eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

Transcriptome regulation network in Multiple sclerosis: Role of circular RNAs.

Leire Iparraguirre Gil

To obtain a doctoral degree in Molecular Biology and Biomedicine from the University of the Basque Country

Supervisors:

David Otaegui Bichot

Maider Muñoz Culla

February 2021

This thesis has been carried out at the Biodonostia Health Research Institute. During this work, I was supported by a fellowship from the Department of Education of the Basque Government (PRE_2017_2_0220, 2017-2020) and by an EMBO Short-Term Fellowship (STF_8188, 2019).

When it comes to RNA,
assume nothing.

Aita eta amari,
Olatzi,

Laburpena

Esklerosi anizkoitza (EA) nerbio sistema zentraleko (NSZ) gaixotasun konplexu eta kronikoa da. Immunitate sistemak gure neuronen axoiak estaltzen eta babesten dituen mielina erasotzen duenean agertzen da gaixotasuna. Hortaz, gaixotasun autoimmune eta desmielinizatzailea ere bada. Desmielinizazio honek, alde batetik, nervio bulkadaren transmisio egokia oztopatu edo eteten du. Bestalde, lesio desmielinizatzaile hauek axoien eta hortaz neuronen kalte eta degenerazioa eragiten dute. Honek guztiak, hainbat sintoma neurologikoen agerpena dakar, denborarekin desgaitasunera eraman dezaketenak. Gainera, EAren sintomak askotarikoak izan daitezke lesioen kokapen, kopuru eta hedaduraren arabera. Ondorioz, ohikoa da paziente bakoitzak sintoma desberdinak edukitzea, eta baita sintoma hauek denborarekin aldatzen joatea ere. Beraz, mila aurpegietako gaixotasuna bezala ere ezagutzen da EA.

Mundu osoan 2,5 milioi pertsonari baino gehiagori eragiten die EAk, eta haren prebalentzia handitzen ari da. Orokorrean, gaixotasuna 20 eta 40 urte bitarteko pertsonetan agertzen da lehen aldiz. Izan ere, trafiko-istripuen ondoren, gazteen artean desgaitasunaren bigarren kausa da. Nabarmentzekoa da EA hiru aldiz ohikoagoa dela emakumeen artean gizonen artean baino, eta hainbat ikerketa lanen arabera proportzio hau, beste gaixotasun autoimmune batzuetan ikusi den bezala, handitzen doa.

Gaixotasunaren fisiopatologia konplexua da eta dagoeneko ezaguna da hainbat faktore genetiko eta inguruneko faktoreek laguntzen dutela gaixotasunaren garapenean. EAren hiru forma kliniko edo mota desberdin bereizi daitezke. Paziente gehienei (%85 inguru) agerraldi neurologiko eta errekuperazioetan garatzen den mota diagnostikatzen zaie EA-RR mota deritzona (ingelesetik, repalsing-remitting). EA-RRaren ezaugarri nabarmenena larriagotze klinikoak edo agerraldiak dira. Agerraldi hauetan, immunitate-sistemako zelulak aktibatu egiten dira eta NSZean sartzen dira bertan sintomak agertzea ekarriko duten desmienilizazioa eta hantura eraginez. Agerraldi batean azaleratzen diren sintomek gutxienez 24 ordu iraungo dute, eta horren ondoren hantura gutxituko da eta birmielinizazio osoa edo partziala emango da. Bigarrengo fase honi erremisiofasea deitzen zaio. Denborarekin ordea agerraldiak ez dira guztiz errekuperatzen, eta ondorioz desgaitasuna

pilatzen joaten da. Paziente askotan, une jakin batetik aurrera desgaitasunaren pilaketa progresiboa izango da, ia agerraldirik gabe. Gaixotasunaren fase hau bigarren azpimota gisa sailkatzen da: EA sekundario progresiboa (EA-SP). Bestalde, pazienteen %10-15ek, desgaitasunaren pilaketa progresiboa pairatzen dute hasieratik, EA primario progresibo edo EA-PP bezala ezagutzen dena.

Gaur egun, EAren diagnostikoa McDonaldek 2010ean proposatutako irizpideei jarraituz egiten da. Irizpide hauek ebidentzia klinikoetan oinarritzen dira eta NSZean lesio desmielinizatzaileak daudela egiaztatu behar dute. Lesio horiek, gainera, denboran eta espazioan sakabanatuta egon behar dute (NSZeko hainbat gunetan agertu behar dira eta momento desberdinetan). Ebidentzia kliniko horien artean daude sintoma neurologikoak, azterketa fisikoan agertutako zeinuak eta proba diagnostiko desberdinak. Bi dira proba diagnostiko ohikoenak: erresonantzia magnetikoak eta likido zefalorrakideoaren (LZR) probak. Alde batetik, erresonantzia magnetikoak, NSZan lesioak aurkitzea ahalbidetzen du. Bestalde, LZR probak, IgG motako immunoglobulinen sintesi intratekala iradokiko luketen banda oligoklonalen presentzia aztertzen du, NSZan erantzun immune bat ematen ari den konprobatuz. Hala ere, gaixotasunaren bilakaera eta fenotipo klinikoa oso aldakorrak dira paciente desberdinen artean, baita paziente berean ere. Horregatik, beharrezkoa da paziente horiek maila desberdinetan karakterizatzea, gaixotasunean parte hartzen duten mekanismo patologikoak hobeto ulertu ahal izateko. Karakterizazio honek gainera, biomarkatzaile berrien aurkikuntza ekarri dezake, eta horietaz baliotuta diagnostikoa hobetu, azpimotak modu errazeagoan bereizi, gaixotasunaren progresioa jarraitu eta tratamenduei emandako erantzuna behartu, aurreikusi eta kontrolatu genitzake.

Kontuan hartuta gaixotasunak NSZri eragiten diola, LZR lagina litzateke biomarkatzaileen iturririk zuzenena. Hala ere, LZR lagina gerriko ziztada baten bitartez lortu behar da eta ez da komeni horrelako ziztada ugari egitea, izan ere, oso prozedura inbaditzailea da eta ondorio kaltegarriak izan ditzake. Horregatik, azken urteotan ahalegin handia egin da hain inbaditzaileak ez diren biomarkatzaileak aurkitzeko, odol-biomarkatzaileak kasu.

Horretarako, eta teknika omikoen etorrerari esker, eskala handiko azterketak egin dira, hainbat maila desberdinetan. Horietako bat adierazpen genikoa edo transkriptomaren azterketa da, biomarkatzaile berriak aurkitzeaz gain, gaixotasunaren fisiopatologiaren atzean dauden mekanismo molekularrak argitzeko aukera ematen duena. Pazienteen odol periferikoan egindako ikerketek erakutsi dute bai RNA kodifikatzaileak eta bai ez-kodifikatzaileak desberdin adierazten direla EAko pazienteetan kontrol osasuntsuekin alderatuta. Gainera, aipatzekoa da desberdintasunak ikusi direla geneen adierazpenean

pazienteen sexuaren arabera. Pazienteetan desberdin adierazten diren transkrito horietako batzuk biomarkatzaile gisa proposatu dira, baina horietako bat ere ez da erabiltzen oraindik praktika klinikoan.

Azken urteotan, deskribatutako RNA ez-kodifikatzaileen zerrenda luzean RNA mota berri bat sartu da, RNA zirkularrak edo circRNAk. RNA hauen ezaugarririk bereizgarriena egitura zirkularra da. Transkripzio osteko erregulazio-mekanismoetan eginkizun garrantzitsua dutela deskribatu da eta era berean, hainbat ikerlanen arabera, circRNAek hainbat prozesu fisiologikotan eta patologikotan parte hartzen dute. Izan ere, hainbat gaixotasunekin lotu izan dira, besteak beste, minbizia, gaixotasun kardiobaskularrak, gaixotasun neurologikoak eta gaixotasun autoimmuneak. Gainera, circRNAk ere etorkizun handiko biomarkatzaile gisa aurkezten dira, egitura zirkularrak ematen dien egonkortasun handiagatik. Hainbat ikerlanek jada proposatu dituzte biomarkatzaile gisa hala nola odolean, serumean, listuan, gernuan eta likido zefalorrakideoan, bai eta zelulaz kanpoko besikuletan (EV, ingelesetik, extracelular vesicles) ere.

Testuinguru honetan, gaixotasunaren prozesu patologikoen ezagutzan sakontzeko eta biomarkatzaile sendo eta erreproduzigarri baten beharrari erantzuteko asmoz, tesi hontan circRNAen adierazpenaren karakterizazioa egin dugu EAn.

Tesi honen lehen kapituluak, lagin mota ezberdinetan circRNAen adierazpena aztertzen duten lau ikerlan ezberdin jasotzen ditu. Bertan gaixotasunaren fenotipo eta lagin mota desberdinak analizatzen dira, hala nola leukozitoak, odol periferikoko zelula mononuklearrak (PBMCak, ingelesetik peripheral blood mononuclear cells) eta EVak.

Lehen ikerlan batean, mikroarray bidez 13.617 circRNAen adierazpena analizatu eta alderatu dugu RR motako EA pairatzen duten 4 emakumeren eta osasuntsu dauden 4 emakumeren leukozitoetan. Ikerlan hau da EAn circRNAen deserregulazioa identifikatzen duen lehena. Gainera, bi circRNAen deserregulazioa balioztatu dugu bigarren kohorte independente batean. Bi circRNA horien adierazpena, gene berdinak sortzen duen RNA mezulariarekin batera, murriztua dago gaixotasunaren hasierako etapetan, eta handitzen doa gaixotasunaren eboluzio-denborarekin. Horregatik, circ_0005402 eta circ_003452_2 proposatzen ditugu, bai eta ANXA2 mRNA ere, gaixotasunaren diagnostikorako biomarkatzaile potentzial gisa. Horiek, gainera, eta oraindik ere azterlan gehiago egitea beharrezkoa den arren, gaixotasunaren lehen faseetan ematen den inflamazio-osagaia monitorizatzeko erabil litezke.

Bigarren ikerlana kohorte handiago batera zabaldu dugu; kohorte horrek gaixotasunaren azpimota desberdinak eta sexu desberdinak barne hartzen ditu. Kasu honetan, RNASeq bidez transkriptoma zirkularraren eta linealaren adierazpena aztertu dugu EA-RR pazienteen 20 laginetan, EA-SP pazienteen 10 laginetan eta 20 kontrol osasuntsuetan. Emaitzen arabera, beste gaixotasun autoimmune batzuetarako deskribatu den bezala, EAn ere circRNAen deserregulazio orokorra ematen da, transkripzio linealari eragiten ez diona. Kasu honetan, pazienteen artean circRNAen adierazpenean gorakada orokorra ikusten dugu, eta horietako sei circRNAen adierazpenaren igotzea balioztatu dugu beste kohorte batean. Gainera, sei circRNA hauek sentsibilitate- eta espezifikotasun-balio onak erakusten dituzte, gaixotasunaren biomarkatzaile gisa oraindik circRNA hauen baliogarritasuna aztertzen jarraitzea beharrezkoa den arren. Aipatzekoa da ere, beste RNA mota batzuetarako deskribatu den bezala, circRNAen adierazpen-profila aldatu egiten dela emakumeen eta gizonen artean. EA daukaten emakumeek are joera handiagoa dute circRNAen gehiegizko adierazpenerako, eta gizonetan, berriz, joera hau ez da hain nabarmena.

Hirugarren azpikapituluak pronostiko desberdina daukaten pazienteen transkriptoma lineala eta zirkularra ezaugarritzea du helburu. Gaur egun, LZRko azterketa egiten da eta IgM banda oligoklonalen presentzia ezagutzeko eta hauek pronostiko txarraren markatzaile gisa hartzen dira. Ikerlan hontan PBMCtako adierazpena aztertuko dugu LZRko IgM markatzailerarekin korrelazionatu dezakeen beste odol-markatzaile bat aurkitzeko helburuarekin. Kasu hontan, aldez aurretik LZR froga eginda daukaten 8 pazienteen adierazpena aztertu dugu PBMCtan, 4 IgM+ eta 4 IgM-. Planteatutako helburua betetzen duten lau transkrito aurkitu eta balioztatu ditugu, bi lineal eta bi zirkular. Beraz, pronostikorako biomarkatzaile kandidatu gisa aurkezten ditugu, haien potentziala etorkizuneko esperimentuetan aztertzen jarraitu beharko den arrren.

Transkriptoma zirkularraren karakterizaziorako azken azterlan gisa, 10 EA-RR pazienteren, 10 EA-SP pazienteren eta 8 kontrol osasuntsuren plasmatik isolatutako EVen transkriptoma ikertu dugu. Aurretik egindako azterlanek erakutsi dute jada pazienteen EVetako transkriptoma eta kontrol osasuntsuena desberdinak direla. Baina bakar batek ere ez zuen ikertu circRNAen presentzia EVtan. Gure emaitzek besikuletan circRNAk daudela baieztatzeaz gain, RNA mezularien ondoren bigarren RNArik ohikoena dela erakusten dute. Nabarmentzekoa da gainera, desberdintasunak daudela besikuletan eta leukozitoetan aurkitutako circRNAen adierazpen-profilaren artean, baita bi lagin mota desberdinetan aurkitzen ditugun circRNAen tamainan ere. Aurkikuntza horrek, EVtan esportatzen diren circRNAk modu erregulatuan aukeratzen dituen mekanismo bat egon daitekela adierazten du. Azkenik, emaitzek adierazten dute ere circRNA eta RNA lineal desberdinak daudela

EA-RR pazienteen, EA-SP pazienteen eta kontrolen EVetan hiru talde hauek desberdintzeko balioko luketenak. Hortaz gain hiru taldeetako besikuletan aurkitu daitezken RNAk ere adierazpen-maila desberdinak dituzte. Aurreko azpikapituluetako ikerlanetan bezala, transkribatu horiek biomarkatzaile gisa erabil daitezke, baina, lehenik, beste kohorte batzuetan balioztatu beharko dira.

Tesi honen bigarren atalean, lehenengo kapituluan egindako esperimentuetan aurkitutako circRNA interesgarrienen funtzioa aztertu nahi izan dugu, gaixotasunarekin duten lotura ulertu ahal izateko.

CircRNA gehienen funtzioa oraindik ezezaguna da. Hala ere, deskribatu zen lehen funtzioa eta gehien ikertu dena miRNAk lotzeko gaitasuna izan da. MiRNAk lotuz, circRNAek miRNAen funtzioa oztopatzen dute hauek erregulatzen dituzten geneen adierazpenean eraginez. Dena den, gero eta azterlan gehiagok adierazten dute funtzio hau betetzeko, circRNAek adierazpen-maila altua eta miRNA berarentzako lotura-gune kopuru handiak eduki behar dituztela. Premisa hori oinarri hartuta, baheketa bat egin dugu, aurrez leukozitoetan eta EVetan identifikatu ditugun circRNA guztien artean miRNAen lotura-guneak identifikatzeko, bai eta hautagai gisa proposatutako circRNA interesgarrienetan ere. CircRNA horietatik oso gutxik dute miRNA lotura-gune kopuru handia, eta horietako bakar batek ere ez du ausaz espero baino lotura-gune gehiago. Hortaz, EArekin lotutako circRNAk miRNAen funtzioa oztopatzearena baino beste funtzioren bat eduki behar dutela ondorioztatzen dugu.

CircRNAetan aztertu diren gainontzeko funtzioak ez daude hain ondo deskribatuta. Hala ere, circRNAk erantzun immunea erregulatu dezakalatela proposatu da. Funtzio hau aurrera eramateko deskribatutako mekanismoetariko bat da R proteina kinasari (PKR) lotu eta bere aktibazioa inhibitzearena horrela erantzun immunea ere inhibituz. EA bezalako gaixotasun batean erantzun immuneak duen garrantzia dela eta, leukozitoetan eta EVetan detektatutako circRNAk PKRri lotzeko izan dezaketen potentziala aztertzea erabaki genuen. CircRNAk PKRri lotzeko urkila motako egiturak sortu behar dituzte, hortaz leukozitoetan eta EVetan detektatutako circRNAen egitura aztertu dugu bai eta hautagai gisa proposatutako circRNArik interesgarrienena ere. CircRNA kopuru handi batek bigarren mailako egiturak osatzen dituztela erakusten dute emaitzek, horien artean lehen kapituluan nabarmendutako hautagaietako batzuk.

Beraz, tesiaren bigarren atal honen bigarren azpikapituluan, circRNAek PKRrekin duten interakzioa aztertu genuen HeLa zeluletan, poly(I:C) bidezko estimulazioaren bidez eragindako erantzun immunearen eredu batean. Emaitzen harabera, poly(I:C) bidezko

estimulazioak circRNAen adierazpena gutxitu, PKRren fosforilazioa eragin eta erantzun immunean inplikatutako geneen adierazpena handitu egiten ditu. Estimulua desagertu eta gero ordea, circRNAen adierazpena berriro berreskuratzen da, eta, aldi berean, gene immuneen adierazpena normaltasunera itzultzen da. Egiaztatu dugu, halaber, zeluletan bigarren mailako egiturak dituzten circRNAen gehiegizko adierazpena induzituz gero, poly(I:C)ak eragiten duen fenotipoa partzialki sahiestu daitekela. Kontrol-zelulekin alderatuta, bigarren mailako egiturak dituen circRNA bat gehiegi adierazten duten zelulek PKR fosforilazio maila txikiagoa dute, bai eta erantzun immunean inplikatutako geneen aktibazio txikiagoa ere. CircRNAetan funtzio hori baieztatzeko, esperimentu gehiago egin beharko direla argi dago. Hala ere, azterlan honek funtzio hori balioztatu eta circRNAek erantzun immunea erregulatzeko daukaten gaitasuna aztertu dezaketen etorkizuneko lanetarako oinarriak finkatzen ditu.

Laburpen gisa esan dezakegu tesi honek EAren transkriptomaren ezagutzan aurrera egiten lagundu duela gaixotasun honetan lehen aldiz circRNAen adierazpena ezaugarritzearekin batera. Egindako ikerketen emaitzek erakusten digute circRNAk nolabait gaixotasunean inplikatuta daudela eta ildo horretan ikertzen jarraitzeko oinarriak ezartzen ditu. Gainera, linea zelularretan egindako azterketek, oso atarikoak izan arren, circRNAek EArekin duten harremana azaltzen duten funtzioetako bat sistema immunearen erregulazioan parte hartzea izan daitekela erakusten dute. Azkenik, tesian zehar hainbat circRNA proposatzen dira gaixotasunaren biomarkatzaile gisa (baita RNA lineal batzuk ere) etorkizuneko azterlanetan kontuan hartu beharko liratekeenak noizbait praktika klinikoan erabiltzera iritsi ahal izateko.

Resumen

La esclerosis múltiple (EM) se define como una enfermedad compleja, autoinmune y desmielinizante que afecta de manera crónica al sistema nervioso central (SNC). En ella el sistema inmunitario ataca a la mielina que recubre y protege los axones de las neuronas. Esta desmielinización entorpece o directamente interrumpe la transmisión del impulso nervioso. Estas lesiones desmielinizantes se asocian además con el daño y la degeneración axonal y neuronal, lo que resulta en la aparición de una serie de síntomas neurológicos y la acumulación con el tiempo de una discapacidad neurológica. Los síntomas en la EM varían en función de la localización, la extensión y la cantidad de lesiones que se presentan. Son además diferentes para cada paciente, por lo que se le conoce también como la enfermedad de las mil caras.

La EM afecta a más de 2,5 millones de personas en todo el mundo y su prevalencia va en aumento. Habitualmente se diagnostica en adultos jóvenes de entre 20 y 40 años, y es de hecho la segunda causa de discapacidad entre los jóvenes tras los accidentes de tráfico. Cabe destacar que la EM es tres veces más frecuente en mujeres que en hombres y hay estudios que demuestran que esta proporción va en aumento, fenómeno que se observa también en otras enfermedades autoinmunes.

La fisiopatología de la enfermedad es compleja con factores tanto genéticos como ambientales que contribuyen al desarrollo de la enfermedad. Se distinguen diferentes formas clínicas siendo para la mayoría de los pacientes (alrededor del 85%) la forma remitente recurrente de la enfermedad o EM-RR el diagnóstico inicial. La EM-RR se caracteriza por exacerbaciones clínicas, o brotes, causados por células inmunes autoreactivas que migran hasta el SNC, dando lugar a una inflamación focal, a la desmielinización y a la consecuente aparición de síntomas. Tras un brote, que dura al menos 24h, la inflamación se resuelve y se produce una remielinización (completa o parcial), entrando en una fase de recuperación denominada remisión. Con el tiempo, para muchos pacientes de EMRR la recuperación comienza a ser incompleta, lo que lleva a la acumulación de discapacidad. Llegado un momento, la acumulación de la discapacidad pasa a ser progresiva sin apenas brotes. Esta fase de la enfermedad se clasifica como un segundo subtipo llamado EM secundaria progresiva (EM-SP). Por otro lado existe un 10-15% de los pacientes, que en lugar de tener

una fase inicial remitente recurrente, presentan una acumulación progresiva de la discapacidad en ausencia de brotes desde el inicio, un curso conocido como EM primaria progresiva o EM-PP.

Hoy en día el diagnóstico de la EM se realiza atendiendo a los criterios propuestos por McDonald en 2010. Los criterios diagnósticos se basan en evidencias clínicas que constaten la presencia de lesiones desmielinizantes en el SNC. Estas lesiones deben además estar diseminadas en el espacio (deben presentarse en distintas zonas del SNC) y el tiempo. Estas evidencias clínicas incluyen síntomas neurológicos, signos revelados durante el examen físico o diferentes pruebas diagnósticas. Las pruebas diagnósticas más comunes son la resonancia magnética, que permite localizar las lesiones, y la prueba del líquido cefalorraquídeo (LCR) donde se busca confirmar la presencia de bandas oligoclonales que sugerirían la síntesis intratecal de IgG y por tanto la existencia de una respuesta immune en el SNC. Sin embargo, tanto el curso de la enfermedad como el fenotipo clínico de los y las pacientes de esclerosis múltiple son muy variables entre pacientes y también a lo largo del tiempo en un mismo individuo. Por esta razón, es necesaria la caracterización de estos pacientes a distintos niveles para así poder tener una mejor comprensión de los mecanismos patológicos implicados en la enfermedad que permitan además descubrir biomarcadores útiles en la mejora del diagnóstico, la diferenciación de los fenotipos de la EM, el seguimiento de la progresión de la enfermedad o en el control de la respuesta al tratamiento.

Teniendo en cuenta que la EM afecta al SNC, el LCR es la fuente más directa de biomarcadores. Sin embargo, el LCR no es una fuente ideal de biomarcadores que vayan a ser utilizados de manera continua, ya que no se recomienda realizar múltiples punciones lumbares debido a su carácter invasivo y a sus posibles efectos adversos. Es por ello que, durante los últimos años, se ha hecho un gran esfuerzo por descubrir biomarcadores menos invasivos, como los biomarcadores sanguíneos.

Para ello, y gracias a la llegada de las técnicas ómicas, se han llevado a cabo estudios a gran escala con distintos abordajes. Uno de ellos es el estudio de expresión génica o transcriptoma que, además de descubrir nuevos biomarcadores, permiten también dilucidar los mecanismos moleculares que subyacen a la fisiopatología de la enfermedad. Estudios de expresión génica realizados en sangre periférica de pacientes han demostrado una desregulación tanto del transcriptoma codificante, así como del no codificante. Además, cabe mencionar que se han observado también diferencias a nivel transcriptómico en función del sexo de los pacientes. Algunos de los transcritos identificados como alterados en la enfermedad se han propuesto

como biomarcadores de la EM, sin embargo, ninguno de ellos está siendo utilizado en la práctica clínica aún.

En los últimos años, se ha incorporado a la larga lista de RNAs no codificantes descritos, un nuevo tipo de ARN, los ARN circulares o circRNAs. Destacan por su estructura circular y se ha descrito que juegan un papel importante en los mecanismos de regulación post-transcripcional. Asimismo, diferentes estudios afirman que los circRNAs participan en varios procesos tanto fisiológicos como patológicos. De hecho, han sido asociados a diversas enfermedades, entre ellas el cáncer, las enfermedades cardiovasculares, las enfermedades neurológicas y las enfermedades autoinmunes. Además, los circRNAs se presentan también como biomarcadores prometedores debida a la gran estabilidad que les confiere su estructura circular. Diversos estudios los han propuesto ya como biomarcadores en muestras como la sangre, el suero, la saliva, la orina y el líquido cefalorraquídeo, así como en las vesículas extracelulares (EVs).

En este contexto, y debido a la creciente necesidad de un biomarcador sólido y reproducible para la EM, esta tesis surge con el objetivo de realizar una caracterización de la expresión de los circRNAs en la esclerosis múltiple con el objeto último de profundizar en el conocimiento de los procesos patológicos de la enfermedad, así como de identificar nuevos potenciales biomarcadores.

A lo largo del primer capítulo de esta tesis presentamos cuatro aproximaciones diferentes que estudian la expresión de los circRNAs en distintos tipos de muestra como son los leucocitos, las células mononucleares de sangre periférica (PBMCs) y las vesículas celulares (EVs) combinados con distintos fenotipos de la enfermedad.

En una primera aproximación hemos comparado la expresión de 13.617 circRNAs en leucocitos de 4 mujeres diagnosticadas de RR-MS y 4 mujeres sanas mediante microarrays de expresión. Este estudio es el primero que identifica una desregulación de los circRNAs en la esclerosis múltiple de entre los cuales hemos podido validar la desregulación de dos circRNAs en una segunda cohorte independiente. La expresión de estos dos circRNAs, junto con el RNA mensajero del gen que los produce se encuentra disminuida en las etapas iniciales de la enfermedad mientras que va en aumento con el tiempo de evolución. Por ello, proponemos los circ_0005402 y circ_003452_2 así como el mRNA ANXA2 como potenciales biomarcadores diagnósticos de la enfermedad que podrían además ser utilizados para monitorizar el componente inflamatorio característico del inicio de la enfermedad a falta de más estudios que confirmen su utilidad.

En una segunda aproximación ampliamos el estudio de expresión en leucocitos a una cohorte más grande que incluye tanto distintos subtipos de la enfermedad como distintos sexos. En este caso analizamos la expresión tanto del transcriptoma circular como del lineal mediante RNASeq en 20 muestras de pacientes EM-RR, 10 pacientes EM-SP y 20 controles sanos. En línea con lo descrito para otras enfermedades autoinmunes, encontramos una desregulación global del transcriptoma circular, que no afecta al transcriptoma lineal. En este caso, observamos un aumento general en la expresión de los circRNAs en pacientes, y validamos la sobreexpresión de seis de estos circRNAs que además muestran buenos valores de sensibilidad y especificidad como biomarcadores de la enfermedad aunque deben seguir siendo estudiados. Cabe destacar que tal y como ya había sido anteriormente descrito para otro tipo de transcritos el perfil de expresión de los circRNAs varía entre mujeres y hombres. Las mujeres muestran una tendencia mayor aún hacia la sobreexpresión de circRNAs mientras que en hombres la proporción entre circRNAs sobreexpresados e infraexpresados está más equilibrada.

El tercer subcapítulo tiene como objetivo caracterizar el transcriptoma lineal y circular en PBMCs de pacientes en los que se ha estudiado previamente la presencia de bandas oligoclonales de IgM en LCR (4 IgM+ y 4 IgM-). La presencia de estas bandas IgM en LCR se ha definido como marcador de mal pronóstico, con una gran utilidad de cara al tratamiento de estos pacientes. Sin embargo, dado que la toma de muestra de LCR presenta serios inconvenientes, en este estudio buscamos encontrar transcritos en sangre periférica que pudieran correlacionar con este marcador y así sustituirlo en la práctica clínica. Encontramos y validamos cuatro transcritos, dos lineales y dos circulares que cumplen el objetivo planteado por lo que su potencial como biomarcadores pronósticos debe seguir siendo estudiado en futuros experimentos.

Como última aproximación a la caracterización del transcriptoma circular en la EM, investigamos el transcriptoma de las EVs aisladas de plasma de 10 pacientes EM-RR, 10 pacientes EM-SP y 8 controles sanos. Estudios previos habían demostrado cambios en la concentración de transcritos en EVs de pacientes frente a los de controles, sin embargo, ninguno había analizado la presencia de circRNAs en el contenido de las EVs. Nuestros resultados no solo confirman la presencia de circRNAs en las EVs si no que los colocan como el segundo transcrito más frecuente después de los RNAs mensajeros. Además, encontramos interesantes diferencias entre el perfil de expresión de circRNAs en EVs y el perfil de expresión descrito en leucocitos. Este hallazgo, unido a una diferencia significativa en el tamaño de los circRNAs presentes en cada uno de los dos tipos de muestra es indicativo de un mecanismo que regula de forma selectiva los circRNAs que son exportados en las EVs.

Por último, pero no menos importante, los resultados indican la presencia de circRNAs y RNAs lineales en EVs específicos tanto de pacientes EM-RR como de pacientes EM-SP y controles. Además, se encuentran también otra serie de transcritos que, aunque son detectados en los tres grupos analizados, también muestran diferencias en sus niveles de expresión. Al igual que en las aproximaciones previas, estos transcritos podrían potencialmente ser utilizados como biomarcadores, sin embargo, deben primero ser validados en otras cohortes.

En el segundo capítulo de esta tesis, hemos realizado una recopilación de los circRNAs más interesantes encontrados en los experimentos incluidos en primer capítulo con el objetivo de hacer una primera aproximación al estudio funcional de los mismos para así poder entender la manera en la que están implicados en la enfermedad.

La primera función descrita para un circRNA fue la de unir miRNAs y secuestrarlos de manera que estos no pudieran realizar su función. Varios estudios posteriores han investigado el potencial de los circRNAs para realizar esta función, y a pesar de que hay evidencias de que unos pocos podrían ejercer esta función son cada vez más los estudios que indican que sólo un número limitado de circRNAs con unos niveles de expresión altos y un gran número de sitios de unión para un mismo miRNA podrían funcionar como esponjas de miRNAs. En base a esta premisa, hemos realizado un cribado para identificar los sitios de unión a miRNAs entre todos los circRNAs que habíamos identificado previamente en leucocitos y EVs, así como en los circRNAs más interesantes propuestos como candidatos. Sin embargo, muy pocos de estos circRNAs presentan un alto número de sitios de unión y ninguno de ellos tiene más sitios de unión de los esperados al azar. Por ello, concluimos que ninguno de los circRNAs relacionados con la EM muestra potencial como para ejercer la función de esponja de miRNA.

Además de la función de esponja de miRNA, que es la más estudiada entre los circRNAs otras funciones han sido propuestas. Sin embargo, dada la importancia de la respuesta inmune en una enfermedad como la EM, decidimos estudiar el potencial de los circRNAs detectados en leucocitos y en EVs así como de los circRNAs más interesantes propuestos como candidatos, para interactuar con la proteína quinasa R o quinasa dependiente de RNAs de doble cadena (PKR). La PKR es una de las proteínas más importantes involucrada en la respuesta inmune innata. Un estudio reciente en pacientes con lupus eritematoso sistémico ha descrito que la gran mayoría de los circRNAs forman estructuras de tipo horquilla que les permiten interaccionar con la PKR e inhibir su activación inhibiendo así la sobreactivación de la respuesta inmune. Motivados por la hipótesis de que los circRNAs

podrían estar regulando la respuesta immune en la EM mediante la interacción con PKR, hemos estudiado la estructura de los circRNAs detectados en leucocitos y EVs para conocer su potencial para formar las horquillas que les permitirían unirse a PKR. Encontramos que un número considerable de circRNAs forman estructuras secundarias, entre ellos varios de los candidatos destacados en el primer capítulo.

Por lo tanto, en la segunda parte de este segundo capítulo de la tesis, estudiamos la interacción de los circRNAs con la PKR en células HeLa en un modelo de respuesta inmune inducida mediante la estimulación con poly (I:C). En este estudio observamos que la estimulación con poly(I:C) provoca una infraexpresión generalizada de los circRNAs, la fosforilación de la PKR y la sobreexpresión de los genes implicados en la respuesta inmune. Pasado un tiempo, sin embargo, la expresión de los circRNAs vuelve a recuperarse a la vez que la expresión de los genes inmunes vuelve a la normalidad. Comprobamos también que la sobreexpresión de circRNAs que contienen estructuras secundarias es capaz de revertir parcialmente el fenotipo observado ante la estimulación con poly (I:C). En comparación con las células control, aquellas células en las que se sobreexpresa un circRNA con estructuras secundarias presentan una menor fosforilación de la PKR, así como una menor activación de los genes implicados en la respuesta inmune. Para confirmar este mecanismo de acción en los circRNAs, será necesario realizar más experimentos, sin embargo, este estudio sienta las bases para futuros trabajos que validen esta función y analicen el potencial de los circRNAs como agentes terapéuticos capaces de regular la respuesta inmune en enfermedades como la EM.

Finalmente, a modo de resumen podemos decir que la presente tesis ha contribuido al avance del conocimiento del transcriptoma en la EM. Hemos caracterizado por primera vez la expresión de los circRNAs en esta enfermedad y hemos descrito la desregulación de los mismos. Los resultados de estos estudios nos muestran que los circRNAs están de alguna manera implicados en la enfermedad y sientan las bases para seguir investigando en esta línea. Además, los estudios preliminares realizados en lineas celulares indican que una de las posibles funciones que explican la relación de los circRNAs con la EM es su participación en la regulación del sistema inmune mediante la interacción con PKR. Gran parte de los resultados han sido o están en proceso de ser trasladados a la comunidad científica a través de publicaciones (4 trabajos aceptados y 2 en preparación). Por último, a lo largo de la tesis se proponen varios circRNAs así como algunos RNAs lineales que han mostrado potencial como biomarcadores de la enfermedad que deberían tenerse en cuenta para futuros estudios que logren acercar alguno de estos transcritos a la práctica clínica.

Table of contents

Abbreviations

eccDNA	Extrachromosomal circular DNA
ncRNA	Non-protein coding RNA
Pol I	RNA polymerase I
rRNA	Ribosomal RNA
Pol II	Polymerase II
mRNA	Messenger RNA
miRNA	MicroRNA
circRNA	Circular RNA
Pol III	RNA polymerase III
tRNA	Transfer RNA
PIC	Preinitiation complex
TF	Transcription factor
CTD	C-terminal domain
PPP	Promoter Proximal Pausing
CBC	Cap binding complex
RBP	RNA binding protein
UTR	Untranslated region
snoRNA	Small nucleolar RNA
RISC	RNA-induced silencing complex
AGO	Argonaute
ecircRNA	Exonic circRNA
EIcircRNA	Exonic-intronic circRNA
CiRNA	Intronic circRNA
BSJ	Back-splicing junction
EV	Extracellular vesicle
m^6A	N^6-methyladenosine
ORF	Open reading frame
IRES	Internal Ribosome Entry Site
PAMP	Pathogen-associated molecular pattern
RIG-I	Retinoic acid inducible gene I protein
PKR	Protein kinase R or interferon induced double stranded RNA activated kinase
OAS	Oligoadenylate synthetase
dsRNA	Double stranded RNA
SLE	Systemic lupus erythematosus
MS	Multiple Sclerosis
CNS	Central Nervous System
HLA	Human leukocyte antigen locus
EBV	Epstein-Barr virus
BBB	Brain blood barrier

CIS	Clinically isolated syndrome
MRI	Magnetic resonance imaging
RR-MS	Relapsing-Remitting MS
SP-MS	Secondary-progressing MS
PP-MS	Primary progressive MS
CSF	Cerebrospinal fluid
NMO	Neuromyelitis óptica
PBMCs	Peripheral blood mononuclear cells
HC	Healthy controls
EDSS	Expanded Dissability Status Scale
AOO	Age of onset
FC	Fold change
RT-qPCR	Quantitative PCR
ROC curve	Receiver Operating Characteristic curve
PHA	Phytohemaglutinin
ddPCR	Droplet digital PCR
FDR	False Discovery rate
DE	Differentially expressed
AUC	Area Under the Curve
GO	Gene ontology
OCMBs	Total oligoclonal IgM bands
LS-OCMBs	Lipid-specific oligoclonal IgM bands
NTA	Nanoparticle tracking analysis
EM	Electron microscopy
BS	Binding site
HS	Highly structured
PS	Poorly structured
ND	Undetermined structure
Poly(I:C)	Polyinosinic:polycytidylic acid

Introduction

1 Transcriptome

As time goes by and we keep studying and investigating the never-ending biological processes underlying all living organisms, the way we understand biology is continuously changing. Nevertheless, some key findings and dogmas have set the ground for the current biology.

This thesis tries to unravel part of the molecular biology implicated in the pathophysiology of Multiple sclerosis and how the gene expression is regulated in the disease. Thus, it is to some extent framed in the central dogma of molecular biology outlined by Francis Crick more than 60 years ago. In the lecture he gave as part of a Society for Experimental Biology symposium on the Biological Replication of Macromolecules, held at University College London, he argued that *"the main function of the genetic material is to control the synthesis of proteins"* although he admitted the speculative nature of this hypothesis at that moment [1]. During this talk, Crick developed a scheme representing the flow of information between genes and proteins that links DNA and proteins through some kind of intermediate, RNA (Figure 1). This scheme is still present in many books and blackboards all over the world and many times its interpretation is mistakenly limited to DNA makes RNA makes protein. This simplified interpretation of the dogma would be in need of many clarifications nowadays, which will be further discussed in this thesis. However, Crick's essential argument "*protein synthesis relies on nucleic acids (not necessarily directly), and once the genetic information has got into the protein, it cannot alter the DNA sequence*" still holds and more importantly, leaves space for many subsequent discoveries such as the reverse transcription, epigenetic modifications or the existence of regulatory non-coding RNAs.

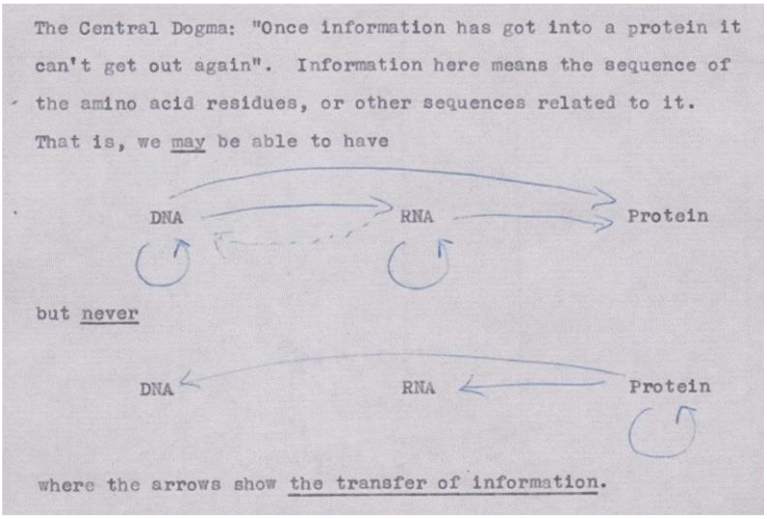

Figure 1. Crick's scheme of the central dogma, from an unpublished note made in 1956. Image from [1]

In the following sections, several of the key concepts, molecules and mechanisms involved in this scheme will be introduced, with particular focus on the collection of all the gene readouts on the cell or RNAs, so called transcriptome, and its regulation.

1.1. Genome and genomic organization

DNA (deoxyribonucleic acid) is a molecule that contains all the biological instructions needed to develop and direct the activities of nearly all living organisms. It had been isolated and characterized in 1868, only a few years after the publication of the laws of heredity by Gregor J. Mendel. Nevertheless, by that time proteins were believed to be the carriers of genetic information. The importance of DNA did not become clear until 1953 thanks to the work of James Watson, Francis Crick, Maurice Wilkins, Rosalind Franklin and her student R.G. Gosling. By studying X-ray diffraction patterns and building models, the scientists figured out the double helix structure of DNA (Figure 2). The discovery of its structure enabled scientists to understand how the DNA was able to carry all the information needed to create a new organism from one generation to the other[2].

This double helix is built up of four chemical units called nucleotides, which comprise the genetic "alphabet". Each nucleotide is at the same time formed by three components: a sugar group, a phosphate group and a nitrogen base. The sugar, which is a 2'-deoxyribose, and the phosphate group, that links the 5' position of one sugar to the 3' position of the next, form the backbone of each DNA strand. Attached to the C1' of the sugar there is a nitrogenous base that interacts via a hydrogen bond with another nitrogenous base in the

complementary strand. The four nitrogenous bases that DNA can contain are classified into two purines, adenine (A) and guanine (G) and two pyrimidines, cytosine (C) and thymine (T). Due to the fact that purines are bigger than pyrimidines, the building of a regular double helix of DNA requires each of the purine bases on one strand to be paired up by a hydrogen bond with a pyrimidine of the complementary strand in a specific manner known as the Watson-Crick base-pairing interactions: guanine with cytosine and adenine with thymine[3]. These pairing between bases that complement one to the other make each of the DNA strands complementary although it is important to note that they run in opposite directions. In addition to the hydrogen bonds, the double helix structure of the DNA is also maintained by hydrophobic interactions between consecutive bases on the same strand that while they try to avoid the interaction with the surrounding water, they also contribute to the winding of the helix[4]. The order, or sequence, of the bases determines what biological instructions are contained in a strand of DNA, and since both strands are complementary there is no information in one of the strands that cannot be deducted from the other. For these reason, all the information and instructions contained in the DNA are represented for convention as the sequence of letters that represent the bases of only one strand, usually the 5'-3' direction strand.

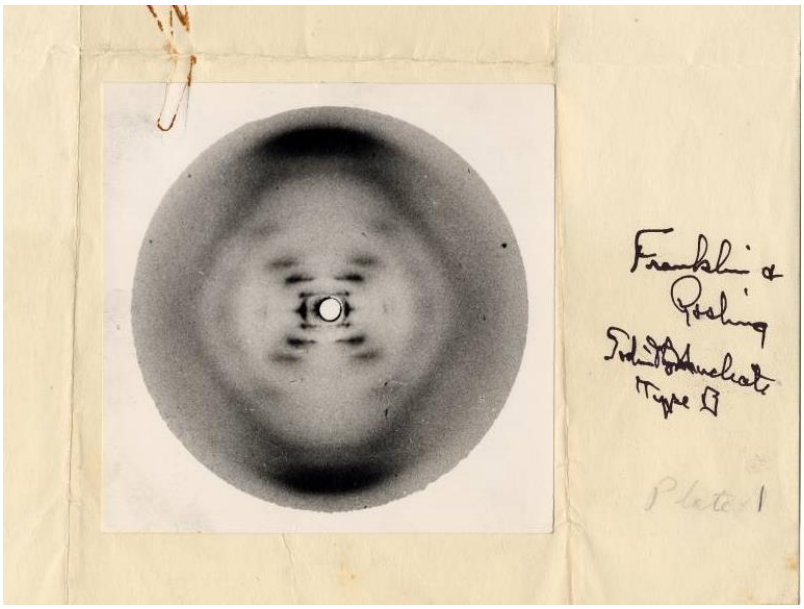

Figure 2. Crystallographic photo of Sodium Thymonucleate, Type B known as the "Photo 51". The photo showing the X-ray diffraction pattern of DNA and was taken by Rosalind Franklin and her student R.G. Gosling in 1952. *Photo obtained fron the Special collection and Archives, Oregon State University Libraries.*

1.1.1. Structural organization of the genome

The complete set of DNA instructions of an organism is called genome and virtually every single cell of the organism contains a complete copy in the nucleus. The human genome contains about 3 thousand million bases. Based on the B form of the DNA double helix model proposed in 1953 and subsequent measurements, it is known that each turn of the DNA molecule contains 10 base pairs (bp) and 36Å, this is 3.6nm[3]. Therefore, the human genome is of approximately 1m long and since most of the human cells are diploids (which means that they contain a complete copy of the genome from each of the two parents, except for gametes), they end up containing 2 m of DNA that have to be extremely compacted in order to fit inside the cell nucleus that is typically 5 to 10 µm in diameter[5]. This compaction is achieved by means of several levels of highly organized packaging that apart from compacting the DNA, plays a key role in the accessibility of the genetic information.

The final packaging depends on an effect known as coiling and supercoiling. Briefly, the double helix of DNA is wounded around a disk of proteins consisting on an octamer of histones and resulting in a DNA-protein complex called chromatin. The basic structural unit of this first level of packaging is the nucleosome that contains 146 bp of DNA compacted into a length of 5.7 nm corresponding to the thickness of the histone disk. Since nucleosomes do not provide enough compaction, this chromatin, is packaged further together with several other proteins (histones and non-histone proteins), into higher-order fiber structures of increasing thickness and compactness. Those fibers will eventually acquire their highest compaction state as chromosomes during the metaphase of the cell division cycle [6,7](Figure 3). The final chromatin structure can condense the original DNA by 10,000 fold of its original length.

