

Shiga, Atsushi et al. (2011). "Cerebral small-vessel disease protein HTRA1 controls the amount of TGF-β1 via cleavage of proTGF-β1." In: *Human Molecular Genetics* 20.9, pp. 1800–1810. DOI: 10.1093/hmg/ddr063.

Shih, Diana M. et al. (2019). "Genetic Deficiency of Flavin-Containing Monooxygenase 3 (Fmo3) Protects Against Thrombosis but Has Only a Minor Effect on Plasma Lipid Levels—Brief Report." In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 39.6, pp. 1045–1054. DOI: 10.1161/atvbaha.119.312592.

Shimizu-Hirota, Ryoko et al. (2004). "Functional characterization of podocan, a member of a new class in the small leucine-rich repeat protein family." In: *FEBS Letters* 563.1-3, pp. 69–74. DOI: 10.1016/s0014-5793(04)00250-9.

Shimizu, Kiyoshi et al. (2001). "Manic Fringe and Lunatic Fringe Modify Different Sites of the Notch2 Extracellular Region, Resulting in Different Signaling Modulation." In: *Journal of Biological Chemistry* 276.28, pp. 25753–25758. DOI: 10.1074/jbc.m103473200.

Shin, HyeRim et al. (2010). "Identification of transcriptional targets of Wnt/β-catenin signaling in dermal papilla cells of human scalp hair follicles: EP2 is a novel transcriptional target of Wnt3a." In: *Journal of Dermatological Science* 58.2, pp. 91–96. DOI: 10.1016/j.jdermsci.2010.02.011.

Shin, Jaehoon et al. (2015). "Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis." In: *Cell Stem Cell* 17.3, 360–372. DOI: 10.1016/j.stem.2015.07.013.

Shin, Minsang et al. (2017). "Recent Insights into Insulin-Like Growth Factor Binding Protein 2 Transcriptional Regulation." In: *Endocrinology and Metabolism* 32.1, p. 11. DOI: 10.3803/enm.2017.32.1.11.

Shin, Wisoo et al. (2020). "Dysfunction of hair follicle mesenchymal progenitors contributes to age-associated hair loss." In: *Developmental Cell* 53.2. DOI: 10.1016/j.devcel.2020.03.019.

Shinjyo, Noriko, Wataru Kagaya, and Marcela Pekna (2021). "Interaction Between the Complement System and Infectious Agents – A Potential Mechanistic Link to Neurodegeneration and Dementia." In: *Frontiers in Cellular Neuroscience* 15. DOI: 10.3389/fncel.2021.710390.

Shook, Brett A. et al. (2018). "Myofibroblast proliferation and heterogeneity are supported by macrophages during skin repair." In: *Science* 362.6417. DOI: 10.1126/science.aar2971.

Shook, Brett A. et al. (2020). "Dermal adipocyte lipolysis and myofibroblast conversion are required for efficient skin repair." In: *Cell Stem Cell* 26.6. DOI: 10.1016/j.stem.2020.03.013.

Siebuhr, A.S., C.S. Thudium, and M.A. Karsdal (2019). "Type XIII collagen." In: *Biochemistry of Collagens, Laminins and Elastin*. Elsevier, pp. 115–120. DOI: 10.1016/b978-0-12-817068-7.00013-6.

Silverman, Justin D. et al. (2020). "Naught all zeros in sequence count data are the same." In: *Computational and Structural Biotechnology Journal* 18, 2789–2798. DOI: 10.1016/j.csbj.2020.09.014.

Simon, M. Celeste and Brian Keith (2008). "The role of oxygen availability in embryonic development and stem cell function." In: *Nature Reviews Molecular Cell Biology* 9.4, 285–296. DOI: 10.1038/nrm2354.

Singh, N. A. (2003). "KCNQ2 and KCNQ3 potassium channel genes in benign familial neonatal convulsions: expansion of the functional and mutation spectrum." In: *Brain* 126.12, pp. 2726–2737. DOI: 10.1093/brain/awg286.

Sinha, Mithun et al. (2018). "Direct conversion of injury-site myeloid cells to fibroblast-like cells of granulation tissue." In: *Nature Communications* 9.1. DOI: 10.1038/s41467-018-03208-w.

Smaldino, Paul E. and Richard McElreath (2016). "The natural selection of bad science." In: *Royal Society Open Science* 3.9, p. 160384. DOI: 10.1098/rsos.160384.

Smerdova, Lenka et al. (2014). "Upregulation of CYP1B1 expression by inflammatory cytokines is mediated by the p38 MAP kinase signal transduction pathway." In: *Carcinogenesis* 35.11, pp. 2534–2543. DOI: 10.1093/carcin/bgu190.

Smith, T J (2005). "Insights into the role of fibroblasts in human autoimmune diseases." In: *Clinical and Experimental Immunology* 141.3, pp. 388–397. DOI: 10.1111/j.1365-2249.2005.02824.x.

Smithey, Megan J. et al. (2014). "Lost in translation: Mice, men and cutaneous immunity in old age." In: *Biogerontology* 16.2, 203–208. DOI: 10.1007/s10522-014-9517-0.

Smola, Hans et al. (1998). "Dynamics of basement membrane formation by keratinocyte–fibroblast interactions in organotypic skin culture." In: *Experimental Cell Research* 239.2, 399–410. DOI: 10.1006/excr.1997.3910.

So, Takanori and Michael Croft (2013). "Regulation of PI-3-Kinase and Akt Signaling in T Lymphocytes and Other Cells by TNFR Family Molecules." In: *Frontiers in Immunology* 4. DOI: 10.3389/fimmu.2013.00139.

Sobanov, Yuri et al. (2001). "A novel cluster of lectin-like receptor genes expressed in monocytic, dendritic and endothelial cells maps close to the NK receptor genes in the human NK gene complex." In: *European Journal of Immunology* 31.12, pp. 3493–3503. DOI: 10.1002/1521-4141(200112)31:12<3493::aid-immu3493>3.0.co;2-9.

Sobolev, Vladimir V. et al. (2022). "Role of the Transcription Factor FOSL1 in Organ Development and Tumorigenesis." In: *International Journal of Molecular Sciences* 23.3, p. 1521. DOI: 10.3390/ijms23031521.

Soliman, Hesham et al. (2021). "Multipotent stromal cells: One name, multiple identities." In: *Cell Stem Cell* 28.10, 1690–1707. DOI: 10.1016/j.stem.2021.09.001.

Solé-Boldo, Llorenç et al. (2020). "Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming." In: *Communications Biology* 3.1. DOI: 10.1038/s42003-020-0922-4.

Son, Hye-Nam et al. (2013). "Multiple FAS1 domains and the RGD motif of TGFBI act cooperatively to bind αvβ3 integrin, leading to anti-angiogenic and anti-tumor effects." In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1833.10, pp. 2378–2388. DOI: 10.1016/j.bbamcr.2013.06.012.

Soneson, Charlotte and Mark D Robinson (2018). "Bias, robustness and scalability in single-cell differential expression analysis." In: *Nature Methods* 15.4, 255–261. DOI: 10.1038/nmeth.4612.

Sontheimer, Richard D. (2014). "Skin is not the largest organ." In: *Journal of Investigative Dermatology* 134.2, 581–582. DOI: 10.1038/jid.2013.335.

Sorci, Guglielmo (2013). "S100B protein in tissue development, repair and regeneration." In: *World Journal of Biological Chemistry* 4.1, p. 1. DOI: 10.4331/wjbc.v4.i1.1.

Sorrell, J. Michael and Arnold I. Caplan (2004). "Fibroblast heterogeneity: More than skin deep." In: *Journal of Cell Science* 117.5, 667–675. DOI: 10.1242/jcs.01005.

– (2009). "Fibroblasts—a diverse population at the center of it all." In: *International Review of Cell and Molecular Biology*, 161–214. DOI: 10.1016/s1937-6448(09)76004-6.

Soundararajan, Avinash et al. (2021). "Cathepsin K Regulates Intraocular Pressure by Modulating Extracellular Matrix Remodeling and Actin-Bundling in the Trabecular Meshwork Outflow Pathway." In: *Cells* 10.11, p. 2864. DOI: 10.3390/cells10112864.

Sowa, Grzegorz (2011). "Novel Insights into the Role of Caveolin-2 in Cell- and Tissue-Specific Signaling and Function." In: *Biochemistry Research International* 2011, pp. 1–9. DOI: 10.1155/2011/809259.

Sprent, Jonathan (1995). "Antigen-presenting cells: Professionals and amateurs." In: *Current Biology* 5.10, 1095–1097. DOI: 10.1016/s0960-9822(95)00219-3.

Spörri, Bernhard et al. (1996). "Juxtacrine stimulation of cytokine production in coculturesof human dermal fibroblasts and T cells." In: *Cytokine* 8.8, 631–635. DOI: 10.1006/cyto.1996.0084.

Squair, Jordan W. et al. (2021). "Confronting false discoveries in single-cell differential expression." In: *Nature Communications* 12.1. DOI: 10.1038/s41467-021-25960-2.

