



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

MEDIKUNTZA
ETA ERIZAINNTZA
FAKULTATEA
FACULTAD
DE MEDICINA
Y ENFERMERÍA

biocruces
bizkaia

osasun ikerketa institutua
instituto de investigación sanitaria

TRABAJO FIN DE MÁSTER

MÁSTER EN INVESTIGACIÓN BIOMÉDICA

IDENTIFICACIÓN DE BIOMARCADORES DE LA PLACA DE ATEROMA EN ARTERIA CARÓTIDA INESTABLE

Autora: Ane Aldekoa

Directora: Iraide Alloza

Leioa, 15 de Junio del 2021

Índice

Introducción:.....	3
Objetivos	6
Materiales y métodos	6
• Ética y selección de pacientes.....	6
• Sujetos.	6
• Extracción y purificación de RNA de placa.	7
• Cuantificación y análisis de calidad del RNA.	7
• Selección de biomarcadores.....	8
• Análisis de expresión génica mediante PCR digital (Fluidigm Biomark).....	8
• Análisis estadístico y selección de variables	10
Resultados	11
• Cuantificación y análisis de calidad del RNA.	11
• Análisis de expresión génica.....	13
• Análisis estadístico.....	14
• Análisis multivariante de los datos.....	15
• Análisis discriminantes.....	16
Discusión	20
Conclusiones.....	23
Bibliografía.....	24

Resumen:

La aterosclerosis es una enfermedad crónica producida por una inflamación persistente de la pared de la arteria además de una acumulación de lípidos. Las placas de ateroma pueden dividirse en dos grupos según la morfología que presenta la placa, estable o inestable, siendo la placa inestable capaz de provocar accidentes cerebrovasculares. En este estudio, se analizó la expresión génica mediante PCR cuantitativa digital de genes anteriormente estudiados en un análisis comparativo basado en RNAseq. En dicho análisis se analizaron las diferencias en el transcriptoma completo de células del músculo liso (CML) aisladas de placas sintomáticas (S) y asintomáticas (A) y a nivel de placa completa. Para el presente estudio, se empleó RNA extraído de placas de ateroma tanto S como A. Se observó que 11 genes de los 116 testados se validaban con los genes ya detectados en CML y a nivel de placa completa. Además, 5 de los genes codificaban proteínas secretadas detectables en suero. Los resultados aportados por este estudio suponen un gran avance en la búsqueda de biomarcadores de aterosclerosis que permitan discriminar A de S realizando un test sérico de detección rápida con el fin de poder detectar a pacientes con alto riesgo de desarrollar accidentes cerebrovasculares.

Introducción:

La aterosclerosis es una enfermedad de carácter crónico que se forma por el resultado de una respuesta inflamatoria de la pared de la arteria, además de una acumulación de lípidos. En consecuencia se forman lesiones focales o en algunos casos la rotura de la placa provocando una complicación trombótica. Estos síntomas son uno de los principales factores de riesgo de sufrir un ictus isquémico. El ictus es la primera causa de muerte en mujeres y la primera causa de discapacidad adquirida en el mundo. Por lo cual tiene un impacto muy alto en los servicios sanitarios (Feigin et al. 2017). Se trata de una enfermedad neurológica cuyas causas se ven determinadas por diferentes factores como, factores genéticos, epigenéticos, ambientales y de estilo de vida. Debido a que se trata de una enfermedad multifactorial, el diagnóstico es difícil, tanto por la variedad de la expresión clínica de la enfermedad, como por la necesidad de hacerlo rápido. Hasta ahora no existe ningún biomarcador con valor pronóstico que pueda acelerar el proceso, por ello, se siguen realizando muchos estudios (Bustamante *et al.* 2017) para entender los mecanismos implicados en la formación y desestabilización de

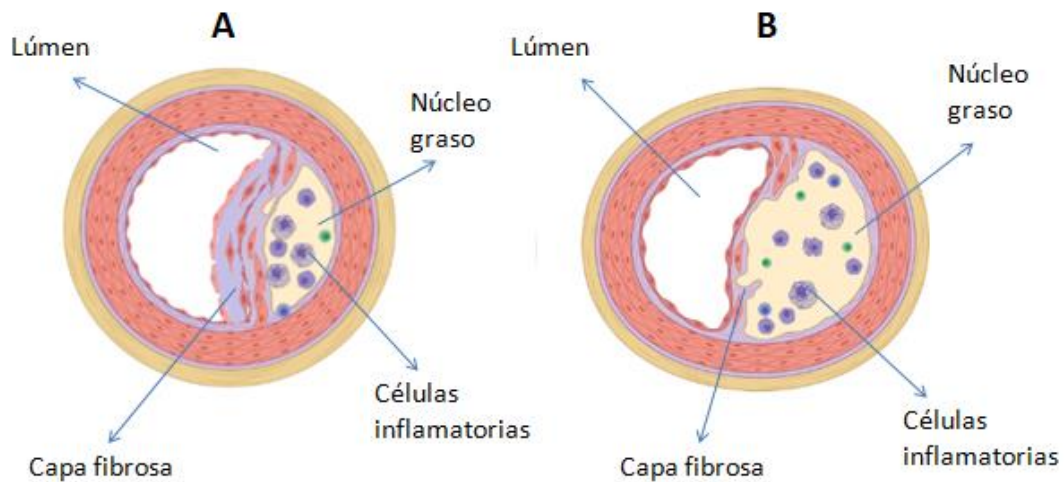
la placa aterosclerótica y así, poder identificar biomarcadores que permitan predecir el riesgo de desarrollo de un ictus en la práctica clínica.

La placa de ateroma de la arteria carótida está formada mayoritariamente por células del sistema inmune, células endoteliales y células del músculo liso (Goikuria et al. 2018). La aterosclerosis tiene lugar en la capa íntima de las arterias de mediano y gran calibre, por ello es tan común que se forme en la arteria carótida. En los estadios iniciales se produce una alteración en la zona endotelial, que provoca una adhesión de los leucocitos permitiendo el traspaso de éstos a la capa íntima junto con lipoproteínas como LDL. La penetración de LDL a la capa íntima desencadena la unión de proteoglicanos a la matriz dando lugar a modificaciones que convierten a la arteria proaterogénica. La placa de ateroma se va desarrollando a medida que se va acumulando LDL, monocitos, linfocitos, células del músculo liso y se produce una cubierta fibrosa. En el centro de la placa se produce una apoptosis de las células dejando un centro necrótico rico en grasa (Carvajal *et al.* 2015). Dependiendo del tamaño del núcleo grasoso y de la capa fibrosa, podemos diferenciar la placa de ateroma en dos tipos, estable e inestable (Figura 1).