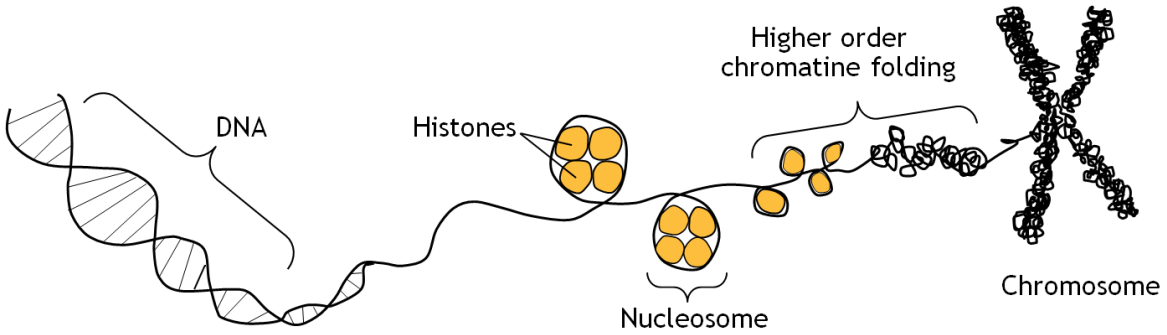

Figure 3. DNA packaging scheme, from double helix structure into an eukaryotic chromosome. *Ilustration by Iparraguirre, L.*

Besides the DNA located in the nucleus, humans also have a small amount of DNA in mitochondria. The mitochondrial DNA contains 16,569bp and unlike the nuclear DNA it is organized into circular molecules. Additionally, different studies published in the last decade have also reported a wide presence of extrachromosomal circular DNAs (eccDNAs) in the nucleus[8–10].

1.1.2. Functional organization of the genome

Functionally, the genome is organized into genes, whose original definition dates back to the mid-19th century when they were defined as pieces of chromosomes that lead to observable traits. This concept was first introduced by Gregor J. Mendel, the father of modern genetics, (although he called them "factors") and since then, it has been continuously evolving.

In 1940, and thanks to the experiments carried out by George W. Beadle and Edward L. Tatum, it became realized that many characteristics were determined by the presence or functions of individual proteins. In fact, they postulated that genes direct the manufacturing of proteins that control the basic metabolic functions and lead to the hypothesis known as "a gene-an enzyme"[7]. Few years later with the progress of genetic and biochemistry studies, this definition was extended to "a gene-a protein" and refined to "a gene-a polypeptide". So, for many years the term "gene" was synonymous of a region of DNA containing the information to create polypeptides that could form proteins, which were believed to be the molecules that do most of the work in our bodies. However, nowadays thanks to the advent of sequencing techniques and the development of projects such as de Human Genome Project and the ENCODE project, it is known that the scope of this definition is too limited. It ignored many other DNA regions, considered for a long time as "Junk DNA" that although not coding for proteins, play very important roles in the cell and influence many phenotypes in other ways[4].

The human genome project was launched in 1990 with the goal of defining and mapping all the genes in the human genome. It was conducted by the National Institute of Health (NIH) in the US with the participation of many research groups all around the world and in parallel by the private company Celera Genomics. Starting from the assumption that each gene gives rise to a protein and thus a function or a characteristic, a direct relation between the size of the genome, the number of genes and the complexity of the organism was expected. For this reason, and in comparison to the number of genes reported for several other organisms that were being sequenced such as the yeast, fruitfly, mouse etc.[11] the estimated number of human genes by the scientists at that time was of 100,000. Nevertheless, after

the first phase of the project, the two articles published in 2001 based on a still incomplete draft of the human genome acknowledged that the number of protein coding genes contained was about 30,000 to 40,000[12,13]. This number was finally reduced to 24,000 protein coding genes by when the project was finished in 2004. The first surprise was the low number of protein coding genes that lead scientist to think that a gene can direct the synthesis of an average of five to six proteins through mechanisms such as the alternative splicing. The second surprise revealed the existence of a big proportion of non-protein coding genes. In fact, they reported 700 additional genes that produced non-coding RNAs as their ultimate product.

The ENCODE (Encyclopedia of DNA Elements) research project launched in 2003 by the National Human Genome Research Institute (NHGRI) with the aim to identify all the functional elements in the human genome further confirmed these results. This project concluded that protein coding genes represent only about the 1.2% of the human genome, while the 93% of the genome is transcribed and has a role in regulating the activity and expression of particular genes[14]. It is also interesting to note that only the 39% of the transcripts map to protein coding exons and that the 54% are mapped outside these regions. The ENCODE project has been running now for 17 years and although more than 30 institutions are participating, it is still far from finished.

In light of this finding, the paradigm of a gene-a protein had to be revised. Therefore, a gene is nowadays defined as a segment of DNA containing the information needed for the synthesis of a biologically functional final product, either protein or RNA. This definition includes both protein coding and non-protein coding RNAs (ncRNAs) but it also gives rise to the question on the features based on which a product is or not considered biologically functional.

After the Human Genome Project, many other genetic studies have tried to create a definitive human gene catalog, and thanks to the development of new techniques for the study of genes and gene expression, the number of coding and non-coding genes is being constantly refined[15]. Nowadays, there are two human gene datasets that are the most widely used although both are continuously evolving and many genes are added and deleted every year. The Ensembl/GENCODE gene set, maintained by the European Bioinformatics Institute (EMBL-EBI) includes 20,438 protein-coding genes and 24,000 non-coding genes, while RefSeq, maintained by the National Center for Biotechnology Information (NCBI) counts 19,934 protein-coding genes and 18,564 non-coding genes (revised in July 2020). The differences between one and the other are mainly due to discrepancies in the classification

of genes, and in fact the total number of transcripts in both datasets rises up to 228,116 and 225,468 genes respectively when all the antisense RNAs, miscellaneous RNAs and pseudogenes are taken into account[16]. In 2018, a study from the John Hopkins University in the USA assembled the sequences from almost 10,000 RNA sequencing experiments and created a new catalog of human genes that included almost 5,000 new genes[17]. However, the study has been controversial since it is relied on computer programs without manual curation[15]. In fact, although it is now widely accepted that a gene does not necessarily have to code a protein, not all the parts of the genome that are transcribed can be taken as genes.

1.2. From genome to transcriptome

As abovementioned, all cells in the human body contain the same instruction book, the same genetic information. However, we consist of a large array of cell types that perform diverse biological functions. This cell type diversity results from differences in gene expression, the process by which the instructions encoded in DNA are transcribed to RNA, and in some cases translated from RNA to peptides. Importantly, this gene expression is different from one cell type to other and it is tightly regulated to produce the right amount of the right protein or functional RNA and at the right time in each cell so they can perform their specific function. In fact, the specialization in a particular function determines the need of different gene products. Only a particular set of genes are transcribed at any one time, and some other DNA portions are never transcribed. Erythrocytes represent a radical example of specialization, in which they accumulate hundreds of copies of hemoglobin that account for more than the 90% of its dry weight.

Similar to the genome concept, the sum of all the RNA molecules produced in a cell under a given set of conditions (independent of whether they are translated into proteins or not) is called the cellular transcriptome and it is unique to each particular type of cell and condition.

The regulatory system that ensures the correct expression of this transcriptome plays a essential role not only in cellular differentiation but in many other biological processes ranging from cell cycle progression, maintenance of intracellular metabolic and physiologic balance and cellular development[18]. Studies comparing the transcriptome of normal and diseased cells have revealed that perturbations of the physiological gene expression are linked to an incorrect cell function in diseased cells although many times it is not clear whether it is cause or consequence of the disease. These abnormal transcript levels can appear for

several reasons such as the presence of extra copies of particular genes or chromosomes, alterations in the RNA degradation rates, response to external stimuli etc. However, many different diseases and syndromes, including cancer, autoimmunity, neurological disorders, diabetes, cardiovascular disease, obesity and rare diseases arise from a breakdown in the regulatory system[19] such as: aberrant chromatin regulation[20], mutated or dysfunctional transcription factors[21–23], non-canonical splicing[24] and changes affecting regulatory ncRNAs[25–27].

Due to the importance of the regulatory system in the maintenance of a healthy transcriptome, in this section, we are going to focus in the description of transcription process and its different levels of regulation.

1.2.1. The eukaryotic transcription process

The transcription process is carried out by an enzyme system that converts the genetic information coded in a double stranded DNA segment into an RNA strand with a sequence complementary to one of the DNA strands. RNA is quite similar to DNA at the chemical level, differing only in a hydroxyl group present at the sugar that converts the deoxyribose into a ribose, and the use of uracil (another nitrogenous base) instead of thymine. However, their biggest difference is structural since unlike DNA, most RNAs carry out their functions as single strands, which gives them the possibility of having a great diversity of secondary structures[28].

Only one of the DNA strands serves as a template that is copied in the 3' to 5' direction, creating the new RNA strand in the 5' to 3' direction following the Watson-Crick base-pairing interactions. RNA polymerase is the enzyme responsible for the transcription, and to do so, it requires of a DNA template, the four ribonucleoside 5' triphosphates (ATP, GTP, UTP and CTP) as precursors for the nascent transcript and Mg^{2+} and Zn^{2+} as cofactors of the enzyme. Three different RNA polymerases have been described in eukaryotes, which are different complexes although they have some subunits in common. RNA polymerase I (Pol I) transcribes preribosomic RNA that will lead to the formation of the 18S, 5.8S and 28S ribosomal RNAs (rRNA). RNA polymerase II (Pol II) is responsible of the synthesis of all the protein-coding RNAs or messenger RNAs (mRNAs) and some other RNAs such as microRNAs (miRNAs) or circular RNAs (circRNAs) that will be later described. The principal function of the RNA polymerase III (Pol III) is to make transfer RNAs (tRNAs), the 5S rRNA and some other small RNAs.

The transcription process is driven by a complex series of ordered events that can be grouped into three main steps: initiation, elongation and termination. Initiation comprises RNA polymerase recruitment to the promoter site, the assembly of the preinitiation complex, the opening of the DNA duplex and the synthesis of a small fragment of RNA until the enzyme escapes the promoter. Once transcription has been initiated, during the elongation the RNA chain in extended until it reaches a termination signal. This signal leads to the termination in which the DNA and RNA are released[29,30].

For the transcription to be initiated, the RNA polymerase must recognize the gene that is going to be transcribed and bind at specific DNA sequences called promoters and located at the beginning of the gene. Importantly, the promoter recognition will only happen provided that the chromatin is partially opened, which comprises the first regulation point of the transcription.

Promoter sequences direct the transcription of adjacent segments of DNA avoiding the initiation of the transcription at random points and they also dictate when, where and at what level are genes transcribed. All this information is not only determined by the core promoter but several other either proximal or distant sequences also take part in the regulation of the transcription[30]. Due to the diversity of transcripts produced by Pol II, it has to recognize thousands of different promoters, however some common features have been recognized in many of them such as the TATA box and the Initiator sequence (Inr). The TATA box refers to a consensus sequence TATA(A/T)A(A/T)(A/G) localized 30 base pairs upstream the point at which the RNA starts to be synthesized whereas the Inr sequence is right at the RNA start site. These promoter sequences serve as the docking site for the preinitiation complex (PIC) formed by the RNA polymerase and several transcription factors (TFs) needed for its assembly and activation. A key function of the PIC is also to open the DNA. The first protein to bind in the assembly of the PIC in the case of a typical Pol II promoter is the TATA-binding protein or TBP that will be followed by many other basal TFs such as TFIIB, TFIIF, TFIIE and TFIIH. Once the PIC is complete, the transcription bubble is formed by the partial unwinding of the DNA. At this point, the TFIIH mediated phosporylation of Pol II C-terminal domain (CTD) is needed in order to break the contacts between Pol II and PIC factors [30], then the polymerase escapes the promoter and the transcription begins (Figure 4).

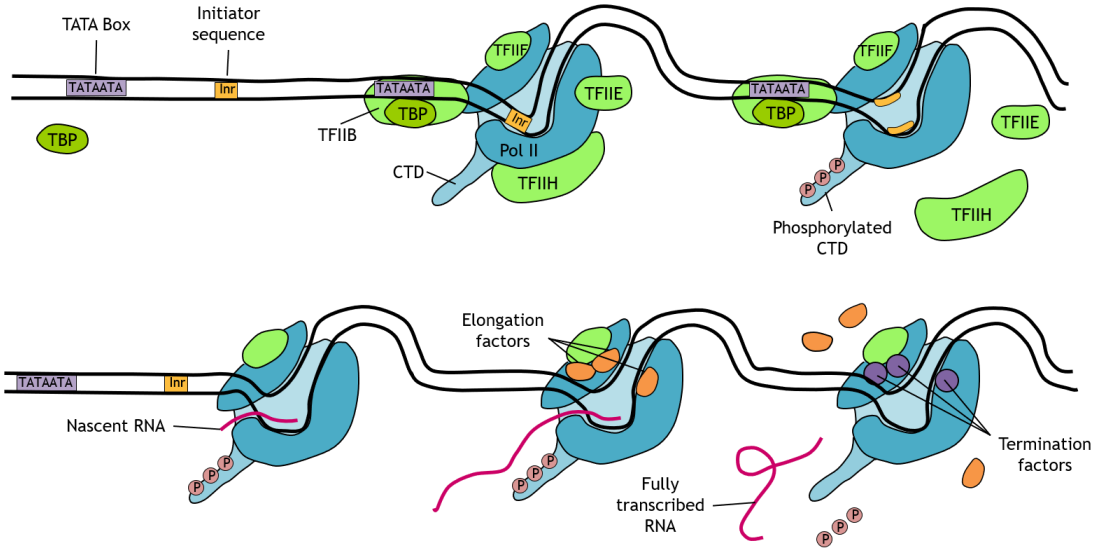

Figure 4. Transcription steps at the RNA polymerase II promoters. The main events occurring during the initiation, elongation and termination steps are represented. *Ilustration by Iparraguirre, L.*

When the RNA growths to a critical length, apart from TFIIF that remains associated with Pol II an elongation complex is formed by the binding of several elongation factors that enhance the polymerase activity. These elongation factors will also coordinate the interactions between proteins involved in the posttranscriptional processing of RNAs. During this phase the RNA extends in a progressive manner. The growing end of the new RNA strand base pairs temporarily with the DNA template to form a short hybrid RNA-DNA and RNA polymerase walks along the template strand in the 3' to 5' direction while it adds a matching RNA nucleotide to the 3' end of the nascent transcript. The elongation continues until it reaches to a termination signal. Shortly after the termination signal is recognized, elongation factors dissociate, the newly synthesized RNA peels off, the DNA duplex reforms and termination factors facilitate the dephosphorylation of Pol II in order to be ready to initiate another transcript[28](Figure 4).

1.2.2. Transcription is regulated at different levels

Various modes of gene regulation are used to fix gene expression programs in time and space. Four different moments for regulation can be distinguished: the regulation occurring before the transcription, regulation of the transcription process itself, regulation by other co-transcriptional processes, and post-transcriptional regulation (Figure 5).

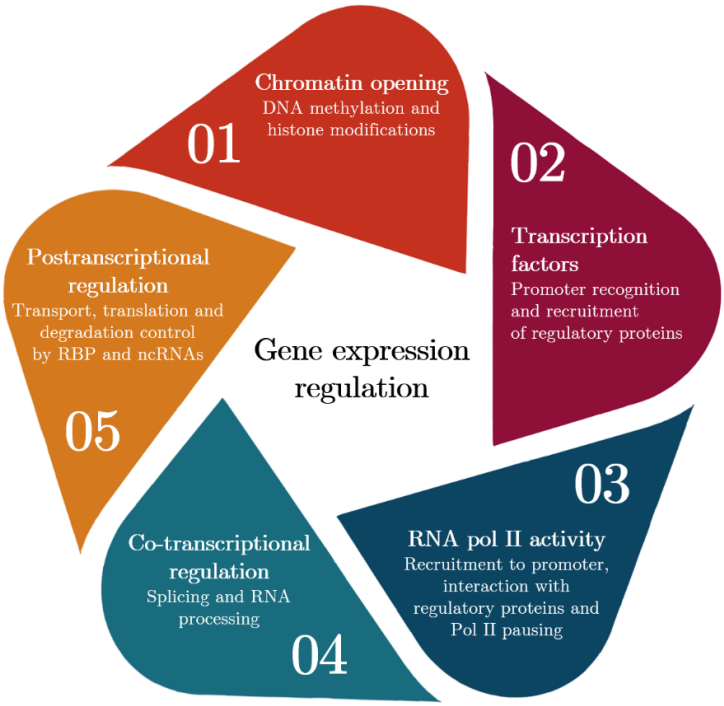

Figure 5. Scheme showing the different regulation levels for gene expression that will be further explained in the sections below. The first level represents pre-transcriptional regulation, levels 2, 3 and 4 refer to regulation during transcription or cotranscriptional regulation and the fifth level gathers postranscriptional regulation. *Scheme by Iparraguirre, L.*

The first complexity layer for the regulation of transcription is provided by the fact that the human genome is tightly packaged into the nucleus and this configuration prevents RNA polymerase to access the DNA sequence. Therefore, some events such as DNA demethylation and post-translational modifications of histone tails have to occur in order to have a opened configuration of the chromatine.

Once the chromatine is opened and DNA is accessible, RNA polymerase can access the genes to be transcribed. The function and interaction between transcription factors, regulatory proteins and the RNA polymerase exert their regulatory role during the transcription process.

The production of a fully functional RNA may require several other processing reactions that occur cotranscriptionally such as the capping of the 5' end, removal of introns by splicing and polyadenylation at the 3' end. Coupling between transcription and RNA processing provides additional points to regulate transcription in many ways.

Finally some of those functional RNAs, mostly ncRNAs together with other proteins can also exert their regulatory functions at the post-transcriptional level.

1.2.3. Pre-transcriptional regulation is mediated by chromatin modifications

As previously mentioned in section 1.1.1., nuclear DNA resides in the nucleus of cells where it is packaged into a highly condensed structure called chromatin which comprises at the same time the first mechanism for regulating different gene expression profiles. There is a spectrum of condensation states of chromatin ranging from the tightest referred with the term heterochromatin to areas with less compaction, euchromatin. Higher compaction levels generally lead to repression of transcription by impairing the promoter access. Heterochromatin is transcriptionally inactive due to the strong presence of nucleosomes that mask promoters. The rest of the chromatin, euchromatin, is less condensed, however, only some, but not all, is transcriptionally active. Nucleosomes, are still present in the euchromatin and must be removed or shifted to allow the promoter access and transcription initiation.

This chromatin opening is highly regulated and involves the participation of different regulatory proteins such as those that recruit ATP-dependent nucleosome remodellers and histone modifying enzymes. DNA bases can also be modified and together with histone modifications give rise to an epigenetic code that can be associated with high or low levels of gene expression[31](Figure 6). This epigenetic code can be influenced by several environmental factors such as diet, physical activity, pollutants, stress or smoking[32].

Analysis of the nucleosome crystal structure has revealed that the N-terminal histone tails protrude outward of the nucleosome core. This disposition facilitates numerous postranslational modifications in histones that influence chromatin compaction and therefore transcription[33]. Histone modification enzymes catalyze different chemical modifications such as methylations that facilitate the binding of histone acetyltransferases that will further acetylate histone specific residues. Acetylation reduces the affinity of the entire nucleosome for the DNA leading to the displacement of nucleosomes with help of remodeling enzymes. In a similar way, these modifications can be reverted by other enzymes as part of a gene-silencing process that restores chromatine inactive state when the transcription of a gene is no longer required[29,34]. Polycomb protein complexes for example are known to mediate gene silencing by inducing chromatin compaction, by ubiquitination of histones that prevent their methylation and by reducing the accessibility of chromatin remodelers[33].

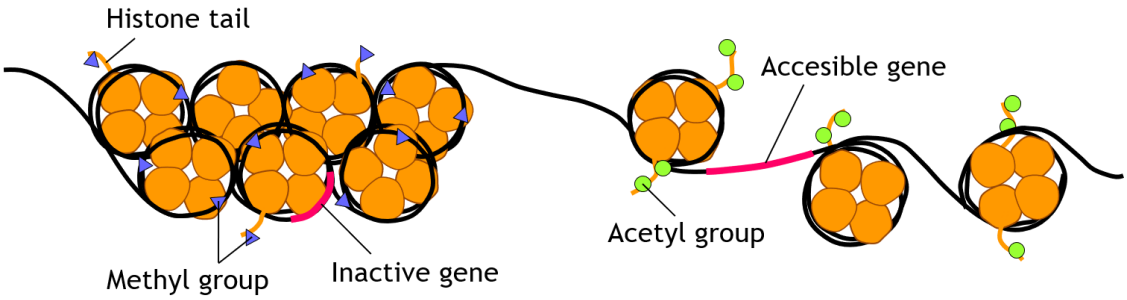

Figure 6. Histone and DNA modifications affect chromatin compaction regulating the accesibility of genes. *Ilustration by Iparraguirre, L.*

Depending on the promoter class that is covered by nucleosomes the chromatin opening is regulated differently. A classical example in case of Pol II promoters are those that contain CpG islands that represent up to the 70% of promoters in vertebrates and can be regulated by methylation. CpG dinucleotides are significantly underrepresented in mammalian genomes due to the fact that the 5' methylation of a cytosine results in its spontaneous deamination to uracil. On the contrary, CpG islands are regions of approximately 1kb in length where the GC dinucleotide content exceeds the expected frequency. The unmethylated state of those CpG islands prevents the deamination of the cytosine but more interestingly, it can impair the assembly of nucleosomes facilitating polymerase access. CpG islands are often found in different regulatory regions such as promoters of housekeeping genes that are transcriptionally active in most of the cells and tissues. Besides, methylation of CpG promoters result in their transcriptional repression [35].

1.2.4. Regulation during the transcription process

Transcription factors and regulatory proteins are essential for the transcription to happen, but they also play a role on its modulation.

1.2.4.1. How transcription factors regulate transcription

Transcription factors (TFs) are proteins that coordinate gene expression in a spatial and temporal manner. About 1,600 human TFs are known accounting for almost 7% of the genes[18].

Most TFs bind to nucleosome-free DNA and regulate promoter activity, but often they do it via interactions with genomic locations that can be distant from the core promoter which are termed enhancers. These sequences are broadly known as gene regulatory regions that generally contain multiple TF binding sites or DNA motifs, which are commonly short.

Similarly, some other TFs can bind nucleosomal DNA and participate in the opening of chromatin locally to enable transcription[29]. Such short DNA motifs frequently occur by chance in the human genome resulting in weak TF-DNA interactions. Nevertheless, TF-DNA interactions must compete with histone-DNA interactions for a stable and productive binding. Moreover TFs often depend on protein-protein interactions with other TFs and regulatory proteins to function[18].

Depending on the TFs and proteins with which the TF interacts, it may activate or repress the gene expression as it will be further exemplified in section 1.2.4.2. TFs may activate transcription through recruitment of coactivators, such as histone methyltransferases, histone aceyltransferases, chromatin modifying enzymes or Pol II Mediator coactivator. By contrast, transcription can become repressed through recruitment of co-repressors such as histone demethylases, histone deacetylases, polycomb complexes and other Pol II co-repressors. Moreover, TFs have the ability to regulate their own expression by binding to their own promoters or enhancers as well as they can also regulate the expression of other TFs creating a TF circuit or network[36].

Based on their function TFs can be divided in two main groups: basal TFs and sequence-specific DNA binding TFs. Basal TFs are components of the PIC and necessary for the recruitment and basal Pol II activity. Sequence specific DNA binding TFs instead bind to specific binding sites and upon binding with other proteins they interact with the core transcription machinery to enhance or repress its activity. They can directly influence Pol II activity via posttranslational modifications of specific Pol II amino acid residues that lead to different conformational changes, each of them linked to different transcriptional states[37].

Human genes are likely to have several regulatory regions either upstream, downstream or within the gene sequence which results in the cooperative action of several TFs that together regulate the expression of a gene. Apart from the action of several TFs, insulator proteins can also stop the action of a distal TF on a promoter if placed between them. If different TFs were required for each gene, the existing repertoire of TF (which accounts for less than the 10% of genes) would not be sufficient to control the expression of the whole human genome. Therefore, the cooperative action of TFs and other regulatory proteins is essential to allow a tight and precise control of the gene expression program of a cell. These regulation is accomplished by utilizing different combinations of a limited number of TF that form homo or heterodimers and interact with different regulatory proteins[34,38]. Studies demonstrate that a big part of the TF that are expressed in cells exert broad effects such as

the chromatin opening or recruitment of the transcription complex while a small subset of the TF are sufficient to reprogram the cell-type specific gene profile when extopically expressed[19](Figure 7).

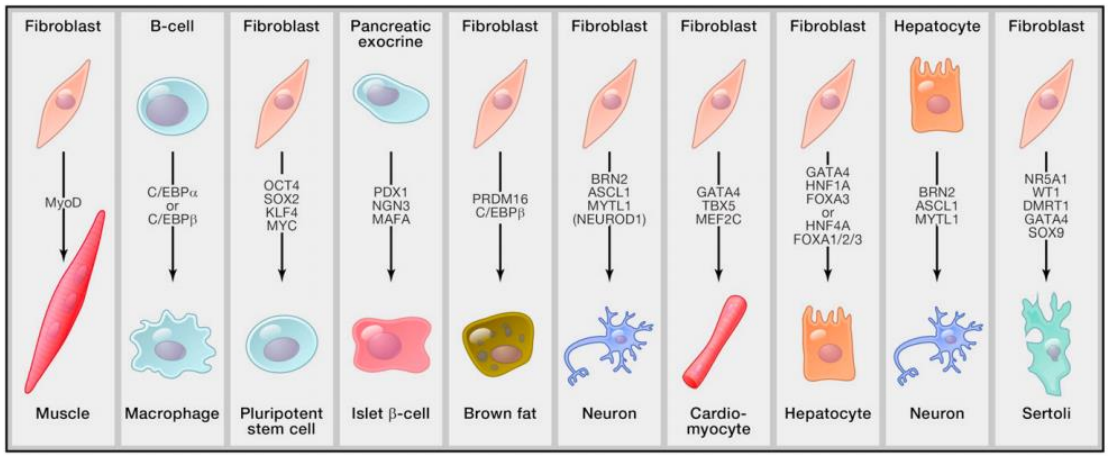

Figure 7. A small subset of transcription factors defines cell-type specific gene expression programs. *Ilustration from [19]*

1.2.4.2. How RNA Pol II activity can be regulated

The activity of the RNA polymerase is quite elaborated and involves the interaction with a wide variety of proteins. It can be regulated at three main points: its recruitment to the promoter sequence, the enhancing or repression of its activity by different regulatory proteins and its pausing/release into elongation.

Traditionally the gene regulation was believed to happen primarily at the recruitment step of the RNA polymerase and the formation of the PIC. Despite we currently know that significant regulation occurs also thereafter, is the best documented part and involves promoters and several regulatory proteins.

Once nucleosomes have been removed or shifted and the promoters are accessible, the RNA polymerases cannot recognize the promoter sequences by themselves. Instead they require transcription initiation factors to form the PIC on the promoter. This recognition step can be influenced by at least three elements. Firstly, promoters differ a lot in their sequence, in fact many of them lack the obvious DNA sequence elements. Differences in sequence lead to different binding affinities to the RNA polymerases affecting the frequency of transcription initiation. At this point regulatory proteins such as specificity factors influence the specificity of the RNA polymerase for a given promoter or set of promoters. Moreover, mutations in the promoter sequences can both increase or decrease their function depending on whether

the mutation makes the promoter more similar or different to the consensus sequence. Secondly, the promoter recognition may also involve sensing of physical features of the DNA shape which explains the formation of a similar PIC in promoters that differ a lot in their sequence. Lastly, a different factor composition of the PIC may also contribute to the recognition of different promoters[29]. Additionally, once the promoter has been recognized, the dynamic of transcription can also be regulated by controlling promoter melting or opening. The opening is mediated by the translocase subunit of the TFIIH factor and therefore promoter melting is influenced by its expression and activity[39].

For some genes, in addition to the previous regulation steps of chromatin opening and promoter binding, the RNA polymerase requires of the combined action of several other proteins for its activity. Activator proteins enhance the interaction between RNA polymerase and promoter by binding near to the promoter when the natural binding is weak. Some of the activators can also exert their function by binding to enhancer sequences found hundreds or even thousands of base pairs upstream or downstream from the transcription start site. Another group of regulatory proteins are essential for the communication between the activators and the RNA polymerase-PIC complex: coactivators. Unlike activators, coactivators do not bind to the DNA but they act as intermediates facilitating the interaction between proteins. Mediator is the mayor eukaryotic coactivator, which consists of more than 30 polypeptides. It provides an assembly surface for the binding of several of the TF and regulatory proteins that have to interact with the RNA polymerase. The interaction between proteins located at enhancers and promoters form DNA loops that are also stabilized by the protein cohesin associated with Mediator[34](Figure 8).

On the contrary, although they are less prevalent than activators, there are also repressor factors that hamper the RNA polymerase activation by a range of mechanisms. They can bind directly to the DNA displacing a protein required for activation or directly interact with activation complexes to prevent their activation. Interestingly, some of these proteins can act both as activators or repressors by adopting different conformations in response to a stimulus. For example, some steroid hormone receptors act as transcription activators of certain genes in presence of a particular steroid hormone and invert their role in absence of the hormone to prevent the transcription initiation.

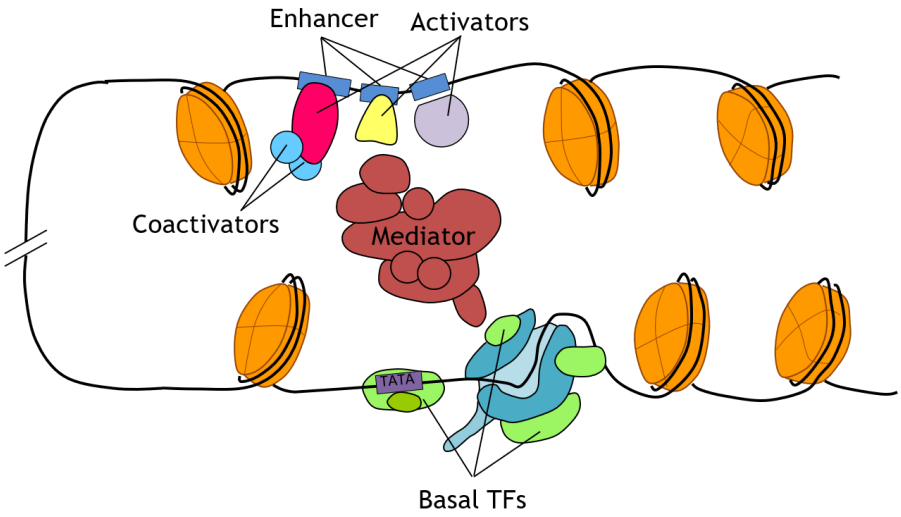

Figure 8. Mediator, the mayor eukaryotic coactivator, acts as a platform that facilitates the interaction between the transcription machinery and regulatory proteins that may be far away from the promoter. *Ilustration by Iparraguirre, L.*

Moreover, a combinatorial control of several regulatory proteins has been reported so that i.e. one activator may facilitate the binding of others. The interaction between several protein complexes can be eased by the action of architectural regulators that promote the looping of the DNA. Most of all these regulatory proteins are DNA binding proteins that generally function as dimers and therefore, any alteration whether on their amino acid sequence or in the DNA sequence can affect their dimerization and/or interaction with the DNA affecting also the frequency of transcription initiation[34].

After initiation and transcription of 20 to 50 nucleotides, the transcriptional regulation has been found to occur by pausing the Pol II activity before its release into elongation by a mechanism called Promoter Proximal Pausing (PPP). The pausing involves the misaligning of the DNA-RNA hybrid with the active site and the binding of factors such as DSIF and NELF that restrict Pol II mobility. The paused polymerases may transition to active elongation through pause release, or they may ultimately terminate transcription with the release of small RNA species. Transcription termination at this point prevents the full-length transcription of some of the genes on which the transcription has been initiated. It has been shown that almost the 30% of genes display hallmarks of transcription initiation without elongation. Some studies suggest that the regulation through pausing and later release of the polymerase enables the rapid reactivation of the transcription in reaction to environmental stimuli[39,40].

Pause release and subsequent elongation occur though recruitment of the elongation factor TFIIS that helps to realign the nascent RNA releasing the RNA and the positive transcription elongation factor b (P-TEFb). P-TEFb phosphorylates the paused Pol II at its CTD enhancing its activity. During elongation phase, P-TEFb can in addition recruit some other activators forming a large complex called super elongation complex. Nevertheless, few it is known about regulation at the elongation and termination steps[30].

1.2.5. Co-transcriptional RNA splicing and processing reactions talk back to transcription

The production of a fully functional transcript is further complicated and requires several processing reactions mainly occurring co-transcriptionally. The set of processing reactions required to be fully functional varies from one to another transcript being the most common: capping of the 5' end, removal of introns by splicing and polyadenylation at the 3' end [41]. The 5' capping and polyadenylation reactions help to prevent RNA degradation, guide the nuclear export and promote translation of mRNAs. Meanwhile, splicing allows creating multiple isoforms from a single newly synthesized transcript increasing a lot the number of functional products produced from a gene.

All those three processing reactions are intimately linked to transcription and therefore they are able to influence and regulate one each other. In fact, the protein complexes that carry out each of the processing steps do not act independently but in association between them and with the CTD of Pol II so that each complex affects the function of the others[28].

1.2.5.1. Splicing and its crosstalk with transcription

In the 1970s, only few years after the DNA structure was discovered, it was observed that vertebrate cells contained in the nucleus a significantly longer 5' caped and 3' polyadelynated RNA compared to the shorter mRNA that was found in the cytoplasm. Besides their size differences the similar 5' and 3' modifications lead to hypothesize that the nuclear RNA could be the precursor of the mRNA. Early studies revealed that most of the eukaryotic protein-coding genes are not continuous but that they contain intervening segments of DNA that are not translated into aminoacids and that were removed at the RNA level during the processing of the longer precursor. Even for non protein-coding genes the primary transcript contains sequences that are later excised to form the mature transcript[42].

Walter Gilbert suggested in 1978 the name of *intr*ons for the *intr*agenic regions and *ex*ons for those regions that will be *ex*pressed[43]. Higher eukaryotic genes typically have much

more DNA dedicated to introns than to exons. An average human gene contains about 8-10 exons and 6-9 introns, however the average exon length is of 170bp while intron length is longer (and more variable) ranging from 4,741 to 23,527 bp[44]. Gilbert also stated that the "one gene, one polypeptide" theory of the moment was no longer valid but that a gene corresponded to a transcription unit that could correspond to many polypeptide chains[43]. We now know that they correspond not only to many polypeptides but that they can also correspond to different functional RNA products. These products can have different or even opposite biological functions with important consequences on cellular processes and phenotypes.

This diversity of products is obtained thanks to alternative splicing. Alternative splicing occurs in about the 95% of multiexonic genes that, instead of giving rise to a single mature transcript containing all the exons, selectively includes or excludes and combines different exons. In addition, some introns may also be retained[23]. Alternative splicing helps to explain the paradox of the obvious higher complexity of humans despite the fact that organisms such as worms have a similar number of genes. Human genes have been estimated to have three times more alternatively spliced forms on average than worm genes[42].

The splicing mechanism can vary from one to other intron but the removal of most mRNA introns is catalyzed by a large protein complex called spliceosome. Spliceosome comprises several small nuclear RNAs (U1, U2, U4, U5 and U6) bound to protein factors that form the small ribonucleoprotein complexes (snRNPs). SnRPs U1 and U2 take part in the identification of the 5'splice site and the branch point within the intron. The addition of U4, U5 and U6 completes the inactive spliceosome complex. Some rearrangements are needed to produce the active spliceosome complex where U1 and U4 have to be removed. The splicing mechanism involves two transesterification reactions. It is initiated by the nucleophilic attack of the 5' splice site by the reactive hydroxyl group of an adenine residue within the intron that comprises the branch point. The reactive OH at the end of the upstream exon performs the second nucleophilic attack to the 3' end of the intron in order to excise it[28,41](Figure 9).

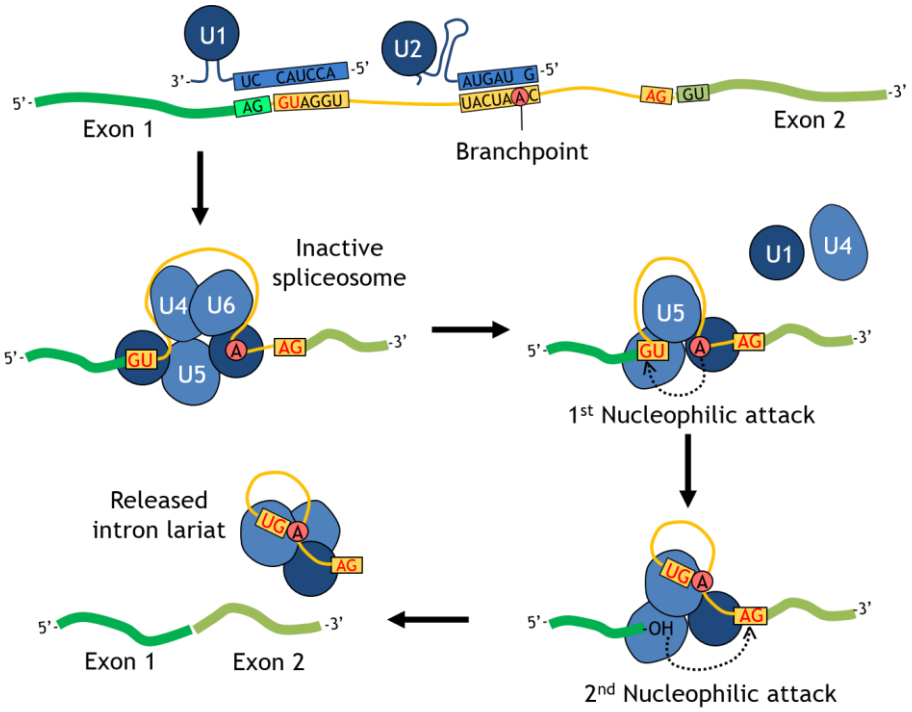

Figure 9. Splicing involves of two transesterification reactions catalyzed by the spliceosome complex. *Ilustration by Iparraguirre, L*

Although this splicing process can occur once the transcript has been fully transcribed, there is increasing evidence showing that a large amount of splicing occurs cotranscriptionally (Figure 10). In fact Carrillo Oesterreich et al. showed that the 80% of the mRNAs are spliced before transcription termination[45]. This connection in time and space explains how splicing reactions are influenced by transcription and what is becoming clearer now, how splicing can in turn represent an additional point to regulate transcription.

Among the cotranscriptionally spliced transcripts also different splicing rates can be observed. Cotranscriptional splicing and the speed of splicing could be defined by a number of different factors such as promoter elements, intron length, splice site length, RNA secondary structure and transcription rate[46]. Transcription can affect splicing by two general mechanisms; through changes in the speed of pol II and through interaction between transcription and splicing machineries. On one hand, a slower elongation rate during transcription results in more time for splice site recognition by the spliceosome. Therefore any factor that could reduce pol II speed during elongation favors a more efficient splicing. Chromatin structure, transcription factors and different regulatory proteins have been previously mentioned among the factors able to regulate pol II transcription rate. It is already known that there are differences in nucleosome positioning and histone modifications between exons and introns resulting in a reduced elongation rate of pol II over spliced exons

relative to introns. Differences between the nucleosome positioning of constitutive exons and alternatively spliced exons have also been reported (reviewed in [41]). Similarly, pol II pausing has been found to be essential for short genes to be cotranscriptionally spliced before termination[45]. On the other hand, many of the factors involved in transcription can act as a platform for the recruitment of the spliceosome favouring the coupling of both processes. Different modifications of the pol II CTD such as the phosphorylation of Serin 5 have been proposed to facilitate the recruitment of splicing factors to the pol II in a timely fashion[47]. Transcription factors such as elongation factors and coactivators such as Mediator also interact with splicing factors, bringing them close to the transcription machinery and stabilizating their association (reviewed in [41]).

Recently it has come to light that splicing does also regulate some of the key steps during transcription. Splicing factors have been for example implicated in pol II pausing, one of the mayor regulatory steps in transcription. The splicing factor SRSF2 can be recruited to the paused pol II complex which helps to the dissociation of P-TEFb from its inhibitory complex facilitating pol II pause release[48]. Spliceosome assembly has also been found to provoke the ubiquitination of pol II, which can slow down the elongation while the splicing takes place. Following splicing, deubiquitination occurs and elongation resumes[49]. In addition, several studies have also reported that the splicing inhibition affects to pol II CTD phosphorylation patterns providing evidence of another mechanism through which splicing regulates pol II activity(reviewed in [41]). Several splicing factors can also directly interact with chromatin and affect histone modifications. It has been suggested that splicing mediated histone modifications can serve as a "chromatin code" that indicates that the gene has been spliced and promoting its transcription in the next round[50].

1.2.5.2. RNA processing factors implicated in 5' capping and 3'polyadenylation can also influence transcription

Besides splicing, many eukaryotic transcripts including mRNAs and some other ncRNAs undergo 5' capping and 3' polyadenylation as part of their post-transcriptional processing. Moreover, evidences of interactions between components of the complexes involved in these two RNA processing steps both with the spliceosome and pol II confirmed the coupling between activities. Together with splicing factors, these other RNA processing factors also regulate transcription[41].

The 5' capping consists in the addition of guanine nucleotide to the 5' end of the nascent transcript right after the first 20 to 30 nucleotides have been transcribed. Next, guanine is methylated at N-7 resulting in a 7' methylguanosine linked to the 5' end that forms the so-

called 5' cap. Additionally, the 2' hydroxyl groups of the second and third nucleotides can also be methylated. Once formed, the cap remains tethered to the pol II CTD through an association with the cap-binding complex (CBC)(Figure 10). The cap defines the first exon of the transcript for the later recruitment of the U1 snRNP, protects the transcripts from degradation by ribonucleases and facilitates translation of protein-coding transcripts by participating in the binding to the ribosome[28]. Moreover, in case the nearest cap-adjacent nucleotide is an adenosine, it can be further methylated not only at the N-7 position but also at the N-6 position to form a N^6-methyladenosine. This modification is one of the most abundant RNA modifications present in most transcript classes and with several functional implications such as regulation of splicing, translation and degradation[51]. As part of the transcription regulation exerted by components of the mRNA capping process, a methyltransferase (RNMT-RAM) has been shown to promote pol II transcription and ARS2, a key protein of the CBC, was associated with decreased promoter proximal pausing and premature termination of pol II[41].

Eukaryotic mRNAs and some other ncRNAs also have a distinctive 3' structure formed by a string of 80 to 250 A residues known as the poly(A) tail. This modification also helps to protect the transcript from enzymatic destruction. Taking into account that the tail is added to the 3' end, it is not rare that the polyadenylation process starts once the RNA has been fully transcribed. Pol II synthetizes beyond the site where the poly(A) is to be added so this extra sequence is first cleaved by an endonuclease that bound to the pol II CTD is able to recognize the consensus sequences defining the cleavage site. With the free 3' hydroxyl group of the cleaved RNA as a primer, the polyadenylate polymerase catalyzes the addition of A residues[28](Figure 10). This process is carried out by a complex of around 85 proteins including the endonuclease and polyadenylate polymerase called cleavage and polyadenylation complex or CPA. This complex also affects transcription and its knockdown was found to result in an increase of pol II pausing indicating that CPA is likely to reduce pol II premature termination[41].