Sriram, Gopu, Paul Lorenz Bigliardi, and Mei Bigliardi-Qi (2015). "Fibroblast heterogeneity and its implications for engineering organotypic skin models in vitro." In: *European Journal of Cell Biology* 94.11, 483–512. DOI: 10.1016/j.ejcb.2015.08.001.

Srivastava, Avi et al. (2019). "Alevin efficiently estimates accurate gene abundances from dscRNA-seq data." In: *Genome biology* 20.1, p. 65.

Ståhl, Patrik L. et al. (2016). "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics." In: *Science* 353.6294, 78–82. DOI: 10.1126/science.aaf2403.

Staton, C.A. et al. (2010). "Angiopoietin-1, Angiopoietin-2 and tie-2 receptor expression in human dermal wound repair and scarring." In: *British Journal of Dermatology* 163.5, 920–927. DOI: 10.1111/j.1365-2133.2010.09940.x.

Staumont-Salle, Delphine et al. (2014). "CX3CL1 (fractalkine) and its receptor CX3CR1 regulate atopic dermatitis by controlling effector T cell retention in inflamed skin." In: *Journal of Experimental Medicine* 211.6, pp. 1185–1196. DOI: 10.1084/jem.20121350.

Stegemann, Agatha and Markus Böhm (2020). "Targeting the α7 nicotinic acetylcholine receptor–A novel road towards the future treatment of skin diseases." In: *Experimental Dermatology* 29.9, pp. 924–931. DOI: 10.1111/exd.14173.

Stegle, Oliver, Sarah A. Teichmann, and John C. Marioni (2015). "Computational and analytical challenges in single-cell transcriptomics." In: *Nature Reviews Genetics* 16.3, 133–145. DOI: 10.1038/nrg3833.

Steiglitz, Barry M., Douglas R. Keene, and Daniel S. Greenspan (2002). "PCOLCE2 Encodes a Functional Procollagen C-Proteinase Enhancer (PCPE2) That Is a Collagen-binding Protein Differing in Distribution of Expression and Post-translational Modification from the Previously Described PCPE1." In: *Journal of Biological Chemistry* 277.51, pp. 49820–49830. DOI: 10.1074/jbc.m209891200.

Steinhoff, Julia S., Achim Lass, and Michael Schupp (2021). "Biological Functions of RBP4 and Its Relevance for Human Diseases." In: *Frontiers in Physiology* 12. DOI: 10.3389/fphys.2021.659977.

Steinle, Alexander et al. (2009). "CLEC2A Is a Novel Skin-Specific, Stimulatory Ligand of Human NK Cells (39.8)." In: *The Journal of Immunology* 182.1_Supplement, pp. 39.8–39.8. DOI: 10.4049/jimmunol.182.supp.39.8.

Stepp, Mary Ann et al. (2015). "Syndecan-1 and Its Expanding List of Contacts." In: *Advances in Wound Care* 4.4, pp. 235–249. DOI: 10.1089/wound.2014.0555.

Stevison, Faith et al. (2015). "Role of Retinoic Acid-Metabolizing Cytochrome P450s, CYP26, in Inflammation and Cancer." In: *Cytochrome P450 Function and Pharmacological Roles in Inflammation and Cancer*. Elsevier, pp. 373–412. DOI: 10.1016/bs.apha.2015.04.006.

Stickels, Robert R. et al. (2020). "Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-SEQV2." In: *Nature Biotechnology* 39.3, 313–319. DOI: 10.1038/s41587-020-0739-1.

Stiles, Ashlee R. et al. (2009). "CYP7B1: One cytochrome P450, two human genetic diseases, and multiple physiological functions." In: *Journal of Biological Chemistry* 284.42, 28485–28489. DOI: 10.1074/jbc.r109.042168.

Stoeckius, Marlon et al. (2017). "Simultaneous epitope and transcriptome measurement in single cells." In: *Nature Methods* 14.9, 865–868. DOI: 10.1038/nmeth.4380.

Stoica, Gerald E. et al. (2002). "Midkine Binds to Anaplastic Lymphoma Kinase (ALK) and Acts as a Growth Factor for Different Cell Types." In: *Journal of Biological Chemistry* 277.39, pp. 35990–35998. DOI: 10.1074/jbc.m205749200.

Stolt, C. Claus et al. (2004). "Transcription factors Sox8 and Sox10 perform non-equivalent roles during oligodendrocyte development despite functional redundancy." In: *Development* 131.10, pp. 2349–2358. DOI: 10.1242/dev.01114.

Street, Kelly et al. (2018). "Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics." In: *BMC Genomics* 19.1. DOI: 10.1186/s12864-018-4772-0.

Strutz, F et al. (1995). "Identification and characterization of a fibroblast marker: FSP1." In: *Journal of Cell Biology* 130.2, 393–405. DOI: 10.1083/jcb.130.2.393.

Stuart, Tim and Rahul Satija (2019a). "Integrative single-cell analysis." In: *Nature Reviews Genetics* 20.5, 257–272. DOI: 10.1038/s41576-019-0093-7.

Stuart, Tim et al. (2019b). "Comprehensive Integration of Single-Cell Data." In: *Cell* 177.7, pp. 1888–1902. DOI: 10.1016/j.cell.2019.05.031.

Stunova, Anna and Lucie Vistejnova (2018). "Dermal fibroblasts—a heterogeneous population with regulatory function in wound healing." In: *Cytokine & Growth Factor Reviews* 39, 137–150. DOI: 10.1016/j.cytogfr.2018.01.003.

Stähli, Alexandra et al. (2016). "Collagen Membranes Adsorb the Transforming Growth Factor-β Receptor I Kinase-Dependent Activity of Enamel Matrix Derivative." In: *Journal of Periodontology* 87.5, pp. 583–590. DOI: 10.1902/jop.2016.150538.

Su, Zao-Zhong et al. (2007). "Central role of interferon regulatory factor-1 (IRF-1) in controlling retinoic acid inducible gene-I (RIG-I) expression." In: *Journal of Cellular Physiology* 213.2, pp. 502–510. DOI: 10.1002/jcp.21128.

Sun, Bill S. et al. (1998). "Identification of a protein isolated from senescent human cells that binds to hepatitis B virus X antigen." In: *Hepatology* 27.1, pp. 228–239. DOI: 10.1002/hep.510270135.

Sun, Ding-Ping et al. (2013). "Interleukin (IL)-19 promoted skin wound healing by increasing fibroblast keratinocyte growth factor expression." In: *Cytokine* 62.3, 360–368. DOI: 10.1016/j.cyto.2013.03.017.

Sun, Hongxian et al. (2022). "Lef1 and Dlx3 May Facilitate the Maturation of Secondary Hair Follicles in the Skin of Gansu Alpine Merino." In: *Genes* 13.8, p. 1326. DOI: 10.3390/genes13081326.

Sun, Mei et al. (2020). "Collagen XI regulates the acquisition of collagen fibril structure, organization and functional properties in tendon." In: *Matrix Biology* 94, pp. 77–94. DOI: 10.1016/j.matbio.2020.09.001.

Sun, Shiquan et al. (2019). "Accuracy, robustness and scalability of dimensionality reduction methods for single-cell RNA-seq analysis." In: *Genome Biology* 20.1. DOI: 10.1186/s13059-019-1898-6.

Suo, Liye et al. (2021). "Dietary Vitamin A Impacts Refractory Telogen." In: *Frontiers in Cell and Developmental Biology* 9. DOI: 10.3389/fcell.2021.571474.

Suomela, Sari et al. (2004). "Interferon α-inducible protein 27 (IFI27) is upregulated in psoriatic skin and certain epithelial cancers." In: *Journal of Investigative Dermatology* 122.3, 717–721. DOI: 10.1111/j.0022-202x.2004.22322.x.

Supuran, Claudiu (2008). "Carbonic Anhydrases An Overview." In: *Current Pharmaceutical Design* 14.7, pp. 603–614. DOI: 10.2174/138161208783877884.

Surowiak, Pawel et al. (2014). "Increase in cyclooxygenase-2 (COX-2) expression in keratinocytes and dermal fibroblasts in photoaged skin." In: *Journal of Cosmetic Dermatology* 13.3, pp. 195–201. DOI: 10.1111/jocd.12103.

Sussmann, M. et al. (2004). "Induction of Hyaluronic Acid Synthase 2 (HAS2) in Human Vascular Smooth Muscle Cells by Vasodilatory Prostaglandins." In: *Circulation Research* 94.5, pp. 592–600. DOI: 10.1161/01.res.0000119169.87429.a0.

Suter, David M. et al. (2011). "Mammalian genes are transcribed with widely different bursting kinetics." In: *Science* 332.6028, 472–474. DOI: 10.1126/science.1198817.

Sutmuller, M., J. A. Bruijn, and E de Heer (1997). "Collagen types VIII and X, two non-fibrillar, short-chain collagens. Structure homologies, functions and involvement in pathology." In: *Histology and histopathology* 12.2, 557–566.