La placa de ateroma estable es aquella que presenta un núcleo grasoso pequeño y una gruesa capa fibrosa pudiendo provocar una obstrucción arterial sin presentarse síntomas cerebrovasculares (asintomática A). Sin embargo, un núcleo grasoso grande y una capa fina de fibras hace que la placa sea más inestable y presenta mayor riesgo de rotura, provocando, como consecuencia, un accidente cerebrovascular (sintomática S). Se ha observado que la inestabilidad de la placa y el riesgo de desarrollar un accidente cerebrovascular está correlacionado (Dagonnier *et al.* 2021). A pesar de esto, no se conoce cuál es el mecanismo por el cual la placa se desestabiliza provocando una rotura. Por ello, la identificación de biomarcadores ayudaría a una detección precoz de aquellas personas con alto riesgo de sufrir un accidente cerebrovascular. En la actualidad, aunque existen terapias para tratar esta enfermedad, no llegan a ser específicas. La Ecografía Doppler es la terapia que se utiliza hoy en día para el diagnóstico de aterosclerosis, así como para otras muchas enfermedades del sistema circulatorio (Fernandez *et al.* 2000). La Ecografía Doppler es una prueba no invasiva que permite detectar el flujo sanguíneo haciendo rebotar ondas sonoras de alta frecuencia en los glóbulos rojos circulantes. Esta prueba es una alternativa a procedimientos más invasivos, como la arteriografía. Sin embargo, no es una prueba que se realiza rutinariamente y por ello, la detección de estas enfermedades puede llegar a ser tardía (Ammirati *et al.* 2013).

La limitación de estas pruebas diagnóstico puede deberse a que no se conoce ningún marcador en suero que se pueda utilizar en la práctica clínica como diagnóstico. Esto posiblemente sea consecuencia de la heterogeneidad inherente a las placas de ateroma de carótida, en las que están involucrados múltiples tipos celulares. Por tanto, para conocer los mecanismos implicados en la inestabilidad de la placa hay que estudiar qué rol juega cada tipo celular en este proceso.

Figura 1. A: placa de ateroma estable con el núcleo lipídico pequeño y la capa fibrosa gruesa correspondiente con pacientes asintomáticos. B: placa de ateroma inestable con el núcleo lipídico grande y una capa fibrosa fina correspondiente con pacientes sintomáticos.



En un estudio reciente realizado en el laboratorio donde he desarrollado mi trabajo, estudiaron las células del músculo liso (CML) de la placa de ateroma y observaron que juegan un papel importante en la desestabilización de la placa (Alloza et al. 2017). El estudio realizó un análisis comparativo basado en RNAseq en el que se encontraron diferencias en el transcriptoma completo de las CML entre las placas S y A. Se observó que había diferencias en los fenotipos de cada placa, siendo el fenotipo de la placa inestable y sintomática de tipo senescencia, mientras que el de la placa estable y asintomática es de tipo osteogénico. Asimismo, se encontraron una serie de genes en los que había una diferencia significativa en su expresión entre los dos tipos de placa. Debido a la relevancia de estos resultados para la comprensión del comportamiento de la placa, se realizaron estudios posteriores a nivel de placa completa, con el objetivo de validar estos biomarcadores. Estos estudios se dividen en dos fases que incluyen una

primera cohorte de validación de muestras de Basurto y una segunda compuesta por muestras de Basurto, Navarra, Albacete y La Fe (Valencia). Con los resultados obtenidos, el objetivo de este trabajo fue analizar si estos genes con expresión diferencial en CML y a nivel de toda la placa estudiados anteriormente, se pueden validar con una cohorte mayor de muestras. Para ello se utilizó una cohorte de muestras del Hospital de Girona y La Paz.

Objetivos

- Obtención y purificación de RNA de la placa de ateroma de pacientes S y A para analizar la expresión génica mediante PCR cuantitativa digital.
- Validación a nivel de placa aterosclerótica completa de los genes diferencialmente expresados en pacientes S y A.
- Identificación de biomarcadores biológicos para detectar más rápido a pacientes con alto riesgo de desarrollar accidentes cerebrovasculares.

Materiales y métodos

Ética y selección de pacientes.

El estudio fue aprobado por el comité ético local (Comité Ético de Investigación del País Vasco (CEIm-E); código de identificación del proyecto PI2018015; Fecha de aprobación: 11 de marzo de 2019). Todas las placas de ateroma carotídeo se recogieron de pacientes que habían firmado un consentimiento informado por escrito. Esta investigación se llevó a cabo de acuerdo con los principios señalados en la Declaración de Helsinki.

Sujetos.

Los pacientes sometidos a endarterectomía carotídea en el Hospital Universitario La Paz (Madrid) y Hospital Universitario de Girona, fueron reclutados para su selección en el estudio actual en base a parámetros clínicos definidos. El grado de estenosis se evaluó con ecografía Eco-Doppler cervical y angiografía tomográfica según los criterios establecidos. Los pacientes sintomáticos (S) fueron identificados como aquellos con una estenosis >70% y que presentaban síntomas de accidente isquémico transitorio o ictus

ipsilateral en los últimos seis meses, mientras que los pacientes asintomáticos (A) eran aquellos con estenosis >80% sin ninguna presencia de enfermedad cerebrovascular a lo largo de su vida. Las muestras de tejido carotídeo se recogieron tras cirugía y se congelan a -8C° hasta el momento de su transporte al Laboratorio de Inflamación y Biomarcadores para la extracción de ARN.

Extracción y purificación de RNA de placa.

Se utilizaron muestras de placa aterosclerótica de pacientes sintomáticos y asintomáticos y se cortaron en fragmentos pequeños. Los fragmentos fueron introducidos en tubos RNasa free Bullet Blender® (Next Advance) y se añadió 1 ml de Tri Reagent (Sigma-Aldrich). A continuación, se empleó el homogeneizador Next Avance realizando tres rondas de 5 minutos con un periodo de reposo en hielo de 10 minutos entre ronda y ronda para realizar la digestión de estos fragmentos. El homogeneizado obtenido fue recogido en nuevos tubos de 1,5 ml en los que se añadió 0,1 ml de 1-bromo-3-cloropropano (BCP; Sigma- Aldrich) por cada mililitro de Tri Reagent. La mezcla obtenida se agitó enérgicamente y se dejó reposar 15 minutos a temperatura ambiente. Posteriormente se centrifugó a 12000 x g durante 15 minutos. Tras la centrifugación la mezcla se separó en tres fases de la cual la fase superior, fase incolora y acuosa es la que contenía el RNA. La fase incolora fue recogida en un nuevo tubo de 1,5mL al que se le añadió un volumen de 1:1 de etanol al 70% y se mezcló bien. Por último, se empleó el PureLink® RNA Mini Kit (Ambion, ThermoFisher) siguiendo los pasos establecidos por el protocolo del mismo. Durante la extracción se realizó un tratamiento con DNasa y se eluyó el RNA de la columna en 50 µl de agua libre de RNasas. Se obtuvo finalmente el RNA purificado que se conservó a -80°C.

Cuantificación y análisis de calidad del RNA.

Para obtener datos sobre la concentración aproximada de RNA presente en cada muestra se realizó una cuantificación espectrofotométrica mediante el espectrofotómetro NanoDrop (Thermo Fisher Scientific). El blanco se calibró con agua libre de RNasas (Thermo Fisher Scientific) empleándose después 1 µl de cada muestra para la cuantificación. A continuación, teniendo en cuenta los valores de la cuantificación por NanoDrop se volvieron a cuantificar mediante QuantiFluor RNA System (Promega). Para ello, se realizaron curvas estándar para RNA de alta concentración y baja

concentración. Se añadió 1 µl de las muestras a cuantificar en cada pocillo de una placa de 96 junto con una mezcla del QuantiFluor RNA Dye y 1X TE Buffer y se midió la fluorescencia a 492 nm EX/540 nm EX . Todas las muestras tenían valores superiores a 10ng/µl.

Selección de biomarcadores.