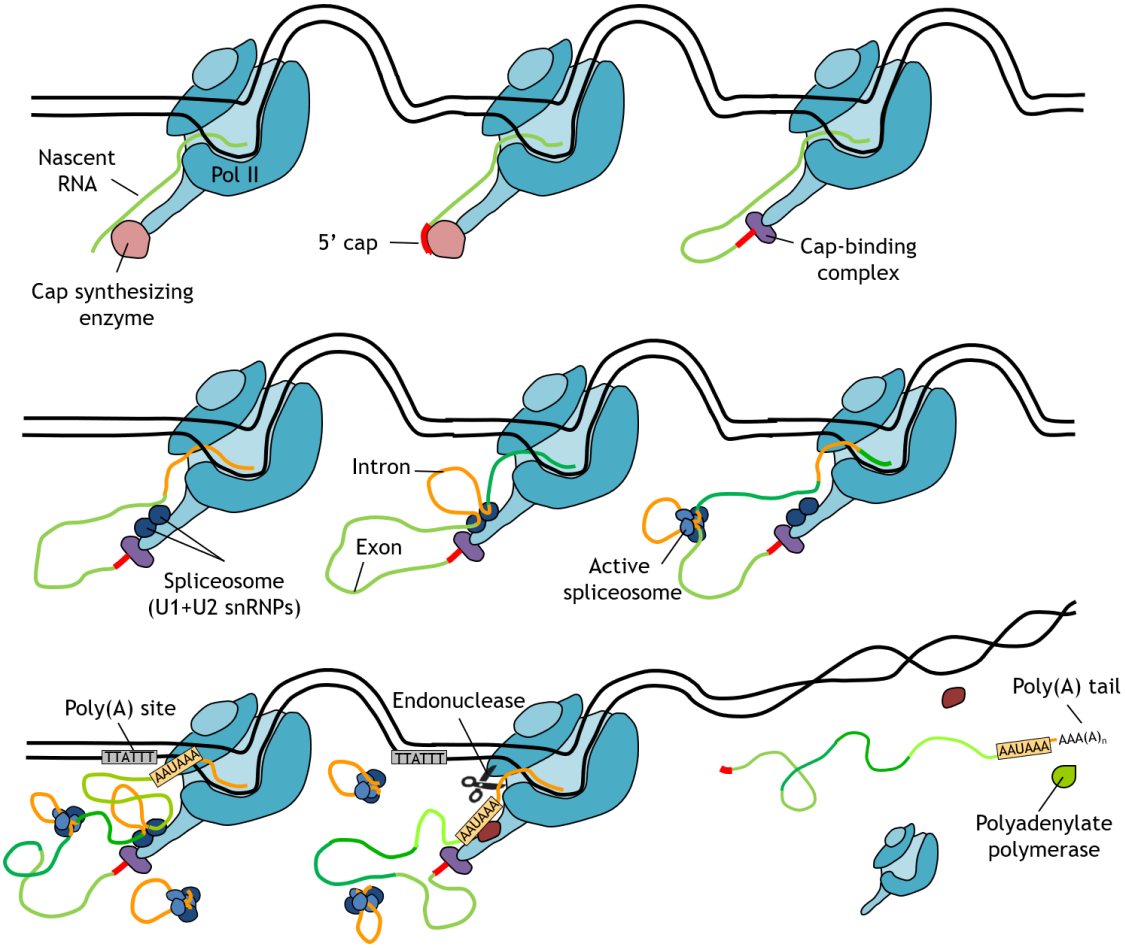

Figure 10. The RNA 5' capping, splicing and polyadenylation reactions that occur co-transcriptionally can regulate back transcription. *Ilustration by Iparraguirre, L.*

1.2.6. Post-transcriptional regulation is mediated by RBPs and ncRNAs

Although each specialized cell in an organism has a characteristic pattern of gene expression or transcriptome, they are all capable of altering this pattern in response to different stimuli. In fact, a transcriptome is not only cell specific but also condition specific. The dynamic change of the transcriptome can be controlled by which genes are or not active to be transcribed (pre-transcriptional control), how often is a given gene transcribed (transcriptional control) or controlling the splicing and processing of the RNA transcripts (co-transcriptional regulation) as it has already been presented.

Nevertheless, once a transcript has been fully transcribed and processed to its mature and functional form there are several other processes through which the transcriptome is

regulated. These post-transcriptional steps select which RNAs are exported to the cytoplasm and where they are localized (RNA transport and localization control), which transcripts are translated by ribosomes (translational control) and how long are they maintained in the cytoplasm (degradation control). Different mechanisms take part in these regulation processes and most of them require the recognition of specific marks in the transcripts to be regulated. The marks include RNA modifications such as the 5' cap and the 3' poly(A) tail, proteins bound to the RNA or specific sequences present in the 3' or 5' untranslated regions (UTR). The task of recognizing and following the information provided by the marks is performed by RNA binding proteins (RBPs) and ncRNAs[52].

As a first post-transcriptional control point, only transcripts that can be recognized as completely transcribed and processed (mRNAs containing the 5' and 3' modifications) will be exported from the nucleus while most of the RNA (introns, damaged transcripts, RNA fragments etc.) is degraded in the nucleus. Upon successful RNA processing RNAs will be bound by a number of RBPs that will load them to export adaptors/receptors to be guided to the nuclear pore complex[53].

Once in the cytoplasm, RNAs will exert their function or be translated at their destination. Specific signals that are located in the 3'UTR regulate how RNAs travel associated to cytoskeletal motors, randomly diffuse through the cytosol or are trapped by anchor proteins at their sites of localization. RBPs are the proteins that bind to these signals and drive RNA localization[54].

Then, translational control mainly depends on the presence of the 5'cap that indicates the site at which the small ribosomal subunit starts scanning the mRNAs for an AUG codon as a translation start site. In addition, the communication between 5'cap and the 3'poly(A) tail is necessary for an efficient translation and therefore, translational repressors that bind one or the other end can inhibit translation.

Lastly, regulation of RNA stability and degradation is an important step to control the time that the transcript has to be translated or exert its function. The more rapid mRNA decays, the less time has the translation to occur. On the contrary, if the mRNA is stable and stays in the cytoplasm for a longer period it will potentially produce more protein. Successfully capped, spliced and polyadenylated transcripts are generally more resistant to decay mechanisms. Nevertheless, the average half-life of a transcript is of less than 30 minutes although some can last up to 10 hours. Degradation typically begins with the gradual shortening of the poly(A)tail to a critical length, followed by either 3' to 5' decay by a large

complex of exonucleases called exosome or decapping and subsequent 5' to 3' decay by exoribonuclease XRN1 [52].

On top of all the above mentioned post-transcriptional regulatory processes, RNAs can also be chemically modified by a process called RNA editing. RNA editing differs from other post-transcriptional processes in that it is a site-specific alteration that can have an impact at different levels. In higher eukaryotes, and regarding mRNAs (although editing of tRNAs,rRNAs and other ncRNAs is also described) two principal types editing occur: the deamination of cytosine to produce uracil (C-to-U editing) and the deamination of adenine to produce inosine (A-to-I editing) being this last one the most widespread type of editing. In A-to-I editing, ADAR (adenosine deaminase acting on RNA) enzymes bind and edit double-stranded RNA (dsRNA) structures formed through base-pairing. Specific changes in the coding region by editing of mRNA can lead to functional alterations of the protein product, while modification of the non-coding regions may globally affect processes as splicing, RNA stability, translational efficiency or interactions with RBPs and ncRNAs[55].

1.2.6.1. ncRNAs in post-transcriptional regulation

At the beginning of the introduction, we have referred to the central dogma of molecular biology according to which RNA can be seen as a mere intermediate of the genetic information flow from DNA to protein. However, during the introduction we have already seen that a big part of the genome is transcribed while very little gives rise to protein suggesting that most of the genome encodes regulatory RNAs from non-coding regions. We have also mentioned the function of some ncRNAs such as the snRNAs taking part of the RNA splicing. Apart from those, the function of some other ncRNAs, the so-called "structural ncRNAs", has been long recognized and extensively studied. Structural ncRNAs include ribosomal (rRNA) and transfer RNAs (tRNAs) involved in protein synthesis or even small nucleolar RNAs (snoRNAs) implicated in modification of rRNAs. Therefore, the ncRNA concept is not new, however, it was not until recently that ncRNAs were revealed as even more prevalent than previously imagined. Since then, they have received increased attention from the scientific community and many different ncRNA types are being discovered. Although initially considered non-functional elements (*the dark matter*) we now know that many are implicated in the post-transcriptional regulation of gene expression and thus they have been enclosed as "regulatory ncRNAs"[56,57].

Regulatory ncRNAs are classified in two general groups based on their size: short ncRNAs (<200bp) and long ncRNAs (>200bp) (Figure 11). Each of the groups consist of several classes of RNAs that exert different but overlapping functions. Small ncRNAs (sncRNAs)

include small interfering RNAs (siRNAs), microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs), endogenous small interfering RNAs (endo-siRNAs or esiRNAs), promoter associate RNAs (pRNAs), YRNAs and small nucleolar RNAs (snoRNAs). On the other hand, long ncRNAs (lncRNAs) include long intergenic ncRNAs (lincRNAs), natural antisense transcripts (NAT), enhancer RNAs (eRNAs), circular RNAs (circRNAs), and promoter upstream antisense transcripts(PROMPTS)[58] (Figure 11).

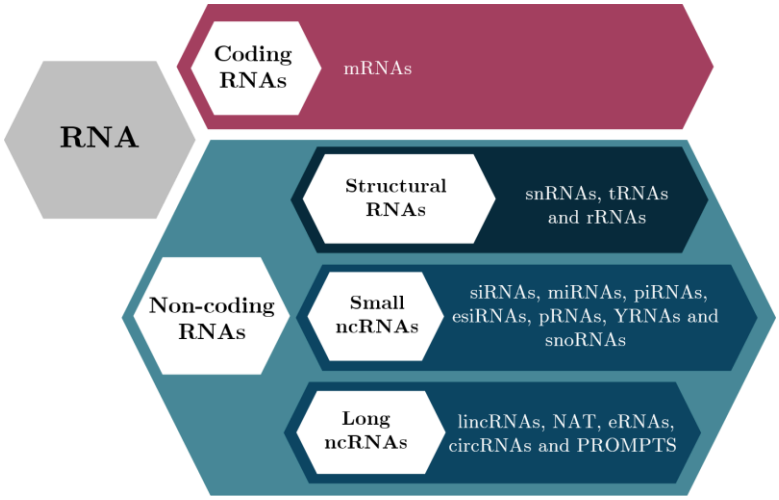

Figure 11. Scheme showing the classification of protein coding and non-coding RNAs. *Ilustration by Iparraguirre, L*

MiRNAs are the best studied ncRNAs since their discovery at the early 1990s. Currently, 2.654 different mature miRNAs are known to be produced from the human genome (miRBase 22.1, revised in July 2020) and since a single miRNA is able to regulate one or multiple mRNAs and a mRNA is also bound by one or different miRNAs, miRNAs appear to fine tune the translation and stability of at least one-third of all human protein-coding genes[52]. This suggests that miRNAs influence essentially all developmental, physiologic and disease processes[59].

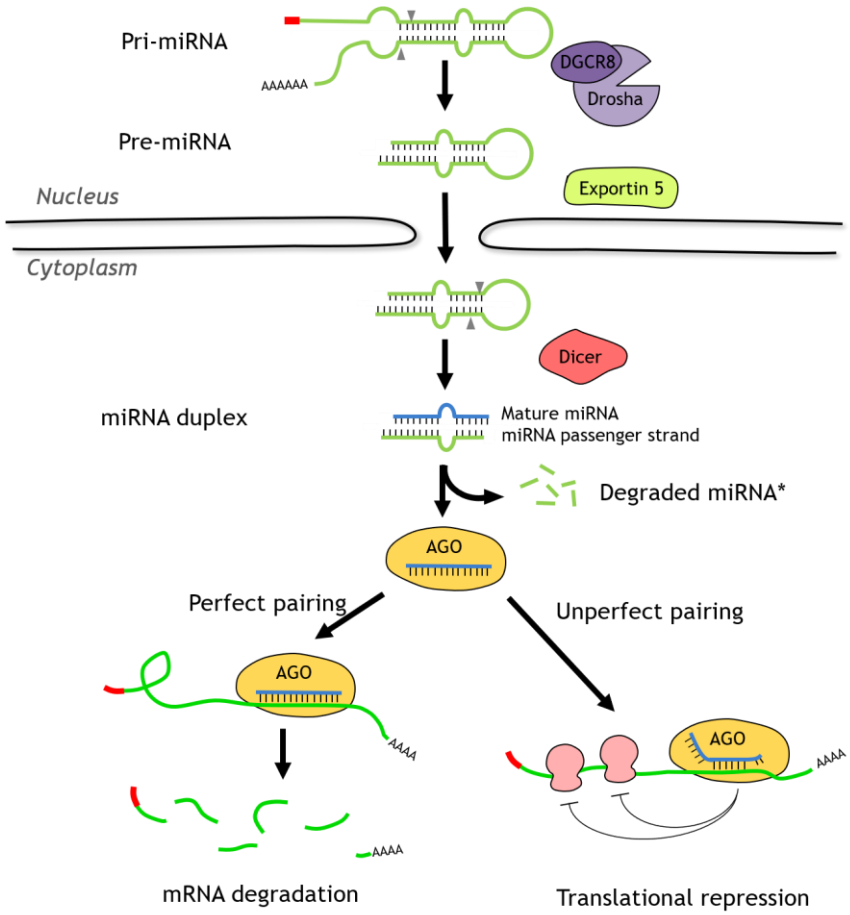

Figure 12. MiRNA biogenesis and function. Unlike other transcripts miRNAs undergo a special type of processing. Once generated and loaded into the silencing complex they post-transcriptionally regulate mRNAs. *Ilustration by Iparraguirre, L.*

These sncRNAs are the shortest ones with about 22 nucleotides in length which are in general transcribed by Pol II as part of a longer precursor with a variable length ranging from 100 to 1,000 nts called "pri-miRNA"[60]. This precursor is normally capped and polyadenylated but it is substrate of a special type of processing (Figure 12). Pri-miRNAs form hairpin structures further recognized and processed by the microprocessor complex, formed by the proteins Drosha and DGCR8. After cleavage by Drosha, pri-miRNAs turn into pre-miRNAs, a stem-loop precursor of about 60 to 120 nts. Pre-miRNAs are then exported to the cytoplasm by the action of Exportin 5 where it will be further processed by the endonuclease Dicer leading to an 18-25nt long duplex. Out of the two miRNA duplex strands the so-called passenger strand is rapidly degraded, while the mature miRNA or guide strand is hereafter incorporated to the RNA-induced silencing complex (RISC) which is formed by different proteins such as Argonaute (AGO) proteins. Once loaded into the silencing complex, RISC seeks out its target mRNA by searching for complementary

nucleotide sequences in the 3'UTR of the targeted mRNA. The miRNA binding to the 3' UTR inhibits further expression of the gene by repressing the translation of proteins, when the base-pairing between the miRNA and the mRNA is not perfect, or by cleavage of the poly(A) tail by the AGO protein which facilitates degradation of the target mRNA, in case of perfect complementarity[52,59](Figure 12).

This miRNA mediated mRNA downregulation process has been widely evidenced, however, some other studies have shown that some miRNAs are able to bind other mRNA regions such as the 5'UTR and can also upregulate the expression of specific mRNAs in some cell types and conditions[61].

Interestingly, one of the most recently discovered family of ncRNAs, circRNAs, have been proposed to regulate miRNAs acting as miRNA sponges which together with a number of other associated regulatory functions has turned the spotlight on them during the last years.

1.3. Circular RNAs

Circular RNAs (circRNAs) are a special and large class of RNAs characterized by the covalently closed loop structure that distinguishes them from the rest of the transcripts in the RNA world. Their unique mechanism of biogenesis and structure endows them special properties and promising applicability that are attracting more and more attention and will be presented in this section.

1.3.1. CircRNAs: Not so new ncRNAs

Although circRNAs are commonly introduced as the last discovered ncRNAs, they are not new. CircRNAs were first identified more than 40 years ago in 1976 by Sanger et al. while studying viroids as plant pathogens. They found that those viroids were unexpectedly resistant to phosphodiesterases, and both electron microscopy and sequencing unequivocally showed their circularity[62]. A second study in 1979 described some other viruses such as hepatitis δ also containing circular structured RNAs[63,64]. Nevertheless, these studies were not followed up at that moment probably due to the overwhelming evidences pointing to the importance of linear RNAs.

In the following decades, sporadic studies revealed that circRNAs were also endogenously produced across the tree of life[65,66]. The first evidence of an endogenously produced circRNA in human arrived in the early 1990s when Nigro et al. found four circular isoforms

of the DCC (Deleted in Colorectal Carcinoma) transcript in human cells, which were called scrambled exons[67]. After this initial report, a few more circular transcripts from different genes were serendipitously discovered, including the transcription factor ETS-1[68], cytochrome P450 2C24[69], dystrophin[70], sodium/calcium exchanger (NCX1)[71] and the sex-determining region Y gene (Sry)[72].

Excepting Sry circRNA that was found to represent up to the 90% of the transcripts formed from this gene in mouse testis[72], circRNAs were generally considered splicing artifacts unlikely to play any biological function due to their low abundance compared to their linear counterparts.

It was not until the advance of RNA sequencing technologies that this perspective changed entirely. High throughput RNA sequencing together with changes in library preparation protocols (without poly(A) selection) and circRNA specific algorithms have allowed to systematically identify thousands of circRNAs widespread in eukaryotes from fruit fly, worm, mice, pigs [73–77] or humans [78–80] to plants[81].

Since then, many other studies have been focused in unravelling the features[82], biogenesis mechanisms [83–85] and functions[86–91] of these intriguing molecules. In the same manner, a big effort is being done towards the discovery and fine tuning of experimental and computational methods for the study of circRNAs[92–95]. Moreover, their regulatory function in many cell functions together with their particular features has also lead to an increasing number of studies investigating their implication and potential application in different diseases such as cancer, cardiovascular diseases, neurological diseases, immune-related diseases, metabolic diseases and others [25,96–98].

1.3.2. Different circRNA types, distinct localizations

CircRNAs arise from different genomic localizations including exons, introns, intergenic or untranslational regions, can be formed following different mechanisms and result in different composition. As a consequence, different circRNA classifications can be found in the literature. The most common classification distinguishes three circRNA types regarding to their composition: exonic circRNAs, exon-intron circRNAs and intronic circRNAs[99].

- Exonic circRNAs (ecircRNAs) represent up to the 80% of circRNAs and consist of one or more exons, usually two or three. It is worth noting that for single-exon containing circRNAs a minimal exon length is needed for its circularization. These ecircRNAs are transported to the cytoplasm in a size dependent manner being short

circRNAs (<400nts) exported by the ATP dependent RNA helicase DDX39A (also called URH49) and long circRNAs (>1,200nts) by the spliceosome RNA helicase DDX39B (also called UAP56). For medium-size circRNAs the export regulation is likely to be affected also by the structure of the circRNA, and it is not well-defined[100]. Their cytoplasmic localization gives ecircRNAs the perfect context to exert their functions in post-transcriptional regulation (Figure 13).

- Circularized exons can sometimes retain introns between them leading to a second type of circRNA, exon-intron circRNAs or EIciRNAs. In a manner similar to incompletely spliced mRNAs, the presence of intronic sequences prevents EIcircRNAs from being exported from the nucleus where they can promote gene transcription by interacting with U1 snRNP and the Pol II transcription complex[101](Figure 13).
- Intronic circRNAs (ciRNAs) are formed by two or more introns and are localized to the nucleus where they have been suggested to function as positive regulators of RNA Pol II transcription[102](Figure 13).

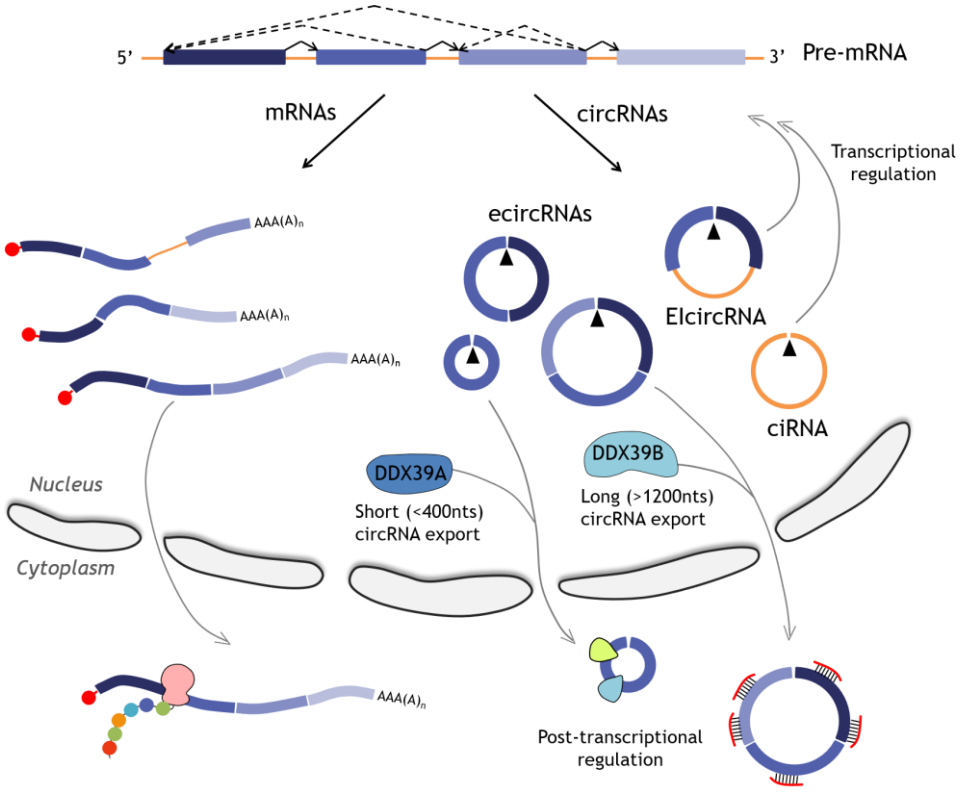

Figure 13. Three different types of circRNAs can be generated. EIcircRNAs and ciRNAs are retained in the nucleus and exert their regulatory functions during transcription. On the contrary, ecircRNAs are exported to the cytoplasm to post-transcriptionally regulate transcription. *Ilustration by Iparraguirre, L.*

1.3.3. The making of a circRNA

As presented in section 1.2.5.1., most eukaryotic genes contain intervening intronic sequences that during the processing of a pre-mRNA have to be removed. These introns are usually spliced in a sequential order resulting in a linear mRNA. This is not the case of circRNAs. They are also generated via splicing of a pre-mRNA catalysed by the canonical spliceosome machinery. However, introns are not removed following the canonical 5′ to 3′ order but a downstream splice donor is joined to an upstream splice acceptor thereby creating a circular transcript with covalently linked ends.

Currently, two major mechanisms of circRNA formation have been proposed for ecircRNAs and EIciRNAs: exon skipping (also called lariat driven circularization) and backsplicing (or intron pairing driven circularization)[79] whereas a third mechanism is being described for the biogenesis of ciRNAs[102](Figure 14).

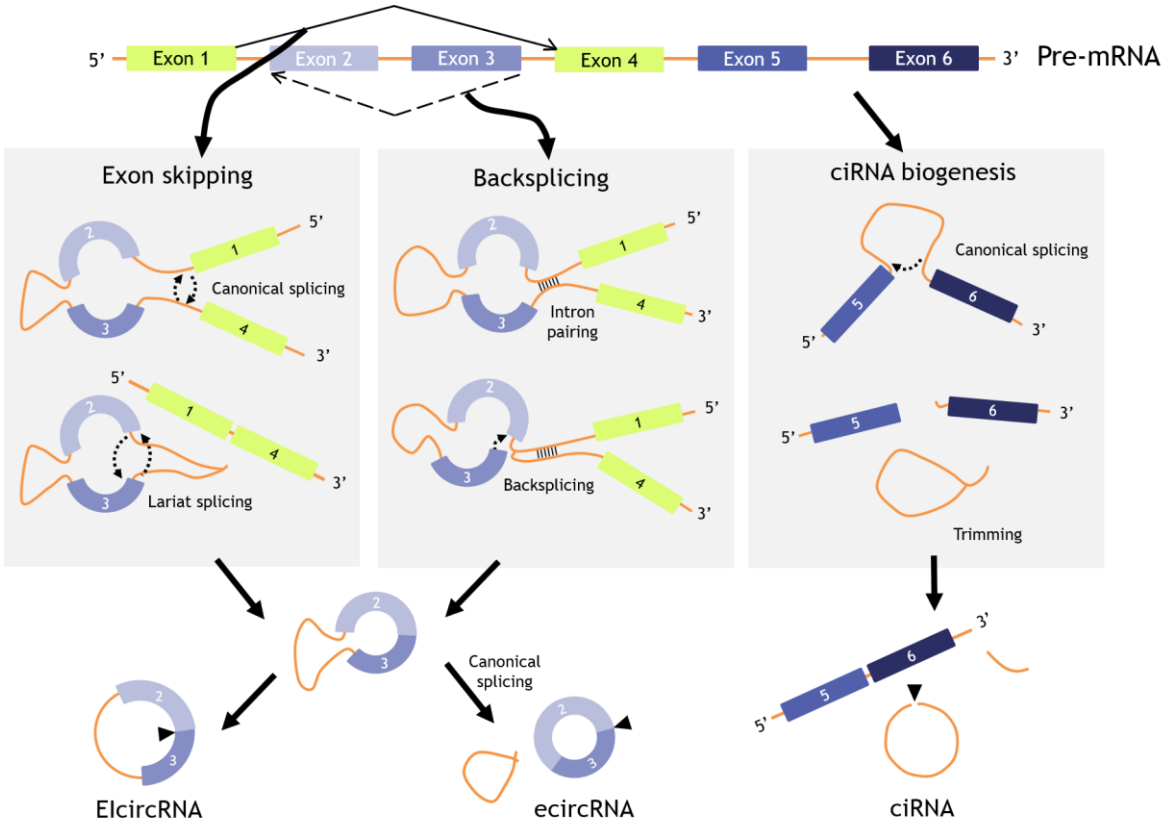

Figure 14. Three different mechanisms have been described for the biogenesis of a circRNA. The first two, exon skipping and backsplicing lead to the formation of EIcircRNAs or ecircRNAs while the third one depicts the biogenesis of a ciRNA. The backspliced junction (BSJ) of each circRNA is indicated by a black arrow. *Ilustration by Iparraguirre, L.*

Exon-skipping involves a canonical splicing reaction between two non-sequential exons resulting in a linear RNA with one or several skipped exons and a lariat intermediate containing the skipped exons. This lariat structure can itself define an EIcircRNA but it can also bring together additional splice sites within the skipped exons which may in some cases lead to a rapid internal splicing of the lariat in order to give rise to an ecircRNA[79] (Figure 14). Not all the lariats are subjected to a second round of splicing, it depends on its topology and on the speed of lariat debranching[103].

Nevertheless, evidence indicates that backsplicing is the most frequent mechanism for circRNA formation[93]. During backsplicing a 5' splice donor attacks and upstream 3' splice site leading to a 3'-5' phosphodiester bond that generates the circular RNA molecule. This backsplicing reaction relies on the looping of the intron sequences flanking the circularizing exons so that the splice sites are brought into close proximity. The pairing can be promoted by the presence of complementary sequences in flanking introns or it can also be RBP mediated which will be later discussed[104,105](Figure 14).

Last, intronic ciRNAs can be produced during canonical splicing when an intron lariat is resistant to debranching due to the presence of C-enriched motifs near the branch point. If it escapes debranching the 3' tail downstream from the branch point is trimmed finally resulting in a stable ciRNA[102](Figure 14).

It is worth noting that these three mechanisms result in the joining of two splice sites that are never joint together during the canonical splicing of a linear RNA, creating what is generally called a head-to-tail junction or bascksplicing junction(BSJ). This BSJ has become the distinctive feature of each circRNAs and most of the circRNA detection methods are based in the detection of this specific sequence.

Even though there are different circRNA types as previously introduced, ecircRNAs are the most common and most studied circRNAs and thus, in literature the general term circRNA is frequently used to refer to ecircRNAs. In order to make the reading easier from now on we will also use circRNAs to refer to ecircRNAs.

At the same time, backsplicing is the most common circRNA biogenesis pathway and thus, also the most studied one. Studies have evidenced that backsplicing mediated circRNA biogenesis can be regulated at three different points described below. First, circRNA biogenesis is coupled to the transcription of the circRNA-producing pre-mRNA by Pol II. Second, back-splicing efficiency depends on cis-regulatory elements and trans-acting factors. And third, circRNA turnover also regulates their levels[106].

1.3.3.1. Coupling between backsplicing, Pol II transcription and canonical splicing

It has already been mentioned that transcription and splicing are tightly coupled and regulate one each other. Similarly, recent studies have revealed that circRNA backsplicing is linked to canonical linear splicing and transcription thus being affected by both[107,108]. There are evidences for circRNAs forming co-transcriptionally, and also for some other circRNAs generated only after the transcription process is finished, therefore, the biogenesis of different circRNAs can be influenced by different events of the transcription and mRNA processing [106].

It has been shown that in general, circRNA producing genes have faster elongation rates allowing the pairing between complementary sequences across introns which results in facilitated backsplicing[107]. In addition, a study performed by Liang et al. found that the inhibition or slowing of co-transcriptional processing events shifts the transcript production from linear to circular isoforms. The processing of the backsplicing reaction requires the spliceosomal machinery and signals described for the canonical splicing although how it is involved remains still incompletely understood. Nevertheless, both splicing mechanisms have different sensitivities to the inhibition of splicing factors so that backsplicing is favored when active spliceosomes are limiting. In a similar manner, the 3'end polyadenylation increases the opportunity for backsplicing to occur and therefore also results in increased circRNA levels[108].

1.3.3.2. Backsplicing is regulated by cis elements and trans factors

Backsplicing is often facilitated by cis regulatory elements consisting of long flanking introns containing complementary sequences[84,109,110]. These sequences often can contain repetitive elements such as the human Alu sequences[79], but short repeats or even non-repetitive sequences are also sufficient to induce circRNA formation[109]. The pairing between these complementary sequences brings the distal splicing sites into close proximity to facilitate circRNA biogenesis (Figure 15). Moreover, for some genes multiple RNA pairs can be formed across different introns which leads to alternative backsplicing resulting in different circRNA isoforms from the same gene[110]. It is worth noting that pairing across different introns leads to circRNAs with different backsplicing junctions, but that circRNAs produced from the same gene locus can also present the same backsplicing junction and differ in the internal canonical splicing sites leading to different exon-intron composition[111,112]. On the contrary, the pairing within individual introns would facilitate

canonical splicing in linear RNA, which competes with backsplicing and leads to reduced circRNA formation[83].

Interestingly, there are circularizing exons that are not flanked by complementary regions and even circRNAs containing the same cis elements can be differently expressed between cells and tissues. These facts indicate that there are some other factors taking part in circRNA regulation as RBPs functioning like regulators in trans[106] (Figure 15). The role of RBPs in circRNA biogenesis is still to be further investigated, however, a few proteins have already been reported to regulate circRNA biogenesis. ADAR1, previously presented as the enzyme responsible for the A-to-I RNA editing, supresses circRNA expression by editing the RNA duplexes formed between flanking introns and thus reducing the complementarity and stability of the RNA pairs[84,113] (Figure 15). DHX9 for its part also inhibits circRNA expression thanks to its helicase activity that unwinds the RNA pairs flanking the exons to be circularized[114]. The splicing factor Quaking (QKI) strikingly increases circRNA formation by binding to flanking introns that may not contain complementary sequences and bringing the circularized exons together via dimerization of two QKI subunits[115]. In *Drosophila* another splicing factor, Musclebind (Mbl) generates a negative feedback loop by controlling the circRNA formation from its own pre-mRNA via binding the flanking introns[83]. Apart from those RBPs, later works have identified some others that are also able to exert regulatory roles in different cells and conditions including the nuclear factors NF90/NF110, epithelial splicing regulatory protein 1 (ESPR1), serine-arginine (SR) rich proteins, the splicing factor FUS, the heterogeneous nuclear ribonucleoprotein L (HNRNPL) or the RNA binding motif protein 20 (RBM20) [105,106].

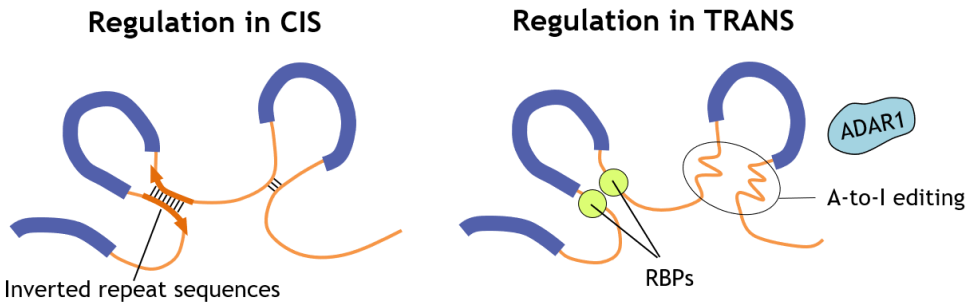

Figure 15. Pairing between flanking introns is essential for backsplicing to occur and it can be facilitated or disrupted by either CIS regulatory elements or trans acting factors. *Ilustration by Iparraguirre, L.*

1.3.3.3. CircRNA turnover also influences circRNA levels

Due to the fact that their circular structure and lack of free 3' or 5' ends circRNAs are not accessible to the canonical decay pathways known to degrade linear RNAs, therefore, for their decay, they should be cleaved internally by endonucleolitic attack. However, specific endonucleases that may cleave circRNAs have not yet been identified (except for the *in vitro* use of RNase H[93]), and thus, how circRNAs are degraded remains poorly understood.

In absence of a circRNA specific endonuclease, AGO2 mediated cleavage has been described for CDR1as (also known as ciRS-7)[116]. CDR1as has an almost fully complementary binding site for the miRNA miR-671, and the binding of miR-671 is able to trigger its cleavage by AGO2 with a mechanism similar to the one described for miRNA targeted mRNAs(Figure 16). In fact, CDR1as levels are directly modulated by miR-671[90]. Whether other circRNAs may undergo a similar miRNA-AGO2 mediated cleavage remains unknown[106]. Studies in human cell lines had also proposed that circRNAs can be actively exported to the extracellular space via release in extracellular vesicles (EVs) as a mechanism of clearance[87] although others suggest that it may be important for cell to cell communication(Figure 16).

In the last two years, three different studies have reported three new possible global circRNA degradation pathways. Park et al. reported a degradation mechanism for circRNAs containing the N^6-methylation of adenosine (m^6A) modification, which has been found to be widely distributed among circRNAs. After being recognized by the m^6A reader protein YTHDF2 circRNAs are endoribonucleolytically cleaved by the complex RNase P/MRP[117]. A study published by Liu et al. described a global degradation of circRNAs containing double-stranded RNA regions by RNase L. The 76% of the analysed circRNAs exhibit those RNA duplexes, however RNase L is only activated upon viral infection, or in autoimmune diseases[118]. Most recently, Fischer et al. uncovered an RNA decay mechanism that selectively degrades highly structured RNAs, both linear and circular, in normal conditions. One third of human circRNAs are predicted to form a highly overall structure which are bound by the RNA dependent helicase and ATPase UPF1 and its associated GTPase-activating protein-binding protein (G3BP1) that contains the endoribonuclease activity[119](Figure 16).

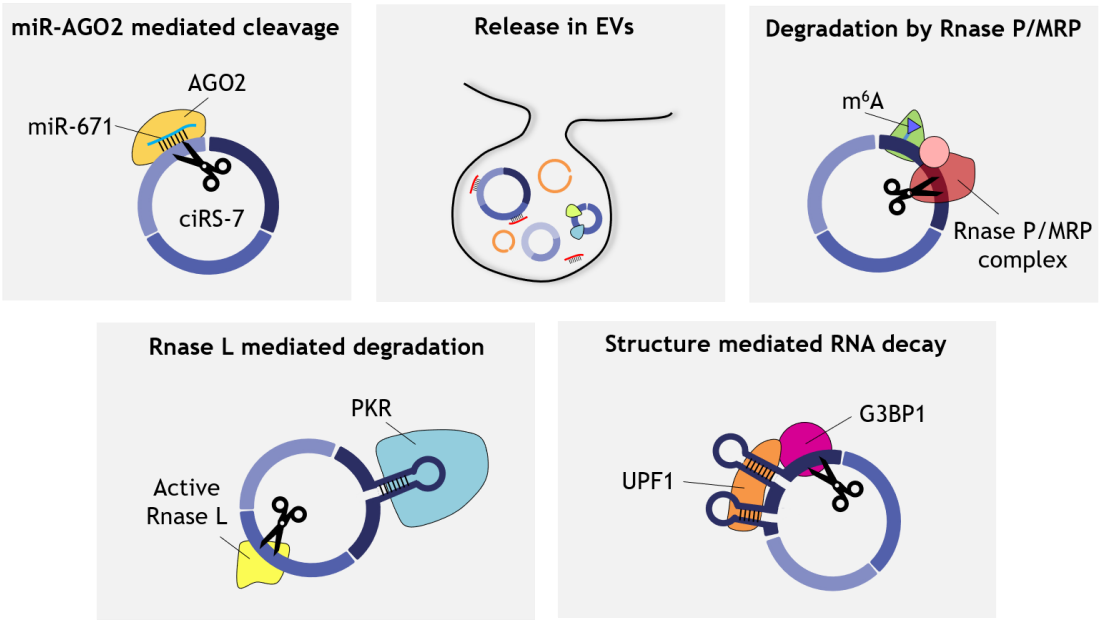

Figure 16. CircRNA turnover is still poorly understood but five degradation mechanisms have been proposed so far. *Ilustration by Iparraguirre, L.*

1.3.4. Properties and features

1.3.4.1. CircRNAs are evolutionary conserved

CircRNAs are evolutionary conserved at multiple levels. First, they have been found in most organisms ranging from archaea, plants, yeast and metazoans[105] in which a percentage of circRNAs are produced from orthologous (same gene in different species) loci[73,79,113]. Different studies have also analysed and confirmed the conservation of one or more splice sites and the intronic inverse complementary repeats that lead to backsplicing[84,113] indicating that circRNAs may have an important adaptative role and thus there is an evolutionary pressure for the generation of circRNAs. The conservation of their functions would imply the presence of their functional elements and structures across different circRNAs and species which still needs to be further investigated in order to gain more evidences[73,89].

1.3.4.2. CircRNA abundance is linked to their stability

The efficiency of backsplicing reactions is much lower than that of canonical splicing, which results in a generally lower circRNA expression when compared to linear RNAs[83,85,109]. These low efficiency may be explained with the sterically unfavorable assembly of the spliceosome when catalysing the ligation of the downstream 5' donor site with an upstream 3' acceptor[107]. However, the abundance of circRNAs is very variable between cells and

conditions ranging from less than one copy to hundreds of copies per cell[80,82]. In fact, studies have revealed more than 50 genes from which circRNAs are expressed more abundantly than their linear isoforms[78]. This suggests that the main function of some protein-coding genes may be to generate circRNAs rather than mRNAs or their expected protein products[120].

Often, circRNA abundance is linked to their stability. CircRNAs have covalently closed loop structures with neither 5'-3' polarity nor a polyadenylated tail that is the reason why they are more stable than linear RNAs[79]. Several studies support this stability[83,108], including a study performed by Enuka et al. showing that circRNAs have longer half-lives (18.8-23.7h) compared with their corresponding linear transcripts (4-7.4h)[121].

This stability and resistance to degradation make possible that slight changes in the efficiency of backsplicing result in the accumulation of circRNAs and consequent increase of their levels. Unlike circRNAs, linear RNAs have a higher decay rate thereby allowing circRNAs to accumulate more easily than linear RNAs and to be the dominant isoform in cells with low division rate [107]. On the contrary, circRNAs seem not to accumulate in tissues with a high proliferation rate leading to a circRNA dilution effect caused when the proliferation is higher than the production[122].

1.3.4.3. CircRNA expression is cell type- and tissue-specific

Another interesting feature is that although some circRNAs are more ubiquitously expressed, most of them show a cell type- tissue- and developmental stage-specific expression pattern [80,82].

Several studies have found circRNAs to be particularly enriched in brain and neural tissues of different species[73,113,123]. This enrichment may be due to the combination of several factors. On one hand, more circRNA host genes are found among brain-specific genes due to the fact that they carry more features that promote circRNA formation such as longer introns. Moreover, neurons display the highest rates of alternative splicing[124]. On the other hand, once circRNAs have been formed, their high stability allows them to accumulate in quiescent cells like neurons, especially with aging[73,75,77]. Interestingly, the fact that although cardiomyocytes usually do not divide and circRNAs are not accumulated in the heart[75], reinforces the idea that circRNAs do not simply accumulate in the brain but that there is a sum of factors that leads to their enrichment in the Central Nervous Sytem (CNS) indicating that they are important functional products.

1.3.4.4. CircRNAs are present in biofluids and EVs

CircRNAs like other ncRNAs and nucleic acids are actively secreted by cells either as a clearance mechanism or for cell to cell communication[87]. Free circRNAs have been detected in different biofluids including blood[88,125], plasma[126,127], saliva[128], semen[129], urine[130] and many others[131] making them ideal candidates as liquid biopsy biomarkers. Moreover, more than a thousand circRNAs have been identified from extracellular vesicles (EVs) in human serum[132,133], and several studies have confirmed that they are enriched in EVs when compared to cells[87,132–134]. Their enrichment together with evidences that point to a selective release of circRNAs suggest that they represent an additional source of information that is transferred from donor to recipient cells[132,133].

1.3.5. CircRNA functions

The function of >99.9% of identified circRNAs remains unknown, which has led to think that they may lack of functional significance [103]. However, there is a growing evidence of circRNAs playing and important role in several physiological and pathological processes and the molecular mechanisms by which they function are also starting to be unravelled[106]. Most of the already known mechanisms indicate that circRNAs shape gene expression at different levels including transcription initiation, elongation and post-transcriptional regulation. The different roles played by circRNAs in each of these stages of transcription are detailed during this section. In addition, some other non-regulatory roles have also been proposed and will be described (Figure 17).

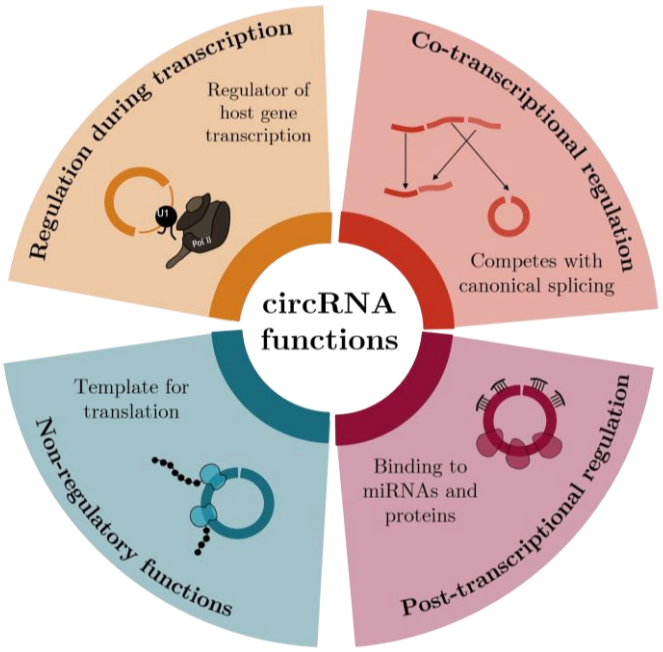

Figure 17. CircRNAs regulate transcription at different levels, and some other non regulatory roles have also been described. *Ilustration by Iparraguirre, L.*

1.3.5.1. Regulation at the initiation and elongation steps: EIcircRNAs interact with pol II at the promoter while ciRNAs are associated with the elongating Pol II

Transcription initiation mainly involves the recruitment of the pre-initiation complex including the Pol II at the promoter. Most of the regulation occurs at this step and is mediated by many transcription factors as abovementioned. Recent studies of crosslinking followed by immunoprecipitation of Pol II have found hundreds of circRNAs associated with it. Most of them are EIcircRNAs containing a retained intron with a binding site for U1 snRNP. This binding is followed by the interaction of the EIcircRNA-U1snRNP complex with the TFIIH transcription factor in the preinitiation complex and with Pol II at its promoter site, which results in the upregulation of transcription. The exact mechanism of action is still poorly understood but studies have revealed that several EIcircRNAs enhance the expression of their parental gene through this mode of action[101].

Regulation also takes place in the nucleus during transcription elongation, and some abundant intronic circRNAs have been found to colocalize with the phosphorylated Pol II complex at their parent gene loci, suggesting that they may regulate the transcription of their parental genes in cis. In support of this hypothesis, the knockdown of a ciRNA(ci-ankrd52) resulted in the production of a new mRNA isoform with retained introns and stop

codons that lead to its non-sense mediated decay and consequent decrease of the parent mRNA level confirming their regulatory role in transcription elongation[102].