Svensson, Valentine (2020). "Droplet scRNA-seq is not zero-inflated." In: *Nature Biotechnology* 38.2, 147–150. DOI: 10.1038/s41587-019-0379-5.

Svensson, Valentine, Eduardo da Veiga Beltrame, and Lior Pachter (2020). "A curated database reveals trends in single-cell transcriptomics." In: *Database* 2020. DOI: 10.1093/database/baaa073.

Svensson, Valentine, Roser Vento-Tormo, and Sarah A Teichmann (2018). "Exponential scaling of single-cell RNA-seq in the past decade." In: *Nature Protocols* 13.4, 599–604. DOI: 10.1038/nprot.2017.149.

Svensson, Valentine et al. (2017). "Power analysis of single-cell RNA-sequencing experiments." In: *Nature Methods* 14.4, 381–387. DOI: 10.1038/nmeth.4220.

Szatmari, Istvan et al. (2004). "Activation of PPARγ Specifies a Dendritic Cell Subtype Capable of Enhanced Induction of iNKT Cell Expansion." In: *Immunity* 21.1, pp. 95–106. DOI: 10.1016/j.immuni.2004.06.003.

Szatmari, Istvan et al. (2006). "PPARγ controls CD1d expression by turning on retinoic acid synthesis in developing human dendritic cells." In: *Journal of Cell Biology* 175.1, pp. i1–i1. DOI: 10.1083/jcb1751oia1.

Szenker-Ravi, Emmanuelle et al. (2018). "RSPO2 inhibition of RNF43 and ZNRF3 governs limb development independently of LGR4/5/6." In: *Nature* 557.7706, pp. 564–569. DOI: 10.1038/s41586-018-0118-y.

T. Sapiens Cons. and Stephen R Quake (2022). "The tabula sapiens: A multiple organ single cell transcriptomic atlas of humans." In: *Science* 376.6594. DOI: 10.1101/2021.07.19.452956.

Tabib, Tracy et al. (2018). "SFRP2/DPP4 and FMO1/LSP1 Define Major Fibroblast Populations in Human Skin." In: *Journal of Investigative Dermatology* 138.4, 802–810. DOI: 10.1016/j.jid.2017.09.045.

Tabib, Tracy et al. (2021). "Myofibroblast transcriptome indicates sfrp2hi fibroblast progenitors in systemic sclerosis skin." In: *Nature Communications* 12.1. DOI: 10.1038/s41467-021-24607-6.

Tabula Muris Consortium (2018). "Single-cell transcriptomics of 20 mouse organs creates a tabula muris." In:

Nature 562.7727, 367–372. DOI: 10.1038/s41586-018-0590-4.

Tadevosyan, Artavazd et al. (2017). "Intracellular Angiotensin-II Interacts With Nuclear Angiotensin Receptors in Cardiac Fibroblasts and Regulates RNA Synthesis, Cell Proliferation, and Collagen Secretion." In: *Journal of the American Heart Association* 6.4. DOI: 10.1161/jaha.116.004965.

Tai, Tzong-Shyuan, Sung-Yun Pai, and I-Cheng Ho (2014). "Itm2a, a Target Gene of GATA-3, Plays a Minimal Role in Regulating the Development and Function of T Cells." In: *PLoS ONE* 9.5. Ed. by Troy A. Baldwin, e96535. DOI: 10.1371/journal.pone.0096535.

Tai, Yifan et al. (2021). "Myofibroblasts: Function, formation, and scope of molecular therapies for skin fibrosis." In: *Biomolecules* 11.8, p. 1095. DOI: 10.3390/biom11081095.

Takahara, K et al. (1994). "Type I procollagen COOH-terminal proteinase enhancer protein: identification, primary structure, and chromosomal localization of the cognate human gene (PCOLCE)." In: *Journal of Biological Chemistry* 269.42, pp. 26280–26285. DOI: 10.1016/s0021-9258(18)47191-8.

Takahashi, Michiko et al. (2008). "Chemerin enhances insulin signaling and potentiates insulin-stimulated glucose uptake in 3T3-L1 adipocytes." In: *FEBS Letters* 582.5, pp. 573–578. DOI: 10.1016/j.febslet.2008.01.023.

Takahashi, Rie et al. (2020). "Defining transcriptional signatures of human hair follicle cell states." In: *Journal of Investigative Dermatology* 140.4. DOI: 10.1016/j.jid.2019.07.726.

Takashima, A et al. (1995). "Colony-stimulating factor-1 secreted by fibroblasts promotes the growth of dendritic cell lines (XS series) derived from murine epidermis." In: *The Journal of Immunology* 154.10, 5128–5135. DOI: 10.4049/jimmunol.154.10.5128.

Takaya, Kento, Toru Asou, and Kazuo Kishi (2023). "Identification of Apolipoprotein D as a Dermal Fibroblast Marker of Human Aging for Development of Skin Rejuvenation Therapy." In: *Rejuvenation Research*. DOI: 10.1089/rej.2022.0056.

Takechi, Tomoki et al. (2021). "Effect of Genetic Polymorphisms of Human SLC22A3 in the 5'-flanking Region on oct3 Expression and Sebum Levels in Human Skin." In: *Journal of Dermatological Science* 101.1, pp. 4–13. DOI: 10.1016/j.jdermsci.2020.10.010.

Taketomi, Yoshitaka et al. (2013). "Mast cell maturation is driven via a group III phospholipase A2-prostaglandin D2–DP1 receptor paracrine axis." In: *Nature Immunology* 14.6, pp. 554–563. DOI: 10.1038/ni.2586.

Takouda, Jun et al. (2021). "SoxE group transcription factor Sox8 promotes astrocytic differentiation of neural stem/precursor cells downstream of Nfia." In: *Pharmacology Research & Perspectives* 9.6. DOI: 10.1002/prp2.749.

Tall, Alan R. and Laurent Yvan-Charvet (2015). "Cholesterol, inflammation and innate immunity." In: *Nature Reviews Immunology* 15.2, pp. 104–116. DOI: 10.1038/nri3793.

Tamura, Kazuhiro et al. (2009). "Insulin-like growth factor binding protein-7 (IGFBP7) blocks vascular endothelial cell growth factor (VEGF)-induced angiogenesis in human vascular endothelial cells." In: *European Journal of Pharmacology* 610.1-3, pp. 61–67. DOI: 10.1016/j.ejphar.2009.01.045.

Tang, Fuchou et al. (2009). "mRNA-seq whole-transcriptome analysis of a single cell." In: *Nature Methods* 6.5, 377–382. DOI: 10.1038/nmeth.1315.

Tang, Fuchou et al. (2010a). "Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-seq analysis." In: *Cell Stem Cell* 6.5, 468–478. DOI: 10.1016/j.stem.2010.03.015.

Tang, Yixin et al. (2010b). "CYP1B1 and endothelial nitric oxide synthase combine to sustain proangiogenic functions of endothelial cells under hyperoxic stress." In: *American Journal of Physiology-Cell Physiology* 298.3, pp. C665–C678. DOI: 10.1152/ajpcell.00153.2009.

Tavolaro, F. M. et al. (2015). "Photoperiodic Effects on Seasonal Physiology, Reproductive Status and Hypothalamic Gene Expression in Young Male F344 Rats." In: *Journal of Neuroendocrinology* 27.2, pp. 79–87. DOI: 10.1111/jne.12241.

Termine, John D. et al. (1981). "Osteonectin, a bone-specific protein linking mineral to collagen." In: *Cell* 26.1, pp. 99–105. DOI: 10.1016/0092-8674(81)90037-4.

Thannickal, Victor J. (2013). "Mechanistic links between aging and lung fibrosis." In: *Biogerontology* 14.6, 609–615. DOI: 10.1007/s10522-013-9451-6.

Thelen, Marcus and Mariagrazia Uguccioni (2016). "Function of Chemokines and Their Receptors in Immunity." In: *Encyclopedia of Immunobiology*. Elsevier, pp. 572–578. DOI: 10.1016/b978-0-12-374279-7.10008-6.

Theocharidis, Georgios and John T. Connelly (2017). "Minor collagens of the skin with not so minor functions." In: *Journal of Anatomy* 235.2, pp. 418–429. DOI: 10.1111/joa.12584.

Theocharidis, Georgios et al. (2020). "Integrated skin transcriptomics and serum multiplex assays reveal novel mechanisms of wound healing in diabetic foot ulcers." In: *Diabetes* 69.10, 2157–2169. DOI: 10.2337/db20-0188.

Theocharidis, Georgios et al. (2022). "Single cell transcriptomic landscape of diabetic foot ulcers." In: *Nature Communications* 13.1. DOI: 10.1038/s41467-021-27801-8.

Thomi, R. et al. (2017). "Interleukin-32 is highly expressed in lesions of hidradenitis suppurativa." In: *British Journal of Dermatology* 177.5, pp. 1358–1366. DOI: 10.1111/bjd.15458.