Los genes a analizar se seleccionaron en base a los resultados obtenidos del estudio basado en RNAseq realizado previamente (Alloza *et al.* 2017). La selección se llevó a cabo teniendo en cuenta el valor de *fold change* (FC) que describe la diferencia de los niveles de expresión génica entre muestras S y A. Es decir, se estudiaron aquellos genes que demostraron tener una expresión diferencial en el transcriptoma completo de las placas S y A. Los FC para la selección fueron FC mayor o igual que 1,2 y menor o igual que -1,2.

Análisis de expresión génica mediante PCR digital (Fluidigm Biomark).

Para poder analizar la expresión génica de los genes seleccionados el primer paso fue retrotranscribir el RNA presente en la muestra en cDNA con Verso cDNA synthesis kit (Thermo Scientific). En esta retrotranscripción del RNA se utilizaron cebadores random y una incubación de 60 minutos en la que se partió de 5 µl de RNA total en un volumen final de reacción de 10 µl. Posteriormente se procedió a amplificar el cDNA obtenido empleando cebadores PrimeTime de Integrated DNA Technologies (IDT). Estos cebadores fueron diseñados para cada uno de los genes cuya expresión interesaba analizar. Los 117 cebadores creados se distribuyeron en 2 pooles: A y B (Tabla 1). Tanto en el pool A como en el pool B se añadió el gen MKLN1 con el objetivo de garantizar que la variabilidad entre los resultados de los pooles era mínima. Por lo tanto, debido a esta duplicación en el gen MKLN1, los genes analizados en este estudio fueron 117 en total. La razón por la que se realizaron dos pooles fue la imposibilidad de migrar todos los genes y muestras en un único chip de 96.96. Se repartieron los 121 cebadores en un pool A con 46 cebadores y un pool B con 71 cebadores. Primero, se realizó una reacción de preamplificación STA (Specific Target Amplification) con el kit Qiagen Multiplex PCR kit (Qiagen) utilizando 1,25 µl de cDNA y concentración final de cebador de 50 mM. Así, se obtuvo un enriquecimiento de la diana de análisis de cada muestra previo al análisis de la expresión génica. Tras la preamplificación se trataron las muestras con Exonucleasa I (Exo I; Thermo Scientific) para eliminar los cebadores no

incorporados. El análisis de expresión se realizó con los arrays de Fluidigm Biomark 96.96 (Fluidigm Corporation) y los ensayos qPCR PrimeTime (IDT) en las 183 muestras por duplicado. Para la reacción se empleó Master Mix SsoFast™ EvaGreen® Supermix (Bio-Rad Laboratories). La concentración final de cebadores en la reacción de qPCR fue de 500 nM para todos los genes. El software empleado para la obtención de los datos de PCR a tiempo real fue Fluidigm Real-Time PCR Analysis Software versión 4.1.3. Los niveles de expresión de los genes housekeeping (G3BP2, BACTIN, MKLN1, EML3, ADCK5) se utilizaron como control endógeno para normalizar las muestras. La razón por la que se utilizó más de un gen housekeeping fue que cada uno correspondía con un intervalo diferente nivel de expresión por lo que la posterior normalización de los datos se realizaría con mayor precisión. Además, como no se migraron todas las muestras en el mismo chip, para eliminar variaciones entre las diferentes placas se migró una muestra de RNA purificado de células embrionarias de riñón humano (HEK) en los dos chips de cada pool. El valor de dicha muestra se normalizó después empleando el software GenEx vs 6.1 (MultiD). Adicionalmente, para evaluar la correcta amplificación de las muestras, se incluyó para cada pool una muestra negativa que carecía de cDNA, sustituyendo éste por agua libre de nucleasas.

Tabla 1. Identidad de los genes de interés a amplificar por los cebadores diseñados.

Genes							
POOL A							
ADCK5	EML3	G3BP2	MKLN1	ABCG2	ADAMTS7	ANXA10	ANXA3
BMP2	C10ORF10	CA12	CHODL	DDR2	DHRS3	DIRAS3	EPHA4
EPHB2	ESM1	FRMD3	GULP	HAPLN1	ID1	ID4	IL17RD
ITGA7	KCNE4	KLF5	MGST1	MOCOS	MYO18B	NPPC-17	NPPC-44
PDK1	PFKFB4	PTHLH	RDH10	RGS2	SMAD9	SOCS3	ST6GANA C5
SYNE3	TBX18	TMEM35	TMTC2	TNFAIP8L3	ZSCAN31		

ADCK5* EML3* G3BP2* MKLN1* BACTIN*

POLL B

ANO6	APOA1	ARID5B	BNIP3L	CADM3	CCL5	CCT6A	CD248
CHI3L1	CHRNA5	CLEC3B	CNTN1	COL14A1	COL18A1	CUL4B	DENND5 A
DKK2	ESYT2	FERMT2	FHL1	FRAS1-51	FRAS1-53	FRMD3	GLRX3
HSPA12A	IFI27L1	IGIP	IL20RB	ITGAE	LPIN2	MAP4	MARCH 8
MMP1	MMP16	MRPL52	MSRB3	NRCAM	NRK	NRXN2	PDLIM4
PI16	PLA2G4A	PLAU-62	PLAU-88	PLXNB2	PPP1R3C	PRKAB2	PTER
PTPRB	RALBP1	REEP2- 65	REEP2- 93	RNF217	SCARB2	SDPR	SESN2
SGCD	SLC26A2	SLC2A3	SLITRK4	SMAP2	SOX17	ST3GAL 5	SYNE3
TCP11L2	TMEM47	TP53BP2	TP53I3	TXN	XG	YTHDC2	

* Los cebadores marcados con un asterisco (*) corresponden a genes housekeeping.

Análisis estadístico y selección de variables

Los niveles de expresión relativa de cada gen en estudio se calcularon utilizando el método comparativo CT también conocido como $2^{-\Delta\Delta CT}$ (Schmittgen y Livak, 2008). Los valores de CT obtenidos para cada muestra y gen fueron normalizados con los genes housekeeping, empleando en cada caso el gen housekeeping con la expresión génica más similar. La significancia estadística de los niveles de expresión entre S (n=68) y A (n=6) se analizó con el software GraphPad Prism 6 usando el test no-

paramétrico Mann–Whitney U–test. Los resultados de las muestras a estudiar fueron integrados con los valores analizados anteriormente. Para la caracterización de los genes con expresión diferencial obtenidos se emplearon las bases de datos Ensembl, (Zerbino et al. 2018) ProteINSIDE (Kaspric et al. 2015) y STRING ([https:// string-db.org](https://string-db.org)) además de bibliografía. Para el análisis multivariante, los datos fueron analizados utilizando el Software SIMCA (software SIMCA P+ 15.0.1). Para poder proceder al análisis los datos fueron previamente transformados probando diferentes combinaciones de transformación. Para el análisis exploratorio se realizó el análisis de componentes principales (PCA) con el que se evaluó la calidad de los datos, presencia de outliers y/o patrones o si las muestras seguían alguna tendencia. Como análisis discriminantes, se utilizaron los datos obtenidos del análisis de expresión FC, su p value, la función que realiza cada gen y si hay conexiones entre varios genes.

Resultados

Cuantificación y análisis de calidad del RNA.