1.3.5.2. Co-transcriptional regulation: backsplicing affects canonical splicing

Some circRNAs may be produced postranscriptionally, however, most of them are generated cotranscriptionally at the same time that the nascent transcript undergoes canonical splicing. CircRNAs are generally formed by the backsplicing of exons in their pre-mRNA, usually from the middle exons that are the ones with longer an less effectively spliced introns[135]. Thus, although backsplicing is less efficient than canonical splicing, they compete for the usage of 5' and 3' splice sites. This competition, whenever a circRNA is formed, can result in lower levels of the mRNA containing the circularized exons. The more an exon is circularized, the less will be present in a mRNA potentially leading to altered expression of the mRNA[83]. As an example, the dystrophin (DMD) gene was observed to produce several circRNAs in the skeletal muscle of patients with dystrophinopathy at the expense of reduced levels of linear in-frame mRNA and a decreased level of functional protein[136]. This example exhibits the important role of circRNA biogenesis in the regulation and manteinance of the right mRNA and protein levels.

1.3.5.3. Post-transcriptional regulation: circRNAs can regulate gene expression by sponging miRNAs

The majority of the circRNAs are exported to the cytoplasm, where they can interact with several key factors involved in transcription or its regulation such as miRNAs and RBPs and post-transcriptionally regulate the gene expression through them.

Although the function of very few circRNAs has been identified, the first demonstration of a functional circRNA came from the circRNA CDR1as which was proposed to act as a miRNA sponge[80,90]. Upon the presence of miRNA binding sites, circRNAs may compete with mRNAs for miRNA binding in the cytoplasm and thus prevent them from repressing the expression of their target genes (Figure 18). CDR1as, also known as ciRS-7 from "circRNA sponge for miR-7", is highly expressed in many tissues and particularly in brain[116]. It contains over 70 conserved miR-7 binding sites, and AGO2-CLIP data indicate that most of these sites are usually occupied by the miRNA[86], resulting in the capacity to bind up to 20,000 miR-7 molecules per cell. Thus, ciRS-7 was found to strongly suppress miR-7 activity, resulting in increased levels of miR-7 targets in cell lines. Moreover, miR-671 was suggested as an indirect regulator of miR-7 activity by its perfect binding and cleavage of ciRS-7, which would release miR-7[90]. A subsequent study created a ciRS-7

knockout mouse in which miR-7 was modestly but significantly decreased and miR-671 increased, resulting in a phenotype associated with neuropsychiatric disorder[86]. This *in vivo* study suggested that ciRS-7 functions as a miR-7 storage stabilizing and potentially releasing it upon a stimulus, rather than sequestering it[105].

Other than ciRS-7 whose miRNA binding function has been extensively demonstrated, whether other circular RNAs function in a similar manner is still being discussed. The miRNA sponging capacity of circRNAs has to be considered critically. It is important to consider the stoichiometric relationship between the miRNA binding sites of the circRNA, the mRNA target sites of the miRNA and the relative abundances of miRNA, circRNA and mRNA. To start with, the abundance of most of the circRNAs is generally low, and therefore, the miRNA might be more abundant than the total number of the available miRNA binding sites within the circRNA and still be able to bind to its target mRNAs. On the other hand, AGO2 PAR-CLIP data revealed that most circRNAs do not contain more miRNA binding sites than expected by chance, indicating that they do not extensively bind to miRNAs [92].

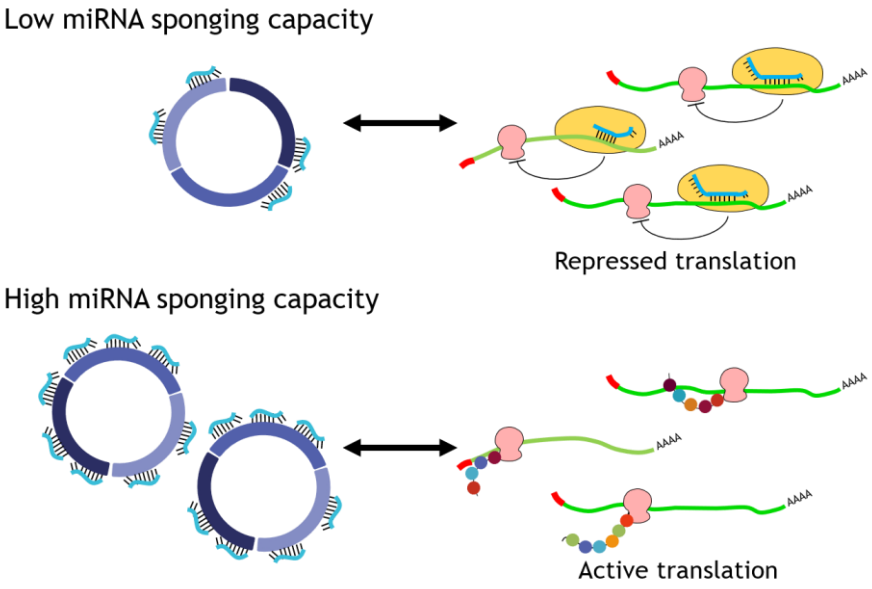

Figure 18. Some circRNAs compete with mRNA for miRNA binding. *Ilustration by Iparraguirre L.*

Therefore, although several studies have proposed different circRNAs acting as miRNA sponges (reviewed in [137,138]), it is unlikely to be a universal circRNA function. Only highly abundant circRNAs containing many competing binding sites such as the previously mentioned ciRS-7, the testis-specific sex-determining region Y circular RNA (circSry) that contains 16 target sites for miR-138 in mice[90], and the circRNA derived from the gene

encoding the zinc-finger protein ZNF91 (circZNF91) with 24 target sites for miR-23b-3p[92] are likely to exert this function[105,137](Figure 18).

1.3.5.4. Post-transcriptional regulation: circRNAs function by interacting with RBPs

Apart from binding miRNAs crosslinking immunoprecipitation data have revealed that circRNAs also interact with many different RBPs[139]. CircRNAs thus can act as reservoirs, scaffolds or transporters of those proteins or they can also have a cross-talk and regulate the protein function[105]. The circular RNA circMbl derived from the gene coding for the musclebind protein in *D. melanogaster* displayed the first example of a circRNA protein sponge. CircMbl is able to bind the mbl protein produced from its own pre-mRNA regulating the expression of both in an autoregulatory loop. When the mbl protein levels are low, the pre-mRNA is spliced to generate the mature mRNA so that the mbl protein is produced. However, as the protein level increases it is able to bind to the pre-mRNA promoting its backsplicing in order to switch the transcript production to the circular form. Once generated, the circMbl is able to sequester the mbl protein to reduce its concentration and consequently slow down the circRNA formation creating a negative feedback mechanism[83].

Other circRNAs also function through interaction with proteins. CircANRIL, circPABPN1 and cia-cGAS post-transcriptionally regulate the action of their respective target proteins (PES1, HUR and cGAS) by sequestering them (reviewed in [105,137]). circFOXO3 and circ-Amotl1 also bind several proteins although they do not supress their function but facilitate the colocalization of several proteins and their substrates (MDM2 with p53 and PDK1 with AKT1 respectively) so they can interact and function(reviewed in [137]).

Nevertheless, the circRNA abundance and RBP target sites also arises the question of whether this RBP sponge function can be a general phenomenon. Interestingly circRNAs have been found to function as a group to modulate protein function. Accumulation of ciRNAs allows to sequester mutated TDP43 protein aggregates reducing its toxicity in neurons of an amyotrophic lateral sclerosis model[140]. Two recent studies have also reported that circRNAs may be able to create a molecular reservoir of proteins involved in the immune response such as NF90/NF110(nuclear factor 90/ nuclear factor 110 complex) and PKR (protein kinase R or interferon induced double stranded RNA activated protein kinase) in order to ensure their availability and thus facilitating a rapid response to a extracellular stimuli such a viral infection[118,141].

1.3.5.5. Other than their regulatory roles, circRNAs can also be translated

CircRNAs are classified as ncRNAs, and in fact initial reports failed to find any evidence of circRNA translation. Both ribosome co-sedimenting experiments [79] and ribosome profiling data [92,123] indicated that circRNAs were not able to recruit ribosomes.

Moreover, circRNAs lack 5'cap and the poly(A)tail that are essential for the classical cap-dependent translation. Nevertheless, thousands of circRNAs include a putative open reading frame (ORF) with an upstream Internal Ribosome Entry Sites (IRES) suggesting that they could undergo a cap-independent translation[142]. CircRNA translation was demonstrated for some IRES containing synthetic circRNAs[143] but it was not until 2017 when three different studies revealed evidences for endogenous circRNA translation[89,91,144,145].

Yang et al. on one hand, reported that the m^6A modification was also sufficient to initiate cap-independent translation on circRNAs. On the other hand, they provided evidence of circRNA translation based on polysome fractionation and RNA-Seq identifying 250 circRNA candidates. They also tried to identify peptides whose sequence matches with translation across the circRNA BSJ and validated 19 of those peptides by mass spectrometry[144]. Pamudurti et al. identified a set of 122 circRNAs associated with active ribosomes in ribosome footprinting data from *D. melanogaster* heads. They also found that the circMbl encodes a protein that is present in fly heads as assessed by mass spectrometry. At the same time Legnini et al. were working on murine and human myoblasts where they evidenced the translation of circZNF609 (derived from the zinc finger protein 609 gene) by polysome fractionation and western blot.

After those, only a few more circRNAs have shown to be templates for translation (reviewed in [137]). This suggests that only a small subset of circRNAs may code for proteins and besides, the functional relevance of most circRNA derived peptides is still unknown. In fact, only a small proportion of circRNAs are associated with polysomes and cap-independent translation is not very efficient, indicating that circRNA derived protein products may be scarce[106]. Nevertheless, circRNA derived proteins display some interesting features. In the case of circZNF609 and circMbl, their translation has shown to occur under specific conditions of cellular stress and strarvation respectively which suggests that they can act as a regulated and inducible polypeptide source[89,91]. In addition, due to the high stability of circRNAs it has been suggested that they could produce significant amounts of protein over their lifespan with obvious biotechnology applications[103]. In the same line, circRNAs often share the start codon with the parental gene and have the stop codon after the BSJ, but it

is tempting to speculate that the circRNA could also lack a stop codon leading to endogenous rolling circle translation that would result in a long and repetitive protein[143].

1.3.6. CircRNAs and disease

The recognition of all the functional roles of circRNAs has given rise to a new perspective of circRNAs boosting their research interest. Studies have linked circRNAs with different cellular processes (autophagy, apoptosis, cell cycle, proliferation etc.) and processes involved in cellular physiology (neurogenesis, metabolism, embryonic development, immune response etc.)but also in disease pathogenesis[96,97,106]. These studies point out that circRNAs might play a key role in different diseases via different mechanisms, and have motivated the advent of studies on the association of circRNAs with different pathologies ranging from cancer, cardiovascular diseases, neurological disorders, immune-related diseases and metabolic diseases to aging[25,96–98]. Moreover, circRNAs have the potential to serve either as therapeutic targets, biomarkers or even as therapeutic agents[146,147].

1.3.6.1. Role as biomarkers

According to the Biomarkers Definitions Working Group (NIH), a biomarker is defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention"[148]. Their use has been extended and allows the early detection and diagnosing of different diseases as well as the classification of different disease subtypes, prediction of the prognosis and measurement of the responses to treatments.

There are several features that a biomarker must have to its suitable use at the clinical setting including stability, sensitivity, specificity, accuracy and reproducibility [147,149].Interestingly, circRNAs have demonstrated several remarkable features that make them very promising biomarkers: they are highly stable with half-lives longer than other transcripts[121], they are abundant in blood[88] and in other biofluids[131] and EVs[132] allowing their detection in samples requiring minimally-invasive procedures, their expression is tissue- and condition-specific[82] which suggests that they can be disease-specific biomarkers, and they are also evolutionary well conserved[79] which holds the potential of being easily translated from animal models to clinical application in humans.

Together with their features, different circRNA expression levels have been found between health and diseased tissues of several diseases, which in many cases also correlate with changes in the levels of circulating circRNAs either in circulating cells, extracellular vesicles

or free. As a result, circRNAs have been suggested as good biomarkers in many diseases(reviewed in [96,147]).

1.3.6.2. Interplay between circRNAs and the immune system

In the last years increasing evidence underscored circRNAs as important modulators in the immune system under physiological and pathological conditions including tumour immunity, activation of inflammation, antibacterial and antiviral responses as well as immunodeficiencies and autoimmune diseases (reviewed in [150–152]).

The general molecular basis of their involvement is still not fully understood, but some of the interactions between circRNAs and the innate immune response have been recently unraveled[153].

The innate immune response is the first line of defence against foreign pathogens and so, it counts with a set of pathogen-associated molecular pattern (PAMP) recognition receptors and effector proteins such as RIG-I (retinoic acid inducible gene I protein), MDA5 (melanoma differentiation associated protein 5), TLR3 (toll like receptor 3), PKR (protein kinase R or interferon induced double stranded RNA activated protein kinase), ADAR1 and NF90/NF110 (nuclear factor 90/ nuclear factor 110 complex)[153,154].

Chen et al. showed that exogenous circRNAs but not endogenous circRNAs are able to elicit a potent antiviral immune response[155]. Several viruses like the hepatitis δ virus and plant viroids have circular RNA genomes[62,64], whereas some others are able to produce circRNAs during their replication. Thus, mammalian cells have acquired the nucleic acid sensor RIG-I to sense and respond to such pathogens. In this study, Chen et al. found that transfection of a purified circRNA triggers the activation of RIG-I independent of its sequence or synthesis method. On the contrary, circRNAs that were transcribed in cells from plasmids were always identified as self presumably thanks to the RBPs that the spliceosome deposited onto the circRNAs as a self-mark[155]. Another recent study shows that RIG-I also discriminates between unmodified and m6A modified circRNAs and is only activated by the first ones. They concluded that m6A and probably some other modifications can also act as circRNA self-marks[51]. Intriguingly, and in contrast with those two studies, another report suggests that exogenous circRNAs may not be immunostimulatory but rather the 5' triphosphorylated linear RNAs resulting from an incomplete RNase R digestion during the preparation of the circRNAs[156].

Beyond their role in recognition and neutralization of exogenous circRNAs, some of these proteins appear to regulate the circRNA biogenesis. The NF90/NF110 complex appears to

promote the backsplicing events by binding to the RNA duplexes formed between flanking introns. Nevertheless, upon viral infection NF90/NF110 is exported from the nucleus to the cytoplasm to bind viral transcripts and inhibit their replication, thereby resulting in a reduction of the backsplicing events and decreased circRNA production[141].

Apart from having the immune system acting on circRNAs for their recognition or biogenesis, an active role of circRNAs modulating the function of the innate immune system has also been described. In fact, as previously mentioned, mature circRNAs are also able to bind NF90/NF110 in the cytoplasm thus creating a stable molecular reservoir of the molecules readily activable but yet maintained in an inactive state to prevent inappropriate reactions until there is a viral infection[141]. Similarly, in a subsequent work Liu et al. found that endogenous circRNAs form intramolecularly imperfect RNA duplexes of about 16-26bp that allow them to bind to PKR stronger than any of the previously described substrates and blocking its activation. Interestingly, they showed that upon a viral infection (simulated by the stimulation with poly(I:C) or EMCV infection) the activation of oligoadenylate synthetase (OAS) and the cytoplasmic endonuclease RNase L leads to a rapid degradation of circRNAs. PKR is then released and activated recognizing pathogenic double strand RNAs (dsRNAs) to aid in the immune response[118](Figure 19).

It is worth noting that a global reduction of circRNAs, increased RNase L activity and enhanced PKR activation have been associated with the autoimmune disease systemic lupus erythematosus (SLE) uncovering the mechanism of a pathophysiological connection between circRNA function and autoimmune diseases. This study also opens a way for future treatment since they found that the overexpression of circRNAs could robustly alleviate aberrant PKR activation in patient cells[118]. Other immune related applications have also been proposed. Chen et al. provided evidence that circRNAs act as vaccine adjuvants facilitating antigen cross presentation and inducing the activation of T and B cells. Moreover, the immunogenicity of circRNA may also be exploited for therapeutic applications such as against cancer[51].

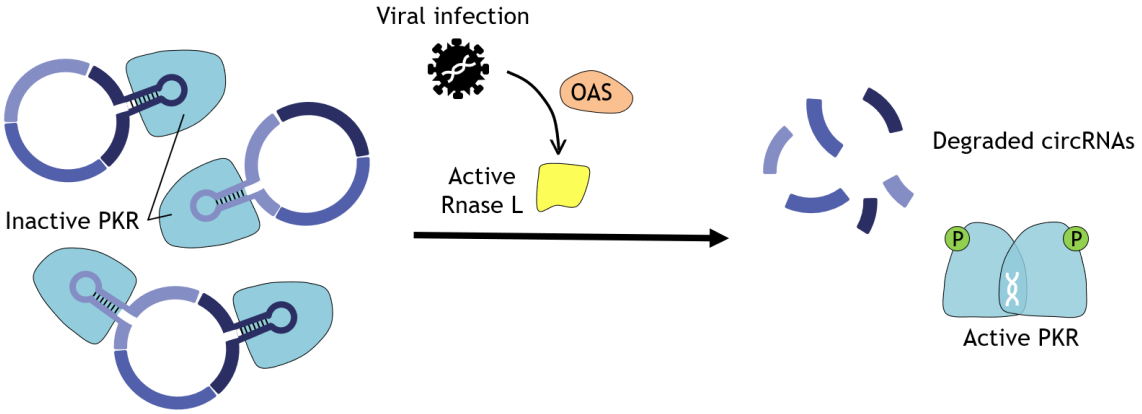

Figure 19. CircRNAs with dsRNA regions are able to bind and inhibit PKR. However, upon viral infection Rnase L is activated, circRNAs are degraded and PKR is released for its activation. *Ilustration by Iparraguirre, L.*

1.3.7. CircRNA databases and nomenclature

Together with different experimental and computational approached that have allowed to identify circRNAs, multiple circRNA databases have been developed as well. At the moment 20 circRNA databases have been reported, divided in curated (based on literature and experimentally validated) and non-curated databases (circRNAs obtained from publicly available datasets). Nevertheless, these databases present several issues and limitations that should be taken into account.

The first one is the limited overlap between databases which could be partially explained by annotation related issues (such as misannotated chromosomal positions) or the use of different samples, detections tools and filters when including circRNAs into the databases. The second limitation is that circRNA annotations such as circRNA-miRNA interactions, circRNA-RBP interactions or protein coding potential are mostly based on computational predictions rather than experimental validation[157]. At this point it should also be noted that most of these annotations rely on the full-length sequence of circRNAs which for most of them is inferred based on the start and end position and on the splicing sites annotated for linear RNAs on the reference genome, but does not take into account alternative splicing. The third major issue is the lack of a consensus circRNA nomenclature. Among the different nomenclature systems that are currently used, the circBase alias is the most widely used naming system which includes a prefix indicating the species where the circRNA was identified and a 7-digit number (eg. hsa_circ_0000001)[158]. Nevertheless, this nomenclature could lead to confusion with other databases that have included a similar but not unified coding. The second most extended naming system includes the circ prefix

followed by the host gene name (eg. circGENE), which is more readable but could also lead to mistakes when naming circRNAs produced from the same host gene or when naming circRNAs produced from genes that lack and official symbol, as is the case for a number of lncRNAs. At the moment, the only way to uniquely identify circRNAs and compare different nomenclatures is by using the BSJ position. Nevertheless, it is worth noting that using the BSJ position is not completely convenient neither, and two key points should be taken into account when using it. The first one is that in order to avoid missanotations, a unique coordinates format should be used (eg. same genome version). The second one is that BSJ coordinates do not take into account alternative splicing of circRNAs. Therefore, there is still an urgent need of a guideline to stablish a uniform and unique naming system to prevent confusion.

In this thesis, different circRNA nomenclatures have been used, but in order to facilitate the identification of each of them for the reader, the BSJ position is indicated for all the circRNAs mentioned (See Appendix Table 1).

1.4. Transcriptome regulation network: interplay between different regulatory levels

In this thesis we focus in unravelling the role, implication and potential application of circRNAs in an autoimmune disease called Multiple sclerosis that will be introduced in the next section. Nevertheless, it is important to bear in mind that the regulatory role of circRNAs is not isolated but through an enormous and complicated transcriptome regulation network.

A transcriptome regulation network, also known as a gene expression regulation network is the collection of molecular regulators in a cell that interact with each other directly or indirectly and are finally responsible of controlling the genes that are expressed in the cell under a given condition. During this section we have overviewed many of these regulators: histone modifications, transcription factors, activator and repressor proteins, splicing factors, several RNA binding proteins and miRNAs among others.

Finally, we have introduced circRNAs that although they are highly interesting molecules by themselves due to their particular features, their implication in the biology and pathology of a disease cannot be understood if we do not take into account their interaction with the rest of mRNAs, miRNAs and proteins in the cell.

2 Multiple sclerosis

Multiple sclerosis (MS) is a chronic disease of the Central Nervous Sytem (CNS) accepted as a demyelinating disorder in which the immune system attacks the myelin that protects the axons leading to neurological disfunction. It affects more than 2.5 million people worldwide, and its prevalence continues to increase[159]. There are evidences of paediatric or juvenile[160] and late-onset MS[161] cases but it typically affects young adults between 20 and 40 years and it is in fact the leading cause of non-traumatic disability in that age range[162]. MS is found to be three times more prevalent in women than men and there is evidence that this ratio may be increasing[163,164], a phenomenon shared with several other autoimmune diseases.

2.1. Etiopathology of the disease

MS has a complex pathophysiology and its causes are still not fully understood; however, it has been described that genetic, epigenetic and environmental factors contribute to the risk of developing the disease. Evidence of a genetic component was assumed from observations indicating family clusters. The first genetic studies back in 1970, revealed the human leukocyte antigen locus (HLA) as the first genetic factor related to the disease[165], being the HLA-DRB1*15:01 haplotype the one conferring the highest risk. Nevertheless, subsequent association studies have confirmed that the HLA gene cluster explains up to the 10.5% of the genetic variance. The rest of the genetic susceptibility is at least partially explained by hundreds of risk alleles present in other non-HLA genes contributing with a small effect to the overall risk[27,166]. Based on the largest meta-analysis of genome wide association studies up to date 200 additional loci have been associated with MS susceptibility[167]. Nevertheless, the concordance rate for monozygotic twins is of about 30%, indicating that environmental and lifestyle factors have also an important role in MS

susceptibility. Several environmental factors have been investigated such as vitamin D levels, distance from the equator, diet, smoking, toxins and viral infections[168]. The strongest evidence of association is related to the infection with Epstein-Barr virus (EBV). The prevalence of seropositivity in the general population is over 90% at the age of 4 years, however infections as young adult have been associated with three times more risk of developing MS (reviewed in [169]). Besides, in the last years, the gut microbiome has been identified as an important factor contributing to the immune system regulation and it has also been linked to MS pathogenesis by modulating the gut–CNS axis[170].

Moreover, the interplay between genes and environment is known to be bridged by epigenetic events that can influence gene expression in response to environmental factors. Therefore, it is easy to foresee that, several ncRNAs and epigenetic alterations have also been characterized and associated with MS(reviewed in [27,171,172]).

The pathological process underlying the disease is believed to start by an inflammatory process mediated by autoreactive T-lymphocytes. These T-lymphocytes are activated in the periphery, however the exact factor or factors that trigger their activation have not been identified yet. Different hypotheses have been proposed, being the most feasible the one suggesting a molecular mimicry between viral antigens (EBV antigens e.g.) and myelin antigens (myelin binding protein e.g.) that would lead to the loss of tolerance to myelin and cross-reaction of T-lymphocytes. The peripheral activation of T-lymphocytes is followed by a clonal expansion and production of proinflammatory cytokines as well as the expression of adhesion molecules that will favour their attachment to endothelial cells of the blood-brain-barrier (BBB). Once attached, T-lymphocytes cross the BBB entering to the CNS where they are reactivated by astrocytes or microglia and start producing different proinflammatory molecules that initiate the neuroinflammation. The neuroinflammation itself can be sufficient for damaging the myelin, but it can also promote the activation of cytotoxic T lymphocytes, macrophages or glial cells that finally damage the myelin sheath[173]. Myelin is produced by oligodendrocytes and does normally insulate axons allowing electrical impulses to transmit quickly and efficiently. In addition, myelin protects axons from being damaged and oligodendrocytes provide trophic support to neurons[174]. Therefore, the immune mediated demyelination leads to abnormally slow action potential transmission axonal damage and neurodegeneration.

2.2. Clinical forms and diagnosis

Based on the symptoms and evolution of patients, different clinical forms of MS are distinguished. However, it is worth noting that even before the appearance of any clinical symptom, during what it has been called the presymptomatic stage, subclinical inflammatory relapses can also be happening (Figure 20).

The clinically isolated syndrome (CIS) is the initial clinical presentation for most of the patients. CIS is recognized as a first episode of neurologic symptoms, lasting at least 24h, that is caused by an acute clinical attack and results in the inflammation and demyelination of one or more CNS sites. Depending on the area of the CNS affected by the lesion it can affect motor, sensory, visual or autonomic systems resulting in different signs and symptoms[169]. Some of the most common are fatigue, walking difficulties, numbness, spasticity, weakness, vision problems, dizziness, bladder or bowel problems, pain and cognitive or emotional changes[175]. This first episode is susceptible of being diagnosed as MS, but it still needs to fulfil some other diagnostic criteria[176]. Among those with CIS, more than the 80% will be finally diagnosed from MS in 20 years[177] usually after a magnetic resonance imaging (MRI) scan that confirms the presence of lessions occurring in different anatomical locations within the CNS (space dissemination) and over time (dissemination in time). [168,176].

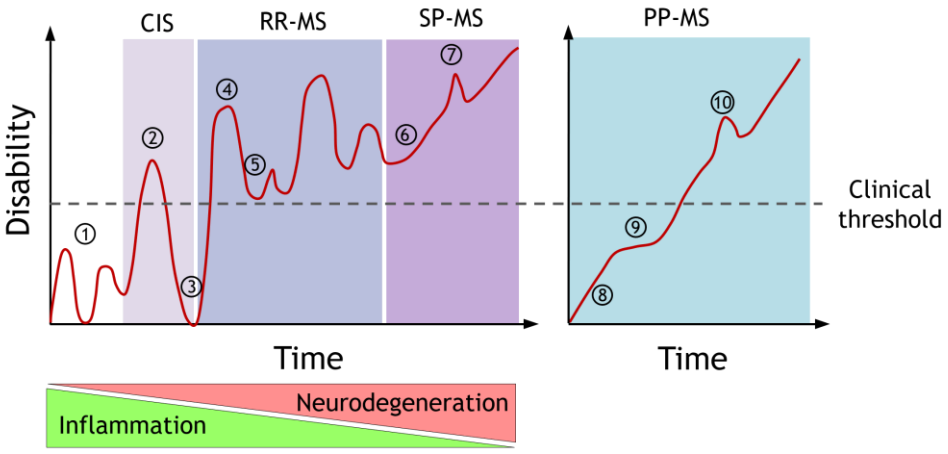

Figure 20. Scheme illustrating the evolution of the different clinical forms in MS assessed by disability progression with time. 1. Presymptomatic stage in which subclinical inflammatory relapses can be happening. 2. Presentation of the first neurological symptoms based on which the patient will probably be classified as a CIS. 3. Complete recovery from relapse. 4. Second relapse probably confirming the diagnose of MS as a RR-MS. 5. Partial recovery and accumulation of disability. 6. Worsening without relapses suggesting the switch to a SP-MS. 7. Puntual relapse. 8. Worsening of the disability from the onset without relapses indicative of a PP-MS subtype. 9. Period without progression. 10. Puntual relapse. *Ilustration by Iparraguirre L.*

For most of the patients (about 85%) the first definitive diagnose is the relapsing- remitting form of MS (RR-MS). RR-MS is characterized by clinical exacerbations, or relapses, caused by a new autoreactive immune attack to the CNS, that produces neurological disability symptoms lasting at least 24h. After these relapses the inflammation resolves and a partial or complete remyelination occurs entering a recovery phase named remission (Figure 20). During early stages of the disease usually there is a good recovery from each clinical episode, but over time, for many patients remyelination starts to fail and the recovery starts to be incomplete leading to disability accumulation. Earlier studies on the natural history of MS reported that 10 to 15 years after the diagnosis of RR-MS up to the 80% of patients converse to a second subtype named secondary progressive MS (SP-MS)(Figure 20). At the moment, with most of the patients under treatment, this percentage could have been reduced[178]. Anyway, this transition is usually gradual, and there are no clear criteria to determine when RR-MS converts to SP-MS. Nevertheless, SP-MS is generally characterized by a gradual worsening in absence or with very scarce exacerbations. In other words, the inflammatory component underlying the earliest phase of the disease loses strength leading to a more neurodegenerative phenotype. For another 10-15% of the patients the disease courses by worsening symptoms from the onset without a preceding relapsing-remitting phase, known as primary progressive or PP-MS[169,176,179–181] (Figure 20).

The diagnosis of MS, mostly relies on clinical symptoms and signs, however, based on the McDonald criteria stablished in 2010 and continuously revised[176,182,183], clinical information has to be integrated with imaging and laboratory findings to provide evidence of dissemination in time and space. MRI scans allow to confirm the presence of lesions at the brain or spinal cord that show up as white spots (Figure 21), and its combination with gadolinium as contrast agent could in addition reveal if the lesions are active. In some cases, CSF examination is also a valuable diagnostic test. This test seeks for evidence of intrathecal antibody synthesis that confirms the neuroinflammation in >90% of MS patients. When observed in a electrophoresis the intrathecal IgGs display a multiple band pattern that gives them the name of oligoclonal bands[184] (Figure 21). In addition to IgG bands, intrathecal IgM synthesis was also reported in MS patients and it is related to a more aggressive course of the disease with higher disability accumulation and risk of converting to SP-MS[185].

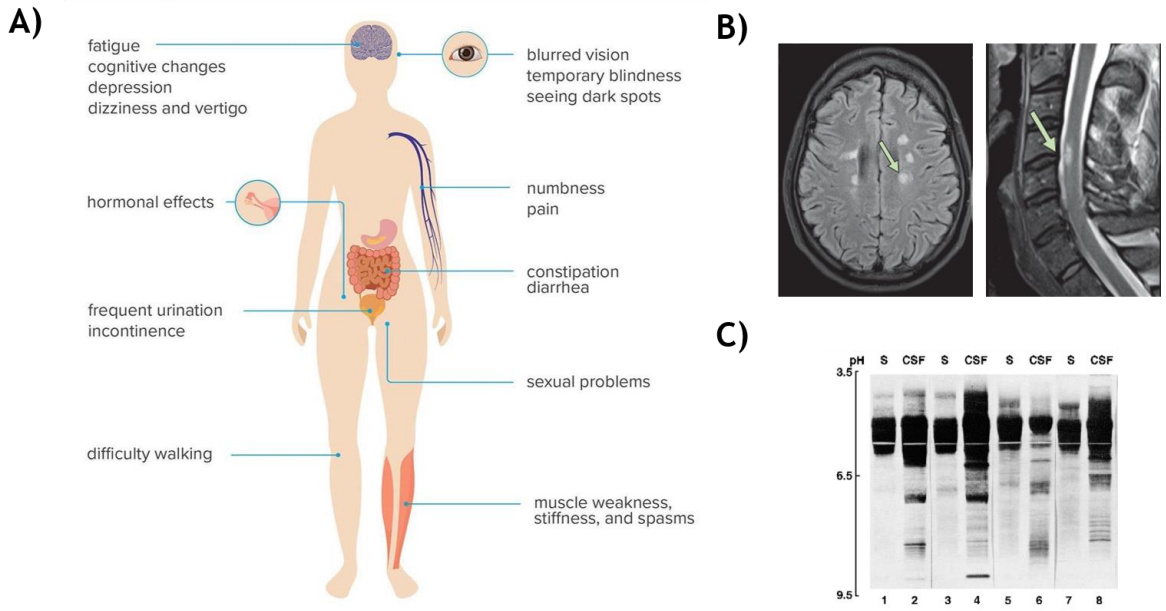

Figure 21. Diagnosis of MS is based on the clinical observation of sign and symptoms (A) and supported by some tests. The two most common tests are (B) a MRI scan that allows to confirm the presence of lesions in brain and spinal cord (marked by an arrow) and (C) the analysis of the CSF by isoelectric focusing electrophoresis to confirm the presence of oligoclonal bands in the CSF (and absence in serum(S)). *Adapted from*[184,186].

2.3. Sexual dimorphism in immune response and MS pathogenesis.

Male and females have the same immune cells and response mechanisms to protect them from pathogens or mistakenly recognize self-antigens, however the kinetics and magnitude of the responses differ dramatically between sexes. These differences result in different incidences of immune-related diseases. While males have a higher risk of suffering X-linked immunodeficiencies and infections, females have a greater risk of developing inflammatory diseases and represent the 80% of patients with autoimmune diseases, including MS [187,188](Figure 22).

In general, females mount higher innate, cell-mediated and adaptive immune responses which helps them to accelerate pathogen clearance but can also lead to the excessive immune responses associated with inflammatory and autoimmune diseases. Regarding innate immune response, it has been reported that the number of monocytes, macrophages and dendritic cells is higher in women as well as their activity and inflammatory responses[189]. In addition, antigen presenting cells are also more efficient in females than in males[190]. They also show a stronger adaptive response with a higher antibody production from B

cells[191,192] and an increased number of T cells (Th1, Th2 and Treg) that are also more active[193–196].

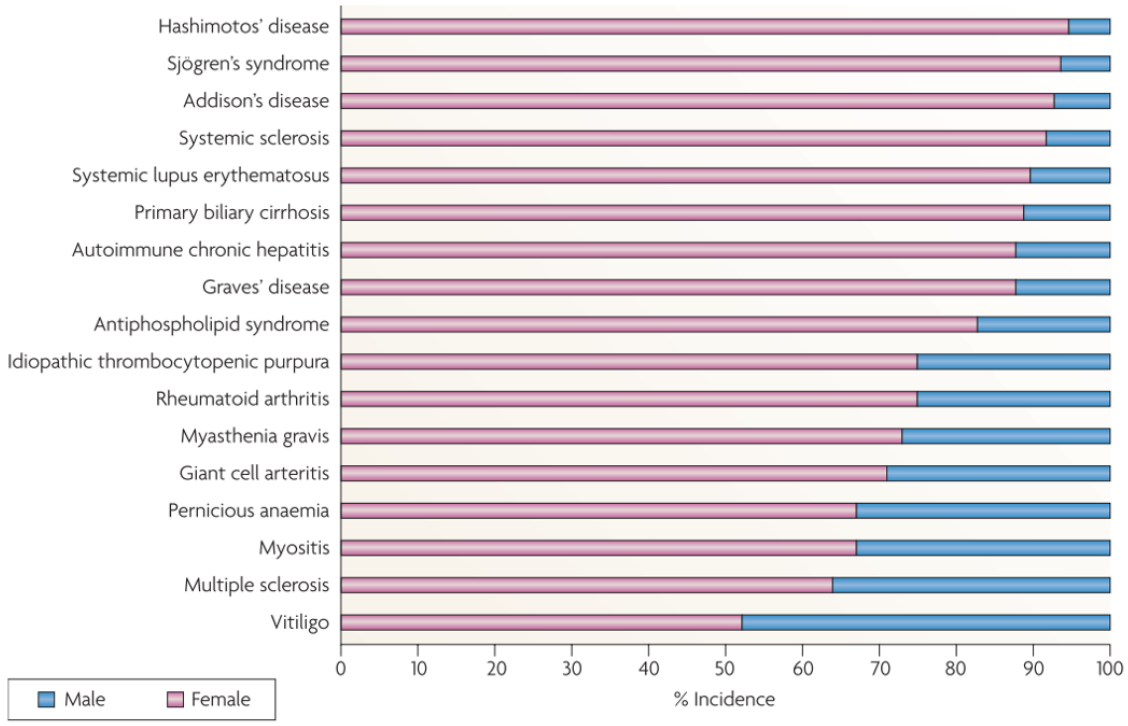

Figure 22. Sex distribution of some of the most important autoimmune diseases showing an increased incidence in females. *Figure from* [188].

In MS, the incidence, prevalence, disease course, severity and prognosis have been reported to be influenced by sex[164]. Remarkably, more females are affected but they usually present a more benign course with predominantly sensory symptoms, and men experience more motor symptoms, cognitive impairment and overall worse progression[197]. The higher prevalence in women has been linked with the stronger immune system responses. It is also considered the reason why women with MS display a more inflammatory phenotype characterized by a higher number of inflammatory relapses and active lesions on MRI[198,199]. In contrast, a higher number of degenerative lesions and increased axonal loss present in men with MS[199], are partially explained by differences in the neuroimmune system, particularly in the microglia. Male microglia has a higher antigen presenting capacity and also reacts more quickly leading to a more neuroinflammatory phenotype that easily triggers neurodegenerative processes[200].

Hormonal, genetic and epigenetic factors have been found to contribute to create these sex differences in immune function and in MS pathogenesis (reviewed in [164,187,201]). The role

of hormones was firstly supported by the observation of changes in the immune and disease phenotype of women during puberty, pregnancy or menopause[201,202]. Further studies have shown that sex hormones are able to exert their immunomodulatory role by binding to nuclear receptors expressed in immune cells such as lymphocytes, macrophages and dendritic cells and directly influencing their signalling pathways and production of cytokines[203]. The role of the three main sex-hormones in the immune system and MS pathogenesis will be summarized below.

Androgens such as testosterone are known to supress the activity of immune cells and the synthesis of proinflammatory cytokines as well as increasing the synthesis of anti-inflammatory cytokines[204]. In addition, a neuroprotective role has been proposed[205]. Accordingly, high testosterone levels during male puberty seem to be protective against immune-mediated diseases while their prevalence increases in older men when testosterone levels naturally decrease[164]. Moreover, low testosterone levels have been reported in the 40% of men with MS correlating with increased disability[206].

Estrogens at low levels enhance proinflammatory phenotype of T cells and proinflammatory cytokine production while sustained high concentrations can promote the opposite phenotype[207]. For this reason, high levels of estrogens, as in the case of pregnancy, are protective for women with MS, and their role decreases after menopause[208]. The role of estrogens in remyelination is still not clear but they have been suggested to participate in recruitment of the oligodendrocytes that produce myelin[208]. Similarly, immunosuppressive roles have also been described for progesterone by modulating the proinflammatory/anti-inflamatory balance and increasing the number of Tregs[209]. Progesterone is also believed to have neuroprotective and promyelinating effects in the CNS[210,211]. Studies in MS mouse model have revealed that it attenuates disease severity and decreases demyelination and axonal damage[212,213]. It has also shown to contribute to the immunosuppressive state during pregnancy[209].

On top of the effects of sex hormones in the immune function and in the demyelinating/myelinating process, the sex-biased expression of some genes also plays a role in modulating sex differences in the physiology and development of immune-related diseases(reviewed in [164,187,188]). On one hand, sexual chromosomes make a difference between males and females, with the X chromosome as the key factor. Its important implication relies in the fact that the X chromosome contains a high proportion of genes that directly or indirectly regulate immune function[214]. Furthermore, X chromosome presents more miRNA genes than any other autosomal chromosome with important

functions in immunity that also have been postulated to contribute to a different gene expression regulation between male and females[215]. Females carry two X chromosomes while males only inherit one from the mother. To compensate the gene dosage between sexes, females undergo the inactivation of one of the X chromosomes however, a number of genes (about the 15-25%) have been found to escape silencing[216]. Taking into account the elevated number of immune-related genes, this inactivation escape as well as other X chromosome specific abnormalities, might result in an elevated number of immune proteins in females that would correlate with the reported gender differences[188].

Nevertheless, the most recent meta-analysis of genome-wide MS data, has only mapped to the X chromosome one out of the 233 loci significantly associated with MS[167] indicating that the X chromosome does not play a crucial role in MS susceptibility. In contrast, several genetic and gene expression studies have evidenced that many other autosomic genes show a sex-specific association or gene expression in patients with MS that have also been associated with different underlying pathogenic events between males and females[217–221]. Further, different and sex-specific ncRNAs profiles have been described for healthy individuals and MS patients indicating that they may also contribute to explain the differences in the biology and pathology between sexes [222–224].

Based on all these studies, it is clear that there is a sex dimorphism affecting biological processes such as the immune response and that it also affects the susceptibility, developing and progression of diseases[225]. Therefore, sex should be taken into account in biomedical research with the aim of obtaining more reproducible observations that could in addition uncover the sex-specific processes and targets that could be otherwise overlooked[217]. Moreover, the incorporation of individual characteristics of female and male biology into the health care would lead us to a more personalyzed medicine that would potentially improve the prevention, management and treatment of both sexes in diseases such as MS[164].

2.4. The need of biomarkers in MS

Although MS diagnosis can be supported by MRI scans to detect the CNS lesions, and the study of CSF to reveal oligoclonal bands, a big part of the evidences come from clinical examination and signs. However, the frequency and duration of symptoms differ between patients and also from time to time in the same patient. This variability together with a high heterogeneity in radiological and pathological features and in treatment response as well as the unpredictable relapses and remissions, make MS diagnosis and clinical management quite challenging [226]. Therefore, new tools such as body-fluid biomarkers

that could reflect all this heterogeneity are urgently needed not only to facilitate the clinical management but also to further understand the etiopathology of the disease. Moreover, it has been observed that an early and appropriate treatment greatly improves the disease evolution and outcome which reinforces the importance of early disease biomarkers[227].

Based on their use, biomarkers can be classified in different types: (1)Predictive biomarkers can be used to identify individuals at risk of developing MS among familiars. (2)Diagnostic biomarkers are useful in individuals presenting neurological symptoms to confirm the diagnose and distinguish it from other MS-mimicking diseases. (3)Disease activity biomarkers can be indicative of an underlying pathological process such as inflammation, demyelination or axonal degeneration. (4) Prognostic biomarkers are intended to predict whether the evolution of the disease will be more or less aggressive. And finally, with the expanding number of available therapies for MS there is also a need of (5) treatment-response biomarkers that can guide the dose selection, allow the monitoring and predict the risk of adverse effects[226,228].

Different types of molecules can serve as biomarkers, being the most common protein, DNA or RNA molecules as long as they satisfy some conditions to be suitable for the clinical use. As it has been previously presented, they must be stable, sensible, specific, accurate and reproducible [147,149] but they also need to be easily detectable in a clinical setting. Thus, molecules should preferentially be present in body fluids. Since MS affects the CNS, cerebrospinal fluid (CSF) is the most direct source of biomarkers, and the one that may reflect more reliably CNS mechanisms. However, the extraction of CSF requires a lumbar puncture which is quite invasive, requires a neurologist and can lead to adverse effects and thus a great effort has been done towards the discovery of minimally-invasive diagnostic biomarkers[169,229–231]. Peripheral blood for example is less invasively collected than CSF and it reflects peripheral immune mechanisms but has also been found to reflect changes in the CNS[231,232].

The advent of omic technologies including genomics, transcriptomics and proteomics have allowed to discover thousands of molecules with potential to be biomarkers of the disease. However, the heterogeneity between patients and low sample numbers together with the variability in the collection/storage conditions and sample processing or differences in study design and analysis have complicated the validation of most of them. Currently only a few biomarkers have been approved for their use in clinical practice. Among them, it stands out the assessment of CSF oligoclonal bands for the confirmation of the diagnosis. The identification of Anti-aquaporin-4 antibodies in the serum as specific markers of

neuromyelitis optica (NMO) was also a step forward towards the differential diagnosis of MS among patients that often show similar symptoms[233]. Apart from those two, the presence in serum of four different antibodies can be used as treatment response biomarkers. Regarding natalizumab, the sustained presence of anti-natalizumab antibodies indicates a suboptimal clinical response[234] and the detection of Anti-JC virus antibodies indicate a high risk of developing progressive multifocal leukoencephalopathy[235], making necessary to discontinue the treatment. In a similar manner, neutralising antibodies in patients treated with a Interferon-beta suggests a loss of efficacy being necessary to consider a switch to a non-interferon-beta treatment[236] and in patients eligible for fingolimod treatment it is important to check for the presence of antibodies against varicella-zoster virus (and vaccinate seronegative individuals) before starting the treatment to avoid the risk of infection by this neurotropic virus that can result in a life-threatening disease in adults[237].

Among the biomarkers that are still in an exploratory phase, and to the purpose of this thesis, it is interesting to highlight the role of miRNAs. In the last ten years, many different studies have identified, first in immune and neural cells and later in biofluids, hundreds of dysregulated miRNAs. Some of them have been found to be implicated in the physiopathology of the disease potentially serving to measure disease processes and treatment responses. They have also emerged as an attractive clinical tool because they are highly stable in blood and easy to collect and quantify in an accurate and cheap manner(reviewed in [27,179,238,239]).