Thrane, Kim et al. (2023). "Single-Cell and Spatial Transcriptomic Analysis of Human Skin Delineates Intercellular Communication and Pathogenic Cells." In: *Journal of In-*

vestigative Dermatology. DOI: 10.1016/j.jid.2023.02.040.

Tiainen, PÃ¤ivi et al. (2008). "Characterization of Recombinant Human Prolyl 3-Hydroxylase Isoenzyme 2, an Enzyme Modifying the Basement Membrane Collagen IV." In: *Journal of Biological Chemistry* 283.28, pp. 19432–19439. DOI: 10.1074/jbc.m802973200.

Tian, Xufan, Dong Soo Kang, and Jeffrey L. Benovic (2013). "β-Arrestins and G Protein-Coupled Receptor Trafficking." In: *Arrestins - Pharmacology and Therapeutic Potential*. Springer Berlin Heidelberg, pp. 173–186. DOI: 10.1007/978-3-642-41199-1_9.

Tikhonova, Anastasia N. et al. (2019). "The bone marrow microenvironment at single-cell resolution." In: *Nature* 569.7755, 222–228. DOI: 10.1038/s41586-019-1104-8.

Tillgren, Viveka et al. (2015). "The Novel Small Leucine-rich Protein Chondroadherin-like (CHADL) Is Expressed in Cartilage and Modulates Chondrocyte Differentiation." In: *Journal of Biological Chemistry* 290.2, pp. 918–925. DOI: 10.1074/jbc.m114.593541.

To, Sarah and Sandeep K. Agarwal (2019). "Macrophages and cadherins in fibrosis and systemic sclerosis." In: *Current Opinion in Rheumatology* 31.6, 582–588. DOI: 10.1097/bor.0000000000000657.

Toma, Jean G. et al. (2001). "Isolation of multipotent adult stem cells from the dermis of mammalian skin." In: *Nature Cell Biology* 3.9, 778–784. DOI: 10.1038/ncb0901-778.

Tomlinson, Matthew J et al. (2013). "Cell separation: Terminology and practical considerations." In: *Journal of Tissue Engineering* 4. DOI: 10.1177/2041731412472690.

Tomoeda, M. et al. (2008). "PLAP-1/asporin inhibits activation of BMP receptor via its leucine-rich repeat motif." In: *Biochemical and Biophysical Research Communications* 371.2, pp. 191–196. DOI: 10.1016/j.bbrc.2008.03.158.

Tong, Xuemei and Pierre A. Coulombe (2006). "Keratin 17 modulates hair follicle cycling in a TNFα-dependent fashion." In: *Genes & Development* 20.10, pp. 1353–1364. DOI: 10.1101/gad.1387406.

Tontonoz, Peter et al. (1998). "PPARγ Promotes Monocyte/-Macrophage Differentiation and Uptake of Oxidized LDL." In: *Cell* 93.2, pp. 241–252. DOI: 10.1016/s0092-8674(00)81575-5.

Toricelli, Mariana et al. (2013). "Timp1 interacts with beta-1 integrin and CD63 along melanoma genesis and confers anoikis resistance by activating PI3-K signaling pathway independently of Akt phosphorylation." In: *Molecular Cancer* 12.1. DOI: 10.1186/1476-4598-12-22.

Townes, F. William et al. (2019). "Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model." In: *Genome Biology* 20.1. DOI: 10.1186/s13059-019-1861-6.

Traag, V. A., L. Waltman, and N. J. van Eck (2019). "From Louvain to Leiden: Guaranteeing well-connected communities." In: *Scientific Reports* 9.1. DOI: 10.1038/s41598-019-41695-z.

Tran, Hoa Thi et al. (2020). "A benchmark of batch-effect correction methods for single-cell RNA sequencing data." In: *Genome Biology* 21.1. DOI: 10.1186/s13059-019-1850-9.

Trapnell, Cole et al. (2014). "The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells." In: *Nature Biotechnology* 32.4, 381–386. DOI: 10.1038/nbt.2859.

Treiber, Nicolai et al. (2012). "The role of manganese superoxide dismutase in skin aging." In: *Dermato-Endocrinology* 4.3, pp. 232–235. DOI: 10.4161/derm.21819.

Tremble, P, R Chiquet-Ehrismann, and Z Werb (1994). "The extracellular matrix ligands fibronectin and tenascin collaborate in regulating collagenase gene expression in fibroblasts." In: *Molecular Biology of the Cell* 5.4, pp. 439–453. DOI: 10.1091/mbc.5.4.439.

Tritschler, Sophie et al. (2019). "Concepts and limitations for learning developmental trajectories from Single Cell Genomics." In: *Development* 146.12. DOI: 10.1242/dev.170506.

Tsitsipatis, Dimitrios et al. (2022). "Proteomes of primary skin fibroblasts from healthy individuals reveal altered cell responses across the life span." In: *Aging Cell* 21.5. DOI: 10.1111/acel.13609.

Tsukui, Tatsuya et al. (2020). "Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis." In: *Nature Communications* 11.1. DOI: 10.1038/s41467-020-15647-5.

Tsuneki, Masayuki and Joseph A. Madri (2015). "CD44 Influences Fibroblast Behaviors Via Modulation of Cell-Cell and Cell-Matrix Interactions, Affecting Survivin and Hippo Pathways." In: *Journal of Cellular Physiology* 231.3, pp. 731–743. DOI: 10.1002/jcp.25123.

Tsutsui, Ko et al. (2021). "Mapping the molecular and structural specialization of the skin basement membrane for inter-tissue interactions." In: *Nature Communications* 12.1. DOI: 10.1038/s41467-021-22881-y.

Tsuyuzaki, Koki et al. (2020). "Benchmarking principal component analysis for large-scale single-cell RNA-sequencing." In: *Genome Biology* 21.1. DOI: 10.1186/s13059-019-1900-3.

Tuckey, John W (1958). "Bias and confidence in not-quite large samples." In: *The Annals of Mathematical Statistics*. 6th ser. 29.2, 614–623. DOI: 10.1214/aoms/1177706647.

Tullai, John W. et al. (2007). "Immediate-Early and Delayed Primary Response Genes Are Distinct in Function and Genomic Architecture." In: *Journal of Biological Chemistry* 282.33, pp. 23981–23995. DOI: 10.1074/jbc.m702044200.

Turnescu, Tanja et al. (2017). "Sox8 and Sox10 jointly maintain myelin gene expression in oligodendrocytes." In: *Glia* 66.2, pp. 279–294. DOI: 10.1002/glia.23242.

Uhlén, Mathias et al. (2015). "Tissue-based map of the human Proteome." In: *Science* 347.6220. DOI: 10.1126/science.1260419.

Ulyanov, Dmitry (2016). *Multicore-TSNE*. https://github.com/DmitryUlyanov/Multicore-TSNE.

Valcourt, Ulrich et al. (2015). "Tenascin-X: beyond the architectural function." In: *Cell Adhesion & Migration* 9.1-2, pp. 154–165. DOI: 10.4161/19336918.2014.994893.

Valenzi, Eleanor et al. (2019). "Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease." In: *Annals of the Rheumatic Diseases* 78.10, pp. 1379–1387. DOI: 10.1136/annrheumdis-2018-214865.

Valiulytė et al. (2019). "The Anti-Tumorigenic Activity of Sema3C in the Chick Embryo Chorioallantoic Membrane Model." In: *International Journal of Molecular Sciences* 20.22, p. 5672. DOI: 10.3390/ijms20225672.

Vallbo, Å. B., H. Olausson, and J. Wessberg (1999). "Unmyelinated afferents constitute a second system coding tactile stimuli of the human hairy skin." In: *Journal of Neurophysiology* 81.6, 2753–2763. DOI: 10.1152/jn.1999.81.6.2753.

Van Linthout, S., K. Miteva, and C. Tschope (2014). "Crosstalk between fibroblasts and inflammatory cells." In: *Cardiovascular Research* 102.2, 258–269. DOI: 10.1093/cvr/cvu062.

VanBuren, Christine A. and Helen B. Everts (2022). "Vitamin A in Skin and Hair: An Update." In: *Nutrients* 14.14, p. 2952. DOI: 10.3390/nu14142952.

Vanbervliet, Beatrice et al. (2002). "Sequential involvement of CCR2 and CCR6 ligands for immature dendritic cell recruitment: possible role at inflamed epithelial surfaces." In: *European Journal of Immunology* 32.1, pp. 231–242. DOI: 10.1002/1521-4141(200201)32:1<231::aid-immu231>3.0.co;2-8.

Vandenbon, Alexis and Diego Diez (2020). "A clustering-independent method for finding differentially expressed genes in single-cell transcriptome data." In: *Nature Communications* 11.1. DOI: 10.1038/s41467-020-17900-3.

Vanderaa, Christophe and Laurent Gatto (2021). "Replication of single-cell proteomics data reveals important computational challenges." In: *bioRxiv*. DOI: 10.1101/2021.04.12.439408.

Vargas, Leonardo et al. (2002). "Functional Interaction of Caveolin-1 with Bruton's Tyrosine Kinase and Bmx." In: *Journal of Biological Chemistry* 277.11, pp. 9351–9357. DOI: 10.1074/jbc.m108537200.