Se extrajo y purificó el RNA de 74 muestras de placa de aterosclerosis completa, siendo 68 muestras sintomáticas y 6 asintomáticas. Después de diluir el RNA de la columna en 50 µl de agua libre de RNasas, se procedió a la cuantificación del RNA presente en cada muestra purificada. Inicialmente, para obtener una concentración orientativa se se utilizó el espectrofotómetro NanoDrop (Tabla 2A). En esta cuantificación se tuvo en cuenta el ratio de absorbancia a 260 nm y a 280 nm (ratio A260/280) considerándose un ratio A260/280 de aproximadamente 2,0 adecuado para una buena pureza del RNA (Wilfinger *et al.* 1997). Los valores de concentración obtenidos mediante NanoDrop sirvieron para determinar cuál era la curva patrón idónea para hacer la cuantificación con el QuantiFluor RNA System (Promega). Por un lado, en aquellas muestras con concentración según NanoDrop mayores de 25ng/µl se utilizó la curva High-Concentration RNA Standars. Por otro lado, las muestras con concentraciones según NanoDrop menores de 25ng/µl se utilizó la curva Low-Concentration RNA Standars (Tabla 2B).

Tabla 2. Concentraciones de RNA total medidas con el espectrofotómetro NanoDrop (A) y QuantiFluor RNA System (B). Se muestran los valores de concentración de 26 muestras de las 74 analizadas.

Identificación muestra	Concentración de RNA total (ng/μl)	
	(A)NanoDrop	(B)Quantifluor
2	28	9,35
3	81,3	37,09
5	13,3	6,21
6	12,2	5,6
7	12,8	8,86
9	42,2	12,8
14	29,5	26,8
16	11	7,61
18	45,7	39,3
20	31	17,8
24	39,6	20,3
25	22,1	21,8
27	11,1	10,7
29	20,6	26
33	18,7	7,17
34	18,9	25,9
63	250	145
64	24,4	11,3
65	88,6	64
67	78	52,2
69	108	39,8
70	45,8	26,6

Análisis de expresión génica.

Los niveles de expresión relativa de cada gen en estudio se calcularon utilizando el método comparativo C_T (Schmittgen y Livak, 2008). En el análisis de expresión génica diferencial se detectaron 14 genes que presentaban valores de fold change (FC) mayores o iguales que 1,2 (sobrexpresión en S frente a A) y valores de FC menores que $-1,2$ (menor expresión en S frente a A) y con un p value inferior o igual a 0,05 (Tabla 3).

Tabla 3. Identidad y características de los genes con expresión diferencial entre muestras S y A.

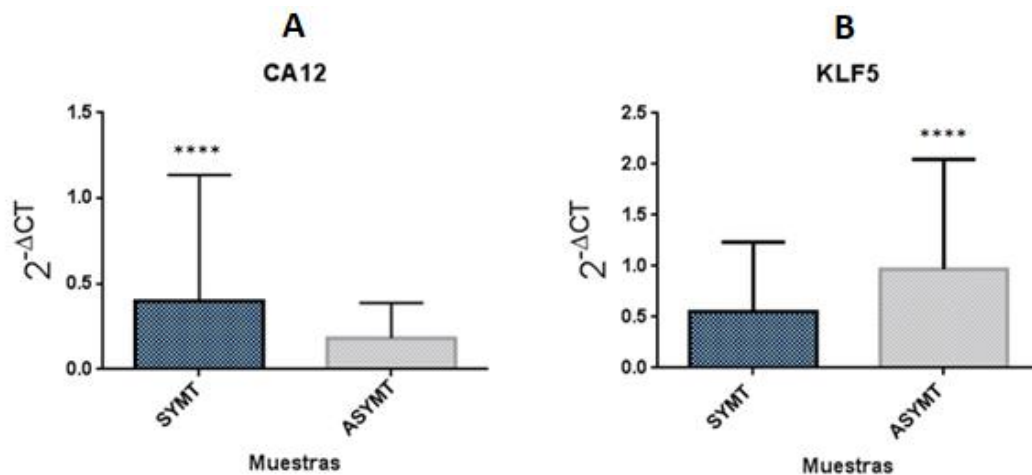
Símbolo del gen	FC	Valor p	Descripción
ADAM7S7	1,12	0,0104	Metaloproteasa que regula la migración de las CML
CA12	2,35	< 0,0001	Anhidrasa carbónica XII
CHODL	5,40	0,0081	Unión de ácido hialurónico
DDR2	3,68	< 0,0001	Dominio discoidina del receptor tirosina quinasa 2
FRMD3	1,43	0,032	Organización de la estructura de la actomiosina
GULP	1,57	0,015	Fagocitosis
IL17RD	-1,5	0,0375	Regulación negativa de la vía de señalización del receptor del factor de crecimiento de fibroblastos
KLF5	-1,49	< 0,0001	Regulación positiva de la transcripción por la ARN polimerasa II
MMP16	1,27	0,0309	Procesamiento de proteínas
PFKFB4	2,42	0,0013	Actividad de la fructosa-2,6-bisfosfato-2-fosfatasa
SLC2A3	2,46	0,0317	Unión de la glucosa
SYNE3	1,32	0,0004	Regulación de la forma celular
TMEM47	1,26	0,0116	Unión célula-célula

ZSCAN31	2,16	< 0,0001	Actividad del factor de transcripción de unión al ADN, específico de la ARN polimerasa II
---------	------	----------	---

Análisis estadístico.

Una vez conocida la presencia de expresión diferencial entre muestras S y A fue necesario comprobar si esta diferencia era significativa. La significancia estadística de los niveles de expresión entre S y A se analizó con el *software* GraphPad Prism 6 usando el test no-paramétrico Mann-Whitney U-test. Se determinó que para considerar una diferencia de expresión significativa se debía cumplir que el valor p fuera menor o igual que 0,05. De esta manera, se garantizó que la variabilidad detectada no se debía a una casualidad sino a la presencia de una expresión diferencial significativa entre muestras S y A. El valor p muestra una expresión génica significativamente diferente en genes que se sobre expresan en S frente a A (Figura 3A) y genes en los que la expresión en S era menor con respecto a A (Figura 3B).

Figura 3. Representación gráfica de la diferencia de expresión génica detectada entre muestras S y A para los genes CA12 y KLF5. Se representa la variación de la expresión génica ($2^{-\Delta CT}$) de las muestras SYMT (S) y ASYMT (A).

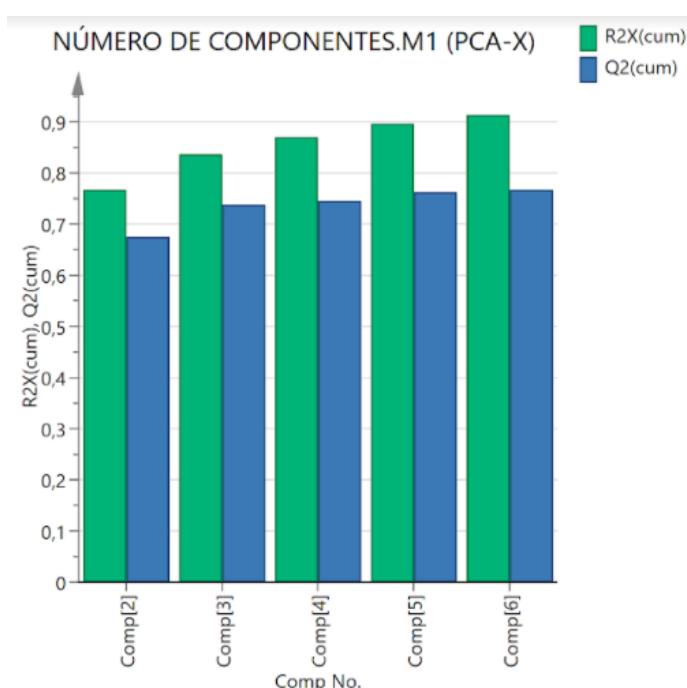


**** p value < 0,001

Análisis multivariante de los datos.