Justification

Since the first case of MS was reported, the joint work of medical doctors and researchers has allowed to advance in the understanding of the disease and also many treatments have been developed. These advances have dramatically changed the natural history of the disease which seems to be generally milder now. Nevertheless, MS is a complex disease, there are still many gaps in the understanding of the etiology and pathological mechanisms of MS and plenty improvements can still be made in terms of the diagnosis and clinical management of patients.

Previous works from our group and others have focused their efforts in the study of gene expression and its regulation as a key factor in the disease as well as like a good approach for the discovery of new targets and biomarkers for MS. However, recent studies in the RNA field keep showing that the regulation network is growing, and have reported circRNAs as new players that have not been studied in MS yet.

In this context, the present project was outlined in 2017, aiming to advance knowledge of the regulation network by including the expression of circRNAs and to contribute to the ultimate goal of improving the clinical management and quality of life of patients with the investigation of new potential biomarkers.

Hypothesis and objectives

Hypothesis

The importance and interest of transcriptomic profiling in several diseases including MS has been widely stablished in literature. In this scenario, during the last decades ncRNAs such as miRNAs stand out as key regulator factors of the transcriptome which are implicated in the physiopathology of MS. In light of the recently reported interactions between miRNAs and circRNAs, as well as all the additional regulatory roles described for circRNAs and their implication in immunity, the initial hypothesis of this thesis is that **circRNAs may be involved in the biology of MS**. Starting from this basis, we postulated several particular hypotheses that lead to the subsequent objectives:

- As long as circRNAs are part of the dysregulated transcriptome regulation network in MS, and leukocytes reflect part of the pathology of the disease, we hypothesize that circRNA expression is also altered in leukocytes from patients when compared to controls.
- Almost all cell types including immune cells and cells from the CNS, produce extracellular vesicles that can be found circulating in blood. CircRNAs together with many other molecules are incorporated in these EVs and thus could be reflecting an additional part of the underlying pathological processes.
- Blood cells and plasma derived EVs are easily accessible, and are interesting sources of biomarkers. Taking into account the intrinsic stability of circRNAs both in cells and in biofluids, circRNAs hold a great potential to be good biomarkers in MS.
- CircRNAs have been revealed to be functional regulatory molecules implicated in many biological processes, therefore we further hypothesize that an altered circRNA profile in MS patients could be linked to the aberrant functioning of immune cells in MS.

Objectives

- To define the differences in the circRNA profile between MS patients and controls and evaluate any clinical or biological characteristic that could be affecting the profile (disability, evolution time, sex...).
- To find one or several circRNAs that could complement the current tests for the early diagnosis of the disease and evaluate their biomarker potential.
- To find one or several circRNAs in blood that could correlate with the presence of IgM oligoclonal bands in CSF and could thus work as prognosis biomarkers.

- To investigate the potential molecular function of the dysregulated circRNAs and their implication in MS.

CHAPTER 1

CircRNA profiling in MS

General introduction

MS is a complex and chronic autoimmune disease of the central nervous system characterized by the demyelination of axons that leads to neurological disability. The disease course and clinical phenotype of MS are highly variable among patients and also with time in the same individual[181]. For this reason, biomarkers that can aid in the diagnosis, the differentiation of MS phenotypes and also in the monitoring of disease progression are needed[226,239].

Currently only a few biomarkers have been approved for their use in clinical practice. Among them, it stands out the assessment of CSF oligoclonal bands for the confirmation of the diagnosis[184]. However, CSF is not an ideal source for biomarkers to be used on an ongoing basis, since multiple lumbar punctures are not recommended due to its invasiveness and potential adverse effects. Therefore, a great effort has been done towards the discovery of minimally-invasive biomarkers such as blood biomarkers[169,229,230].

To this aim, and since the early 2000s with the arrival of advanced omic techniques, different approaches have been carried out. One of them is the gene expression profiling that, apart from uncovering new biomarkers[240–243], allows also to elucidate the molecular mechanisms underlying the pathogenic processes occurring in the disease[244,245]. The first works were focused in the profiling of the coding transcriptome, but from 2009, the non-coding transcriptome has gained great interest, and several authors, including our group, have demonstrated that it has a role in MS pathogenesis[220,246–249].

Recently, circRNAs have emerged as a new player in the transcriptome and its regulation network by interacting with several other transcripts. They have also been found to participate in several processes such as tumor, metabolism and immune-related pathways. Consequently, they have also been linked to various diseases including cancer[25], cardiovascular diseases[250], neurological diseases[251] and autoimmune diseases[151,152]. Moreover, circRNAs have also been regarded as promising source of biomarkers in samples such as blood, serum, saliva, urine and CSF, as well as in extracellular vesicles, due to their unique stability in body fluids[147]. And several circRNAs have been postulated as biomarkers in different diseases[25,147]. However, to our knowledge they had not been studied in MS before the beginning of this thesis.

Two main different approaches have been described for the profiling of circRNAs, microarrays and RNASeq. Both techniques can or can not be preceded by a sample

treatment for the enrichment of circRNAs. Such treatments include depletion of ribosomic RNA, removal of polyadenylated RNAs and/or treatment with RNase R that selectively degrades linear RNAs[94]. In the microarray based profiling, circRNAs are detected by their hybridization to probes designed spanning the backspliced junction. At the moment, Arraystar is the only company offering expression arrays with probes for 13,617 circRNAs. On the contrary, RNASeq allows to also detect and quantify the abundance of new RNA species, including circRNAs. In this case, the quantification relies on the detection of reads that are mapped to the backsplicing junction that requires of specialized pipelines and algorithms[95,252].

The aim of our study is to characterize the expression of circRNAs in MS patients and healthy controls to determine whether their expression is changed with the disease. In addition, we also want to investigate the biomarker potential of the circRNAs and their involvement in the biology of the disease. To that end, we have studied different sample types (leukocytes, Peripheral blood mononuclear cells (PBMCs) and extracellular vesicles (EVs)) and conditions (RR-MS, SP-MS, healthy controls, and patients with IgM oligoclonal band characterization) by using the two different screening methods previously mentioned. Each of the experiments is presented in a separated subchapter:

1. Microarray based circRNA profiling of leukocytes from RR-MS patients and healthy controls.
2. RNASeq based transcriptome profiling of leukocytes from different MS types patients and healthy controls.
3. RNASeq based transcriptome profiling of PBMCs from patients with IgM oligoclonal band characterization.
4. RNASeq based transcriptome profiling of plasma derived EVs from MS patients and ealthy controls.

1.1 Microarray profiling of circRNAs in leukocytes

Part of this chapter has been published in *Human Molecular Genetics:*

Iparraguirre L, Muñoz-Culla M, Prada-Luengo I, Castillo-Triviño T, Olascoaga J, Otaegui D. *Circular RNA profiling reveals that circular RNAs from ANXA2 can be used as new biomarkers for multiple sclerosis*. Hum Mol Genet. 2017 Sep 15;26(18):3564-3572. doi: 10.1093/hmg/ddx243. PMID: 28651352.

Introduction

Blood transcriptome has been extensively studied in Multiple sclerosis (MS). It has been widely shown that both coding and non-coding transcripts are differentially expressed in patients and healthy controls.

MiRNAs have been the most widely studied ncRNAs, and in fact, miRNA expression studies in blood and plasma have generated a significant amount of data about circulating dysregulated miRNAs. There is evidence of at least 62 circulating miRNAs significantly dysregulated in MS patient's blood while profiling results are still continuously emerging[238,249]. Several of those candidates have already been proposed as biomarkers for disease diagnosis, severity or response to therapy, but still none of them has been incorporated to the list of biomarkers in clinical practice [179,226].

These studies, together with others support that the blood transcriptome network is changed with the disease and holds a great potential to be used as minimally invasive biomarkers,

however, there are still some shortcomings that have to be addressed. In this regard, circRNAs due to their circularity, present several characteristics for which they have been reported as promising candidate for biomarkers of human diseases. Indeed, several circRNAs have already been suggested as biomarkers for a potential non-invasive diagnosis of several diseases such as vascular diseases, neurological diseases or cancers (reviewed in [149,253]) but not in MS.

In this context, the current study aimed to investigate the expression of circRNAs and their validity as biomarkers that could in a future complement the current diagnostic methods. To that end, and as a first approach we performed a profiling of circRNAs in leukocytes from RR-MS female patients and healthy controls by using microarrays (Arraystar Human Circular RNA Microarray V2.0).

Materials and methods

Blood sample collection

Whole blood was collected from a total of 45 patients with RR-MS, 10 CIS cases and 26 healthy donors in the Department of Neurology at Donostia University Hospital. It is known that gender affects the expression of the genes, and as the prevalence of MS is higher in women (around 65%), we decided to study only women in the discovery cohort and in the validation cohort, in order to avoid background noise coming from gender differences. Recruited individuals were distributed into five different cohorts.

For the discovery cohort we selected four RR-MS untreated patients in remission, while in the validation cohort 20 RR-MS untreated patients in remission were included. For each cohort four and 18 age matched healthy donors were selected respectively. Moreover, we included a second validation cohort of 21 RR-MS patients (twelve females and nine males). In this cohort, we wanted to analyse the phenomena occurring during relapse and to do that, two blood samples were drawn from those 21 patients : one during a relapse and another during a remission. A relapse was defined as an episode of new neurological symptoms of at least 24-hours of duration, not associated with fever or infection. Relapse blood samples were collected before giving any corticosteroid treatment. A third cohort of 10 patients in CIS and 10 age-matched controls was also analysed (out of those 10 controls, 6 have been also included in the first validation cohort). CIS refers to a first episode of neurologic symptoms that may or may not go on to develop MS. In this case, all the patients in CIS selected for the study have been later diagnosed from RRMS in order to assess the

predictive value of the circRNAs at this stage. Finally, 17 patients diagnosed from SP-MS were also recruited in order to investigate the expression of circRNAs in a more advanced stage of the disease.

The main clinical and demographical characteristics of both patients and healthy donors are summarised in Table 1. Samples from all donors were collected after receiving written informed consent. The study was approved by the hospital's ethics committee (UEM-IMN-2017-01) and samples have been processed and stored at the Basque Biobank (http://www.biobancovasco.org).

Table 1. Main clinical and demographical characteristics of the individuals enrolled in the study. A total of 45 MS patients and 21 healthy controls were studied after being separated into five different cohorts.

		Sex	Age	EDSS	Evolut. time	AOO
Discovery cohort	RR-MS (n=4)	Female	43,8 (±6,8)	3,5 (2-4)	19,3 (±3,8)	24,5 (±5,1)
	HC (n=4)	Female	34,3 (±6,7)	-	-	-
1st validation cohort	RR-MS (n=20)	Female	48,5 (±14,3)	1,5 (0-6,5)	15,2 (±15,2)	33,3 (±10,9)
	HC (n=18)	Female	44,9 (±19,9)	-	-	-
2nd validation cohort	Relapse (n=21)	Female(n=12) Male (n=9)	40,1 (±10,1)	2,5 (0-7,5)	9,4 (±8,8)	30,1 (±11,5)
	Remission (n=21)	Female(n=12) Male (n=9)	40,4 (±10,3)	2,5 (0-7,5)	9,4 (±8,6)	
CIS cohort	CIS (n=10)	Female	35,5 (±10,4)	1,6 (0-4)	0,5 (±0.9)	35,0 (±10,1)
	HC (n=10)	Female	36 (±9,0)	-	-	-
SP cohort	SP (n=17)	Female	57,4 (±13,6)	6.5 (4-8)	23,4 (±8,7)	34,0 (±10,5)

* Abreviations: RR-MS= Relapsing-remitting multiple sclerosis; CIS=Clinically Isolated Syndrome; HC= healthy control; Evolut. Time=Evolution time; AOO= age of onset; EDSS=Expanded Disability Status Scale. Sex, age, evolut. time and AOO data are presented as "average (standard deviation)", EDSS data is shown as "median(range)"

RNA isolation

Total RNA, including small RNAs, was isolated from leukocytes. Briefly, erythrocytes from peripheral whole blood samples were lysed with Buffer EL. Subsequently, leukocytes were washed several times with DPBS (GIBCO, Thermo Fisher Scientific) and pelleted. Total RNA was isolated from the leukocytes with the miRNeasy Mini Kit (Qiagen) following the manufacturer's instructions for all the cohorts except for the second validation cohort.. For the second validation cohort the LeukoLOCK kit (Ambion) was used for total RNA extraction, using the alternative protocol to capture small RNAs following the instructions. RNA concentration was measured using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific).

Microarray analysis

RNA samples from four RR-MS patients and four controls were prepared and hybridised to the Arraystar Human Circular RNA Microarray V2.0, which covers 13,617 circRNA probes, based on Arraystar's standard protocols. Briefly, total RNAs were treated with RNase R (Ribonuclease R) (Epicentre, Inc.) to remove linear RNAs and enrich circular RNAs. They were then amplified and transcribed into fluorescent cRNA (complementary RNA) utilising a random priming method (Arraystar Super RNA Labeling Kit; Arraystar). The labelled cRNAs were hybridised onto the array and after having washed the slides they were scanned by the Agilent Scanner G2505C (Agilent Technologies).

After quantile normalisation of the raw data, low intensity filtering was performed, and the circRNAs that had at least one out of four samples with flags in "P" or "M", meaning Present (fluorescence intensity above the background) and Marginal (Fluorescence intensity above the background, but background intensity was not uniform), were retained for further analyses. Using Bioconductor gplots package a hierarchical clustering was created with differentially expressed circRNAs calculating Euclidean genetic distances and a complete linkage algorithm for clustering.

Differentially expressed circRNAs between the disease and control groups were identified using FC and its statistical significance was estimated by t-test using the R software limma package. CircRNAs having FC≥1.5 and p-values≤ 0.05 were selected as significantly differentially expressed.

Divergent primer design for circRNA amplification

CircRNA sequences were obtained from UCSC (University of California Santa Cruz) Genome Browser (GRCh37/hg19 assembly) or circInteractome [139] and divergent primers were designed to amplify the circular transcripts so that the qPCR amplicon spans the backspliced junction (BSJ) [93]. Divergent primers were located within the exons that formed the BSJ, except for the cases in which these primers could result in off-target amplification of another circRNA also containig these exons. In those cases divergent primers were designed so that one of them hybridizes right on the BSJ. Primer3Plus software[254] and BLAST (Basic Local Alignement Search Tool form the National Centre of Biotechnology Information, NCBI) were used to assist the primer design and to ensure the target specificity of the primers respectively.

Commonly used convergent primers detecting a sequence downstream of the circRNA producing exons, were also designed in order to amplify the linear transcripts. Primer sequences are available in Appendix Table 1.

Quantitative PCR (RT-qPCR) for technical and biological validations

Five upregulated and five downregulated circRNAs, those showing the highest fold change, were selected for technical validation performed in the same samples as those studied in the microarray analysis (discovery cohort) by real-time quantitative PCR, namely, hsa_circ_0000518, hsa_circ_0000517, hsa_circ_0000519, hsa_circ_0000520, hsa_circ_0001400, hsa_circ_0056731, hsa_circ_0064644, hsa_circ_0005402, hsa_circ_0024892 and hsa_circ_0035560.

In addition, four of those circRNAs (two upregulated: hsa_circ_0000518 and hsa_circ_0001400 and two downregulated: hsa_circ_0056731 and hsa_circ_0005402) were chosen for a second technical validation performed by Arraystar (Arraystar, Inc., USA) in order to confirm the reliability of our primer design.

For the lab technical validation, a two-step RT-qPCR approach was carried out. Total RNA was reverse transcribed into cDNA with random primers using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Inc., USA). The RT-PCR was peformed in a Verity Thermal cycler (Applied Biosystems, Inc., USA) with the following program: 25^{o}C for 10 mins, 37^{o}C for 120 mins and 85^{o}C for 5 mins. For the qPCR, the cDNA was amplified using Power SYBRGreen Master Mix (Applied Biosystems) in a 7900HT thermal cycler (Applied Biosystems, Inc., USA) or a CFX96/CFX384 Touch Real-Time PCR Detection System (Bio-Rad laboratories, Inc.) following the manufacturer's instructions. Each sample was run in triplicate (starting from 10ng of cDNA in 10 µl total reaction volume). Q-PCR cycling conditions were: 50^{o}C for 2 mins, 95^{o}C for 10 mins, 40 cycles of 95^{o}C for 15 sec and 60^{o}C 1 min followed by a dissociation curve analysis. The presence of a single-peak in the melting curve indicated the specificity of the amplification.

The expression level of the circular transcript, represented as Fold change (FC) was calculated using the $2^{-\Delta Cq}$ method, where ΔCq was calculated for the divergent amplicon (Cq Multiple sclerosis group- Cq Control group). In addition, we further calculated the FC after normalisation of the circular RNA expression by the expression level of its corresponding linear form (convergent amplicon), where $\Delta\Delta$Cq= [(divCq – conCq) Multiple sclerosis group - (divCq – conCq) Control group].

In order to asses the statistical significance of the differences between all groups except for the SP cohort, a Wilcoxon test was used (paired in relapse vs remission) due to the fact that some of the circRNAs do not follow normal distribution. This was assessed using the Shapiro-Wilk test. For the contrast between SP-MS and RR-MS the distribution of the expression values followed a normal distribution as assessed by Kolmogorov-Smirnov Test and then the expression difference was assessed by a Student's T-test. Analyses were performed in Excel ,R 3.3.2 in RStudio v1.0.44 and SPSS. P-values$<$0.05 were considered significant, exact p-values are indicated during the results section. qPCR results plots were created in RStudio v1.0.44 with R 3.3.2 and "beeswarm" or "ggplot" packages.

For the technical validation performed by Arraystar, divergent primers spanning across the backspliced junction were used to amplify the circRNAs and β-actin was selected as the reference gene.

In addition, four circRNAs for which the trend of the FC in both technical validations was in agreement with the microarray results (circ_0005402, circ_0056731, circ_0024892 and circ_0035560) were further selected for validation in a second cohort of 20 untreated RR-MS patients in remission and 18 controls (first validation cohort) by qPCR as described above for the lab validation. Based on the results of the biological validation performed in the first validation cohort, the expression of two of these circRNA candidates (circ_0005402 and circ_0035560) was studied by qPCR in 21 matched samples of patients in relapse and remission (second validation cohort) and in a third cohort of 10 patients with CIS and 10 age-matched controls. Finally the expression of circ_5402 and circ_0035560 was also assessed in a cohort of 17 SP-MS patients and compared to the expression of the 20 RR-MS patients from the first validation cohort.

Receiver Operating Characteristic (ROC) curve

MedCalc 14.8 was employed to plot the ROC curve using DeLong et al methodology[255].

Correlation analysis

To assess whether any of the most relevant transcript canidates expression (has_circ_0005402, hsa_circ_0035560 and ANXA2) was further associated with any clinical variable we performed a correlation analysis. For this analysis, Pearson's correlation was calculated between the normalized DCq value of each transcript for the 20 RR-MS and

17 SP-MS patients and different clinical variables such as the time of evolution, EDSS and age of individuals. Correlated transcripts were considered those with a p-value<0.05

PBMC isolation and storage

Peripheral blood from 6 RR-MS patients and 4 healthy controls was collected in heparin tubes (16 ml) after receiving informed consent. PBMCs were isolated by density gradient centrifugation with Lymphoprep™ (Abbott), within one hour of sampling and following the manufacturer's instructions. Cells were frozen in RPMI medium 1640 with L-Glutamine (Gibco, Thermo Fisher) supplemented with 10% foetal bovine serum, 10,000 U/ml penicillin, 10,000 µg/ml streptomycin and 10% DMSO and stored in liquid nitrogen until used. For cell culture experiments PBMCs were thawed, immediately washed and resuspended in the fresh RPMI medium to remove DMSO.

PBMC culture and stimulation with PHA

To investigate the expression of circRNAs in response to T cell activation and inflammation PBMC cells from 6 RR-MS patients and 4 healthy controls were cultured and stimulated. T cell activation, circRNA expression and cytokine production were measured. Clinical and demographical data from these individuals is shown in Table 2.

Table 2. Clinical and demographical data of individuals from which PBMCs were cultured for the PHA stimulation experiment.

	Sex	Age	EDSS	Evol. time	AOO
RR-MS patients	Female (n=6)	39.4 (±6.9)	2.3 (0-3.5)	13.0 (±7.1)	26.4 (±3.7)
HC	Female (n=4)	43.5 (±2.4)	-	-	-

* Abreviations: RR-MS= Relapsing-remitting multiple sclerosis; HC=healthy control; Evolut. Time=Evolution time; AOO= age of onset; EDSS=Expanded Disability Status Scale. Sex, age, evolut. time and AOO data are presented as "average (standard deviation)", EDSS data is shown as "median(range)"

Thawed cells were cultured in 24-well flat-bottom plates in 500 µl RPMI medium supplemented with 10% exosome-depleted FBS (Gibco, Thermo Fisher), 10,000 U/ml penicillin and 10,000 µg/ml streptomycin, at a final density of $8{\cdot}10^5$ cells/ ml and incubated for 3 h at 37 °C and 5% CO_2. Then, activation of cells was induced by stimulating them with 15 µg/ml phytohemagglutinin (PHA) (Sigma-Aldrich) in corresponding wells. For control non stimulated cells, DPBS was added. PHA was chosen to induce a polyclonal, nonspecific and significant lymphocyte activation, similar to the one produced against infection agents [256]. All cultured cells were incubated for 72 h at 37 °C and 5% CO_2. For

each condition, we performed the culture in duplicate. After the 72h of stimulation, both duplicates were taken from culture, and transferred to 1.5ml tubes. From a final volume of 1ml, 400µl of cell suspension were stained for flow cytometry. The remaining cells were harvested and centrifuged (400g for 5min) so that the supernatant was aliquoted and frozen for interleukin analysis and QiaZOL was added to the cell pellet for RNA isolation with miRNAeasy mini kit as previously described for subsequent ddPCR. A schematic representation of the culture protocol is presented in Figure 23.

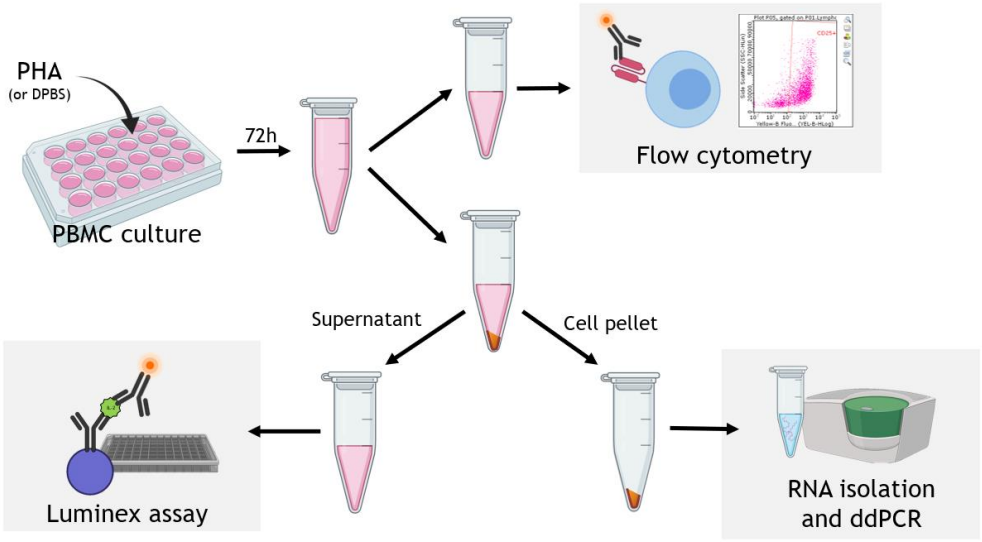

Figure 23. PBMC culture protocol to test the influence of inflammation in circRNA expression. $8{\cdot}10^5$ cells/ml were cultured and stimulated with PHA for 72h. Subsequently cells were harvested and prepared to analyse their activation by flow cytometry, to measure the *in vitro* cytokine production by Luminex assay and to assess the expression of some transcript candidates by ddPCR. *Ilustration by Iparraguirre, L. Some icons were obtained from biorender.com*

Flow cytometry

After the 72h of incubation, cells harvested for cytometry were washed with DPBS with 5% bovine serum albumin (BSA)(Sigma-Aldrich) to block Fc receptor before staining. Corresponding antibodies were added and samples were incubated for 20 minutes at room temperature in the dark. Cell viability was assessed with 7-aminoactinomycin D (7-AAD)(Molecular Probes) and PE-conjugated anti-human CD25 (BD Pharmingen^{TM}) was used to determine cell activation. For identification of CD4 and CD8 T lymphocytes FITC conjugated anti-CD8, and PE-Cy7 conjugated anti-CD4 andtibodies (BD Biosciences) were used. After antibody staining cells were washed with DPBS to remove unbound antibodies and acquired in a Guava EasyCyte 8HT flow cytometer (Millipore, Merck). Single staining and fluorescent minus one control tubes were used to adjust compensations and set the

gating strategy. After gating for singlets, lymphocytes were gated based on FSC and SSC and 10,000 lymphocytes were acquired for each sample. Then, lymphocyte populations were distinguished based on fluorescence and analysis of obtained results was performed with the InCyte 3.1 software (Millipore, Merck).

In vitro cytokine production measurement

Cell-culture supernatants from PHA stimulated conditions (6 RR-MS patients and 4 healthy controls) were recovered to measure *in vitro* cytokine production by PBMCs. The concentration of IL-1β, IL-2, IL-6, IL-10 and TNFα were measured in a luminex assay with Human Magnetic Luminex Assay LXSAHM (R&D systems) following manufacturer's instructions. RPMI culture medium was applied as background. A MAGPIX device with xPONENT software (Luminex Corporation) was used for fluorescence measurement and median fluorescent intensity data were analysed using the 5-paramenter logistic method for calculating cytokine concentrations.

Droplet digital PCR (ddPCR)

RNA was reverse transcribed as previously described for the RT-qPCR. The ddPCR was performed using QX200 EvaGreen ddPCR SuperMix (BioRad) starting from 10ng of cDNA. Droplets were generated in the QX200 Droplet Generator (BioRad) according to the manufacturer's protocol and then transferred into a PCR plate and sealed. The PCR cycling protocol was carried out in a C1000 Touch™ Thermal Cycler (BioRad) as follows: 95^{o}C for 5 min and 40 cycles of 95^{o}C for 30 sec and 60^{o}C for 1 min, finally samples were kept at 4^{o}C for 5 min and 90^{o}C for 5 min for signal stabilization. Each sample was run in duplicates. Droplets were finally counted on the QX200 Droplet Reader (BioRad). Analysis was carried out using QuantaSoft 1.6.6. software (BioRad) and data was expressed as the number of copies per nanogram (ng) of total RNA. Fold changes were calculated in Excel and statistical analysis was performed in Graphpad Prism. Unpaired T-test was used to compare RR-MS and HC groups and paired T-test to compare unstimulated and PHA activated conditions.

Results

Differentially expressed circRNA profile by microarray

Out of the 13,617 probes interrogated in eight blood samples (four RR-MS untreated patients in remission and four healthy controls, discovery cohort), 10,183 (74.78%) are

present in at least one sample. Among them, 406 are differentially expressed (FC≥1.5, p≤0.05) in patients, compared to healthy controls. 324 are downregulated and 82 upregulated (Figure 24A). Hierarchical clustering performed with differentially expressed circRNA, shows that this circRNA expression pattern is able to classify individuals according to their status (Multiple sclerosis or Controls) except one healthy control, who is grouped with multiple slcerosis patients (Figure 24B).

Validation of candidate circRNAs by RT-qPCR

Among the 406 differentially expressed circRNAs, the five upregulated circRNAs (circ_0000518, circ_0000517, circ_0000519, circ_0000520, circ_0001400) and the five downregulated circRNAs (circ_0056731, circ_0064644, circ_0005402, circ_0024892 and circ_0035560) showing the highest fold change were chosen in order to perform a technical validation of microarray results by qPCR. Divergent primers were designed for each circRNA and 9/10 were successfully amplified whereas hsa_circ_0064644 could not be amplified and was therefore excluded from this study.

The expression trend of five circRNAs is consistent with microarray results whereas in four circRNAs, it is opposite to that found in the microarray (Figure 24C). After a second technical validation performed by Arraystar (Arraystar, Inc., USA), we find out that circ_0056731, circ_0005402, circ_0024892 and circ_0035560 are confirmed by at least one of the validations or show consistent results for both qPCRs (Figure 24C). Therefore, those four circRNAs are further selected for validation in a larger cohort of 20 RR-MS untreated patients in remission and 18 healthy controls (first validation cohort).

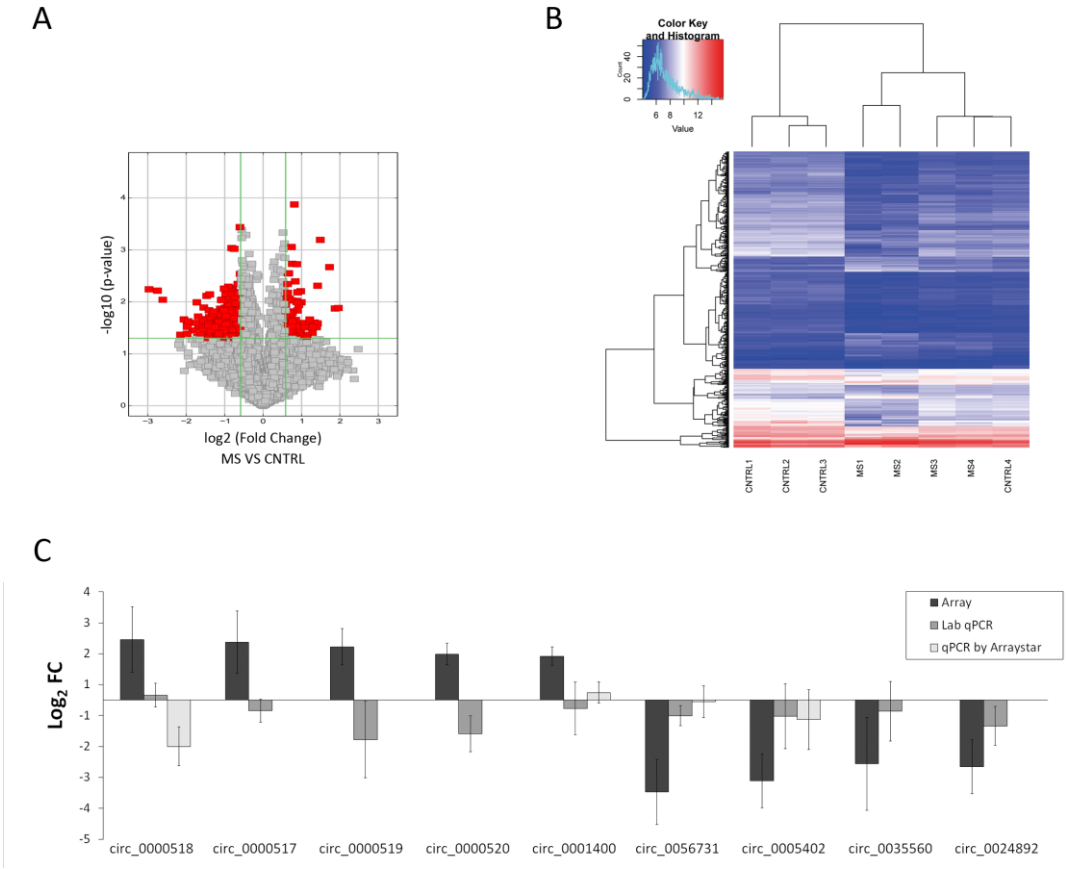

Figure 24. Differentially expressed circRNAs in MS. (A) Volcano plot showing the results for the 10.183 circRNAs expressed in at least one of the four samples studied in the array. The vertical lines correspond to 1.5-FC up and down, and the horizontal line represents a p-value of 0,05, so red points depict the 82 upregulated and the 324 downregulated circRNAs with statistical significance. (B) Heatmap showing the differentially expressed circRNAs which quite efficiently clusters control vs patient samples. The colour scale reflects the log2 signal intensity and runs from blue (low intensity) to red (high intensity). The dendrogram shows the relationships among the expression levels of samples (C) Comparison of circRNA expression levels between microarray and the technical validation results by qPCR in which four circRNAs were validated by at least one of the validations or with consistent results for both validation qPCRs. Error bars represent the standard deviation for the patients group (n=4) for which the fold change is calculated (MS= multiple sclerosis; CNTRL=control).

The Wilcoxon test shows that the expression level of both circ_0005402 (FC=0.395, p=0.00007) and circ_0035560 (FC = 0.737, p = 0.0304) is significantly lower in patients compared to controls. For circ_0056731 and circ_0024892 the same downregulation trend is maintained, however the difference is not statistically significant (FC=0.814, p=0.111 and FC=0.861, p=0.266 respectively). When we measure the relative abundance of circRNAs compared to the linear transcripts derived from their respective host genes, only circ_0035560 (FC = 0.736, p = 0.0412) remains significantly deregulated while

circ_0005402 is no longer significant (FC=0.677, p=0.109) (Figure 25). Note that the expression of circ_0035560 is higher compared to the expression of the linear RNA isoform.

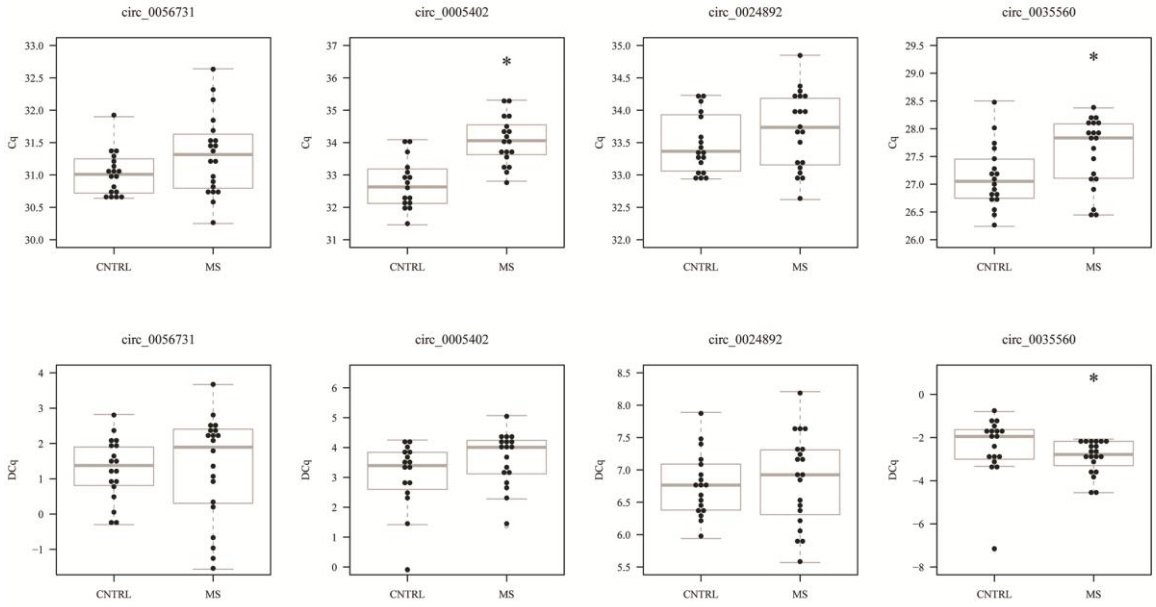

Figure 25. Candidate circRNA expression in the first validation cohort. Both the expression of the circular isoform (above) and its expression after normalisation by the expression level of the corresponding linear counterpart (below) are depicted through dot and boxplots for the MS and control groups. Note that the expression is represented by the Cq (above) and DCq (below) value obtained in the qPCR where lower values correspond to higher expression values. *Indicates $p<0.05$ assessed by Wilcoxon test. (MS= multiple sclerosis; CNTRL=control).

Based on these results, the expression of circ_0005402 and circ_0035560 is studied in 21 relapse and remission paired samples in order to assess their potential implication in the biology of the disease. None of them passed the significance threshold in this second cohort (FC=0,932, p=0,392 and FC=1,358, p=0,633 respectively), neither do they in the analysis stratified by sex (FC=0,745, p=0,109 and FC=1,091, p=0,850 for females and FC=1,256, p=0,734 and FC=1,819, p=0,250 for males) (Figure 26A). Moreover, those two circRNAs were also studied in a third cohort of 10 patients with clinically isolated syndrome (CIS) for which no significant expression differences are found as assessed by the Wilcoxon test. Nevertheless, the same downregulation trend found for RR-MS patients is maintained (FC=0,764, p=0,352 and FC=0,926, p=0,911) (Figure 26B).

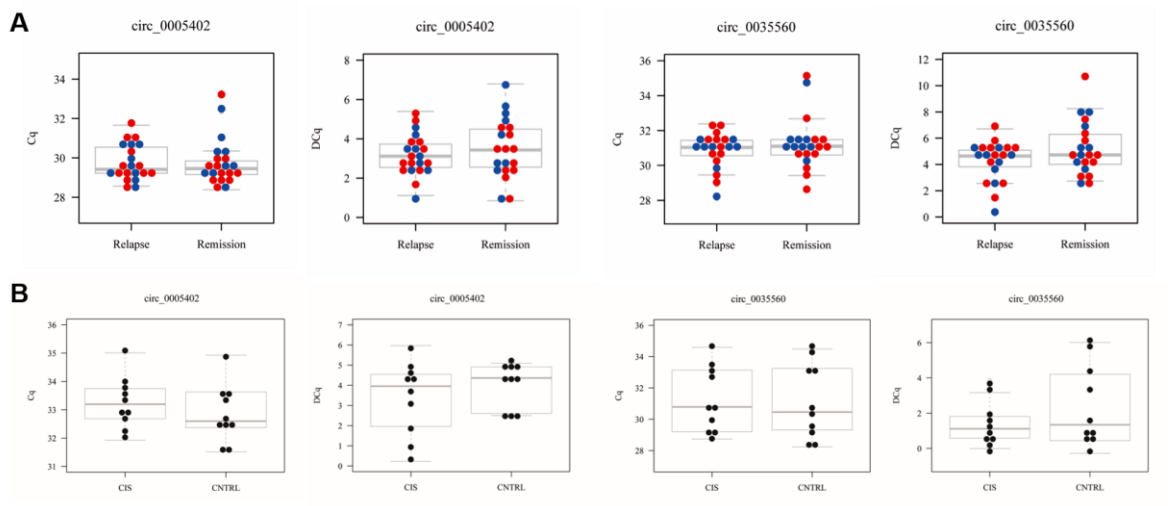

Figure 26. Circ_0005402 and circ_00035560 raw expression (Cq, left)and normalized expression (DCq, rigth) in relapse, remission and CIS. (A) Both the expression of the circular isoform (left) and its expression after normalisation by the expression level of the corresponding linear counterpart (rigth) are depicted through dot and boxplots for the relapse and remission groups. Blue dots represent male patients whereas red dots represent female patients. (B) Both the expression of the circular isoform (left) and its expression after normalisation by the expression level of the corresponding linear counterpart (rigth) are depicted through dot and boxplots for the CIS and healthy control groups. Note that the expression is represented by the Cq (lefth) and DCq (rigth) value obtained in the qPCR where lower values correspond to higher expression values. According to Wilcoxon test, there are not expression differences between groups.

Two ANXA2 derived circRNAs and mRNA show potential to be MS biomarkers

In order to assess whether the two differentially expressed circRNA or even the linear form of the parental gene could be candidate biomarkers for the disease, the predictive value of circ_0005402, circ_0035560 and ANXA2 is evaluated separately by employing the ROC curve analysis using the Cq value for each of them. The results show that the area under the curve (AUC) is significant for circ_0005402 with a value of 0.899 (95% CI: 0.747-0.976), a p-value < 0.001, a sensitivity of 94.4 % and a specificity of 75.0% and also for circ_0035560 with a value of 0.706 (95% CI:0.536-0.842), with a sensitivity of 55% and a specificity of 88.9% and for the linear form of the gene with a value of 0.831 (95% CI: 0.674-0.932) with a sensitivity of 80% and a specificity of 83.33 (Figure 27).

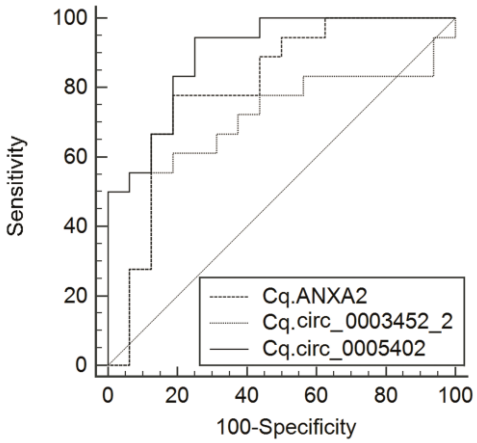

Figure 27. ROC curve for ANXA2 mRNA, circ_0003452_2 and circ_0005402 indicating their potential to be biomarkers. The AUC values are 0.831, 0.706 and 0.899 respectively.

The expression of ANXA2 derived circRNAs and the mRNA is changed between MS types and correlates with the disease evolution time

With the aim of understanding the implication of the two ANXA2 derived circRNAs (hsa_circ_0005402 and hsa_circ_0035560) and the mRNA in the disease we studied their expression in a second group of patients that have already switched from a RR-MS phenotype to be diagnosed from SP-MS. Results indicate that the expression of circ_0005402 (FC=1,44, p=0,048), circ_0035560 (FC=1,39, p=0,008) and ANXA2 (FC=1,36, p=0,029) is significantly increased in patients with SP-MS when compared to RR-MS patiens (Figure 28).

SP-MS and RR-MS patient groups differ in several of their demographical and clinical characteristics. The evolution time, EDSS and age values are higher for SP-MS patients when compared to RR-MS (Figure 29) although only the difference in EDSS is statistically significant (p=0,0574, p<0,001 and p=0,0618 respectively). In light of these differences, we decided to perform a correlation analysis to investigate whether the change in expression found between groups is related to a longer evolution time, a higher age or a higher disability (measured by EDSS).

Interestingly, we found that the expression of the three transcripts correlate significantly both with evolution time and age (circ_0005402 p=0.003 and p=0.008 respectively) (circ_0035560 p=0.018 and p=0.003 respectively) (ANXA2 p=0.004 and p=0.014 respectively) whereas there is no significant correlation with EDSS (Figure 29).

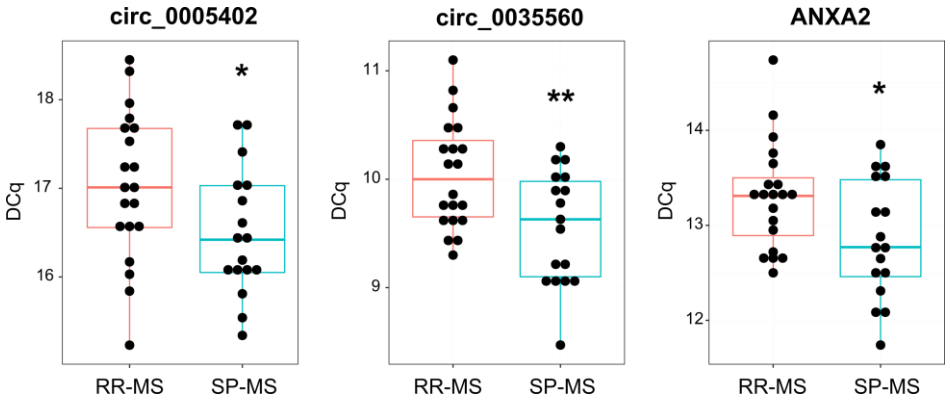

Figure 28. Circ_0005402, circ_0035560 and ANXA2 expression in SP-MS (n=17) and RR-MS (n=20) patients. Expression of the transcripts was normalized with an endogenous gene (EEF1A1) and depicted through dot and boxplots. Note that the expression is represented by DCq value obtained in the qPCR where lower values correspond to higher expression values. According to Students T-test, significant differentes are found between groups.