Varlet, Béatrice Le et al. (1991). "Human Epidermal Langerhans Cells Express Integrins of the β1 Subfamily." In: *Journal of Investigative Dermatology* 96.4, pp. 518–522. DOI: 10.1111/1523-1747.ep12470229.

Venkatachalam, Kaliyamurthi et al. (2009). "WISP1, a Pro-mitogenic, Pro-survival Factor, Mediates Tumor Necrosis Factor-α (TNF-α)-stimulated Cardiac Fibroblast Proliferation but Inhibits TNF-α-induced Cardiomyocyte Death." In: *Journal of Biological Chemistry* 284.21, pp. 14414–14427. DOI: 10.1074/jbc.m809757200.

Verzijl, Nicole et al. (2000). "Effect of collagen turnover on the accumulation of advanced glycation end products." In: *Journal of Biological Chemistry* 275.50, 39027–39031. DOI: 10.1074/jbc.m006700200.

Vialou, Vincent et al. (2004). "Organic Cation Transporter 3 (Slc22a3) Is Implicated in Salt-Intake Regulation." In: *The Journal of Neuroscience* 24.11, pp. 2846–2851. DOI: 10.1523/jneurosci.5147-03.2004.

Vickovic, Sanja et al. (2019). "High-definition spatial transcriptomics for in situ tissue profiling." In: *Nature Methods* 16.10, 987–990. DOI: 10.1038/s41592-019-0548-y.

Viegas, Carla S. B. et al. (2017). "Gla-rich protein function as an anti-inflammatory agent in monocytes/-macrophages: Implications for calcification-related chronic inflammatory diseases." In: *PLOS ONE* 12.5. Ed. by Pablo Garcia de Frutos, e0177829. DOI: 10.1371/journal.pone.0177829.

Vieth, Beate et al. (2017). "powsimR: Power analysis for bulk and single-cell RNA-seq experiments." In: *Bioinformatics* 33.21, 3486–3488. DOI: 10.1093/bioinformatics/btx435.

Villani, Rehan et al. (2017). "Dominant-negative Sox18 function inhibits dermal papilla maturation and differentiation in all murine hair types." In: *Development* 144.10, pp. 1887–1895. DOI: 10.1242/dev.143917.

Villapol, Sonia, Trevor T. Logan, and Aviva J. Symes (2013). "Role of TGF-β Signaling in Neurogenic Regions After Brain Injury." In: *Trends in Cell Signaling Pathways in Neuronal Fate Decision*. Ed. by Sabine Wislet-Gendebien. Rijeka: IntechOpen. Chap. 1. DOI: 10.5772/53941.

Villard, Pierre H et al. (2002). "Increase of CYP1B1 Transcription in Human Keratinocytes and HaCaT Cells after UV-B Exposure." In: *Toxicology and Applied Pharmacology* 178.3, pp. 137–143. DOI: 10.1006/taap.2001.9335.

Vincze, Orsolya et al. (2006). "Tubulin Polymerization Promoting Proteins (TPPPs): Members of a New Family with Distinct Structures and Functions." In: *Biochemistry* 45.46, pp. 13818–13826. DOI: 10.1021/bi061305e.

Virchow, Rudolf (1859). *Die Cellularpathologie in ihrer Begründung auf physiologische und pathologische Gewebelehre*. Berlin : Verlag von August Hirschwald.

Vistejnova, Lucie et al. (2014). "Low molecular weight hyaluronan mediated CD44 dependent induction of IL-6 and chemokines in human dermal fibroblasts potentiates innate immune response." In: *Cytokine* 70.2, pp. 97–103. DOI: 10.1016/j.cyto.2014.07.006.

Viswanathan, Sivakumar et al. (Sept. 2021). "Critical Conditions for Studying Interleukin-11 Signaling In Vitro and Avoiding Experimental Artefacts." In: *Current Protocols* 1.9. DOI: 10.1002/cpz1.251.

Vogel, Christoph F.A. and Thomas Haarmann-Stemmann (2017). "The aryl hydrocarbon receptor repressor – More

than a simple feedback inhibitor of AhR signaling: Clues for its role in inflammation and cancer." In: *Current Opinion in Toxicology* 2, pp. 109–119. DOI: 10.1016/j.cotox.2017.02.004.

Vollmar, Johanna et al. (2019). "Deletion of organic cation transporter Oct3 promotes hepatic fibrosis via upregulation of TGFβ." In: *American Journal of Physiology-Gastrointestinal and Liver Physiology* 317.2, G195–G202. DOI: 10.1152/ajpgi.00088.2019.

Vorstandlechner, Vera et al. (2020). "Deciphering the functional heterogeneity of skin fibroblasts using single-cell RNA sequencing." In: *The FASEB Journal* 34.3, 3677–3692. DOI: 10.1096/fj.201902001rr.

Vorstandlechner, Vera et al. (2021). "The serine proteases dipeptidyl-peptidase 4 and urokinase are key molecules in human and Mouse Scar Formation." In: *Nature Communications* 12.1. DOI: 10.1038/s41467-021-26495-2.

Wagner, Florian (2020a). "Monet: An open-source Python package for analyzing and integrating scRNA-Seq data using PCA-based latent spaces." In: *bioRxiv*. DOI: 10.1101/2020.06.08.140673. eprint: https://www.biorxiv.org/content/early/2020/06/10/2020.06.08.140673.full.pdf.

– (2020b). "Straightforward clustering of single-cell RNA-Seq data with t-SNE and DBSCAN." In: *bioRxiv*. DOI: 10.1101/770388. eprint: https://www.biorxiv.org/content/early/2020/04/09/770388.full.pdf.

Wagner, Florian, Yun Yan, and Itai Yanai (2017). "K-nearest neighbor smoothing for high-throughput single-cell RNA-seq data." In: *bioRxiv*. DOI: 10.1101/217737.

Waise, Sara et al. (2019). "An optimised tissue disaggregation and data processing pipeline for characterising fibroblast phenotypes using single-cell RNA sequencing." In: *Scientific Reports* 9.1. DOI: 10.1038/s41598-019-45842-4.

Wajapeyee, Narendra et al. (2008). "Oncogenic BRAF Induces Senescence and Apoptosis through Pathways Mediated by the Secreted Protein IGFBP7." In: *Cell* 132.3, pp. 363–374. DOI: 10.1016/j.cell.2007.12.032.

Walko, Gernot, Maria J. Castañón, and Gerhard Wiche (2015). "Molecular architecture and function of the hemidesmosome." In: *Cell and Tissue Research* 360.3, pp. 529–544. DOI: 10.1007/s00441-015-2216-6.

Wang, Dan et al. (2021a). "CCL13 is upregulated in alopecia areata lesions and is correlated with disease severity." In: *Experimental Dermatology* 30.5, pp. 723–732. DOI: 10.1111/exd.14293.

Wang, Dingzhi, Jiqing Sai, and Ann Richmond (2003). "Cell Surface Heparan Sulfate Participates in CXCL1-Induced Signaling." In: *Biochemistry* 42.4, pp. 1071–1077. DOI: 10.1021/bi026425a.

Wang, Dongdong et al. (2021b). "GDF15: Emerging biology and therapeutic applications for obesity and cardiometabolic disease." In: *Nature Reviews Endocrinology* 17.10, 592–607. DOI: 10.1038/s41574-021-00529-7.

Wang, Hong-Jun et al. (2004). "Stimulation of skin repair is dependent on fibroblast source and presence of extracellular matrix." In: *Tissue Engineering* 10.7, 1054–1064. DOI: 10.1089/1076327041887709.

Wang, Jia et al. (2013). "VEGF expression is augmented by hypoxia-induced PGIS in human fibroblasts." In: *International Journal of Oncology* 43.3, pp. 746–754. DOI: 10.3892/ijo.2013.1994.

Wang, JianFei et al. (2008). "Deep dermal fibroblasts contribute to hypertrophic scarring." In: *Laboratory Investigation* 88.12, 1278–1290. DOI: 10.1038/labinvest.2008.101.

Wang, Jing et al. (2001). "The Critical Role of LIGHT, a TNF Family Member, in T Cell Development." In: *The Journal of Immunology* 167.9, pp. 5099–5105. DOI: 10.4049/jimmunol.167.9.5099.

Wang, Ke-Yong et al. (2018a). "Critical in vivo roles of WNT10A in wound healing by regulating collagen expression/synthesis in WNT10A-deficient mice." In: *PLOS ONE* 13.3. Ed. by Robert W. Dettman, e0195156. DOI: 10.1371/journal.pone.0195156.

Wang, Renying et al. (2022). "Construction of a cross-species cell landscape at single-cell level." In: *Nucleic Acids Research*. DOI: 10.1093/nar/gkac633.

Wang, Sheila C. et al. (2015a). "Tumour endothelial marker-8 in wound healing and its impact on the proliferation and migration of keratinocytes." In: *International Journal of Molecular Medicine* 37.2, pp. 293–298. DOI: 10.3892/ijmm.2015.2434.