Se efectuó un PCA para encontrar outliers e identificar patrones entre las muestras S y A. El objetivo de este análisis es detectar variables latentes a medida que se agregan o se quitan los componentes. En la figura 4 se muestran los parámetros de R2 y Q2 de los 6 componentes principales. Para evaluar cuántos son los componentes necesarios es necesario que los valores de R2 y Q2 estén cerca del 1 (Abdi and Williams, 2010).

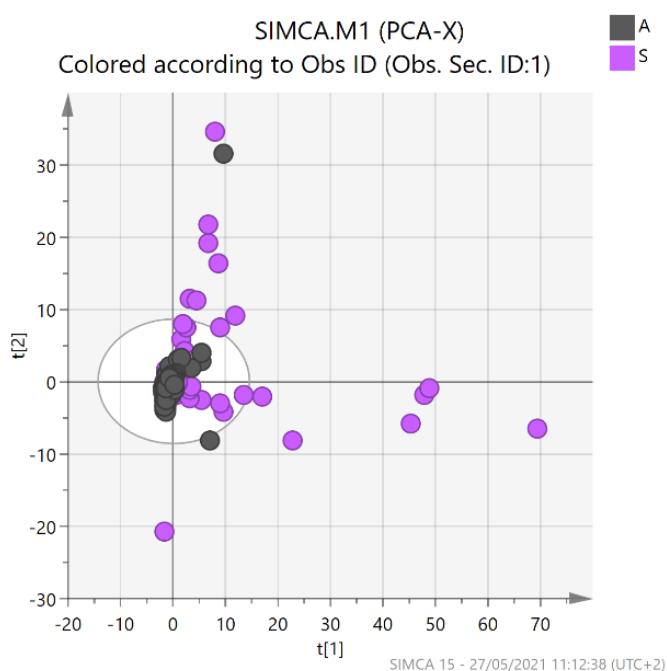
Figura 4. Variabilidad explicada (R2) y capacidad de predicción (Q2) de las primeras 6 componentes principales del PCA realizado con los datos obtenidos de la PCR digital. R2 (cum): R2 acumulada, Q2 (cum): Q2 acumulada.



En las **figuras 5 y 6** se muestran los gráficos de dispersión de los scores. Para la evitar un desajuste de los resultados se utiliza una elipse con 95% de nivel de confianza a partir del estadístico T2 de Hotelling. Las muestras pertenecientes a los pacientes A se encuentran agrupadas y centradas en la elipse mientras que las muestras de los S caen fuera de la elipse que forman parte del 5% de los casos extremos. Los outliers se evaluaron individualmente para la detección de algún error en el análisis de las muestras, el cual se concluyó que no hubo ningún problema. Además se realizaron los análisis con diferentes grupos de genes, teniendo en cuenta los datos estadísticos obtenidos, así como la función de cada gen. Los scores de la gráfica están nombrados según la sintomatología de cada individuo.

En la figura 5 se observa un análisis de todos los genes analizados y la 1ª con la 2ª componente en el que se muestra un patrón de discriminación entre las muestras de pacientes S y A.

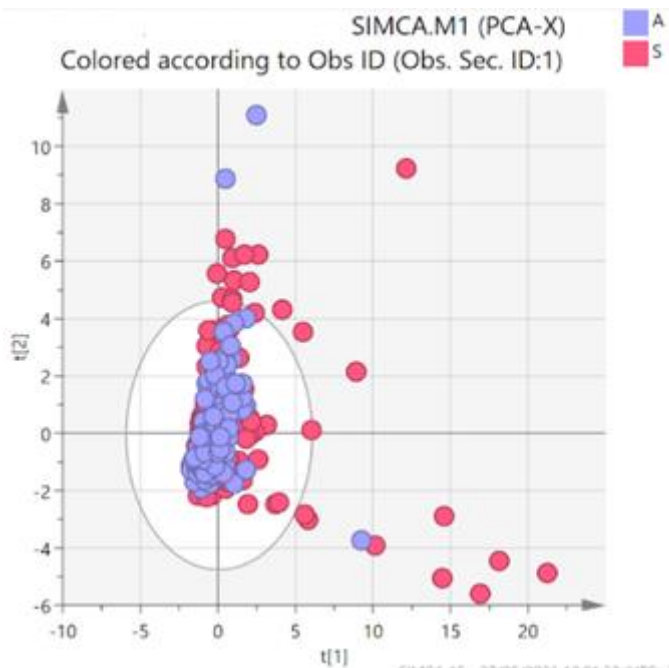
Figura 5. Gráfico de dispersión de los scores de la 1ª vs 2ª componente principal para los resultados de PCR digital. Elipse T2 de Hotelling (95%). Cada punto corresponde a una muestra. T1: componente principal; T2: componente 2; S: pacientes sintomáticos; A: pacientes asintomáticos.



Análisis discriminantes.

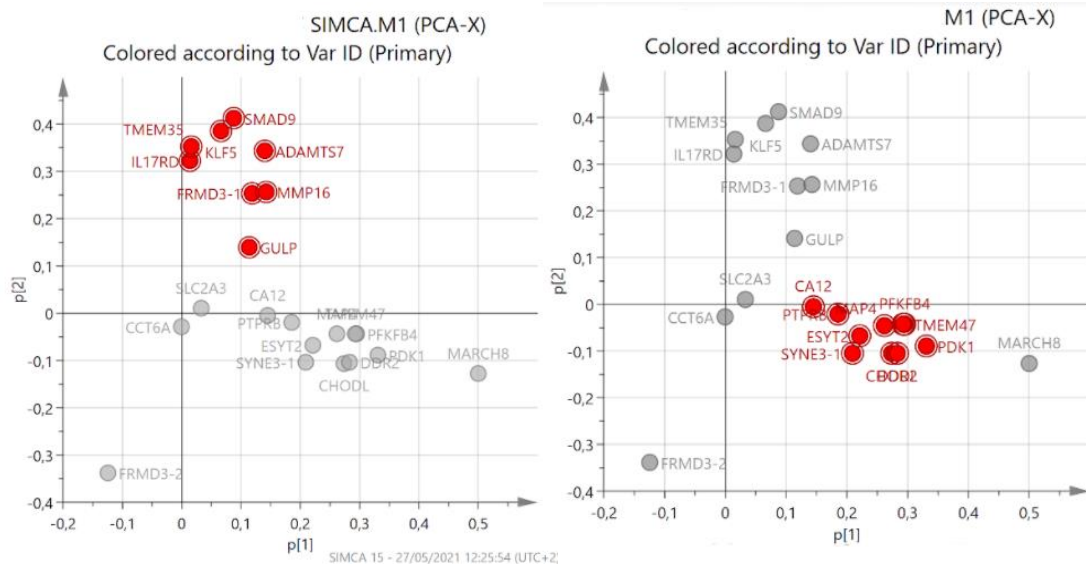
Para estos análisis se redujeron los genes a analizar teniendo en cuenta cuáles habían variado en su expresión según si procedían de pacientes S o A. Al igual que en el PCA, para su representación se construyó un elipse con un nivel de confianza del 95% a partir del estadístico T2 de Hotelling. En la **figura 6** se muestra un análisis de las muestras que en el análisis estadístico dieron un p value igual o menor de 0,05. Se observa un plano formado por la 1ª y 2ª componente en los que se encuentran la mayoría de las muestras de A dentro del score, aunque se puede observar algunas muestras que no entran dentro. Sin embargo, en las muestras S hay una tendencia de salirse de ese score pudiéndose observar una separación de los grupos. En el análisis de los planos con el resto de componentes se vio que no había una variabilidad muy grande con el análisis del plano formado por la 1ª y 2ª componente.