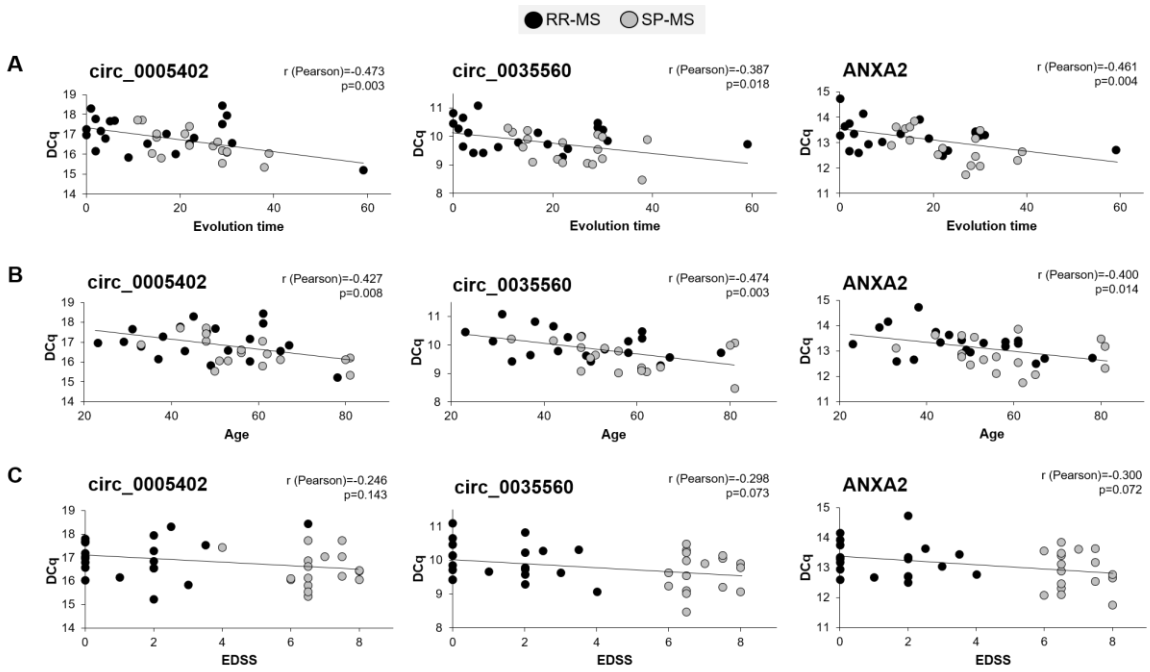

Figure 29. Correlation analysis between circ_0005402, circ_0035560, ANXA2 mRNA and clinical characteristics from SP-MS and RR-MS patients. Expression of the transcripts was normalized with an endogenous gene (EEF1A1) and Pearson coefficient was calculated to assess the correlation with evolution time (A), age (B) and EDSS (C). Expression values from RR-MS patients were depicted by black dots and SP-MS patients data with gray dots. Pearson coeficients and pvalues are shown at the upper right corner of each graph. Note that the expression is represented by DCq value obtained in the qPCR where lower values correspond to higher expression values.

Taking into account that higher disease evolution times go usually together with patients that are older, we hypothesize that the correlation found between age and the expression of these three transcripts may not be independent from the correlation found with the evolution time. To address this question, we studied the expression of circ_0005402, circ_0035560 and ANXA2 mRNA in a cohort of healthy controls that included 6 individuals aged between 21-30 years, 6 individuals aged between 31-40 years and another 6 aged between 41 and 50 years. In contrast with what we found for MS patients, the expression of none of the three transcripts correlated with their age suggesting that the age-correlation found for patients is influenced by their disease evolution time (Figure 30).

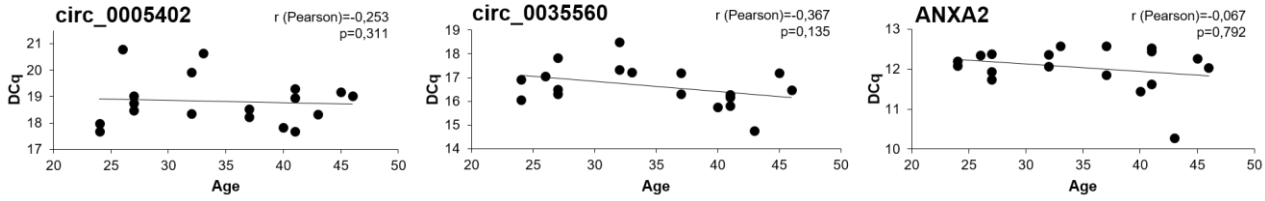

Figure 30. Correlation análisis between circ_0005402, circ_0035560 and ANXA2 mRNA with age from 18 healthy controls. Expression of the transcripts was normalized with and endogenous gene (EEF1A1) and Pearson coefficient was calculated to assess the correlation with age. Pearson coeficients and pvalues are shown at the upper right corner of each graph. Note that the expression is represented by DCq value obtained in the qPCR where lower values correspond to higher expression values.

PBMCs activation with PHA results in decreased levels of ANXA2 derived circRNAs and mRNA.

Based on the results that show a decreased expression of the ANXA2 derived circRNAs and mRNA in RR-MS patients when compared to controls and also with SP-MS, we considered that their expression could be related with the inflammatory state associated with RR-MS.

In order to check this hypothesis, PBMC samples from 6 MS patients and 4 healthy controls were cultured and stimulated with PHA to induce their activation thus mimicking an inflammatory response. We confirmed with the measurement of CD25 by flow cytometry that for healthy controls, the 44.1% of lymphocytes were activaded and that for MS patients the response was even bigger with a 60,8% of the lymphocytes activated (p=0.005)(Figure 31A). On the other hand, the levels of 5 different cytokines were measured in the medium were cells had been cultured. All of them could be detected both in RR-MS samples and HC demonstrating that PHA has also induced the cytokine production of cells. Nevertheless, in contrast with the results obtained by flow cytometry, no significant difference was found

between patients and controls (Figure 31B). Even so, it is worth noting that IL-10 is the only antiinflamatory cytokine tested and its concentration is lower than the concentrations found for healthy controls in five out of the six RR-MS patients (Figure 31B).

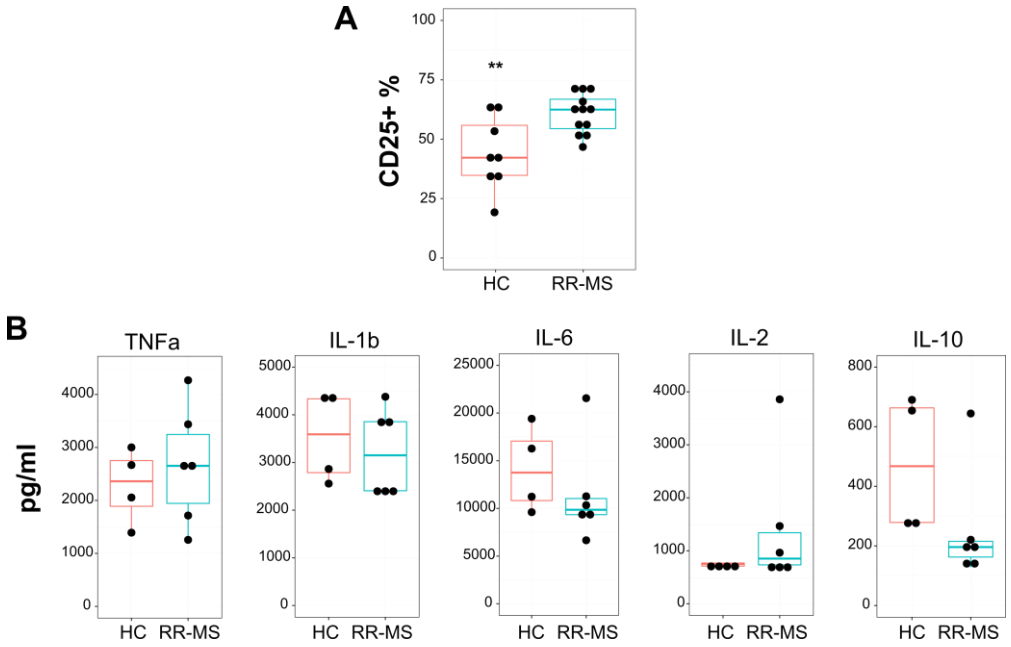

Figure 31. Flow cytometry and Luminex results confirm that PBMCs have been activated upon PHA stimulation. (A) Boxplot showing the percentage of CD25+ lymphocytes in HC an RR-MS samples assessed by flow cytometry. Results show a significant difference in the activation between groups (T-test). (B) Concentration of the different interleukins was shown in pg/ml. None of them showed a significant difference in concentration between groups.

After demonstrating that PBMCs have been correctly activated, the expression levels of circ_0005402, circ_0035560 and ANXA2 mRNA were measured both in unstimulated and PHA stimulated cells by ddPCR (Figure 32). Results show that, in line with what was found for leukocytes, less copies of the three transcripts are found in RR-MS patients when compared to healthy controls in the unstimulated condition. Nevertheless, only ANXA2 for which 2.1 times copies less were amplified in RR-MS was statistically significant (p=0.049). On the other hand, when the expression in unstimulated and PHA activated cells was compared, we found that all the three transcripts are downregulated upon activation. Interestingly, this differences are bigger in healthy controls with significative differences for circ_0005402 (FC=-4.6 p=0.0421) and ANXA2 (FC=2.8, p=0.0246) (Figure 32).

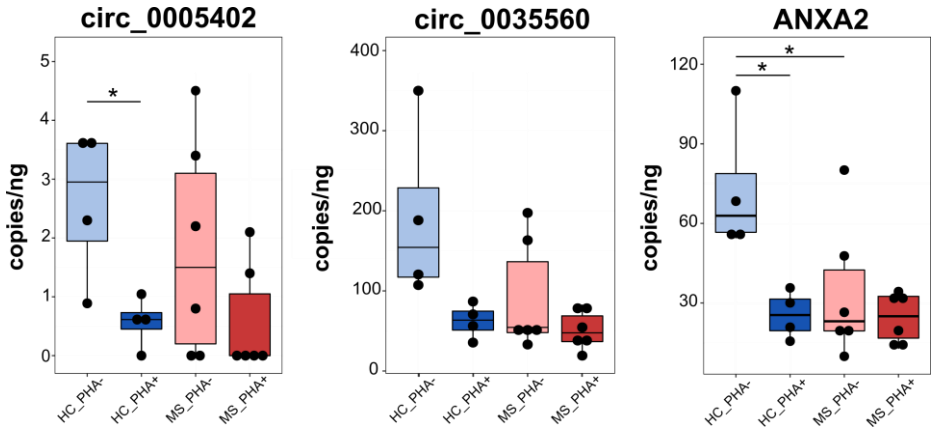

Figure 32. Absolute quantification of circ_0005402, circ_0035560 and ANXA2 mRNA in PBMC samples from RR-MS patients (n=6) and healthy controls (n=4) assessed by ddPCR. Each sample was studied in two conditions: unstimulated and activated by PHA. Unpaired T-test was used to asses the difference between RR-MS and HCs and paired T-test for comparing unstimulated(PHA-) and PHA activated (PHA+)conditions. Significant differences are indicated by star code: *, p-value<0.05.

Discussion

CircRNAs are a recently discovered class of ncRNA[257] with some special characteristics that make them good candidates to be biomarkers. They are highly conserved and are resistant to degradation by RNAses, with a half-life over 48 hours [79,258]. In fact, in the last two years several papers have proposed this role for specific circRNAs in some diseases.

In thissubchapter we report a first approach to the study of the circular transcriptome profile in RR-MS patients and controls, showing a distinctive blood circRNA expression profile for RR-MS patients. To our knowledge, this is the first study revealing a dysregulated profile of circRNAs in MS. During the preparation of our manuscript, which was published in Human Molecular Genetics in September 2017[259], Cardamone et al presented a paper in which they studied the alternative splicing abnormalities in the GSDMB gene and found novel isoforms of this gene and a circRNA (hsa_circ_0106803) that its upregulated in RR-MS[260]. In our dataset, the microarray analysis reveals the presence of 406 differentially expressed circRNA, most of them (nearly 80%) showing a trend of downregulation in patients. From these 406, two circRNA have been found to be downregulated in the MS population after all the validations; circ_0005402 which is 6.11 times downregulated in arrays, and 2.53 times in the first validation cohort, and circ_0035560 which is 4.17 times downregulated in arrays, and 1.36 times in the first validation cohort. Moreover, our ROC curve analysis results indicate that both circRNA expression and even *ANXA2* expression could be used as biomarkers that could help in the diagnosis of RR-MS with good values

of specificity and sensitivity. Indeed, this downregulation trend is also maintained for patients with CIS, pointing to an early dysregulation of their expression which could be very informative in the clinic.

To further characterize the expression of these transcripts in the disease, we also studied the expression of the two circRNAs and the linear ANXA2 in a more advanced stage of the disease were patients have already entered a secondary progressive form characterized by a less inflammatory and more neurodegenerative phenotype. We observed that in these patients the expression of circ_0005402, circ_0035560 and ANXA2 mRNA is increased compared to RR-MS. In line with this, we could also confirm that the expression of the three transcripts increases with the time of evolution suggesting that they could be used to monitor the decrease of the inflammatory component associated with the early stages of the disease.

In cultured PBMCs we also confirm that, as it has been observed in leukocytes, RR-MS patients, which present a more activated basal state, have reduced levels of the three ANXA2 derived transcripts when compared to HCs. Moreover, once the immune activation is induced with PHA, which could be used to mimic a relapse *in vitro*, we observe that both patients and healthy controls experience a decrease in the levels of the two circRNAs and ANXA2 mRNA. Nevertheless, the decrease in expression upon immune activation is only significant for circ_0005402 and ANXA2 in HCs were the effect is more pronounced. For RR-MS patients in contrast, the difference is smaller, which could be partially explained by the fact that RR-MS patients do already start from a lower number of transcript copies in the basal state. It is also worth noting that the same effect has been observed when analysing the expression form leukocytes between relapse and remission phases were there was no difference.

Both circRNAs are located in chromosome 15, inside the *ANXA2* gene[158] and have not been previously related to any pathology or biological process. Interestingly, the linear form of *ANXA2* is also downregulated. Even though the expression of these circRNAs might not be completely independent of the linear RNA isoform abundance, they are still notably downregulated as can be seen when the expression of circRNAs is normalised with *ANXA2* expression. This gene is predicted by circBase to give rise to 30 circRNAs, from which only five have been included in the Arraystar platform (due to the fact that the rest did not fulfil the company's inclusion criteria). All of them are expressed in at least one sample, but only circ_0005402 and circ_0035560 are differentially expressed.

Up to now, based on the literature, *ANXA2* gene had not been related to MS, neither in Genome Wide Association studies nor in expression studies, however, it has been related to several cancer types [261–263] and immune-mediated diseases such as arthritis or antiphospholipid syndrome [264,265]. Thus, *ANXA2* presumably has an immune function, however, whether it has a pro or anti-inflammatory role could be context dependent[264,266,267]. Interestingly, *ANXA2* has been proposed to play a role in leukocytes passing the blood-brain barrierthat protects the CNS[268]. *ANXA2* has also been reported to be a target of miR-155, a critical miRNA in neuroinflammation at the blood-brain barrier and relevant in Th1 and TH17 cell differentiation and myeloid cell polarisation in MS. MiR-155 has also been shown to be significantly increased in both peripheral blood mononuclear cells and active lesions and to correlate with disease severity in MS patients and mice with EAE[269–271] which could be in agreement with the downregulation of *ANXA2* suggesting a possible complex interaction between miRNA, mRNA and circRNAs. Moreover, *ANXA2* shows an important role in post-transcriptional regulation of gene expression and in the transport of mRNA and vesicles [272]. These processes have been proposed as clues to understanding the network of transcriptome regulation in MS.

Circ_0005402 has a size of 480 bp and includes exons three, four, five, six and seven (referred to the exons of the NM_004039 mRNA transcript) (Figure 33) and the presence of the backspliced junction between exons three and seven has been confirmed in our data by Sanger Sequencing. For circ_0035560, the structure proposed in circBase includes exons three, four and five (Figure 33), however when we checked the sequence of our amplicon, we found a junction that includes exon two between exons five and three, but not the proposed structure. CircBase already includes a circRNA formed by the exons two, three, four and five, hsa_circ_0003452 (which is not included in the Arraystar system). However, its sequence does not totally overlap with the amplicon we are obtaining since it includes an 11 nt shorter sequence of exon five than hsa_circ_0003452. In a nascent field such as circRNAs, nomenclature and databases need to be refined after more data is generated. Therefore, we propose calling this circRNA hsa_circ_0003452_2. So, from now on, we will refer to what we first called circ_0035560 as circ_0003452_2.

To sum up, this subchapter comprises the first study of the expression profile of more than 13.000 circRNAs in RR-MS patients. Our results report the first evidences of the implication of circRNAs in the disease and moreover, we suggest the implication of two circRNAs in the disease associated inflammatory process although it should be further investigated.

To finish, we propose two particular circRNAs, circ_0005402 and circ_0003452_2, and their linear counterpart, ANXA2 as new biomarkers in RR-MS.

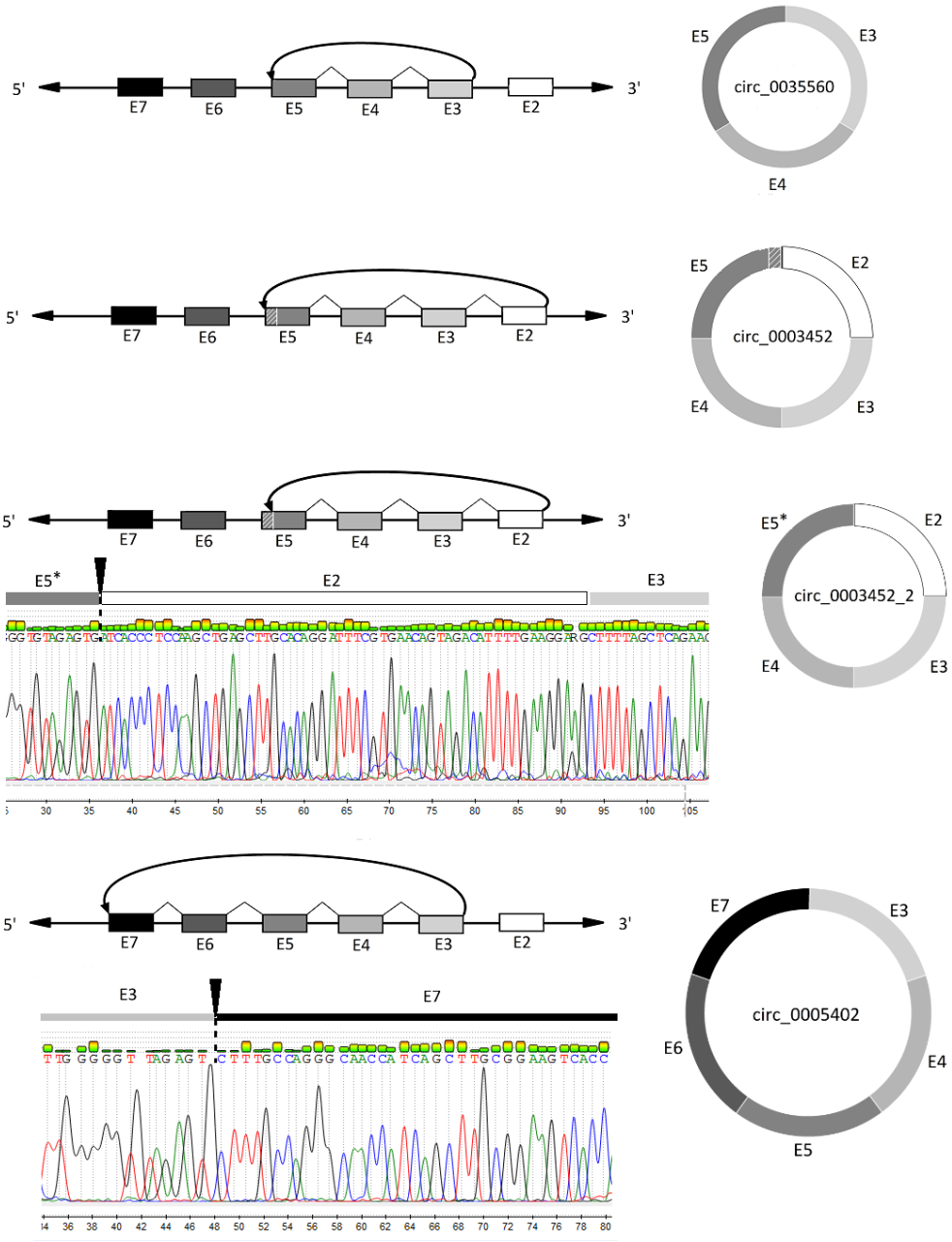

Figure 33. Schematic representation of four different backsplicing events of ANXA2 leading to four circRNAs. Exons are indicated with E and the corresponding number numbered starting from the 3' since they are transcribed from the minus strand. E5* refers to shortened version of the exon five. For circ_0005402 and circ_0003452_2 the Sanger-sequencing electropherogram showing the backspliced junction is presented below the diagram.

1.2 RNASeq profiling of circRNAs and linear RNAs in leukocytes

This subchapter has been published in *Human Molecular Genetics:*

Iparraguirre L, Alberro A, Sepúlveda L, Osorio-Querejeta I, Moles L, Castillo-Triviño T, Hansen TB, Muñoz-Culla M, Otaegui D. *RNA-Seq profiling of leukocytes reveals a sex-dependent global circular RNA upregulation in multiple sclerosis and 6 candidate biomarkers.* Hum Mol Genet. 2020 Oct 8:ddaa219. doi: 10.1093/hmg/ddaa219. Epub ahead of print. PMID: 33030201.

Introduction

The first circRNA profiling approach in leukocytes from RR-MS patients presented in the subchapter one reported that hundreds of circRNAs are dysregulated when compared to controls, and more interestingly, that some of them hold a great potential to work as biomarkers of the disease.

Based on those preliminary results, in this subchapter, we wanted to extend the profiling to a larger discovery cohort that could better gather the heterogeneity of the disease. To that end, we decided to include SP-MS patients which represent the second most common MS subtype[179,181] but whose definitive identification is usually delayed. The transition from RR-MS to SP-MS is usually defined retrospectively after a sustained period of neurologic worsening during which patients experience a diagnostic uncertainty with several

implications related to patient care. During this period patients may remain on therapies that are no longer effective with unnecessary adverse effects and costs[273]. Some studies have previously described different transcriptional profiles at different disease stages[274,275] which together with the different expression of circ_0005402 and circ_0003452_2 reported in subchapter 1.1., indicate that circRNA profiling of these patients could reveal useful information to better characterize the RR-MS to SP-MS transition. On the other hand, we also decided to include both female and male individuals so that the sex variable could be studied. In fact, it is known that the immune system is strongly affected by sex-specific differences[187] and regarding MS a clear gender bias has already been detected in gene expression studies of both coding and non-coding genes [217,246,276,277]. Thus, taking MS-types and sex into account, the discovery cohort consisted of 30 MS patients including 20 RR-MS diagnosed and 10 SP-MS diagnosed patients and 20 healthy controls among which individuals from both sexes were included.

On the other hand, we changed the profiling technique from microarray to RNASeq in order to have a more complete transcriptome profiling of the samples. RNASeq allowed us to study a higher number of circRNAs that includes not only the annotated circRNAs but also to discover and quantify new circRNAs. Additionally, we could also study the expression of many of the linear RNAs.

Materials and methods

Blood sampling and RNA extraction

Whole blood was collected from a total of 120 MS patients and 66 healthy controls (HC) and paired by age and sex in the Department of Neurology at the Donostia University Hospital. Samples were divided into three cohorts. The discovery cohort was formed by 30 MS patients (20 RR-MS and 10 SP-MS) and 20 HC and were analyzed by RNASeq. Next, 70 MS (62 RR-MS and 8 SP-MS) and 46 HC samples were included in the RNASeq validation cohort and subjected to real-time PCR (RT-qPCR). Lastly, the expression of several circRNA candidates was studied in a third cohort that includes the same paired samples in relapse and remission from 20 MS patients that were also used in the subchapter 1. The main clinical and demographical characteristics of both patients and healthy donors are summarised in (Table 3). Samples from all donors were collected after receiving written informed consent. The study was approved by the hospital's ethics committee (UEM-IMN-

2017-01) and part of the samples have been processed and stored at the Basque Biobank (www.biobancovasco.org).

Table 3. Main clinical and demographical characteristics of the individuals enrolled in the study. A total of 120 MS patients and 66 healthy controls were studied in three different cohorts.

		Sex	Age	Type of MS	EDSS	Evol. Time	AOO
Discovery cohort	MS (n=30)	Female (n=23)	44.3 (±10.0)	16RR-MS/7SP-MS	3.7 (0-7)	14.7 (±10.4)	29.6 (±7.4)
		Male (n=7)	44.7 (±8.5)	4RR-MS/3SP-MS	4.6 (2-8)	15.7 (±7.6)	29.0 (±10.1)
	HC (n=20)	Female (n=14)	49.6 (±7.9)	-	-	-	-
		Male (n=6)	38.5 (±4.5)	-	-	-	-
Validation cohort	MS (n=70)	Female (n=58)	41.3 (±9.5)	50RR-MS/8SP-MS	2.3 (0-8)	10.2 (±7.6)	31.1 (±8.9)
		Male (n=12)	40.0 (±8.2)	12RR-MS	2.4 (0-4)	5.9 (±3.7)	28.9 (±9.8)
	HC (n=46)	Female (n=14)	33.2 (±8.7)	-	-	-	-
		Male (n=18)	34.9 (±10.2)	-	-	-	-
Relapse-Remission cohort	Relapse (n=20)	Female (n=12)	38.6 (±10.7)	12RR-MS	2.75 (0-7.5)	9.9 (±7.5)	28.2 (±14.4)
		Male (n=8)	40.1 (±10.0)	8RR-MS	2.75 (0-6)	8.6 (±10.9)	29.6 (±7.6)
	Remission (n=20)	Female (n=12)	38.3 (±11.3)	12RR-MS	3 (0-7.5)	9.4 (±7.4)	28.2 (±14.4)
		Male (n=8)	40.9 (±9.7)	8RR-MS	2.75 (0-6)	9.1 (±10.6)	29.6 (±7.6)

Abbreviations: MS, multiple sclerosis; HC, healthy control; RR-MS, relapsing-remitting MS; SP-MS, secondary progressive MS; EDSS, Expanded Disability Status Scale; Evol.Time: Evolution time; AOO, Age of Onset. Sex, age, evolution time and AOO data are presented as 'average (standard deviation)', EDSS data are shown as 'median(range)'.

Sample processing and leukocyte and RNA isolation was performed as described in the subchapter 1.1..

RNASeq

Library preparation and Next Generation Sequencing was performed at CD Genomics (USA). The concentration and quality of the RNA samples was again measured using Bioanalyzer 2100 instrument (Agilent) before library preparation. After normalization, rRNA was depleted from the total RNA sample using the Ribo-Zero rRNA removal kit and followed by purification and fragmentation steps. To construct the sequencing libraries, a strand-specific cDNA synthesis was performed, the 3' ends were adenylated and adaptors were ligated. The resulting libraries were subjected to quality control and normalization process. Paired-end sequencing was performed with Illumina HiSeq X Ten and an average of 10Gb of data were obtained per sample.

CircRNA detection and quantification in RNA-seq data

Sequencing reads were quality checked and mapped to the hg19 using BWA[278] or Bowtie[279]. Subsequently, circRNA prediction was performed by find_circ, version 1.0[80], and CIRI2[280] adhering to the recommendation by the authors. For find_circ[80], an increased stringency threshold was used requiring that both adaptor sequences map with

highest possible mapping quality (mapq = 40). Moreover, only circRNAs supported by at least two reads in a given sample and found by both algorithms were used in subsequent analyses. CircRNA expression was based on back-spliced junction (BSJ)-spanning reads according to CIRI2 quantification. Differential expression analysis was performed using DESeq2[281] in R-studio and specifying a design formula (~condition + sex) that controls the effect of the sex while contrasting the different conditions (MS vs HC or SP-MS vs RR-MS). Sex was not included as a covariable when the analyses were limited to individuals of one sex.

CircRNAs with an absolute fold-change (FC) value higher than 1,5 (FC >|1.5|) and a p-value less than 0.05 ($p<0.05$) were considered differentially expressed circRNAs (from now on, DE circRNAs). Further filtering criteria were applied to these DE circRNAs for the validation by RT-qPCR. These circRNA candidates were selected among the high confidence DE circRNAs defined as those detected in at least the 50% of the samples in each group and with a base mean, defined as the mean of normalized counts of all samples, above 10. All the steps explained above are schematically shown in Figure 34, from sample preparation to differential expression.

Linear transcript detection and quantification in RNA-seq data

After the quality check, the sequencing reads were mapped to the hg19 with STAR[282] using default parameters. HTSeq[283] was used to quantify expression of genecode annotated genes (gencode v28) in a strand-specific manner. Read counts were then used for gene-level differential expression analyses using DESeq2 in R-studio and, as defined for circRNAs, differentially expressed linear RNAs (DE linear RNAs) were defined as those with a FC >|1.5|and p-value<0.05 (Figure 34).

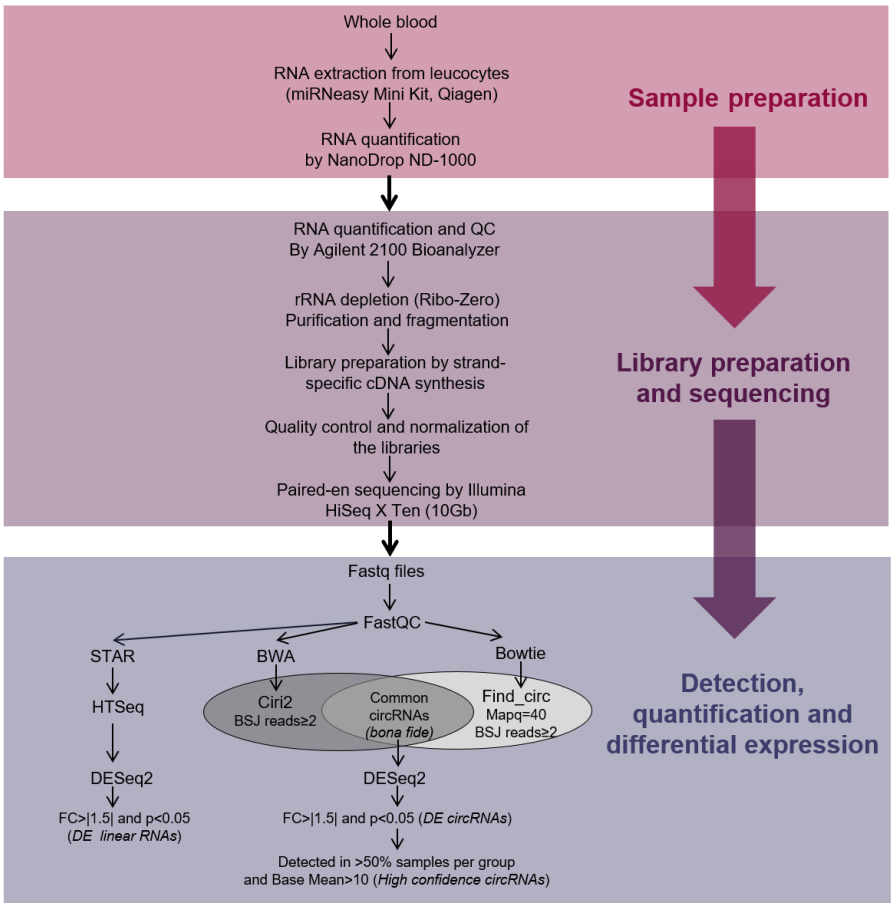

Figure 34. Diagram showing the workflow followed for the RNASeq from sample preparation to differential expression.

cDNA synthesis and Quantitative PCR

A two-step RT-qPCR approach was carried out as described for the Lab technical validation in the subchapter 1. Primer sequences are available at Appendix Table 1. In this case, a CFX384 Touch Real-Time PCR Detection System (BioRad) was used for all the PCRs and EEF1A1 was used as the reference gene for normalization. The raw Cq values and melting curves were analysed in CFX Maestro 1.0 (BioRad) where a single-peak in the melting curve indicated the specificity of the amplification. The change in the expression level of the circular transcript, represented as FC was calculated in Microsoft Excel 2010 using the 2_DDCq method. Graphpad Prism 6.01 was used for the normality and Student's T-tests.

The amplification products of all the circRNAs were subsequently purified with the ExoSAP-IT™ PCR Product Cleanup Reagent (Applied Biosystems) following the manufacturer's instructions and Sanger sequenced (ABIprism 3130) in order to check for the presence of the BSJ. In addition, the PCR products were also subjected to electrophoresis on agarose gel to confirm the presence of a single amplification product.

Gene ontology analysis

The GOrilla web-based gene ontology tool (http://cbl-gorilla.cs.technion.ac.il/) was used to perform gene enrichment analysis for biological processes[284]. For each of the contrasts (All MS patients vs HC, female patients vs HC and male patients vs HC) two gene lists were analyzed. The first list comprised the differentially expressed linear RNAs and the list of all the genes quantified by HTSeq was used as background. For the enrichment of biological processes in which DE circRNAs could be involved, a second list of all the genes that give rise to one or more DE circRNAs was used. All the genes from which a bona-fide circRNA has been detected were set as the background gene list. Based on their False Discovery Rate (FDR) value, the top 25 processes were defined as the most enriched processes.

Receiver operating characteristic (ROC) curve

SPSS v20 was employed to plot the ROC curve using DeLong et al. methodology[255].

Results

Identification and characterization of circRNA profile in MS.

For the profiling of circRNAs in MS in an unbiased and genome-wide manner, we performed a RNA-seq of rRNA depleted total RNA from leukocytes of MS patients (n=30) and HCs (n=20). Find_circ and CIRI2 detected a total of 65,726 and 31,314 unique circRNAs supported by at least two BSJ spanning reads, respectively. A good detection overlap was observed between both algorithms with 22,835 *bona fide* circRNAs in common and with a good read count correlation (Figure 35A).

Most of the circRNAs are modestly expressed, the 96.5% of circRNAs were detected with less than 10 reads, but there is also a small group of 92 circRNAs (0.4%) that are highly abundant, detected with more than 50 read counts (Figure 35B). The distribution of the BSJ mapping read counts from all the samples after normalization was visualized, showing that there is not any outlier sample in the dataset (Figure 35C).

In order to analyze whether there was any bias in the distribution of circRNAs along the chromosomes, the number of circRNAs in each chromosome was displayed. The number of detected circRNAs was in general consistent with the size of the chromosome for all the autosomes while fewer circRNAs are produced from sexual chromosomes. In this regard it is worth noting that only 19 circRNAs are produced from the Y chromosome, and none of them is consistently detected in at least half of the samples of each group. The proportion of downregulated and upregulated circRNAs in MS patients with respect to HCs is also maintained for each of the chromosomes accounting for approximately 26-33% downregulated and 66-73% upregulated of the total number of circRNAs (Figure 35D).

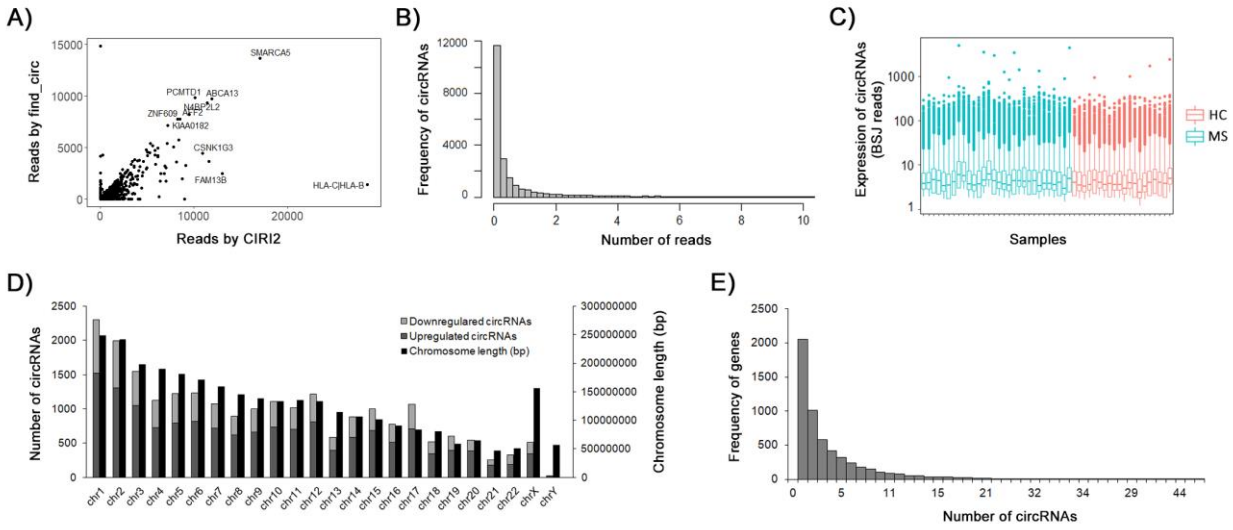

Figure 35. General features of the circRNA profile. A) Correlation of the quantification performed by find_circ and CIRI2. B) Histogram showing the number of reads per circRNA (up to 10 reads). C) Overall circRNA expression for each sample based on the mean number of reads spanning a BSJ. D) CircRNA distribution in chromosomes compared to the chromosome length. E) Histogram showing the number of circRNAs produced from each gene (up to 50 circRNAs).

Out of the 22,835 circRNAs detected, 22,197 are located in a known gene locus, while 638 circRNAs are intergenic. Interestingly, only 5,599 different genes have been identified within the genic regions producing circRNAs, revealing that most of the genes produce several circular transcripts. In fact, the 63,3% of the genes are producing several circRNAs with a mean of 5,6 circRNAs produced by each gene. Nevertheless, 2,053 genes give rise to a single circRNA and at the other end, there are also 533 genes producing 10 or more circRNAs (Figure 35E).

CircRNAs are more abundant in MS patients relative to controls and a sexual imbalance is observed.

Among the 22,835 *bona fide* circRNAs detected in total, 60% were detected both in MS and HC groups, while the rest were unique to one of the groups. Interestingly, more circRNAs were detected in patients (19,781 circRNAs) when compared to controls (16,772 circRNAs). In addition, regarding the overall expression of the bulk of circRNAs a subtle difference between groups is reported, indicating that circRNAs are slightly more abundant in MS patients than in HCs (Mean expression of the normalized read count MS=1.973 ± 10.811; HC=1.854 ± 7.526; $p<0.0001$) (Figure 35C).

In line with this, 464 circRNAs were found to be differentially expressed with a p-value<0.05 and a FC >|1.5| (DE circRNAs) and a clear trend to upregulation is observed with more than 96.1% of circRNAs upregulated in patients (446 up and 18 down) (Figure 36A).

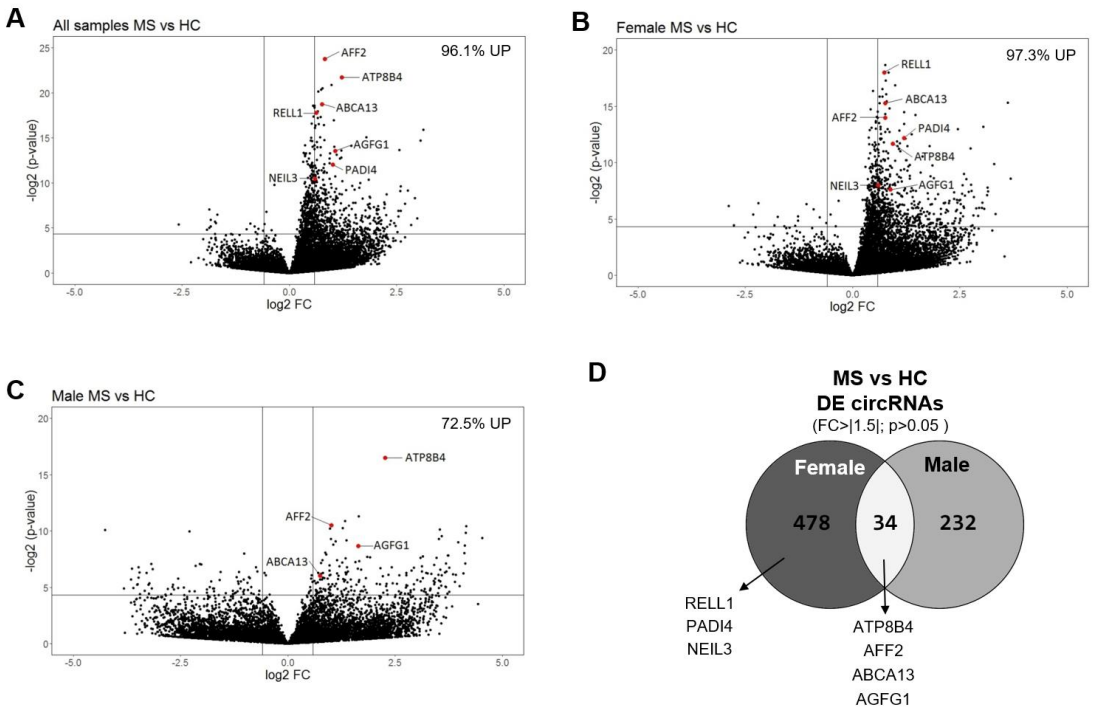

Figure 36. DE circRNAs between MS patients and HCs show a clear upregulation trend and sex bias. Volcano plots displaying the differential expression analysis results for circRNAs between A) all MS patients and HC (30 vs 20) B) Female MS patients and HC (23 vs 14) C) male MS patients and HC (7 vs 6). The percentage of significantly upregulated circRNAs was calculated for each of the comparisons and it is shown in the upper-right corner of each volcano plot. D) Venn diagram comparing the differentially expressed circRNAs for each of the sexes showing that there is a sex specific signature. 7 circRNA selected for RT-qPCR validation are indicated. Abbreviations: DE circRNAs, differentially expressed circRNAs; MS, multiple sclerosis; HC, Healthy controls; UP, upregulated; FC, fold change.

To investigate whether this upregulation was influenced by sex, we subdivided the cohort in female (20 MS vs 14HC) and male (10MS vs 6 HC) individuals and performed a second differential expression analysis. Using the same criteria, the upregulation trend was again observed with 498 circRNAs significantly upregulated and only 14 downregulated between patients and controls when the comparison was limited to females (97.3% upregulated) (Figure 36B). For males, in general fewer circRNAs are deregulated, with a total of 266 circRNAs differentially expressed between patients and controls. It can also be observed in the volcano plot that the upregulated to downregulated ratio is not as marked as it is for females. In fact the 72.5% of the circRNAs are upregulated (193 up and 73 down) and the upregulation is also less significant (Figure 36C). Interestingly, only 34 circRNAs are differentially expressed both in females and males while 93.4% and 87.21% of the deregulated circRNAs respectively, comprise a sex-specific signature (Figure 36D).

In order to reveal the biological processes in which the DE circRNAs could be involved a gene enrichment analysis was performed with the genes producing the DE circRNAs. However, none of the processes reached the significance criteria established indicating that DE circRNAs were derived from genes with very different activities.

Among the 464 DE circRNAs between all the MS patients and controls, 68 were high confidence circRNAs (Base Mean>10 and detected in >50% of the samples) and based on their abundance, fold changes in expression and statistical significance, 7 of those circRNAs were selected as candidates for their subsequent validation (Table 4). All the candidates happenned to be upregulated in MS patients.

Table 4. Table showing the fold changes and p-values for the seven candidate circRNAs both in the discovery (RNASeq) and validation (RT-qPCR) cohorts. The genomic location, parental gene and circBase ID are also shown for each of the circRNAs. Significant values are highlighted in bold.

			MS vs HC				**MS vs HC (Female)**				**MS vs HC (Male)**			
			FC RNASeq (p-value)		**FC RT-qPCR (p-value)**		**FC RNASeq (p-value)**		**FC RT-qPCR (p-value)**		**FC RNASeq (p-value)**		**FC RT-qPCR (p-value)**	
Genomic location (GRCh37)	**Gene**	**circBase ID**	MS N=30	HC N=20	MS N=70	HC N=46	MS N=23	HC N=14	MS N=58	HC N=28	MS N=7	HC N=6	MS N=12	HC N=18
chr15:50294349-50311173	ATP8B4	hsa_circ_0141241	2.39 (<0.0001)		1.63 (<0.0001)		1.92 (0.0003)		1.58 (0.0003)		4.79 (<0.0001)		1.40 (0.1028)	
chr2:228356262-228389631	AGFG1	hsa_circ_0058514	2.10 (<0.0001)		1.70 (<0.0001)		1.84 (0.0050)		1.75 (<0.0001)		3.11 (0.0024)		1.35 (0.1591)	
chr1:17668437-17668897	PADI4	circPADI4	2.02 (0.0002)		2.95 (<0.0001)		2.29 (0.0002)		2.17 (<0.0001)		1.43 (0.3392)		3.61 (<0.0001)	
chrX:147743428-147744289	AFF2	hsa_circ_0001947	1.78 (<0.0001)		1.67 (<0.0001)		1.70 (<0.0001)		1.68 (<0.0001)		2.02 (0.0006)		1.48 (0.0472)	
chr7:48541721-48542148	ABCA13	hsa_circ_0001707	1.69 (<0.0001)		1.80 (<0.0001)		1.69 (<0.0001)		1.62 (0.0002)		1.68 (0.0154)		1.91 (0.0088)	
chr4:37633006-37640126	RELL1	hsa_circ_0001400	1.55 (<0.0001)		1.21 (0.1664)		1.67 (<0.0001)		1.21 (0.2230)		1.27 (0.1956)		1.12 (0.6786)	
chr4:178274461-178274882	NEIL3	hsa_circ_0001459	1.51 (0.0006)		1.65 (<0.0001)		1.52 (0.0039)		1.81 (<0.0001)		1.50 (0.0857)		1.19 (0.4032)	

Abbreviations: MS, multiple sclerosis; HC, healthy control; FC, Fold change.