Wang, Shuxiong et al. (2020). "Single cell transcriptomics of human epidermis identifies basal stem cell transition states." In: *Nature Communications* 11.1. DOI: 10.1038/s41467-020-18075-7.

Wang, Wan et al. (2017). "SOD2 Facilitates the Antiviral Innate Immune Response by Scavenging Reactive Oxygen Species." In: *Viral Immunology* 30.8, pp. 582–589. DOI: 10.1089/vim.2017.0043.

Wang, Xiao et al. (2018b). "Three-dimensional intact-tissue sequencing of single-cell transcriptional states." In: *Science* 361.6400. DOI: 10.1126/science.aat5691.

Wang, Yaming et al. (2015b). "Nonredundant roles of keratinocyte-derived IL-34 and neutrophil-derived CSF1 in Langerhans cell renewal in the steady state and during inflammation." In: *European Journal of Immunology* 46.3, pp. 552–559. DOI: 10.1002/eji.201545917.

Wang, Yan, Judith A. Mack, and Edward V. Maytin (2019). "CD44 inhibits α-SMA gene expression via a novel G-actin/MRTF-mediated pathway that intersects with TGFβR/p38MAPK signaling in murine skin fibroblasts." In: *Journal of Biological Chemistry* 294.34, pp. 12779–12794. DOI: 10.1074/jbc.ra119.007834.

Wang, Yan et al. (2014). "Hyaluronan Synthase 2 Protects Skin Fibroblasts against Apoptosis Induced by Environmental Stress." In: *Journal of Biological Chemistry* 289.46, pp. 32253–32265. DOI: 10.1074/jbc.m114.578377.

Warrier, Manya et al. (2015). "The TMAO-Generating Enzyme Flavin Monooxygenase 3 Is a Central Regulator of Cholesterol Balance." In: *Cell Reports* 10.3, pp. 326–338. DOI: 10.1016/j.celrep.2014.12.036.

Watanabe, Eiji et al. (2000). "Nav2/NaG Channel Is Involved in Control of Salt-Intake Behavior in the CNS." In: *The Journal of Neuroscience* 20.20, pp. 7743–7751. DOI: 10.1523/jneurosci.20-20-07743.2000.

Watson, J. D. and F. H. Crick (1953). "Molecular structure of Nucleic Acids: A structure for deoxyribose nucleic acid." In: *Nature* 171.4356, 737–738. DOI: 10.1038/171737a0.

Weckbach, Ludwig T. et al. (2012). "Midkine acts as proangiogenic cytokine in hypoxia-induced angiogenesis." In: *American Journal of Physiology-Heart and Circulatory Physiology* 303.4, H429–H438. DOI: 10.1152/ajpheart.00934.2011.

Weckbach, Ludwig T. et al. (2014). "The cytokine midkine supports neutrophil trafficking during acute inflammation by promoting adhesion via β2 integrins (CD11/CD18)." In: *Blood* 123.12, pp. 1887–1896. DOI: 10.1182/blood-2013-06-510875.

Weddell, G. (1945). "The anatomy of cutaneous sensibility." In: *British Medical Bulletin* 3.7-8, 167–172. DOI: 10.1093/oxfordjournals.bmb.a071903.

Wei, Jianlu (2014). "ADAMTS-18: A metalloproteinase with multiple functions." In: *Frontiers in Bioscience* 19.8, p. 1456. DOI: 10.2741/4296.

Weinbaum, Justin S. et al. (2008). "Deficiency in Microfibril-associated Glycoprotein-1 Leads to Complex Phenotypes in Multiple Organ Systems." In: *Journal of Biological Chemistry* 283.37, pp. 25533–25543. DOI: 10.1074/jbc.m709962200.

West, Nathaniel R. (2019). "Coordination of Immune-Stroma Crosstalk by IL-6 Family Cytokines." In: *Frontiers in Immunology* 10. DOI: 10.3389/fimmu.2019.01093.

Weyemi, Urbain et al. (2012). "SOD2 deficiency promotes aging phenotypes in mouse skin." In: *Aging* 4.2, pp. 116–118. DOI: 10.18632/aging.100433.

Whittaker, Charles A. and Richard O. Hynes (2002). "Distribution and Evolution of von Willebrand/Integrin A Domains: Widely Dispersed Domains with Roles in Cell Adhesion and Elsewhere." In: *Molecular Biology of the Cell* 13.10. Ed. by Thomas D. Pollard, pp. 3369–3387. DOI: 10.1091/mbc.e02-05-0259.

Widjaja, Anissa A. et al. (2022). "IL11 Stimulates IL33 Expression and Proinflammatory Fibroblast Activation across Tissues." In: *International Journal of Molecular Sciences* 23.16, p. 8900. DOI: 10.3390/ijms23168900.

Wilhelm, Dafne et al. (2023). "Tissue-specific collagen hydroxylation at GEP/GDP triplets mediated by P4HA2." In: DOI: 10.1101/2023.01.25.524868.

Wilhelm, Klaus-P. and Howard I. Maibach (1990). "Factors predisposing to cutaneous irritation." In: *Dermatologic Clinics* 8.1, 17–22. DOI: 10.1016/s0733-8635(18)30514-x.

Wilhelmsen, Kevin, Sandy H.M. Litjens, and Arnoud Sonnenberg (2006). "Multiple Functions of the Integrin α6β4 in Epidermal Homeostasis and Tumorigenesis." In: *Molecular and Cellular Biology* 26.8. PMID: 16581764, pp. 2877–2886. DOI: 10.1128/MCB.26.8.2877-2886.2006. eprint: https://doi.org/10.1128/MCB.26.8.2877-2886.2006.

Wilkinson, Holly N. et al. (2019). "Elevated local senescence in diabetic wound healing is linked to pathological repair via CXCR2." In: *Journal of Investigative Dermatology* 139.5. DOI: 10.1016/j.jid.2019.01.005.

Williams, Andrew, C. Donnelly, and J.A. Baugh (2022). "Macrophage Migration Inhibitory Factor (MIF)." In: *Encyclopedia of Respiratory Medicine*. Elsevier, pp. 384–388. DOI: 10.1016/b978-0-08-102723-3.00262-6.

Winnemöller, M., G. Schmidt, and H. Kresse (1991). "Influence of decorin on fibroblast adhesion to fibronectin." In: *European Journal of Cell Biology* 54.1, 10–7.

Wipff, J. et al. (2010). "Dermal tissue and cellular expression of fibrillin-1 in diffuse cutaneous systemic sclerosis." In: *Rheumatology* 49.4, pp. 657–661. DOI: 10.1093/rheumatology/kep433.

Wittamer, Valerie et al. (2003). "Specific Recruitment of Antigen-presenting Cells by Chemerin, a Novel Processed Ligand from Human Inflammatory Fluids." In: *Journal of Experimental Medicine* 198.7, pp. 977–985. DOI: 10.1084/jem.20030382.

Wolf, F. Alexander, Philipp Angerer, and Fabian J. Theis (2018). "Scanpy: Large-scale single-cell gene expression data analysis." In: *Genome Biology* 19.1. DOI: 10.1186/s13059-017-1382-0.

Wolf, F. Alexander et al. (2019). "PAGA: Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells." In: *Genome Biology* 20.1. DOI: 10.1186/s13059-019-1663-x.

Wolfe, Stephen L. (1985). *Introduction to cell biology*. Jin ming.

Wolock, Samuel L., Romain Lopez, and Allon M. Klein (2019a). "Scrublet: Computational identification of cell doublets in single-cell transcriptomic data." In: *Cell Systems* 8.4. DOI: 10.1016/j.cels.2018.11.005.

Wolock, Samuel L. et al. (2019b). "Mapping distinct bone marrow niche populations and their differentiation paths." In: *Cell Reports* 28.2. DOI: 10.1016/j.celrep.2019.06.031.

Woo, Jiwon, Wonhee Suh, and Jong-Hyuk Sung (2022). "Hair Growth Regulation by Fibroblast Growth Factor 12 (FGF12)." In: *International Journal of Molecular Sciences* 23.16, p. 9467. DOI: 10.3390/ijms23169467.

Woo, Seung-Hyun et al. (2015). "Piezo2 is the principal mechanotransduction channel for proprioception." In: *Nature Neuroscience* 18.12, pp. 1756–1762. DOI: 10.1038/nn.4162.

Worthen, Christal A. et al. (2020). "CD26 identifies a subpopulation of fibroblasts that produce the majority of collagen during wound healing in human skin." In: *Journal of Investigative Dermatology* 140.12. DOI: 10.1016/j.jid.2020.04.010.

Wu, Angela R et al. (2013). "Quantitative assessment of single-cell RNA-sequencing methods." In: *Nature Methods* 11.1, 41–46. DOI: 10.1038/nmeth.2694.

Wu, Jason, Amanda H. Lewis, and Jörg Grandl (2017). "Touch, Tension, and Transduction – The Function and Regulation of Piezo Ion Channels." In: *Trends in Biochemical Sciences* 42.1, pp. 57–71. DOI: 10.1016/j.tibs.2016.09.004.