Figura 6. Gráfico de dispersión de los scores de la 1ª vs 2ª componente principal para los resultados de PCR digital con genes significativos. Elipse T2 de Hotelling (95%). Cada punto corresponde a una muestra. T1: componente principal; T2: componente 2; S: pacientes sintomáticos; A: pacientes asintomáticos.



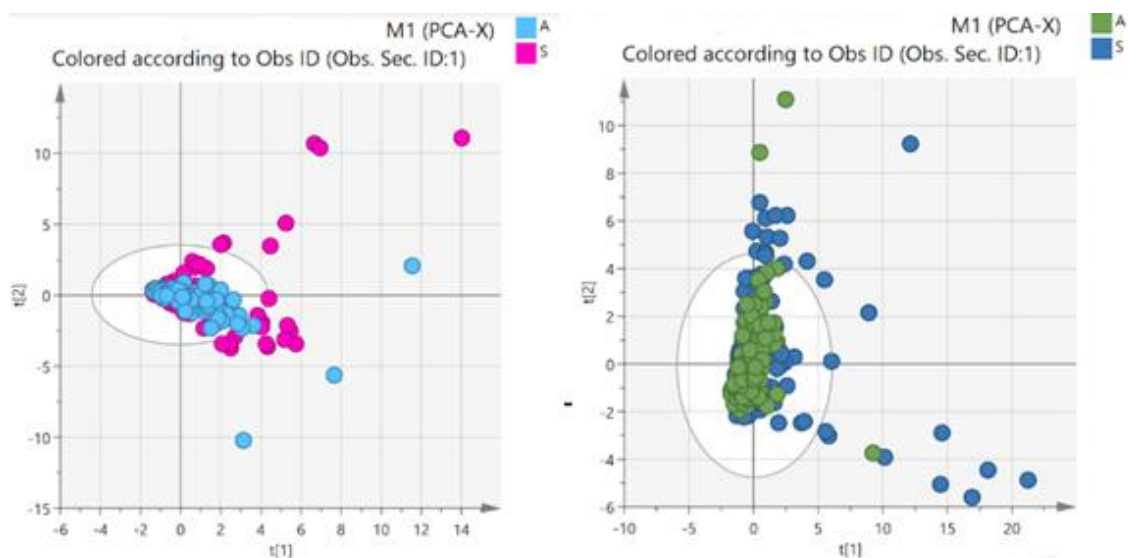
Por otro lado, se analizaron mediante un loading scatter plot (**figura 7**) para analizar la asociación de dos variables como son la 1ª componente con la 2ª. Se analizaron los genes significativos y se realizó un análisis grupal con aquellos genes que seguían una misma dispersión seleccionando aquellos que se encontraban agrupados.

Figura 7. Gráficos loading Scartter plot de la 1ª vs 2ª componente principal para los resultados de PCR digital con genes significativos. Elipse T2 de Hotelling (95%). Cada punto corresponde a un gen a analizar. P1: componente principal; P2: componente 2. A: selección de grupo de genes de la parte superior. B: selección de grupo de genes de la parte inferior.



Una vez realizado ese análisis se procedió a estudiar los grupos individualmente para ver si había correlación entre los genes seleccionados y los pacientes S y A. Para ellos se realizó un PCA, para su representación se construyó un elipse con un nivel de confianza del 95% a partir del estadístico T2 de Hotelling. En la **figura 8** se muestra como las muestras A tienden a formar parte de la elipse mientras que las S no se encuentran dentro de ésta.

Figura 8. Gráfico de dispersión de los scores de la 1ª vs 2ª componente principal para los resultados de PCR digital con genes significativos y agrupados. Elipse T2 de Hotelling (95%). Cada punto corresponde a una muestra. T1: componente principal; T2: componente 2; S: pacientes sintomáticos; A: pacientes asintomáticos.



Por último, se realizó un análisis de las muestras con el programa STRING (<https://string-db.org>), una base de datos biológica y un recurso web de interacciones proteína-proteína conocidas y previstas. Como se muestra en la **figura 9**, las proteínas PFKFB4, PDK1, SLC2A3 y PTPRB presentan una serie de interacciones entre ellas. Teniendo en cuenta esta información, se realizó un análisis de datos de PCA, para su representación se construyó un elipse con un nivel de confianza del 95% a partir del estadístico T2 de Hotelling (**figura 10**). Como se puede observar, en este caso aunque la mayor parte de las muestras se encuentran dentro del elipse, gran parte de las muestras de pacientes S se encuentran fuera de este, creando una separación mayor entre pacientes S y A.

Figura 9. Interacciones de las proteínas secretadas por los genes analizados con PCR digital y con un p value inferior o igual a 0,05.

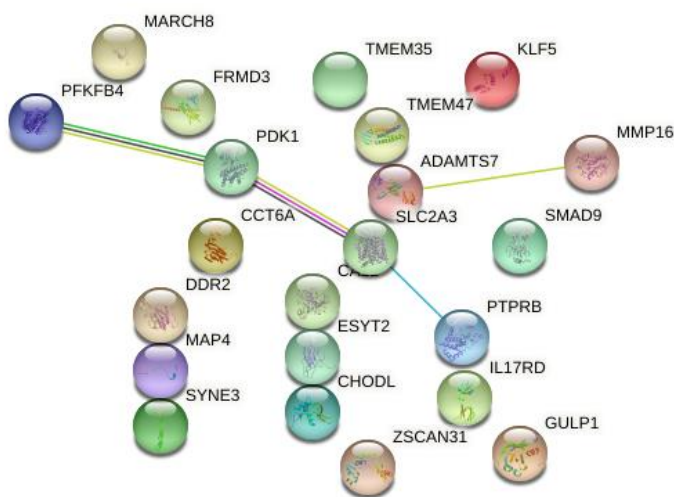
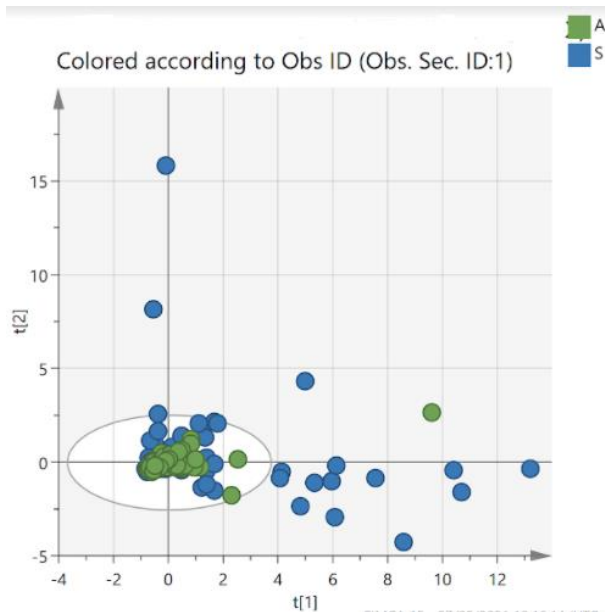


Figura 10. Gráfico de dispersión de los scores de la 1ª vs 2ª componente principal para los resultados de PCR digital con genes significativos que tienen interacción entre ellos. Elipse T2 de Hotelling (95%). Cada punto corresponde a una muestra. T1: componente principal; T2: componente 2; S: pacientes sintomáticos; A: pacientes asintomáticos.