For the validation, RT-qPCRs were conducted using divergent primers spanning the respective backspliced junctions for each of the 7 circRNA candidates. The correct amplification of the BSJs was confirmed by agarose gel electrophoresis and Sanger sequencing (Figure 37). Six out of the 7 circRNAs were significantly upregulated in a second independent cohort of 70 MS patients and 46 controls, and RELL1 circRNA (chr4:37633006-37640126) showed also an upregulation trend although statistical significance was not reached (Figure 38).

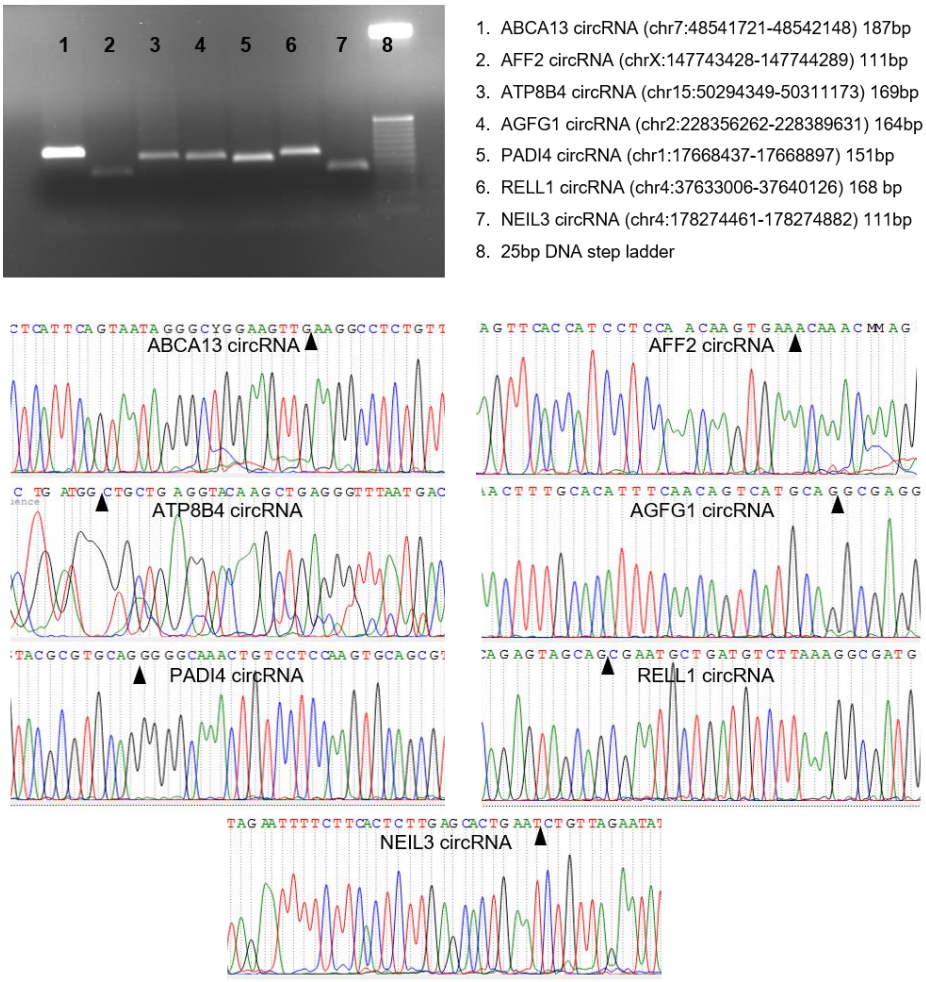

Figure 37. Agarose gel and sanger sequencing results confirming a single amplicon spanning the BSJ. The BSJ is indicated by and arrow for each of the 7 circRNAs.

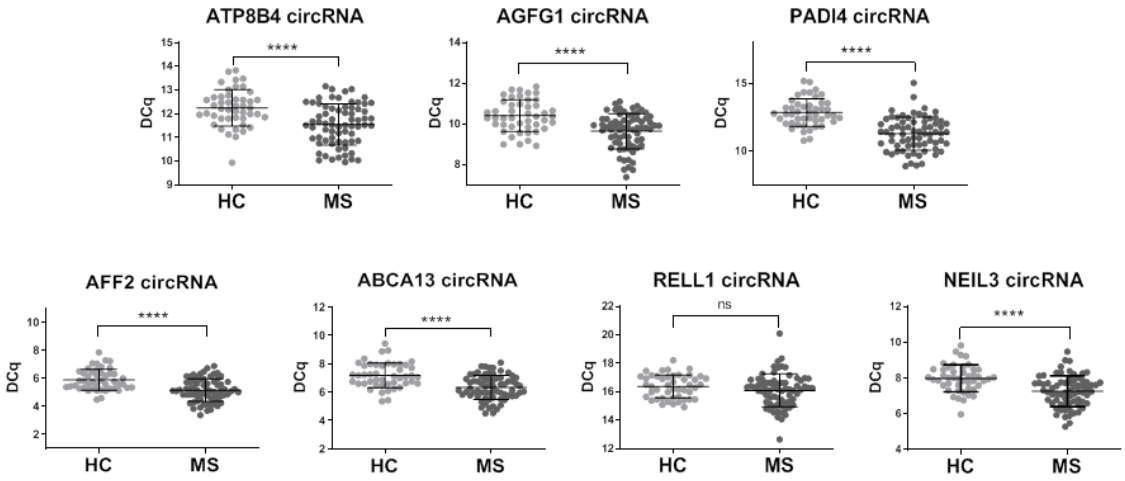

Figure 38. Six of the circRNA candidates, are found to be upregulated in a second independent cohort. 70 MS patients and 46 HC were included in this validation cohort. T-test was performed for the statistical analysis. Abbreviations and symbols: ns, not significant; ****, p-value ≤0.0001.

Based on the sex-bias observed for the RNASeq results, circRNA expression was also evaluated for women and men separately in the validation cohort (Figure 39). The 6 circRNAs whose upregulation was validated for the total of the samples, were also significantly upregulated for women. For men, only 3 circRNAs remain significantly upregulated, but it is worth noting that the FC for PADI4 circRNA (chr1:17668437-17668897) and ABCA13 circRNA (chr7:48541721-48542148) are increased in comparison with the results obtained both for females and for the whole cohort (Figure 39).

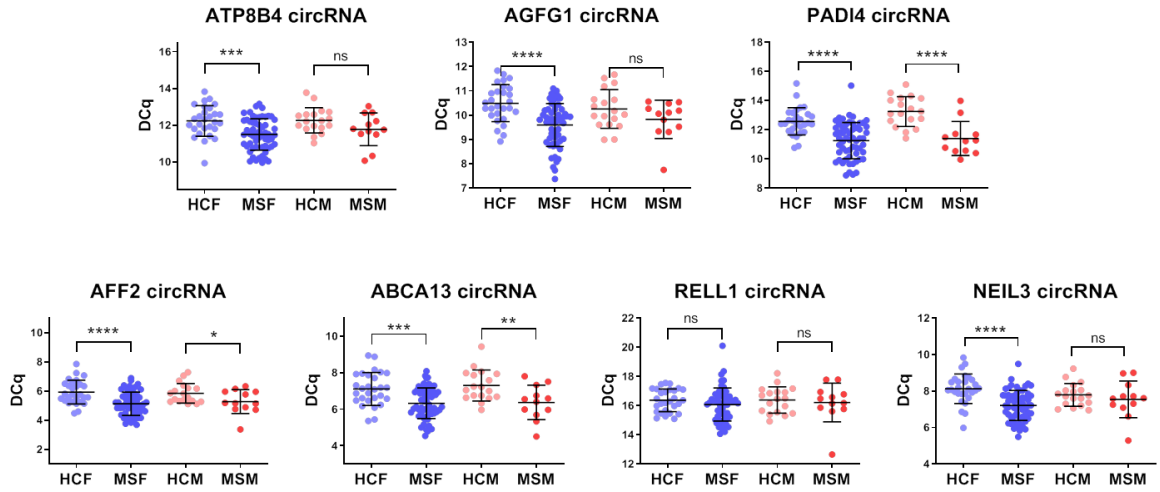

Figure 39. The upregulation is validated for 6 circRNAs in females but only for 3 in males. T-test was performed for the statistical analysis. Abbreviations and symbols: HCF, Female Healthy control; MSF, Female Multiple slcerosis patient; HCM, Male Healthy control; MSM, Male Multiple sclerosis patient; ns, not significant; ****, p-value≤0.0001; *** p-value≤0.001; **p-value≤0.01; *p-value≤0.05.

A small subset of circRNAs correlate with the MS type.

Among the 30 MS patients included in the RNASeq experiment, 20 were diagnosed from RR-MS and 10 from SP-MS at the moment of blood sampling. The 19,871 circRNAs detected in MS patients were subjected to a differential expression analysis between both MS types resulting in 121 DE circRNAs (See supplementary material in [285]). These DE circRNAs account for only the 0.61% of the circRNAs detected in patients, indicating that the general circRNA signature does not change with the disease progression (Figure 40). Moreover, only 7 of those circRNAs were high confidence circRNAs.

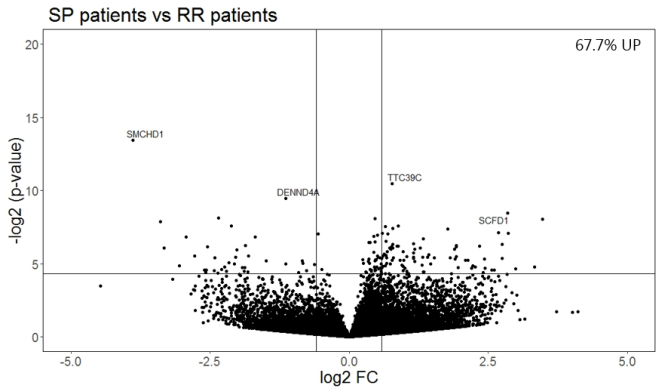

Figure 40. Volcano plot displaying the differential expressed circRNAs between SP-MS (n=10) and RR-MS (n=20) patients. The percentage of significantly upregulated circRNAs was calculated and it is shown in the upper-right corner of the volcano plot. Abbreviations: SP-MS, secondary progressive multiple sclerosis; RR-MS, relapsing remitting multiple sclerosis; UP, upregulated; FC, fold change.

CircRNA expression changes during relapses.

All the patient samples included both in the RNASeq and validation cohorts were drawn during the remission phase of the disease. To investigate whether the circRNA upregulation is general to the disease or it is associated with an underlying process such as the inflammation or the immune response, we have studied by RT-qPCR the expression of the 7 circRNA candidates in a third cohort that includes paired samples of 20 patients in relapse and remission as examples of two different inflammatory states of the disease.

For 6 out of the 20 patients there is no change (FC<|1.5|) for the circRNA expression between the relapse and remission phases. The other 14 patients show a differential expression of most, if not all the circRNAs studied during relapse. 7 patients have at least half of the circRNAs downregulated in relapse with respect to the remission, while the other 7 patients show the opposite trend, with at least half of the measured circRNAs upregulated (FC>|1.5|) (Figure 41).

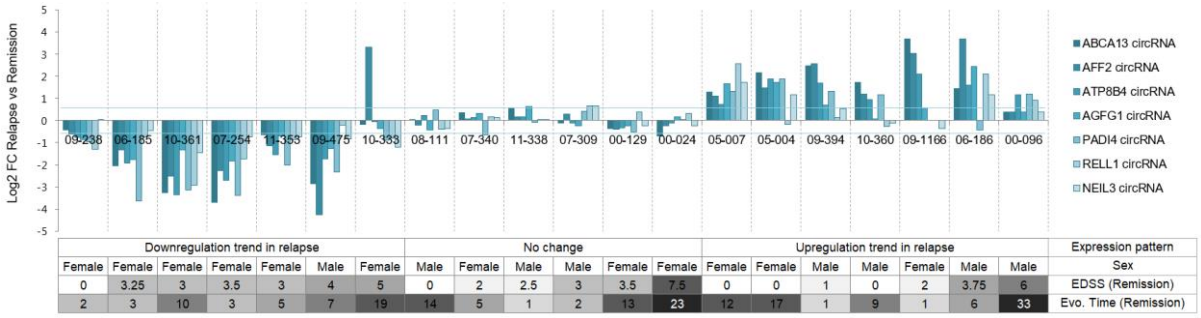

Figure 41. CircRNA expression changes between relapse and remission phases of the disease. The expression change of the 7 circRNA candidates between relapse and remission is shown by the log2 FC for each of the patients. A blue line is depicted in order to represent the FC>|1.5| threshold. A table including the sex, EDSS and Evolution time data for each of the patients is shown below the graphic.

None of the clinical details of the patients is able to gather the patients based on their deregulation trend. Nevertheless, slight differences (although not statistically significant) can be noted between those who experiment an upregulation of the circRNAs during relapse when compared with the ones experimenting a downregulation of circRNAs during relapse. Patients with an upregulation trend show a less severe phenotype reported by less accumulated disability (Median $EDSS_{upregulated\ group}$=1 vs Median $EDSS_{downregulated\ group}$==3.3) together with a longer evolution time (Median Evol. $time_{upregulated\ group}$=9 vs Median Evol. $time_{downregulated\ group}$=5). The EDSS and evolution time value taken for these comparisons are the ones observed during the remission period. Additionally, 6 out of the 7 patients in the group of those with an upregulation trend in relapse are men while they represent the 57.1% of the group showing a downregulated trend.

The linear transcriptome profile does not show the same upregulation trend found for circRNAs.

Taking advantage of the RNASeq dataset obtained for the circRNA profiling, the linear transcriptome profile was also studied. 45,736 linear RNAs were detected using HT-Seq and the differential expression analysis between MS patients and HCs reported a total of 1,372 DE linear RNAs described as those with a FC >|1.5|and a p-value<0.05. Interestingly, 689 linear RNAs were upregulated and 683 downregulated in patients, revealing that, in contrast with circRNAs (96.1% upregulated), a global upregulation trend is not found for linear RNAs (50.2% upregulated) (Figure 42).

No trend for upregulation was neither found for circRNAs when the differential expression analysis was limited to female (20 MS vs 14HC) and male (10 MS vs 6 HC) individuals.

Among the 1,598 DE linear RNAs found between female MS patients and controls, 718 were upregulated accounting for the 44.9% of the DE linear RNAs while, as presented before, the 97.3% of circRNAs were upregulated in the same samples. Similarly, 1,510 linear RNAs were differentially expressed between male patients and controls and only the 58.5% were upregulated (884 linear RNAs) (Figure 42).

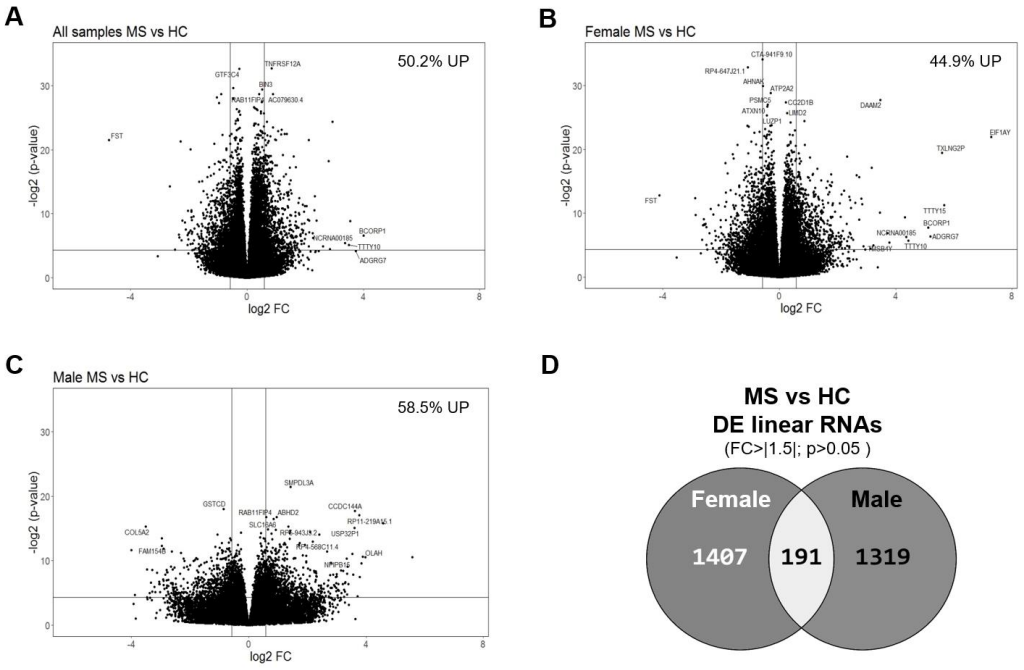

Figure 42. DE linear RNAs between MS patients and HCs do not show any general deregulation trend. Volcano plots displaying the differential expression analysis results for linear RNAs between A) all MS patients and HC (30 vs 20) B) Female MS patients and HC (23 vs 14) C) male MS patients and HC (7 vs 6). The percentage of significantly upregulated linear RNAs was calculated for each of the comparisons and it is shown in the upper-right corner of each volcano plot. D) Venn diagram comparing the differentially expressed linear RNAs for each of the sexes showing that there is a sex specific signature. Abbreviations: DE linear RNAs, differentially expressed linear RNAs; MS, multiple sclerosis; HC, Healthy controls; UP, upregulated; FC, fold change.

The gene enrichment analysis reported that the DE linear RNA dataset is enriched in genes coding for proteins involved in 178 different processes, most of them immune related processes, being the most enriched ones myeloid leukocyte activation (enrichment score=4.27, FDR-value $3.59 \cdot 10^{-23}$), neutrophil activation involved in immune response (enrichment score=4.43, p-value=$3.24 \cdot 10^{-21}$), neutrophil degranulation (enrichment score=4.47, FDR-value=$3.34 \cdot 10^{-21}$) and granulocyte activation (enrichment score=4.37, p-value=$3.87 \cdot 10^{-21}$). The same processes are also enriched in the datasets from females and males although the enrichment score is smaller (See supplementary table at [285]).

CircRNAs as potential biomarkers for MS diagnosis.

To explore the diagnostic potential of the aforementioned circRNA candidates, we performed the ROC curve analysis. As shown in Figure 43, the 6 circRNAs showed a good performance with and area under the curve (AUC) to discriminate patients from healthy controls of 0.843 for PADI4 circRNA (chr1:17668437-17668897), 0.748 for ABCA13 circRNA (chr7:48541721-48542148), 0.753 for AFF2 circRNA (chrX:147743428-147744289), 0.741 for NEIL3 circRNA (chr4:178274461-178274882), 0.733 for AGFG1 circRNA (chr2:228356262-228389631) and 0.722 for ATP8B4 circRNA (chr15:50294349-50311173). PADI4 circRNA outstands as the circRNA with the biggest diagnostic potential, and its combination with the other 5 circRNAs improves the obtained values implying that the whole signature of the 6 upregulated circRNAs could have the potential to diagnose the 85.2% of the MS cases.

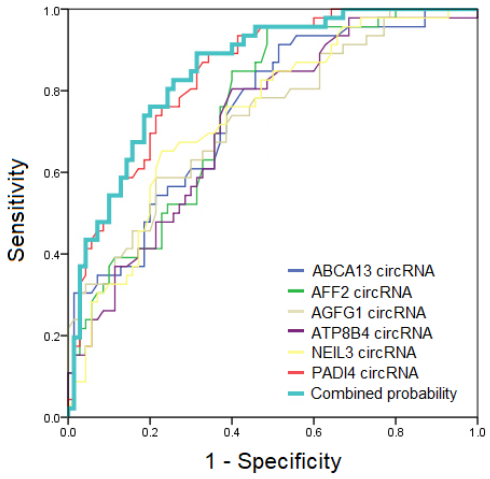

Figure 43. ROC curves showing the potential of the 6 validated circRNAs to be biomarkers in MS. A curve for each of the circRNAs as well as a curve representing the combined probability of the 6 of them are shown.

Discussion

MS is a complex disease, whose physiopathology is still not completely understood. Nevertheless, growing evidence supports a deregulation of the immune system[27]. Non-coding RNAs such as microRNAs participate in the regulation of various immune responses. Thus, they have been proposed not only as disease targets, but also as biomarkers. In fact, hundreds of miRNAs have been found deregulated in MS and many of them have shown potential to predict the conversion of CIS patients to MS, to discriminate between MS subtypes, to predict disease progression, severity and disability [228,231,239] although none of them have reached to the clinic yet. Circular RNAs have recently emerged as new players

in the non-coding RNA family and as miRNAs, recent works have shown their participation in different immune responses and immune-related diseases[152] among others. Nevertheless, compared to other RNA biomarkers such as miRNAs, their particular structure endows them several remarkable characteristics that potentially make them better biomarkers[147]. Unlike linear RNAs, they are highly resistant to RNases resulting in higher stability and longer half-lifes in body fluids [121] making easier their detection and reducing the impact of preanalytical factors such as sample collection and storage. In addition, they are expressed in a tissue- and developmental stage- specific manner[80,82] which potentially overcomes the low specificity of some of the current biomarkers. Notably, circRNAs also present better cross-species conservation [105] favoring an easier translation of biomarkers from animal models to human beings. On top of that, given the bigger size of circRNA compared to miRNAs, they could be tested in RNA samples isolated with kits that do not recover small RNA fraction, theoretically expanding their usage to samples that have been stored for many years.

Taking all together, circRNAs have been regarded as good non invasive biomarkers in several diseases and also hold a great potential as biomarkers for MS, as shown by some studies in the field (reviewed in [286]). Nevertheless, to our knowledge, the global expression of circRNAs in MS has only been studied in a previous work published by our group where we studied circRNAs by expression arrays in female RR-MS patients and healthy controls[125]. Based on the interesting results obtained in that study, we wanted now to extend the analysis by increasing the number of samples and by changing the screening method to RNASeq. This time, the discovery cohort comprised 50 samples including individuals of two different sexes and two different phenotypes of the disease as two of the major factors affecting the heterogeneity of the disease. Besides, the RNASeq technique has allowed us to study a higher amount of circRNAs including those that have not previously been described and included in the array (Arraystar Human Circular RNA Microarray V2.0).

Concerning the study of circRNAs by RNASeq, several circRNA prediction algorithms have been developed but a gold standard pipeline has still not been defined. In fact, different comparative studies have shown that the algorithms yield highly divergent results and a high level of false positives based on RNase R resistance data[95,252,287]. The pairing up of different algorithms has been proposed as an appropriated approach that provides a more reliable output[287], and for this reason, we decided to combine the prediction results from find_circ and CIRI2 algorithms. A final number of 22,835 *bona fide* circRNAs detected with this approach.

When the circRNAs detected for MS patients were compared to those detected in HC samples we found that in MS patients there is a higher diversity and abundance of circRNAs and also a pronounced upregulation trend. We investigated whether the upregulation was general to the transcriptome, which could be affecting other types of transcripts such as linear RNAs, or whether it was specific for circRNAs. We observed that although many linear RNAs were dysregulated, there was not an evident trend to upregulation with the 50.2% of the transcripts upregulated compared to the 96.1% of the circRNAs upregulated (Figure 36) (Figure 42). Moreover, in contrast with the DE linear RNAs that are found to be related to a immune response, the DE circRNAs are not particularly coming from immune related genes but from genes with very different and diverse functions. These results demonstrate that the circRNA transcriptome of MS patients is specifically changed, independently of the linear transcriptome.

The drastic and specific upregulation of circRNAs could be driven by an alteration of the circRNA biogenesis. The circRNA biogenesis requires the spliceosomal machinery and it is regulated by *cis*-regulatory elements and several RNA binding proteins acting in *trans*[288]. Several different RNA binding proteins (RBPs) take part in the regulation of circRNA biogenesis, such as ADAR1 and DHX9 known to suppress the circRNA formation and NF90, QKI, FUS and HNRNPL that have been reported to be involved in the circRNA production by enhancing it[106]. Nevertheless, based on the RNA-seq data, none of these RBPs are deregulated in MS patients at least at RNA level, indicating that in this case, they are unlikely to be the main drivers of circRNA upregulation in MS patients.

CircRNAs have also been shown to be cell and tissue specific, therefore, and since we have analysed the total RNA from leukocytes with no separation of cell types, differences in the blood cell proportions associated with MS could be responsible for the observed change in circRNA expression. In order to indirectly make out whether there was any difference in the cell count between patients and healthy controls, we compared the expression of several genes reported to correlate with the abundance of B-lymphocytes, T-lymphocytes and granulocytes[289]. Interestingly, all the 10 genes reflecting the abundance of granulocytes were significantly upregulated in MS, with 6 of them being expressed at least 1,5 times more than in healthy individuals. Meanwhile, only 1 of the B-lymphocyte genes was significantly upregulated and 2 T-lymphocyte genes were downregulated. These preliminary results are in line with other studies that have found an increased neutrophil to lymphocyte ratio in MS patients[290].

Taking all together, and having confirmed that the upregulation is circRNA-specific but that it is unlikely to be related to an alteration of their biogenesis or in the leukocyte cell proportion, we hypothesize that circRNAs are upregulated because they are playing a role in the disease. Interestingly, the involvement of circRNAs in the regulation of different immune responses and immune-related diseases has previously been studied[151,152]. In fact, studies performed in other autoimmune diseases such as systemic lupus erythematosus (SLE) and psoriasis also reported a global change in the circRNA expression levels, found to be downregulated in SLE patients and lesional psoriatic skin[118,291]. These results, together with our findings showing a global change of circRNAs both between MS patients and controls and between MS patients in relapse and remission conditions reinforce the implication of circRNAs in the regulation of immune responses.

The molecular function of most of the circRNAs remains unknown. So far, circRNAs have shown to exert their regulatory functions by sponging miRNA or RBPs[90,139,292], regulating the transcription of their parental genes, competing with the linear splicing[106] or even coding for proteins[91]. Most of these functions are sequence-specific, however, the global change in the expression of circRNAs points to a function that could also be global. Interestingly, a recent study has reported a structure-related function. They provided evidence that many endogenous circRNAs tend to form imperfect RNA duplexes and can regulate the innate immune response via their interaction with the double stranded RNA - activated protein kinase[118]. Whether the global circRNA upregulation observed in MS and other autoimmune diseases is related to this conformation and function should be further investigated.

Moreover, it is known that the immune system is strongly affected by sex specific differences[187], which could explain the different circRNA profile between females and males. In fact, regarding MS, a clear gender bias has already been detected in gene expression studies of both coding and non-coding genes[217,220,221,246]. Therefore, the circular transcriptome data described in this study grow the existing evidence indicating that sex may be a fundamental factor not only when diagnosing MS but also for designing and administering personalized treatments.

To date few works have addressed the role of circRNAs in MS, and thus, very few circRNAs have been related to the disease. Cardamone et al. studied the alternative splicing abnormalities in the GSDMB gene and found that a circular isoform of this gene (chr17:38065210-38066177) is upregulated in RR-MS[260]. Our group also found in a microarray study, several differentially expressed circRNAs in MS patients when compared

to controls and proposed 2 circRNAs derived from the ANXA2 gene as potential biomarkers of the disease (chr15:60648117-60674640 and chr15:60653142-60678285)[125]. Nevertheless, in the present study none of the two ANXA2 derived circRNAs were detected and GSDMB circRNA (chr17:38065210-38066177) was not differentially expressed in our samples (FC=1.05 and p-value=0.92). Out of the 464 DE circRNAs detected in this study, 7 circRNAs were also found to be differentially expressed in the study by microarrays (FC >|1.5|and p<0.05)[125]. Nevertheless, out of these 7 circRNAs only three circRNAs coming from the genes *RELL1* (chr4:37633006-37640126), *RNF149* (chr2:101898320-101911643) and *CHD9* (chr16:53288349-53308214) respectively had the same deregulation trend in both studies. Interestingly, in the study performed in 2017 the upregulation of the RELL1 circRNA (chr4:37633006-37640126) could not be confirmed in the validation cohort (20MS and 18 HC) but this time it was selected again as a candidate due to its high expression and its upregulation was confirmed in a bigger validation cohort by RT-qPCR (70 MS and 46 HC). In addition, Paraboschi et al. carried out a bioinformatical analysis showing that the MS genome-wide associated loci are enriched in circRNAs suggesting their possible involvement in the susceptibility of the disease. 482 circRNAs were listed in the work by Paraboschi et al. and interestingly 4 of them were differentially expressed in our samples although their expression is low (BaseMean<10).

To the best of our knowledge, this study is the first to investigate, together with the linear transcriptome, the circRNAome in blood from MS and HC individuals by RNASeq. It is also the first reporting a global and specific increase of circRNA expression levels in leukocytes from patients. Based on our data, the circRNA expression differences could not be explanied by an alteration of the circRNA biogenesis nor by changes in the blood cell proportion. In light of previous studies that have reported the role of circRNAs in several immune processes, we point to their possible implication in the physiopathology of MS. Moreover, we validated the upregulation of 6 circRNAs in a large cohort of patients and controls concluding that they hold a great potential to be used as minimally invasive biomarkers in MS.

1.3

RNASeq profiling of circRNAs and linear RNAs in PBMCs from patients with IgM characterization

This subchapter has been published in *Biomedicines:*

Iparraguirre L, Olaverri D, Blasco T, Sepúlveda L, Castillo-Triviño T, Espiño M, Costa-Frossard L, Prada A, Villar L-M, Otaegui D, Muñoz-Culla M. *Whole-transcriptome analysis in peripheral blood mononuclear cells from patients with lipid-specific oligoclonal IgM band characterization reveals two circular RNAs and two linear RNAs as biomarkers of highly active disease.* Biomedicines. Accepted the 23 of November 2020

Introduction

Disease activity biomarkers can be associated with the different pathophysiological processes of the disease and they could ideally help to distinguish between patients with an aggressive course and those with a more benign form[226]. In this context, the presence restricted to cerebrospinal fluid (CSF) of total oligoclonal IgM bands (OCMBs) and particularly, the presence of lipid-specific oligoclonal IgM bands (LS-OCMBs) has been defined as an accurate predictor of an aggressive evolution[293]. However, for this test, cerebrospinal fluid sample is needed, a quite invasive liquid biopsy. Therefore, a great effort has been done towards the discovery of minimally invasive biomarkers such as blood biomarkers in order to avoid lumbar punctures[230].

In this context, due to the increasing need for a solid, reproducible and accessible biomarker for MS, and in light of the evidences showing that circRNAs are promising disease biomarkers, we hypothesized that a blood transcriptome study of PBMCs from patients with LS-OCMBs characterization, could reveal new biomarkers that correlate with this stablished CSF marker. Such a biomarker could be used more easily to monitor patients on an ongoing basis, since blood sampling is less invasive and with less adverse effects than doing a lumbar puncture.

With this in mind, in the present study we analyzed the expression of circRNAs and linear RNAs in peripheral blood mononuclear cells (PBMCs) from 88 patients with positive (n=42) and negative (n=46) LS-OCMBs characterization.

Materials and methods

Patients, sample collection and RNA isolation

MS patients were recruited in the Hospital Ramon y Cajal after giving informed consent. Blood samples were obtained and PBMCs were isolated following a standard Ficol gradient separation protocol and frozen in liquid nitrogen until used. Total RNA was isolated using miRNeasy mini kit (Qiagen) following the manufacturer's instructions. RNA concentration was measured using NanoDrop ND-1000 spectrophotometer and the quality of the samples included in RNA-seq experiments was assessed using Agilent 2100 Bioanalyzer, obtaining a RNA integrity number higher than 6 in all the samples. CSF was obtained to analyze the status of LS-OCMBs as previously described [293,294].

The main clinical and demographical characteristics of patients are summarised in Table 5. A total of 88 patients with positive (n=42) and negative (n=46) LS-OCMBs characterization were included distributed in three different cohorts. Some patients were included in two cohorts. Out of the 88 patients 8 were selected for the profiling cohort by RNASeq, circRNA expression was validated in all of them and only 60 were included in the linear RNA validation cohort due to sample volume limitations. The study was approved by the hospital's ethics committee (MMC-UEM-2018-01). The profiling cohort includes only female samples, aiming at reducing the source of variability and considering that MS is three times more prevalent in females than men in this disease[182].

Table 5. Table summarizing clinical and demographical data of patients enrolled in the study.

	LS-OCMB status	Sex	Age
Profiling cohort	Positive (n=4)	Female (n=4)	38 (±14.5)
	Negative (n=4)	Female (n=4)	35 (±8.3)
CircRNA validation cohort	Positive (n=42)	Female (n=24) Male (n=18)	34 (±9.3)
	Negative (n=46)	Female (n=38) Male (n=8)	37 (±8.0)
Linear RNA validation cohort	Positive (n=30)	Female (n=18) Male (n=12)	33.4 (±8.7)
	Negative (n=30)	Female (n=24) Male (n=6)	36.9 (±7.6)

RNASeq

Library preparation and RNA sequencing was performed at CD Genomics (USA) as previously described in subchapter 2.

CircRNA detection and quantification in RNASeq data

First, quality of the sequencing was checked and then, reads were mapped to the human genome (hg19, downloaded from UCSC Genome Browser [295]) using STAR [282] or BWA [296]. Subsequently, circRNA prediction was performed by CIRCexplorer2 [111] and CIRI2 [280] adhering to the recommendation by the authors. Moreover, only circRNAs supported by both algorithms were considered as *bona fide* circRNAs and used in subsequent analyses, a method that has been described in the literature by other authors [95]. CircRNA expression was based on back-spliced junctions-spanning reads according to CIRI2 quantification. After filtering out circRNAs with low expression (sum of reads < 10), differential expression analysis was performed using DESeq2 [281] package for R (version 3.6.3) in R-studio (Version 1.0.136), considering as differentially expressed those circRNA showing a p-value <0.05 and a fold-change higher than 2. To select a group of candidates for validation purposes, two additional filters were applied: 1) BaseMean value higher than five (BM$>$ 5) and 2) the transcripts having a read count value of zero in any of the samples were removed. Finally, we selected ten candidate circRNAs having the highest absolute fold-change value.

Linear RNA detection and quantification in RNA-seq data

After sequencing and quality control, Kallisto [297] algorithm was used for transcript pseudoalignment, identification and quantification. Differential expression analysis was

performed using DESeq2 package. Before running DESeq2 algorithm, we applied a filter of a minimum number of reads (sum of number of reads higher than or equal to 10) to remove low expression transcripts. For subsequent analysis, due to the large number of transcripts detected and that we have a small sample size, we added a filter to keep only the most robust transcripts regarding their read count. Therefore, we created a detection filter, as follows; 1) Transcripts detected in both groups: transcripts having at least two reads in ¾ of samples from each group. 2) Transcripts detected only in one of the groups: transcripts having at least two reads in ¾ of samples in the negative group but only in 1 sample from the positive group (detected only in negative group) and transcripts having at least two reads in ¾ of positive group but only in 1 from negative group (detected only in positive group). Transcripts selected with this detection filter were considered as consistent linear RNAs. Finally, differentially expressed transcripts were considered those showing a p-value$<$0.05 and an absolute fold-change value higher than two. For validation purposes, we selected ten candidate linear RNAs having the highest absolute fold-change value and having an assigned gene name.

All these profiling experiments and data analysis steps, tools and filters are summarized in Figure 44.

cDNA synthesis and Quantitative PCR

For the validation of candidate circRNAs, a two step RT-qPCR was carried out as described in subchapters 1 and 2. In this case, EEF1A1 and B2M were used as the reference genes, using the mean value of both genes for normalization purposes. Primer sequences are available at the Appendix Table 1. The presence of a single-peak in the melting curve indicated the specificity of the amplification.

For the validation of candidate linear RNAs, total RNA (500 ng) was reverse transcribed into cDNA with oligo-dT and random primers using the miScript II RT (Qiagen) kit with the HiFlex buffer, as indicated in manufacturer's protocol. PCR reaction was set up using 10 ng of cDNA as template and using miSscript SYBRGreen PCR kit (Qiagen) with the following thermal cycling program: 95^{o}C for 15 minutes, 40 cycles of 94^{o}C for 15 seconds, 55^{o}C for 30 seconds and 70^{o}C for 30 seconds followed by a dissociation curve analysis. The amplification of each linear RNA was carried out with QuantiTect primer Assays (Table S2) and B2M was used as a reference gene for normalization. The presence of a single-peak in the melting curve indicated the specificity of the amplification.

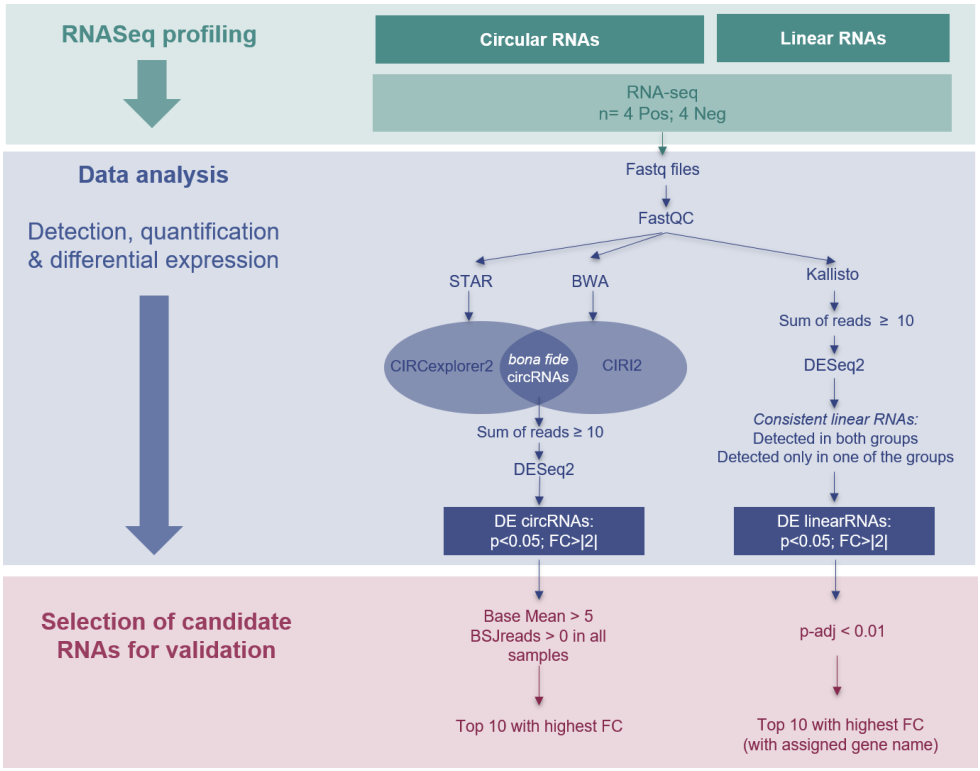

Figure 44. Summary of the study indicating profiling experiments, data analysis tools, filters and candidate selection criteria.

All retrotranscription reactions were run in a Veriti Thermal Cycler (Applied Biosytems) and quantitative PCR reactions were performed in a CFX384 Touch Real-Time PCR Detection System (BioRad).

Prior to run all the validation experiments, a technical validation of the circRNA amplification was carried out. As described in previous subchapters, circRNA amplicons were checked for the presence of the BSJ by Sanger sequencing and subjected to electrophoresis.

The Raw Cq values and melting curves were analysed in CFX Maestro 1.0 (BioRad). Expression levels represented as Fold change (FC) were calculated using the 2^DDCq method. Statistical analysis was done by R in RStudio on DCq data. Distribution of data was tested with Shapiro-Wilk test and the difference of the distribution was assessed by Student t-test or non-parametric Wilcoxon Rank sum test. In each circRNA and linear RNA dataset, outlier samples were removed before the statistical analysis, identifying those values using boxplot.stat function in R.

Gene overrepresentation test

In order to describe the function of differentially expressed transcripts, PANTHER overrepresentation test was applied (released 2019-07-11, Panther [298]), based on Complete Biological Process from Gene Ontology database (released 2019-12-09). As a reference background, we used the list of genes giving rise to detected transcripts defined above. Fisher test was carried out and false discovery rate (FDR) correction was applied to raw p-values, taking as a significant threshold FDR values below 0.05.

ROC curve analysis

Based on DCq values obtained by RT-qPCR we performed a Receiver Operating Characteristic curve (ROC curve) analysis in R environment in RStudio. Three different datasets were used: 1) circRNA validation data, 2) linear RNA validation data and 3) samples in both datasets to test different combination of transcripts. This last dataset was smaller given that we needed to include only samples included in both cohorts and with expression value for all tested transcripts. The combination of different transcripts was computed with ROC function of "Epi" package.

Results

The expression of some circRNAs is different between LS-OCMB^{+} and LS-OCMB^{-} patients.

RNA-seq analysis pipeline of circular transcriptome detected 27,630 *bona fide* circRNAs and 5,431 circRNAs (19.6%) with sum of reads ≥ 10 (A). Differential expression analysis revealed that 124 circRNAs are altered ($p<0.05$; $FC>|2|$) between the two groups, 72 of them being upregulated while 52 are downregulated.

We selected ten candidate circRNAs to confirm their differential expression by RT-qPCR in a larger cohort. First, the technical test for primer amplification confirmed a single and specific circRNA amplification for 9 out of the 10 circRNA candidates. circMETRNL was the one that did not show a good amplification and a single band in agarose gel, so we discarded it in subsequent validation experiments. Additionally, Sanger sequencing of the PCR product confirmed that the amplicons span the BSJ (Figure 45). As it can be observed in Figure 46C, the lower expression of hsa_circ_0000478 and hsa_circ_0116639 was

confirmed by qPCR in PBMCs from patients with positive LS-OCMBs (FC = -1.5 and FC=-1.65, respectively; $p<0.01$).

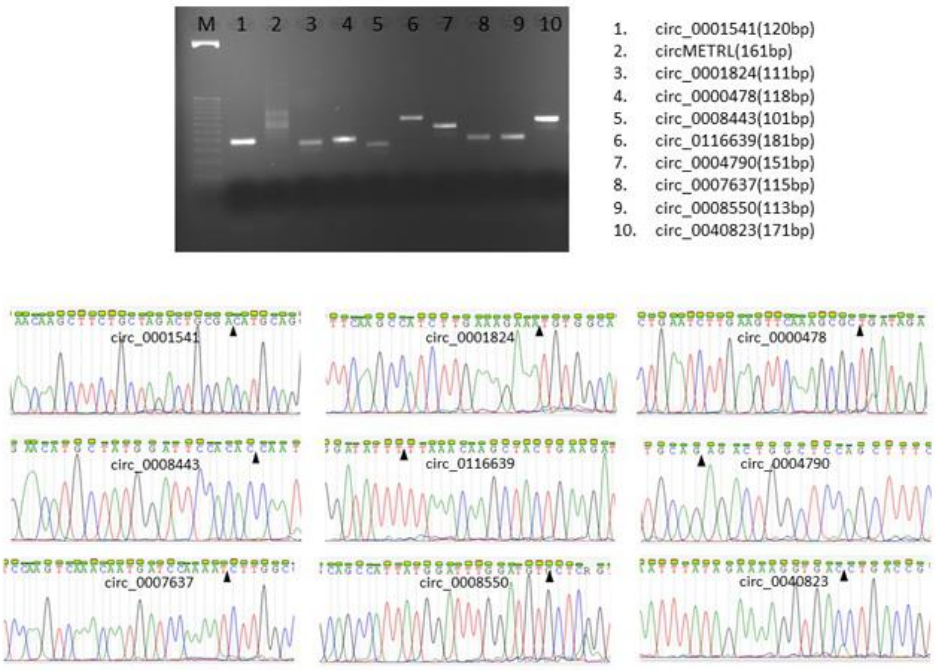

Figure 45. Agarose gel electrophoresis image of PCR product and Sanger Sequencing results showing the backspliced junction of circRNAs.