Wu, Sijie et al. (2022). "Single-cell transcriptomics reveals lineage trajectory of human scalp hair follicle and informs mechanisms of hair graying." In: *Cell Discovery* 8.1. DOI: 10.1038/s41421-022-00394-2.

Wu, Zewen et al. (2021). "Fibroblast-like synoviocytes in rheumatoid arthritis: Surface markers and phenotypes." In: *International Immunopharmacology* 93, p. 107392. DOI: 10.1016/j.intimp.2021.107392.

Xiang, Ruizhi et al. (2021). "A comparison for dimensionality reduction methods of single-cell RNA-seq data." In: *Frontiers in Genetics* 12. DOI: 10.3389/fgene.2021.646936.

Xiao, Zhengtao, Ziwei Dai, and Jason W. Locasale (2019). "Metabolic landscape of the tumor microenvironment at single cell resolution." In: *Nature Communications* 10, p. 3763. DOI: 10.1038/s41467-019-11738-0.

Xie, Leike et al. (2011). "Caveolin-2 is a negative regulator of anti-proliferative function and signaling of transforming growth factor-β in endothelial cells." In: *American Journal of Physiology-Cell Physiology* 301.5, pp. C1161–C1174. DOI: 10.1152/ajpcell.00486.2010.

Xie, Ting et al. (2018). "Single-cell deconvolution of fibroblast heterogeneity in mouse pulmonary fibrosis." In: *Cell Reports* 22.13, 3625–3640. DOI: 10.1016/j.celrep.2018.03.010.

Xu, Hao et al. (2021a). "Pcpe2, a Novel Extracellular Matrix Protein, Regulates Adipocyte SR-BI–Mediated High-Density Lipoprotein Uptake." In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 41.11, pp. 2708–2725. DOI: 10.1161/atvbaha.121.316615.

Xu, Zijian et al. (2021b). "Anatomically distinct fibroblast subsets determine skin autoimmune patterns." In: *Nature* 601.7891, 118–124. DOI: 10.1038/s41586-021-04221-8.

Yamada, Haruyoshi et al. (1997). "Stimulation of collagen expression and glycosaminoglycan synthesis by midkine in human skin fibroblasts." In: *Archives of Dermatological Research* 289.7, pp. 429–433. DOI: 10.1007/s004030050216.

Yamaguchi, Yuji et al. (2005). "Mesenchymal–epithelial interactions in the skin: Aiming for site-specific tissue regeneration." In: *Journal of Dermatological Science* 40.1, 1–9. DOI: 10.1016/j.jdermsci.2005.04.006.

Yamaguchi, Yukie (2014). "Periostin in skin tissue skin-related diseases." In: *Allergology International* 63.2, 161–170. DOI: 10.2332/allergolint.13-rai-0685.

Yan, Min et al. (2004). "15-Hydroxyprostaglandin dehydrogenase, a COX-2 oncogene antagonist, is a TGF-β-induced suppressor of human gastrointestinal cancers." In: *Proceedings of the National Academy of Sciences* 101.50, pp. 17468–17473. DOI: 10.1073/pnas.0406142101.

Yan, Yan et al. (2019). "CCL19 and CCR7 Expression, Signaling Pathways, and Adjuvant Functions in Viral Infection and Prevention." In: *Frontiers in Cell and Developmental Biology* 7. DOI: 10.3389/fcell.2019.00212.

Yang, Hanseul et al. (2017). "Epithelial-mesenchymal microniches govern stem cell lineage choices." In: *Cell* 169.3. DOI: 10.1016/j.cell.2017.03.038.

Yang, Ruifeng et al. (2014). "Generation of folliculogenic human epithelial stem cells from induced pluripotent stem cells." In: *Nature Communications* 5.1. DOI: 10.1038/ncomms4071.

Yang, Zhigang et al. (2020). "CD49fhigh Defines a Distinct Skin Mesenchymal Stem Cell Population Capable of Hair Follicle Epithelial Cell Maintenance." In: *Journal of Investigative Dermatology* 140.3, 544–555.e9. DOI: 10.1016/j.jid.2019.08.442.

Yannas, Ioannis V. (2014). "Regeneration of skin." In: *Tissue and Organ Regeneration in Adults*, 89–136. DOI: 10.1007/978-1-4939-1865-2_5.

Yellin, Michael J. et al. (1995). "Ligation of CD40 on fibroblasts induces CD54 (ICAM-1) and CD106 (VCAM-1) upregulation and IL-6 production and proliferation." In: *Journal of Leukocyte Biology* 58.2, 209–216. DOI: 10.1002/jlb.58.2.209.

Yeo, I.-K. (2000). "A new family of power transformations to improve normality or symmetry." In: *Biometrika* 87.4, 954–959. DOI: 10.1093/biomet/87.4.954.

Yi, Li et al. (2017). "Inflammation-mediated SOD-2 upregulation contributes to epithelial-mesenchymal transition and migration of tumor cells in aflatoxin G1-induced lung adenocarcinoma." In: *Scientific Reports* 7.1. DOI: 10.1038/s41598-017-08537-2.

Yi, Tangsheng et al. (2012). "Oxysterol Gradient Generation by Lymphoid Stromal Cells Guides Activated B Cell Movement during Humoral Responses." In: *Immunity* 37.3, pp. 535–548. DOI: 10.1016/j.immuni.2012.06.015.

Yin, Yanlin, Weiwei Liu, and Yalei Dai (2015). "SOCS3 and its role in associated diseases." In: *Human Immunology* 76.10, pp. 775–780. DOI: 10.1016/j.humimm.2015.09.037.

Yokose, Urara et al. (2012). "The Endogenous Protease Inhibitor TIMP-1 Mediates Protection and Recovery from Cutaneous Photodamage." In: *Journal of Investigative*

Dermatology 132.12, pp. 2800–2809. DOI: 10.1038/jid.2012.204.

Yokota, Masaya et al. (2021). "Staphylococcus aureus impairs dermal fibroblast functions with deleterious effects on wound healing." In: *The FASEB Journal* 35.7. DOI: 10.1096/fj.201902836r.

Yon Rhee, Seung et al. (2008). "Use and misuse of the Gene Ontology Annotations." In: *Nature Reviews Genetics* 9.7, 509–515. DOI: 10.1038/nrg2363.

Yonezawa, Atsushi and Ken ichi Inui (2011). "Importance of the multidrug and toxin extrusion MATE/SLC47A family to pharmacokinetics, pharmacodynamics/toxicodynamics and pharmacogenomics." In: *British Journal of Pharmacology* 164.7, pp. 1817–1825. DOI: 10.1111/j.1476-5381.2011.01394.x.

Yoon, Sun-Young et al. (2014). "A role of placental growth factor in hair growth." In: *Journal of Dermatological Science* 74.2, pp. 125–134. DOI: 10.1016/j.jdermsci.2014.01.011.

Yoshida, Yuzo, Tsutomu Soma, and Jiro Kishimoto (2019). "Characterization of human dermal sheath cells reveals CD36-expressing perivascular cells associated with capillary blood vessel formation in hair follicles." In: *Biochemical and Biophysical Research Communications* 516.3, 945–950. DOI: 10.1016/j.bbrc.2019.06.146.

Young, Matthew D and Sam Behjati (2020). "SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data." In: *GigaScience* 9.12. DOI: 10.1093/gigascience/giaa151.

Yu, Hui et al. (2020). "Targeting NF-κB pathway for the therapy of diseases: mechanism and clinical study." In: *Signal Transduction and Targeted Therapy* 5.1. DOI: 10.1038/s41392-020-00312-6.

Yu, Kyung-Rok et al. (2012). "CD49f Enhances Multipotency and Maintains Stemness Through the Direct Regulation of oct4 and SOX2." In: *STEM CELLS* 30.5, pp. 876–887. DOI: 10.1002/stem.1052.

Yu, Mei et al. (2007). "Interleukin-6 cytokine family member oncostatin M is a hair-follicle-expressed factor with hair growth inhibitory properties." In: *Experimental Dermatology* 0.0, 0711170316070006–??? DOI: 10.1111/j.1600-0625.2007.00643.x.

Yu, Sihui et al. (2022). "Cancer-associated fibroblasts-derived FMO2 as a biomarker of macrophage infiltration and prognosis in epithelial ovarian cancer." In: *Gynecologic Oncology* 167.2, pp. 342–353. DOI: 10.1016/j.ygyno.2022.09.003.

Yuan, Li et al. (2002). "Abnormal lymphatic vessel development in neuropilin 2 mutant mice." In: *Development* 129.20, pp. 4797–4806. DOI: 10.1242/dev.129.20.4797.

Yun, Haesung et al. (2022). "The Chemerin-CMKLR1 Axis is Functionally important for Central Regulation of Energy Homeostasis." In: *Frontiers in Physiology* 13. DOI: 10.3389/fphys.2022.897105.