Discusión

La aterosclerosis es una enfermedad heterogénea, provocada por una combinación de factores biológicos y ambientales que hasta el momento se desconocen (Goikuria *et al.* 2018a). La formación de la placa aterosclerótica conduce a un estrechamiento de la arteria provocando un mayor riesgo de obstrucción, especialmente cuando la placa es inestable y se rompe, provocando accidentes cerebrovasculares (Silvestre- Roig *et al.* 2014). Actualmente existen diferentes tratamientos para tratar y prevenir el desarrollo de las enfermedades cerebrovasculares, pero, aunque se conoce mayor información y se han mejorado estos tratamientos, siguen siendo limitantes, ya que su detección en algunos casos sigue siendo tardía (Ammirati *et al.* 2013). La heterogeneidad de la enfermedad hace que un único biomarcador sea insuficiente para reflejar la complejidad de esta enfermedad, por ello, la combinación de los biomarcadores es una buena opción como estrategia más sensible y específica (Jickling *et al.* 2015). Para la detección de la combinación de biomarcadores es necesaria una caracterización de los diferentes procesos implicados en la formación de la placa de ateroma. Una correcta caracterización implicaría poder realizar una predicción de los pacientes con mayor riesgo de desarrollo de una placa inestable y así encontrar un tratamiento específico para cada caso. En este estudio, se han conseguido validar genes que se encontraban diferencialmente expresados en estudios anteriores realizados a nivel de placa completa y a nivel de células el músculo liso. De los 116 genes testados se ha sido capaz de replicar 11 genes que presentan expresión diferencial tanto en CML aisladas de placa

como en la placa completa. Estos genes podrían tratarse de aspirantes a ser biomarcadores útiles para diferenciar los grupos de pacientes S de A.

Por otro lado, se han conseguido validar 3 genes con expresión diferencial significativa entre muestras S y A que están implicados en el desarrollo de la aterosclerosis a nivel de placa completa pero que en CML no. Así como, 7 genes que si se encontraban en las CML y que no se validaron en el estudio anterior a nivel de placa completa. De modo que, un total de 14 genes de los 116 genes totales estudiados han demostrado tener una expresión mayor o menor en muestras de pacientes S frente a pacientes A (**tabla 3**). La validación de estos genes determina que podrían estar implicados en el heterogéneo proceso del desarrollo de la placa de ateroma y en su inestabilidad. Además, la detección de éstos, podría ser un gran paso hacia el desarrollo de un test diagnóstico de la placa inestable. Para poder utilizar estos biomarcadores se debe tener en cuenta que se trata de genes que presentan una mayor o menor expresión y que están implicados en la producción de proteínas intracelulares. En consecuencia, para su aplicación en la clínica sería un proceso poco rentable ya que habría que extraer el RNA de las muestras y cuantificarlo para poder realizarle un análisis de expresión de los biomarcadores. Tras ello, habría que realizar un estudio para determinar el riesgo del paciente a desarrollar una placa de ateroma inestable. Para evitar este proceso tan largo y poco rentable, se podría utilizar la estrategia de detección de proteínas secretadas codificadas por los biomarcadores para el desarrollo de test rápidos. El test consistiría en la relación entre la presencia o ausencia de proteínas codificadas por genes biomarcadores con la estabilidad de la placa de ateroma. En este estudio, de un total de 14 genes validados, 5 de ellos son genes secretores de proteínas, por lo que estos resultados se podrían implementar en la práctica clínica. Dichos genes son: ADAMTS7, DDR2, IL17RD, CA12 y CHODL.

El gen ADAMTS7 codifica la metaloproteasa ADAM con el motivo 7 de la trombospondina tipo 1. Esta enzima cataliza la degradación de la proteína oligomérica de la matriz del cartílago (Qin *et al.* 2017). Estudios recientes han demostrado que una afección en la maduración de ADAMTS7 afecta a la migración de las CML produciendo riesgo de enfermedad arterial coronaria. (Pereira *et al.* 2016). Además, se ha demostrado que ADAMTS7 es eficaz inhibiendo la osificación (Mead *et al.* 2018). Tal y como se muestra en la tabla 3, el gen ADAMTS7 muestra una menor expresión en las muestras A (FC= 1,12) que en las S. Por ello, en las placas de las muestras A se ve

favorecida la osificación ya que no estaría inhibida por ADAMTS7. Este resultado concuerda con estudios previos en la estabilidad de la placa ya que se ha detectado una correlación entre la estabilidad de las placas de los pacientes A y el fenotipo osteogénico de las mismas (Alloza *et al.* 2017). Por todo ello, este gen se ha convertido en un candidato idóneo como biomarcador para la enfermedad arterial coronaria (Patel *et al.* 2013).

El gen DDR2 codifica una tirosina quinasa que funciona como receptor de superficie celular para el colágeno fibrilar y regula la diferenciación celular, la remodelación de la matriz extracelular, la migración celular y la proliferación celular, necesaria para el desarrollo normal del hueso (GeneCards; <http://www.genecards.org>). Además, regula la diferenciación de los osteoblastos y la maduración de los condrocitos.

Las enfermedades obstructivas de los vasos sanguíneos como la aterosclerosis se caracterizan por la degradación y la síntesis de nuevos componentes de la matriz extracelular además de una acumulación excesiva de células de músculo liso (Ferri *et al.* 2004). Los receptores de dominio de discoidina (DDR) 1 y 2 representan una familia de receptores de colágeno de tirosina quinasa que son activados por colágenos fibrilares. La sobreexpresión de DDR2 aumenta la degradación *in vitro* de colágeno y elastina mediada por CML (Ferri *et al.* 2004). En este estudio se ha visto una mayor expresión de este gen en aquellas muestras que presentan una placa inestable (S) frente a las muestras de la placa estable (A). Por ello, una infra expresión de este gen favorece un aumento de colágeno y elastina mediada por CML para la estabilización de la placa de ateroma. Por lo tanto, se podría decir que el gen DDR2 podría utilizarse como un posible biomarcador de estabilidad de la placa.

El gen IL17RD codifica una proteína de membrana perteneciente a la familia de proteínas del receptor de interleucina-17. Además, el producto del gen también afecta a la señalización del factor de crecimiento de fibroblastos, inhibiendo o estimulando el crecimiento a través de la señalización MAPK/ERK (GeneCards; <http://www.genecards.org>). Como ya se ha mencionado anteriormente, la inestabilidad de la placa viene dada por una menor capa fibrosa en la arteria y un mayor núcleo graso. En el presente estudio el gen IL17RD tiene una mayor expresión en las muestras A que en las S (FC= -1,5), esto puede deberse a que el gen cuando se encuentra sobre expresado estimula el crecimiento de fibroblastos. En consecuencia, esta mayor

expresión de IL17RD se traduce en una mayor estabilidad de la placa A y en un menor riesgo de desarrollar accidentes cerebrovasculares.

Otro de los genes detectados que codifican proteínas secretadas es CA12. Las anhidrasas carbónicas son una gran familia de metaloenzimas de zinc que catalizan la hidratación reversible del dióxido de carbono. Además, participan en diversos procesos biológicos, como la respiración, la calcificación, el equilibrio ácido-base, la resorción ósea y la formación de líquido cefalorraquídeo, saliva y ácido gástrico (GeneCards; <http://www.genecards.org>). En el análisis realizado, el gen CA12 se encuentra menos expresado en muestras A (FC=2,35). Esto puede deberse a que cuando se reduce la expresión de CA12 hay una mayor calcificación estabilizando la placa de ateroma.

Por último, el gen CHODL desempeña un papel en el desarrollo del sistema nervioso, como en el crecimiento y elongación de las neuritas y puede estar implicado en el crecimiento y la orientación de los axones motores (GeneCards; <http://www.genecards.org>). En relación con los resultados de expresión génica obtenidos, se observa que la expresión de CHODL es menor en muestras A (FC= 5,4). Es decir, hay una mayor síntesis de CHODL en placas de ateroma inestables. Este hecho puede suponer una desregulación en el proceso de mantenimiento de la placa de ateroma S, lo que promueve su inestabilidad e integridad.

En resumen, se han encontrado un conjunto de genes con una expresión diferencial que nos permiten discriminar entre las muestras de A y S de las placas de la arteria carótida. La detección de estos genes supone un gran avance en el desarrollo tecnológico de nuevos estudios e identificación de biomarcadores biológicos. En este estudio se han encontrado 5 genes que codifican proteínas secretadas lo que permite crear un test de detección rápida y sencilla con muestras de suero de pacientes. El objetivo de este test sería poder discriminar entre pacientes con placas estables e inestables y realizar un tratamiento específico para cada caso.

Conclusiones

Esta investigación proporciona información útil para el desarrollo de herramientas rápidas y sencillas para la detección temprana de la aterosclerosis carotídea, especialmente la placa inestable. Se ha verificado la expresión diferencial de 14 genes en 116 pruebas. Además, 11 de estos genes mostraron expresión diferencial tanto en

placas aisladas como en CML estudiados anteriormente. La identificación de estos biomarcadores representa un gran paso adelante en el desarrollo de métodos que pueden implementar resultados en la práctica clínica. También puede ayudar a evaluar a los pacientes con alto riesgo de accidente cerebrovascular, lo que lo convierte en una importante medida preventiva.

Asimismo, hay que considerar que 5 de los genes con expresión diferencial codifican proteínas secretadas (ADAMTS7, DDR2, IL17RD, CA12 y CHODL) lo cual representa una excelente oportunidad para desarrollar técnicas serológicas de detección temprana de aterosclerosis carotídea.

Bibliografía

Abdi, H., and L. J. Williams. *et al.* (2010) Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. 2:433–459. doi:10.1002/wics.101.

Ammirati, E., Magnoni, M. y Camici, P.G. (2013) Need for new non-invasive imaging strategies to identify high-risk asymptomatic patients with carotid stenosis. *International Journal of Cardiology* 168: 4342–4343.

Alloza, I., Goikuria, H., Idro, J.L., Triviño, J.C., Fernández Velasco, J.M., Elizagaray, E. *et al.* (2017) RNAseq based transcriptomics study of SMCs from carotid atherosclerotic plaque: BMP2 and IDs proteins are crucial regulators of plaque stability. *Scientific Reports* 7: e3470.

Bustamante, Alejandro, et al. "Blood biomarkers for the early diagnosis of stroke: the stroke-chip study." *Stroke* 48.9 (2017): 2419-2425

Carvajal Carvajal, Carlos. (2015). LDL oxidada y la aterosclerosis. *Medicina Legal de Costa Rica*, 32(1), 161-169. Retrieved June 13, 2021, from http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S1409-00152015000100020&lng=en&tlng=es.

Feigin, V.L., Norrving, B. y Mensah, G.A. (2017) Global Burden of Stroke. *Circulation Research* 120: 439– 448.

Goikuria, H., Freijo, M.D.M, Vega Manrique, R., Sastre, M., Elizagaray, E., Lorenzo, A. *et al.* (2018a) Characterization of carotid smooth muscle cells during phenotypic transition. *Cells* 7: e23. doi:10.3390/cells7030023.

Goikuria, H., Vandebroek, K. y Alloza, I. (2018) Inflammation in human carotid atheroma plaques. *Cytokine Growth Factor Reviews* 39: 62–70.

Dagonnier, M., Donnan, G. A., Davis, S. M., Dewey, H. M., & Howells, D. W. (2021). Acute Stroke Biomarkers: Are We There Yet?. *Frontiers in Neurology*, 12, 73.

Fernandez, V, Bellmunt- Montoya, S, Escribano- Ferrer, JM, et al. (2000) "Indications for carotid endarterectomy without arteriography. A validation study using eco-Doppler." *Revista de neurologia* 31.5: 412-416.

Ferri, N., Carragher, N. O., & Raines, E. W. (2004). Role of discoidin domain receptors 1 and 2 in human smooth muscle cell-mediated collagen remodeling: potential implications in atherosclerosis and lymphangiomyomatosis. *The American journal of pathology*, 164(5), 1575–1585. [https://doi.org/10.1016/S0002-9440\(10\)63716-9](https://doi.org/10.1016/S0002-9440(10)63716-9)

Jickling, G.C., y Sharp, F.R. (2015) Biomarker panels in ischemic stroke. *Stroke* 46: 915–920.

Kaspric, N., Reichstadt, M., Picard, B., Tournayre, J. y Bonnet, M. (2015) Protein function easily investigated by genomics data mining using the ProteINSIDE online tool. *Genomics and Computational Biology* 1: e16.

Mead, T.J., McCulloch, D.R., Ho, J.C., Du, Y., Adams, S.M., Birk, D.E. *et al.* (2018) The metalloproteinase proteoglycans ADAMTS7 and ADAMTS12 provide an innate, tendon-specific protective mechanism against heterotopic ossification. *JCI Insight* 3: e92941.

Patel, R.S. y Ye, S. (2013) ADAMTS7: a promising new therapeutic target in coronary heart disease. *Expert Opinion on Therapeutic Targets* 17: 863–867.

Pereira, A., Palma dos Reis, R., Rodrigues, R., Sousa, AC., *et al.* (2016) Association of ADAMTS7 gene polymorphism with cardiovascular survival in coronary artery disease. *Physiological genomics*, 48(11), 810–815.
<https://doi.org/10.1152/physiolgenomics.00059.2016>

Qin, W., Cao, Y., Li, L., Chen, W. y Chen, X. (2017) Upregulation of ADAMTS-7 and downregulation of COMP are associated with aortic aneurysm. *Molecular Medicine Reports* 16: 5459–5463.

Silvestre-Roig, C., de Winther, M.P., Weber, C., Daemen, M.J., Lutgens, E. y Soehnlein, O. (2014) Atherosclerotic plaque destabilization: mechanisms, models, and therapeutic strategies. *Circulation Research* 114: 214–226.

Schmittgen, T.D. y Livak, K.J. (2008) Analyzing real-time PCR data by the comparative CT method. *Nature Protocols* 3: 1101–1108.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J. *et al.* (2018) Ensembl 2018. *Nucleic Acids Research*, 46: 754–761.