Yunna, Chen et al. (2020). "Macrophage M1/M2 polarization." In: *European Journal of Pharmacology* 877, p. 173090. DOI: 10.1016/j.ejphar.2020.173090.

Zakrzewicz, Anna et al. (2019). "SLPI Inhibits ATP-Mediated Maturation of IL-1β in Human Monocytic Leukocytes: A Novel Function of an Old Player." In: *Frontiers in Immunology* 10. DOI: 10.3389/fimmu.2019.00664.

Zand, Maryam and Jianhua Ruan (2020). "Network-based single-cell RNA-seq data imputation enhances cell type identification." In: *Genes* 11.4, p. 377. DOI: 10.3390/genes11040377.

Zappia, Luke, Belinda Phipson, and Alicia Oshlack (2018). "Exploring the single-cell RNA-seq analysis landscape with the scrna-tools database." In: *PLOS Computational Biology* 14.6. DOI: 10.1371/journal.pcbi.1006245.

Zeltz, Cédric et al. (2020). "Cancer-associated fibroblasts in desmoplastic tumors: Emerging role of integrins." In: *Seminars in Cancer Biology* 62, 166–181. DOI: 10.1016/j.semcancer.2019.08.004.

Zhang, Allen W. et al. (2019a). "Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling." In: *Nature Methods* 16.10, 1007–1015. DOI: 10.1038/s41592-019-0529-1.

Zhang, Cheng et al. (2020a). "Caveolin-1 upregulates Fut8 expression by activating the Wnt/β-catenin pathway to enhance HCC cell proliferative and invasive ability." In: *Cell Biology International* 44.11, pp. 2202–2212. DOI: 10.1002/cbin.11426.

Zhang, Haihua et al. (2018). "Balance between fibroblast growth factor 10 and secreted frizzled-relate protein-1 controls the development of hair follicle by competitively regulating β-catenin signaling." In: *Biomedicine & Pharmacotherapy* 103, pp. 1531–1537. DOI: 10.1016/j.biopha.2018.04.149.

Zhang, Hong-Liang, Jiang Wu, and Jie Zhu (2010). "The Immune-Modulatory Role of Apolipoprotein E with Emphasis on Multiple Sclerosis and Experimental Autoimmune Encephalomyelitis." In: *Clinical and Developmental Immunology* 2010, pp. 1–10. DOI: 10.1155/2010/186813.

Zhang, Jesse M., Govinda M. Kamath, and David N. Tse (2019b). "Valid post-clustering differential analysis for single-cell RNA-seq." In: *Cell Systems* 9.4. DOI: 10.1016/j.cels.2019.07.012.

Zhang, Lian et al. (2007). "Discovery of common human genetic variants of GTP cyclohydrolase 1 (GCH1) governing nitric oxide, autonomic activity, and cardiovascular risk." In: *Journal of Clinical Investigation* 117.9, pp. 2658–2671. DOI: 10.1172/jci31093.

Zhang, Ling-juan et al. (2019c). "Age-related loss of innate immune antimicrobial function of dermal fat is mediated by transforming growth factor beta." In: *Immunity* 50.1. DOI: 10.1016/j.immuni.2018.11.003.

Zhang, Min et al. (2021a). "Inhibition of fibroblast IL-6 production by ACKR4 deletion alleviates cardiac remodeling after myocardial infarction." In: *Biochemical and Biophysical Research Communications* 547, pp. 139–147. DOI: 10.1016/j.bbrc.2021.02.013.

Zhang, Rui et al. (2021b). "ITM2A as a Tumor Suppressor and Its Correlation With PD-L1 in Breast Cancer." In: *Frontiers in Oncology* 10. DOI: 10.3389/fonc.2020.581733.

Zhang, Xinjun et al. (2015). "Notum Is Required for Neural and Head Induction via Wnt Deacylation, Oxidation, and Inactivation." In: *Developmental Cell* 32.6, pp. 719–730. DOI: 10.1016/j.devcel.2015.02.014.

Zhang, Y et al. (1997). "Secretory leukocyte protease inhibitor suppresses the production of monocyte prostaglandin H synthase-2, prostaglandin E2, and matrix metalloproteinases." In: *Journal of Clinical Investigation* 99.5, pp. 894–900. DOI: 10.1172/jci119254.

Zhang, Yi et al. (2013). "Lef1 Contributes to the Differentiation of Bulge Stem Cells by Nuclear Translocation and Cross-Talk with the Notch Signaling Pathway." In: *International Journal of Medical Sciences* 10.6, pp. 738–746. DOI: 10.7150/ijms.5693.

Zhang, Yun-Long et al. (2020b). "Chemokine Receptor CXCR-2 Initiates Atrial Fibrillation by Triggering Monocyte Mobilization in Mice." In: *Hypertension* 76.2, pp. 381–392. DOI: 10.1161/hypertensionaha.120.14698.

Zhang, Zhe et al. (2014). "Increased periostin expression affects the proliferation, collagen synthesis, migration and invasion of keloid fibroblasts under hypoxic conditions." In: *International Journal of Molecular Medicine* 34.1, pp. 253–261. DOI: 10.3892/ijmm.2014.1760.

Zhao, Jilai et al. (2021). "SOX9 and RBPJ differentially regulate endothelial to mesenchymal transition and wound scarring in murine endovascular progenitors." In: *Nature Communications* 12.1. DOI: 10.1038/s41467-021-22717-9.

Zhao, Jin et al. (2020). "Multifaceted Functions of CH25H and 25HC to Modulate the Lipid Metabolism, Immune Responses, and Broadly Antiviral Activities." In: *Viruses* 12.7, p. 727. DOI: 10.3390/v12070727.

Zhao, S et al. (2015). "NKD2, a negative regulator of Wnt signaling, suppresses tumor growth and metastasis in osteosarcoma." In: *Oncogene* 34.39, pp. 5069–5079. DOI: 10.1038/onc.2014.429.

Zhao, Yun et al. (2013). "Cyp1b1 Mediates Periostin Regulation of Trabecular Meshwork Development by Suppression of Oxidative Stress." In: *Molecular and Cellular Biology* 33.21, pp. 4225–4240. DOI: 10.1128/mcb.00856-13.

Zheng, Grace X. Y. et al. (2017). "Massively parallel digital transcriptional profiling of single cells." In: *Nature Communications* 8.1. DOI: 10.1038/ncomms14049.

Zheng, Mei et al. (2022). "CXCL12 inhibits hair growth through CXCR4." In: *Biomedicine & Pharmacotherapy* 150, p. 112996. DOI: 10.1016/j.biopha.2022.112996.

Zheng, Zhong et al. (2014). "Fibromodulin-Deficiency Alters Temporospatial Expression Patterns of Transforming Growth Factor-β Ligands and Receptors during Adult Mouse Skin Wound Healing." In: *PLoS ONE* 9.3. Ed. by Zhongjun Zhou, e90817. DOI: 10.1371/journal.pone.0090817.

Zhou, Cefan et al. (2019). "Integral membrane protein 2A inhibits cell growth in human breast cancer via enhancing autophagy induction." In: *Cell Communication and Signaling* 17.1. DOI: 10.1186/s12964-019-0422-7.

Zhou, Yangyan et al. (2010). "Structure-function analysis of human l-prostaglandin D synthase bound with fatty acid molecules." In: *The FASEB Journal* 24.12, pp. 4668–4677. DOI: 10.1096/fj.10-164863.

Zhu, Lingxue et al. (2018). "A unified statistical framework for single cell and bulk RNA sequencing data." In: *The Annals of Applied Statistics* 12.1. DOI: 10.1214/17-aoas1110.

Ziegenhain, Christoph et al. (2017). "Comparative analysis of single-cell RNA sequencing methods." In: *Molecular Cell* 65.4. DOI: 10.1016/j.molcel.2017.01.023.

Ziegler, Ernst (1895). *General Pathology: Or, The Science of the Causes, Nature and Course of the Pathological Disturbances which Occur in the Living Subject*. William Wood and Company.

Zlotogorski, A. (1987). "Distribution of skin surface ph on the forehead and cheek of adults." In: *Archives of Dermatological Research* 279.6, 398–401. DOI: 10.1007/bf00412626.

Zomer, Helena D. and Andrea G. Trentin (2018). "Skin wound healing in humans and mice: Challenges in translational research." In: *Journal of Dermatological Science* 90.1, 3–12. DOI: 10.1016/j.jdermsci.2017.12.009.

Zouboulis, Christos C. (2004). "Acne and sebaceous gland function." In: *Clinics in Dermatology* 22.5, 360–366. DOI: 10.1016/j.clindermatol.2004.03.004.

Zwick, Rachel K. et al. (2018). "Anatomical, physiological, and functional diversity of adipose tissue." In: *Cell Metabolism* 27.1, 68–83. DOI: 10.1016/j.cmet.2017.12.002.

d'Ovidio, Roberto et al. (2007). "Alterated integrin expression in lichen planopilaris." In: *Head & Face Medicine* 3.1. DOI: 10.1186/1746-160x-3-11.