LS- OCMB+ and LS-OCMB- patients show different linear RNA profiles in PBMCs

In the present study, we also analyzed the linear transcriptome by RNA-seq, identifying 115,869 transcripts with ≥ 10 reads in total. We applied an additional filter using a detection criterion, which permitted to identify 84,863 linear RNAs with a consistent expression pattern, from which 92.8% were detected in both groups. Among them, 2,441 transcripts were differentially expressed ($p<0.05$ and FC $>|2|$), from which 1421 were detected in both groups (Figure 46B). The rest of the transcripts are specific to one of the groups, 382 being expressed only in PBMCs from patients with negative LS-OCMBs status while 638 transcripts are only expressed in patients with positive LS-OCMBs.

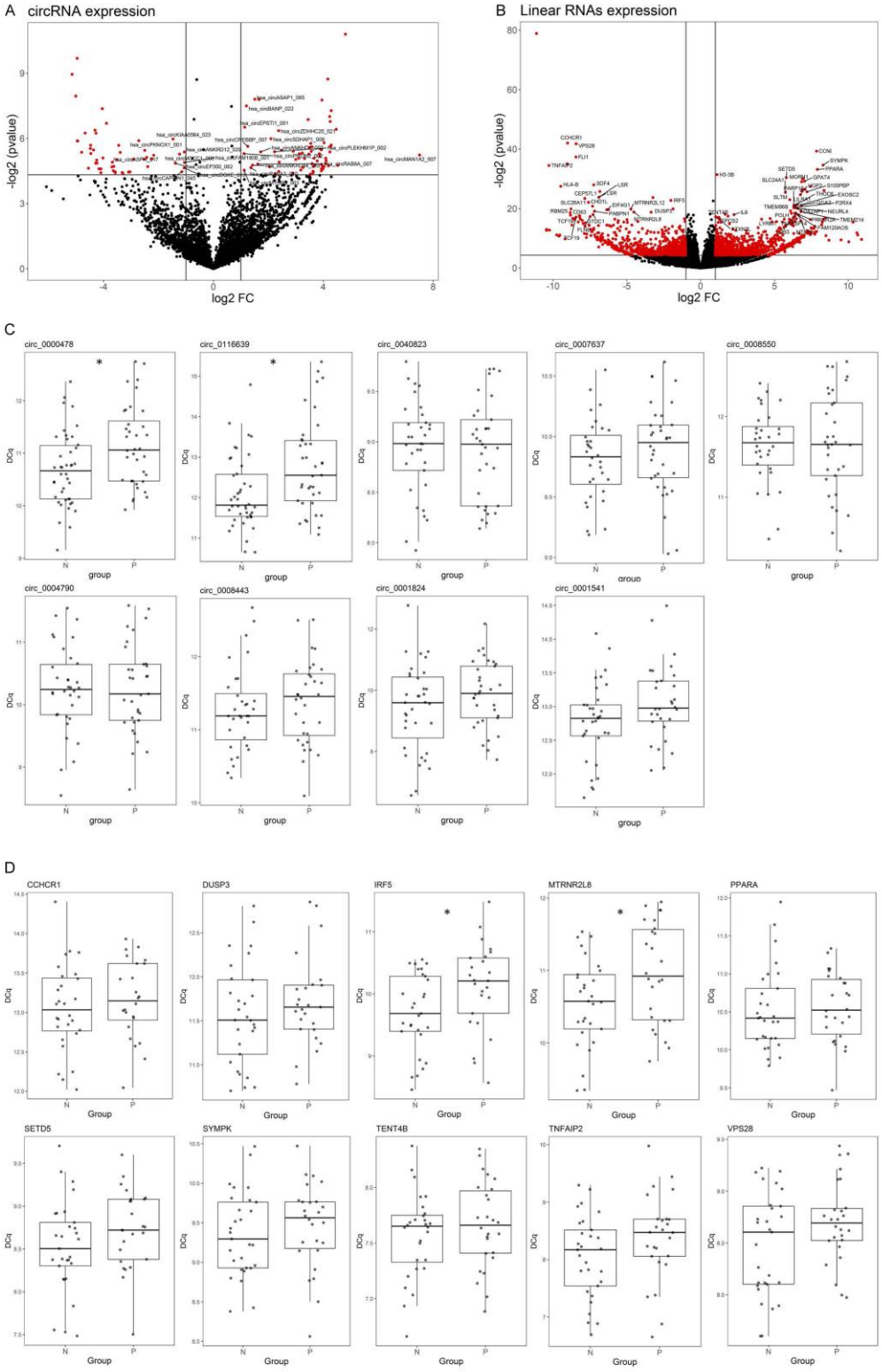

Figure 46. Results of circRNAs and linear RNAs expression profile by RNA-seq. Volcano plots showing the expression difference between postive and negative group of circRNAs with a sum read>10 (A) and linear RNAs (B) DE circRNAs and linear RNAs are highlighted in red. C) RT-qPCR validation results of selected candidate circRNAs. Asterisk indicates a statistically significant difference ($p<0.01$). D) RT-qPCR validation results of selected candidate linear RNAs. Asterisk indicates a statistically significant difference ($p<0.05$). P: positive LS-OCMB status. N: negative LS-OCMB status.

We selected ten candidate linear RNAs to confirm their differential expression by RT-qPCR in a larger cohort. As it can be observed in Figure 46D, the lower expression of IRF5 and MRNR2L8 was confirmed in PBMCs from patients with positive LS-OCMBs (FC = -1.33 in both transcripts; $p < 0.05$). Differentially expressed linear RNAs are enriched in biological processes regarding mainly the immune system. The most significant and enriched terms (FDR <0.01 and fold-enrichment > 2) are shown in Figure 47. Complement activation, humoral immune response and type I interferon signaling pathway appear among the top enriched terms.

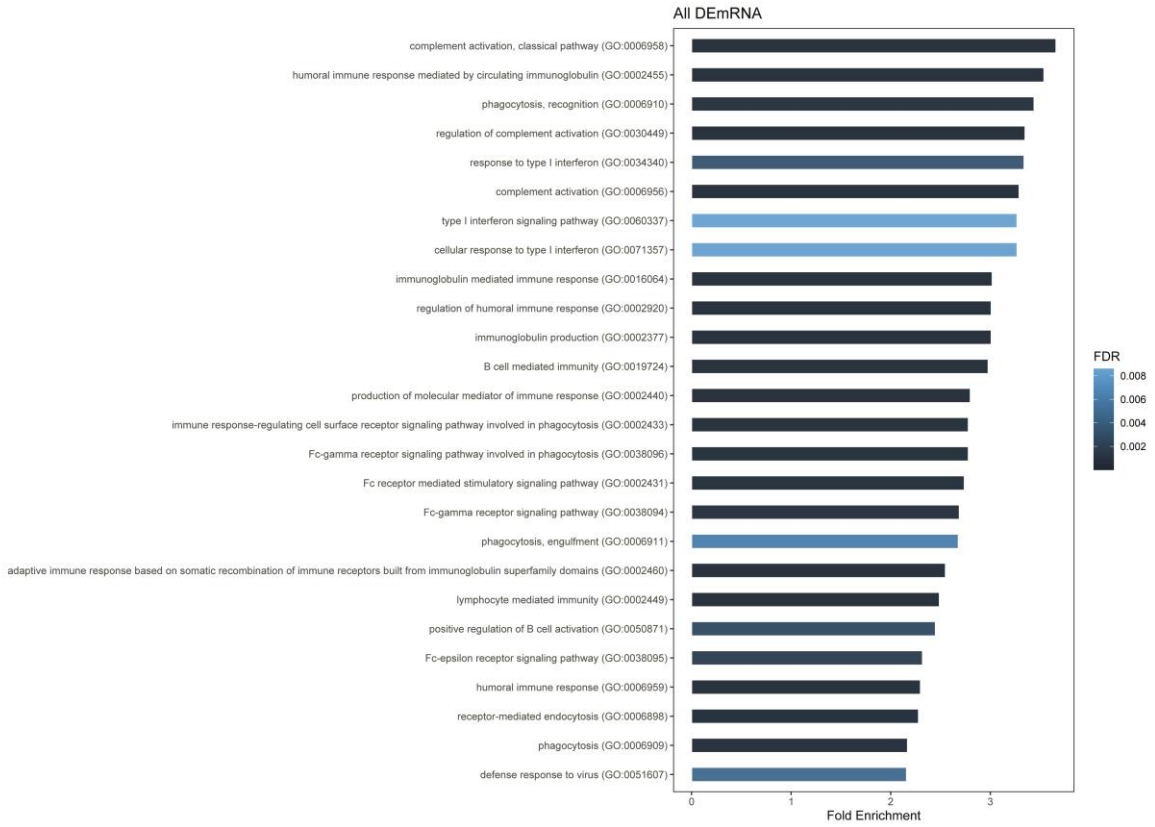

Figure 47. Results from gene overrepresentation test of 2441 DEmRNA. The most significant (fold-enrichment > 2) and enriched (FDR <0.01) GO biological processes are shown. Bars are coloured according to their FDR value.

Evaluation of circRNA and linear RNAs as biomarkers of a highly active disease

In order to assess the potential of the four validated RNAs as blood biomarkers of the LS-OCMB status we performed the ROC curve analysis. As shown in Table 6, different combination of RNAs have been tested to find the best performance in discriminating both

groups. We found that different combination of both circRNAs and linear RNAs improves the performance, reaching AUC values of almost 70%.

Table 6. ROC analysis results of the four candidate transcripts and different combinations.

Marker/Combination	Transcripts	Sample size	AUC (%)	CI (%)
circRNA-1	circ_0000478	80 (43 LS-OCMB-; 37 LS-OCMB+)	65.9	53.9-77.9
circRNA-2	circ_0116639		66.4	54.2-78.5
Combination_circ	circ_0000478-circ_0116639		68.9	57.2-80.5
linearRNA-1	IRF5	55 (29 LS-OCMB-; 26 LS-OCMB+)	65.6	50.7-80.6
linearRNA-2	MTRNR2L8		67.0	52.4-81.7
Combination_linear	IRF5-MTRNR2L8		68.6	54.1-83.1
Combination_All	circ_0000478-circ_0116639-IRF5-MTRNR2L8	51 (26 LS-OCMB-; 25 LS-OCMB+)	69.8	55.3-84.4
Combination 1	circ_0000478-IRF5		68.3	52.8-83.8
Combination 2	circ_0000478-MTRNR2L8		68.6	53.7-83.6
Combination 3	circ_0116639-IRF5		69.8	55.0-84.7
Combination 4	circ_0116639-MTRNR2L8		67.8	52.8-82.9
Combination 5	circ_0000478-circ_0116639-IRF5		69.8	55.0-84.6
Combination 6	circ_0000478-circ_0116639-MTRNR2L8		68.2	53.3-83.0
Combination 7	circ_0000478-IRF5-MTRNR2L8		68.8	53.7-83.8
Combination 8	circ_0116639-IRF5-MTRNR2L8		68.8	53.8-83.7

Discussion

In this work, we have characterized the circular and linear transcriptome of PBMCs from MS patients with distinct LS-OCMB status. This profiling and validation experiments have revealed that the global transcriptome of PBMCs from patients with positive LS-OCMBs is different from those patients with negative LS-OCMB status.

RNA-seq results reveal that there are 124 circRNAs and 2,441 linear RNAs differentially expressed between the two groups. Interestingly 58% of linear RNAs are detected only in one of the groups, highlighting that there is a specific expression pattern in PBMCs from patients with different LS-OCMB status. It is true, however, that our RNA-seq sample size is limited and these observations should be taken with care.

To the best of our knowledge, none of the circRNAs have been previously related with MS. Circ_0000478 is located in chromosome 13 and its host gene is von Willebrand factor A domain containing 8 (VWA8). Interestingly, the linear transcript of this gene is also among the differentially expressed linear RNAs in our dataset, having a fold-change of -3.23

(p=0.047). This transcript has been recently described and codes a mitochondrial protein with still uncertain function [299]. Regarding circ_0116639, it is located in chromosome 22, overlapping EP300 gene. EP300 is a histone acetyltransferase, which is detected in our study but is not differentially expressed. These differences in expression pattern of the circRNA with their host genes reveals that the biogenesis of circRNAs is independently regulated from that of their host genes, as it has been previously described [291].

Of note, interferon regulatory factor 5 (IRF5) is one of the confirmed downregulated transcripts, and polymorphisms in this gene has been associated to the risk of developing MS in last genome-wide association analysis performed by the International Multiple Sclerosis Genetics Consortium as well as in replication studies in Spanish cohorts [167,300]. Additionally, SNPs in this gene have been related to increased levels of CXCL13 in CSF, a chemokine related to highly active disease [301].

On the other hand, MT-RNR2 like 8 transcript (MTRNR2L8), the other validated linear transcript showing lower expression in LS-OCMB+ group is coded in chromosome 11, and interestingly, it spans the location of the miR-4485 stem-loop location, which has also been found to be downregulated in LS-OCMB+ patients. These results point at a possible interaction between these two RNAs that might be related to a more active disease, but further research is needed to confirm this hypothesis.

Gene overrepresentation test shows that biological processes related to complement activation, humoral immune response and type I interferon response are among the most enriched term. Interestingly, intrathecal synthesis of IgM antibodies has previously been correlated with intrathecal complement activation[302] and its role in demyelination and axonal damage has been clearly demonstrated[303]. In addition, the plasma levels of different components of the complement system have been proposed as MS disease state biomarkers [304]. Of note, these processes are enriched among altered genes in PBMC from patients with a marker of a high activity disease. This analysis may reveal that those patients may have an altered immune response that could explain their higher disease activity, but further and other kind of studies will be needed to explore this hypothesis.

The main aim of this study was to identify in blood a biomarker or a group of them that could differentiate between positive and negative LS-OCMB patients and thus, could be used as a more accessible marker of highly active disease. ROC curve analysis revealed that these markers in combination have around 70% of accuracy, which is considered in general an acceptable performance [305]. Taking into account that these are measured in blood, they may help in the clinics to perform repeated measurements, while taking serial samples of

CSF has serious drawbacks. Efforts are being made to define patients with a highly active disease course, given that their effective therapeutic window might be narrower than a more benign or less active disease form [306]. In line with this, several biomarkers have been proposed to identify individuals with aggressive disease course, such as CSF levels of neurofilament light chain, CXCL13 or CHI3L1 [307–309]. Other authors have also reported differences in non-coding transcriptome between patients with different disease activity. Quintana and colleagues studied the expression in CSF of a panel of miRNAs patients with MS with LS-OCMB characterization and patients with other neurological diseases (OND) [242]. They found differences in the expression of miR-30a-5p, miR-150, miR-645 and miR-191 between OND group and LS-OCMB+ group, but they could not find any difference between patients with positive and negative LS-OCMBs. Another study reported that serum levels of miR-24-3p correlate with disease progression index, calculated dividing EDSS value by disease duration, but the status of LS-OCMB is not measured [310]. More recently, the expression in blood of long non-coding RNAs has also been described to show the capacity of discriminating highly active MS patients from others with less active disease course, based on a age-related MS severity score [311].

All these studies highlight the need of biomarkers that are able to identify patients with highly active disease and the effort that scientific community is doing in this direction. In addition, is also important to have a consensus on the definition of what is an aggressive disease course, given that each study uses different parameters to measure and classify patients in each of the groups, which makes the comparison between them really challenging [312,313]. An early recognition of a patient in this group could benefit from different therapeutic advice and therefore, the possibility to improve the long-term outcome of the disease.

In summary, in this subchapter we have identified two circRNAs and two linear RNAs which are downregulated in patients with positive LS-OCMB status. Remarkably, an important advantage of these transcripts compared to other biomarkers is that they are detected in blood, allowing serial tests to monitor their levels, with almost no drawbacks. Future studies in larger cohorts will be needed to confirm their utility but we consider that they may serve as a first screening tool to decide which patients should go into a lumbar puncture to confirm their LS-OCMB status.

1.4 RNASeq profiling of circRNAs and linear RNAs in extracellular vesicles

Introduction

Extracellular vesicles (EVs) are membrane-coated particles of endosomal or plasma membrane origin that are secreted to the extracellular environment. Almost all cell types release EVs both in physiological and pathological conditions and they can be isolated from plasma and other body fluids. They play an essential role in indirect cell-to-cell communication as their membrane and cytosolic proteins, lipids and nucleic acids can be transferred between cells[314].

This EVs mediated communication has been shown to be implicated in the regulation of a number of biological functions including the immune response[315,316]. Moreover, they have been related to inflammatory and autoimmune diseases including MS[315,317]. Studies on EVs in MS patients have revealed a general change in their number and a different cargo when compared to controls, but also in relapse and in response to treatments [318–320]. In

this context, EVs and their cargo have gained attention as potential biomarkers in MS in the last years.

Recently, circRNAs have been discovered to be enriched in EVs[87,132]. Exosomal circRNAs have been detected in body fluids from different pathological conditions and have been found to play a role in processes such as drug resistance, metastasis or cell proliferation[134]. Most of the studies have been focused in exosomal circRNAs in cancer, however still few studies have investigated the role of circRNAs in EVs in other diseases.

In this subchapter we will characterized the transcriptomic profile of EVs in MS, including for the first time the characterization of circRNAs.

Materials and methods

Blood sampling

For the study of EV transcriptome, whole blood was obtained from a total of 10 MS patients classified as RR-MS patients, 10 SP-MS patients, and 8 sex and age matched healthy controls (HC) in the Department of Neurology at the Donostia University Hospital. All donors provided written informed consent prior to sample extraction. The main clinical and demographical characteristics of both patients and healthy donors are summarised in Table 7. The study was approved by the hospital's ethics committee (UEM-IMN-2017-01) and part of the samples have been processed and stored at the Basque Biobank (www.biobancovasco.org).

Table 7. Main clinical and demograpgical characteristics of the individuals enrolled in the study classified by disease status.

	Sex	Age	EDSS	Evol. Time	AOO
RR-MS (n=10)	Male (n=5)	39.7 (±13.7)	2 (0-3.5)	16.1 (±9.8)	23.6 (±9.1)
	Female (n=5)	42.0 (±19.8)	0 (0-2)	13.2 (±12.2)	28.8 (±9.9)
SP-MS (n=10)	Male (n=5)	54.8 (±6.2)	6 (6-8.5)	20.7 (±8.7)	34.1 (±9.5)
	Female (n=5)	48.7 (±6.7)	7 (4-8)	22.5 (±5.0)	26.2 (±6.6)
HC (n=8)	Male (n=4)	56.1 (±11.9)	-	-	-
	Female (n=4)	53.6 (±11.8)	-	-	-

*Abreviations: RR-MS, relapsing-remitting multiple sclerosis; SP-MS, secondary-progressive multiple sclerosis; HC, healthy control; EDSS, Expanded Disability Status Scale; Evol.Time: Evolution time; AOO, Age of Onset. Age, evolution time and AOO data are presented as 'average (standard deviation)', EDSS data are shown as 'median (range)'

Peripheral blood was colleted by venipuncture in EDTA tubes (Vacutainer, BD Biosciences) and centrifuged at 1258g for 20 min to separate the plasma from the cellular fraction. Plasma was carefully collected, aliquoted and stored at -80^{o}C.

EV isolation and RNA extraction

EVs were isolated following a differential centrifugation step protocol as previously described by our group[321]. Briefly, plasma aliquots were centrifuged at 13,000 g for 2 min, supernatant was transferred to a new tube and centrifuged again at 20,000 g for 20 min to pellet EVs. The pellet was then resuspended with 100µl of DPBS (GIBCO, ThermoFischer). The DPBS had previously been double filtered through a 0.22µm-pore filter in order to remove any other particle. RNA was directly isolated from the EV samples by using Trizol LS (ThermoFisher) following manufacturers' protocol. RNA was quantified by NanoDrop ND-1000 spectrophotometer.

Nanoparticle tracking analysis (NTA)

The size distribution and concentration of isolated plasma EVs was measured using a NanoSight LM10 device (Malvern) as described elsewhere[322]. Samples were diluted to appropriated levels to get accurate acquisitions. Camera settings were fixed and maintained for all samples. Filtered DPBS was tested and no background signal was detected. For each sample, two videos of 1 min were recorded and analysed with NanoSight NTA software 2.2 (Malvern).

Cryo-electron microscopy (Cryo-EM)

For visualization of the isolated EVs, samples were vitrified following standard protocols[323]. Glow-discharged Quantifoil holey carbon film grids (Orthogonal Array of 2µm Diameter Holes - 2µm Separation, mounted on a 300M Cu grid, #657-300-CU, Ted Pella) were vitrified in liquid ethane in Vitrobot (FEI) after deposition of 3 µL of sample. Cryo-transfer sample holders of the type GATAN Model 626 were used to keep the sample vitrified during electron microscopy analysis. The sample was observed in a JEM-2100F UHR (80-200kV, JEOL, Ltd.) field emission gun (FEG) transmission electron microscope at different magnifications. Micrographs were recorded on TVIPS F216 CMOS camera (2k x 2k).

RNASeq

For library preparation and sequencing, a minimum input of 2µg RNA is required. In order to fulfil this criterium RNA samples were pooled in pairs or trios of samples with the same disease status and sex. A total of 12 pools were obtained, 4 RR-MS pools, 4 SP-MS pools and 4 HC pools, the extended clinical and demographical information for each pool was included in Table 8.

The concentration and quality of the RNA pools was measured using Bioanalyzer 2100 instrument (Agilent) before library preparation at CD Genomics (USA). After normalization, rRNA was depleted from the total RNA sample using the Ribo-Zero rRNA removal kit and followed by purification and fragmentation steps. To construct the sequencing libraries, a strand-specific cDNA synthesis was performed, the 3' ends were adenylated and adaptors were ligated. The resulting libraries were subjected to quality control and normalization process. Paired-end sequencing was performed with Illumina HiSeq Ten X and an average of 10Gb of data were obtained per sample.

Table 8. Main clinical and demograpgical characteristics of the individuals enrolled in the study classified by pools.

	Status	Sex	Age	EDSS	Evol. Time	AOO
Pool1 (n=3)	RR-MS	Male	43 (±7.1)	0 (0-3.5)	15 (±2.9)	29 (±7.2)
Pool2 (n=2)	RR-MS	Male	34 (±23.3)	2.5 (2-3)	19 (±18.7)	16 (±4.6)
Pool3 (n=3)	RR-MS	Female	39 (±23.7)	0 (0-2)	11 (±10.7)	28 (±13.9)
Pool4 (n=2)	RR-MS	Female	46 (±19.6)	1 (0-2)	16 (±18.4)	30 (±1.2)
Pool5 (n=3)	SP-MS	Male	55 (±7.2)	6.5 (6-8.5)	20 (±4.1)	36 (±11.3)
Pool6 (n=2)	SP-MS	Male	54 (±6.8)	6 (6-6)	22 (±16.1)	32 (±9.3)
Pool7 (n=3)	SP-MS	Female	45 (±2.7)	7 (6.5-7.5)	21 (±6.1)	24 (±7)
Pool8 (n=2)	SP-MS	Female	54 (±7.7)	6 (4-8)	25 (±0.8)	29 (±6.9)
Pool9 (n=2)	HC	Male	62 (±15.5)			
Pool10 (n=2)	HC	Male	50 (±7.4)			
Pool11 (n=2)	HC	Female	60 (±14.7)			
Pool12 (n=2)	HC	Female	47 (±4.7)			

*Abreviations: RR-MS, relapsing-remitting multiple sclerosis; SP-MS, secondary-progressive multiple sclerosis; HC, healthy control; EDSS, Expanded Disability Status Scale; Evol.Time: Evolution time; AOO, Age of Onset. Age, evolution time and AOO data are presented as 'average (standard deviation)', EDSS data are shown as 'median (range)'

CircRNA detection and quantification in RNA-seq data

Sequencing reads were quality checked and mapped to the hg19 using BWA[278] or Bowtie[279]. Subsequently, circRNA prediction was performed by find_circ, version 1.0[80], and CIRI2[280] adhering to the recommendation by the authors. For find_circ[80], an

increased stringency threshold was used requiring that both adaptor sequences map with highest possible mapping quality (mapq = 40). Moreover, only circRNAs supported by at least two reads in a given sample and found by both algorithms were used in subsequent analyses. CircRNA expression was based on back-spliced junction (BSJ)-spanning reads according to CIRI2 quantification. Differential expression analysis was performed using DESeq2[281] in R-studio and specifying a design formula (~condition + sex) that controls the effect of the sex while contrasting the different conditions (RR-MS vs HC, SP-MS vs HC or SP-MS vs RR-MS).

CircRNAs detected in at least one pool for each of the groups contrasted, with an absolute fold-change (FC) value higher than 1,5 (FC >|1.5|) and a p-value less than 0.05 ($p<0.05$) were considered differentially expressed circRNAs (from now on, DE circRNAs). CircRNAs detected in at least one pool for a given group and absent in all the pools comprising the other group included in the contrast were considered as group-exclusive circRNAs.

Linear transcript detection and quantification in RNA-seq data

After the quality check, the sequencing reads were mapped to the hg19 with STAR[282] using default parameters. HTSeq[283] was used to quantify expression of genecode annotated genes (gencode v28) in a strand-specific manner. Read counts were then used for gene-level differential expression analyses using DESeq2 in R-studio and, as defined for circRNAs, differentially expressed linear RNAs (DE linear RNAs) were defined as those detected in at least one pool for each of the groups contrasted, with a FC >|1.5| and p-value<0.05. Linear RNAs detected in at least one pool for a given group and absent in all the pools comprising the other group included in the contrast were considered as group-exclusive transcripts.

Classification of transcript types

CircRNAs were defined as those detected with at least 2 reads by CIRI2 and find_circ. For linear transcripts, gene type information was obtained from Biomart database. For transcripts that were not classified in Biomart, gene type information was manually completed based on GeneCards, Ensembl and Lncipedia. Twelve transcript type categories were finaly defined: protein coding transcript, long non-coding RNA (lncRNA), pre-microRNA (pre-miRNA), miscellaneous RNA (such as RNAs involved in important ribonucleoprotein complexes implicated in the transcription and translation processes), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), small Cajal body-specific RNA

(scaRNA), ribosomic RNA (rRNA), transfer RNA (tRNA), circRNA, other RNAs (including pseudogenes, rybozimes, mitochondrial RNAs etc.) and non defined RNAs.

For each transcript type the mean number of reads was calculated for each condition (RR-MS, SP-MS and HC) by calculating the mean for the raw readcount of the samples comprising each group.

Results

Characterization of plasma isolated EVs

NTA results shows a particle size distribution that confirms efficient isolation of small EVs and removal of larger particles such as plasma platetets, with most vesicles ranging between 50 and 300nm (Figure 48). Moreover, the presence of EVs and their mean size, as well as their rounded shape could be confirmed by cryo-EM (Figure 48).

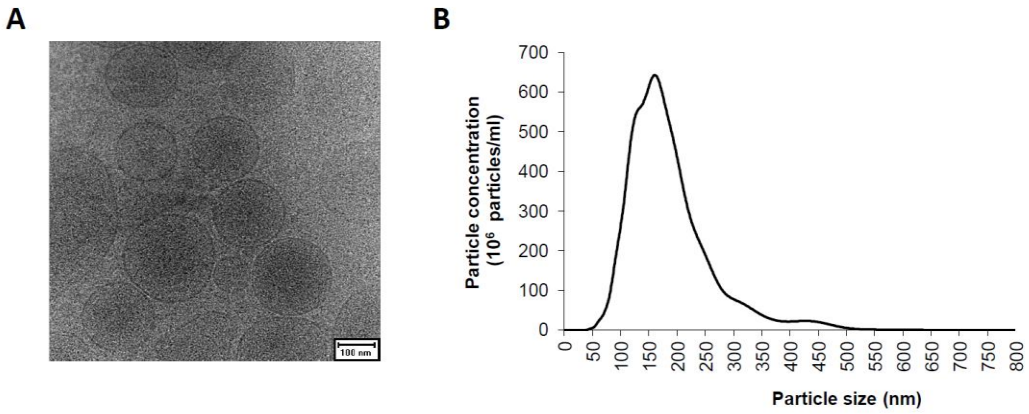

Figure 48. Plasma derived EV characterization. (A) Representative cryoEM image of EVs isolated following a differential centrifugation step protocol. (B) Representative figure of particle size distribution of EVs obtained by NTA.

CircRNAs are abundantly detected in plasma derived EVs.

For the RNA profiling of EVs from MS patients and HC, we performed a RNA-seq of rRNA depleted total RNA from plasma derived EVs of 20 MS patients and 8 HCs divided in 4 RR-MS pools, 4 SP-MS pools and 4 HC pools. Find_circ and CIRI2 detected a total of 10,906 and 17,542 unique circRNAs supported by at least two BSJ spanning reads, respectively. A good detection overlap was observed between both algorithms with 6,575 circRNAs in common which were classified as *bona fide* circRNAs (Figure 49A).

The 60.7% of the circRNAs were modestly expressed, with a mean of less than 10 reads, whereas the 39,3% of the *bona fide* circRNAs are quite abundant (Base Mean ≥ 10) out of which 819 circRNAs could be considered as highly abundant with more than 50 read counts (Figure 49B). Among those there is a group of 72 circRNAs that stands out with more than 500 reads. The distribution of the BSJ mapping read counts from all the samples after normalization shows that all the samples have a similar distribution of the BSJ read numbers (Figure 49C).

The number of detected circRNAs was in general homogenously distributed along the automosomes and consistent with the chromosome length. As an exception, both sexual chromosomes account a fewer circRNAs than expected based on their size. Chromosome X has a size that places it between chromosomes 7 and 8, however it gives rise to only 142 circRNAs while chromosomes 7 and 8 produce 310 and 267 circRNAs, respectively. Similarly, chromosome Y is the one producing less circRNAs with a total of 5 whereas 135 circRNAs are produced from a chromosome with a similar length such as chromosome 19 (Figure 49D).

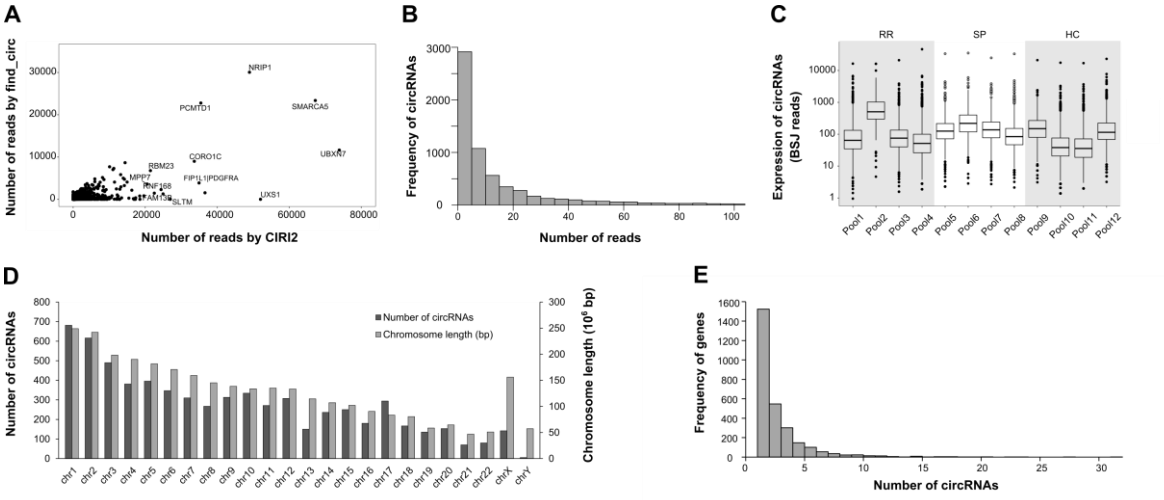

Figure 49. General features of the circRNA profile in EVs. (A) Correlation of the quantification performed by find_circ and CIRI2. (B) Histogram showing the number of reads per circRNA (For clarity, read number has only been ploted up to 100 read counts). (C) Overall circRNA expression for each sample based on the mean number of reads spanning a BSJ. (D) CircRNA distribution in chromosomes compared with the chromosome length. (E) Histogram showing the number of circRNAs produced from each gene.

Out of the 6,575 *bona fide* circRNAs detected, 6,369 are located in a known gene locus, while 206 circRNAs are intergenic and have not been assigned to a gene. Among the 6,575 circRNAs detected in the plasma derived EVs, only 2,814 different genes have been identified, revealing that some genes produce more than one circular transcript. 1,523

(54.1%) of those genes give raise a single detectable circRNA transcript, while 1,247 (44.3%) produce between 2 and 10 circRNAs. There is a small subset of genes (1.5%) from which more than 10 different circRNAs are generated and incorporated into EVs, with a maximum of 31 circRNAs produced from the gene coding for the serine/threonine-protein kinase TAO1 (TAOK1). Interestingly TAOK1 is located in the chromosome 17, which presents the highest density of circRNAs/bp (Figure 49D and E).

Linear transcripts are also detected in EVs.

Apart from the circRNA profiling, the linear transcriptome profile in EVs was also studied. 18,372 linear RNAs were detected using HTSeq.

Similar to circRNAs, the 65.6% of the linear RNAs were expressed with a mean of less than 10 reads, whereas the 34.4% can be considered as quite abundant (Base Mean ≥ 10). Among those, it is worth noting that 3,214 linear transcripts (17.5%) were detected with a Base Mean ≥ 50 and the 4.1% are highly abundant transcripts that account more than 500 reads (Figure 50A). The normalized BSJ mapping read counts were displayed for the different pools showing that all of them are similar with no outlier pools (Figure 50B).

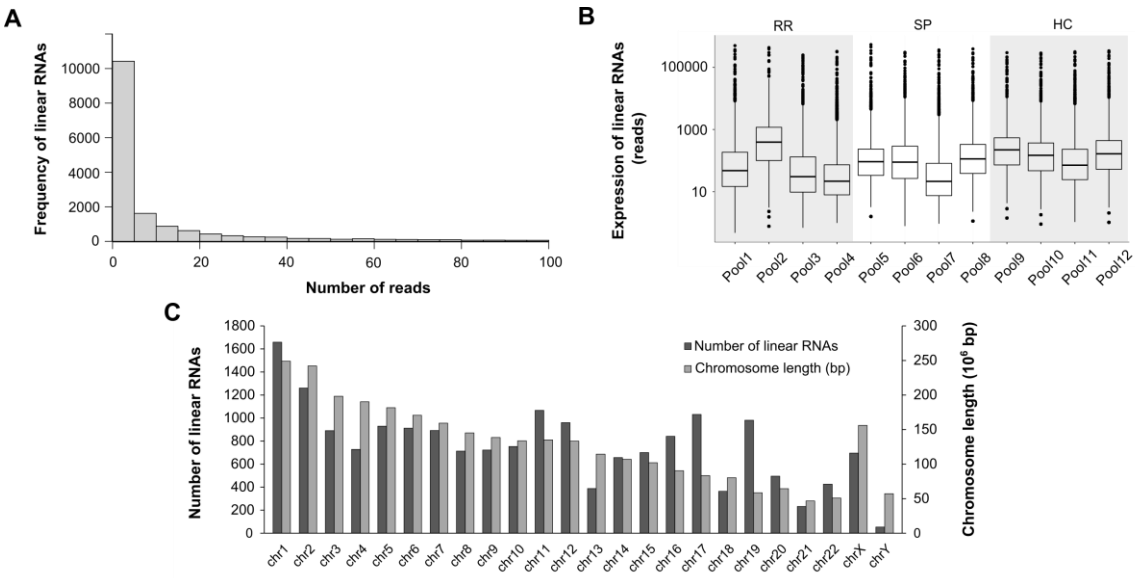

Figure 50. General features of the linear RNA profile in EVs. (A) Histogram showing the number of reads per linear RNA (For clarity, read number has only been ploted up to 100 read counts). (B) Overall linear RNA expression for each sample based on the mean readcount for each transcript. (C) CircRNA distribution in chromosomes compared with the chromosome length.

Each chromosome produces approximately twice as much linear RNAs as circRNAs, but generally it is also well correlated with the size of the chromosome except for the sexual

chromosomes. Among the autosomes, four of them are also worth mentioning for having a more active transcriptional activity. Chromosomes 10, 11 and 12 have very similar lengths, however, EVs contain 1,4 and 1,3 times more linear RNAs from chromosomes 11 and 12 produce when compared to chromosome 10. In a similar maner, EVs contain more linear RNAs from chromosomes 16, 17 and 19 expected for their sizes (Figure 50C).

CircRNAs are the second more abundant RNA transcript in EVs but not in leukocytes

Taking together all the linear and circular transcripts we have analized more than seven million reads over 15,000 different transcripts per sample pool. In order to characterize the complete transcriptome in EVs we categorized those transcripts in 12 different types as defined in materials and methods. All the 12 transcript types were present in RR-MS and SP-MS patients samples as well as in controls. The three groups have a very similar distribution of the transcript types even if SP-MS samples have a lower number of reads. Obtained data show that protein coding transcripts are the most prevalent in plasma-derived EVs acounting for more than the 90% of the total reads per sample (Figure 51A). Among the non-protein coding RNAs (Figure 51B), circRNAs are the most prevalent type of transcripts and represent the 5.6% and 3.9% of the RNAs for RR-MS and SP-MS patients respectively whereas the 2,9% of the RNAs in HC samples are circRNAs. The rest of the transcripts that could be categorized, including lncRNAs, pre-miRNAs and snoRNAs among others, represent less than the 1% respectively.

Interestingly, when the transcriptomic profile found for EVs was compared to the one found for leukocyte samples several differences were revealed. First, the mean number of reads in leukocytes is $19.7 \cdot 10^6$ reads, higher than in EVs with a mean of $6.6 \cdot 10^6$ per sample (almost 3 times higher). In both cases protein coding transcripts are the most prevalent but they represent a smaller proportion in leukocytes (76.6% in leukocytes compared to 92.9% in EVs)(Figure 51C). In contrast with what we found for EVs, circRNAs are not the second most prevalent RNA type in leukocytes were they only account for the 0.2% of the reads. In leukocytes miscellaneous RNAs are the ones that are more abundant among the non-protein coding transcripts (13.2%) a surprisingly high proportion that is not reflected in the EVs where they only account for the 0.1% of the reads (Figure 51D) Among those miscellaneous RNAs, it is worth noting the presence of 7SK RNA that plays and important cellular role by regulating the transcription elongation factor P-TEFb and 7SL RNA or signal recognition particle RNA that recognizes the signal peptide in the newly sintetized proteins and binds to the ribosome stoping the protein synthesis.

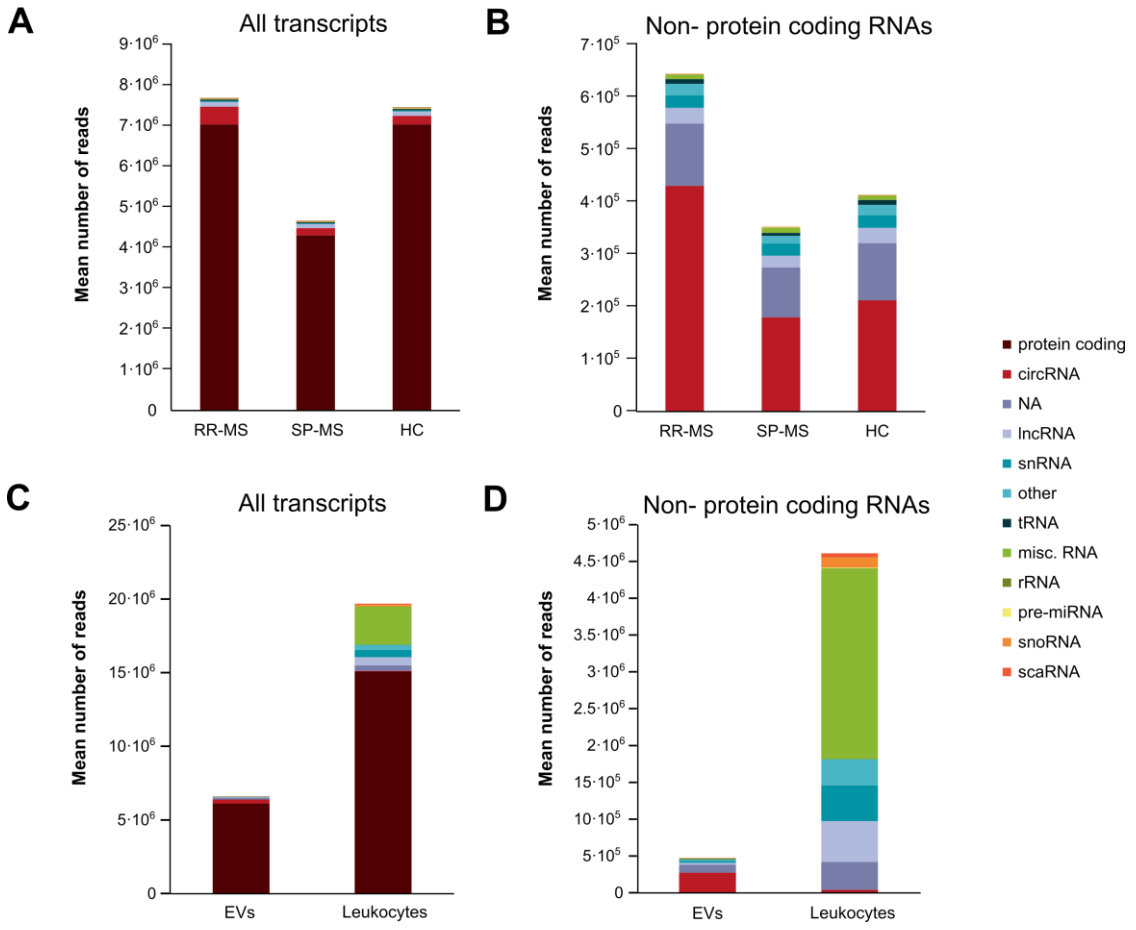

Figure 51. Distribution of the different RNA types in the EV transcriptome profile of MS patients and controls. (A) All transcript types are represented by their mean number of reads. (B) For clarity, the fraction corresponding to protein coding genes, which are the most abundant, is omitted so that the distribution of the non-protein coding RNAs can be appreciated.

Moreover, apart from the change in the relative abundance of circRNAs between EVs and leukocytes, the size of the circRNAs present in each of the samples types was also different (p-value=0.002). The median size for EV circRNAs is of 386bp while in leukocytes circRNAs are slightly larger with a median size of 476bp. In fact the length of the 65% of the circRNAs in EVs is <500bp while for leukocytes they represent the 53%. Nevertheless, very large circRNAs are found in both samples with circRNAs up to 90,000bp (Figure 52).

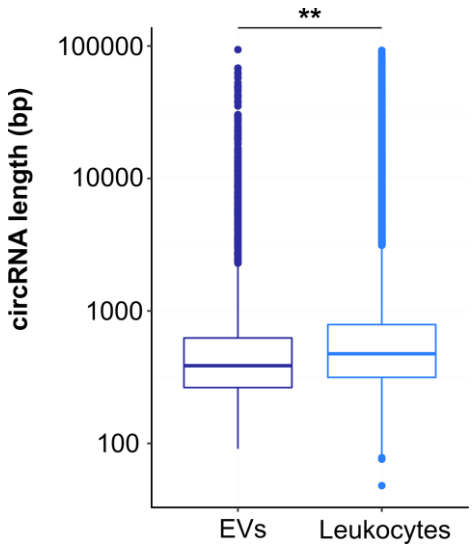

Figure 52. CircRNA size distribution for EVs and Leukocytes.

The circRNA profile in EVs from different MS types patients is different to the profile from controls.

With the hypothesis that patients diagnosed from RR-MS and SP-MS could present a different EV transcriptomic profile, we performed three different contrasts: RR-MS vs HC, SP-MS vs HC and SP-MS vs RR-MS.

When we compared the circRNA profile between RR-MS patients and HC, we found that among the 6,575 *bona fide* circRNAs detected in total, only 2,942 circRNAs (48.3%) have been detected in RR-MS and HC simultaneously. The remaining 3,144 of circRNAs comprise a group specific profile. 1,731 circRNAs (28.4%) were exclusively detected in RR-MS samples whereas 1,413 circRNAs (23.2%) were exclusive to HCs (Figure 53A). We performed a differential expression analysis including all the group-specific circRNAs as well as those detected in both groups. However, taking together the fact that many of the circRNAs have been abundantly detected in EVs and the fact that a high percentage of circRNAs are unique to one group, the differential expression analysis has reported a high number of circRNAs with extremely high FCs that could have been overestimated. Therefore, we decided to take into consideration only the FC and p-values calculated for circRNAs detected in at least one pool of each group. Among the 2,942 common circRNAs between RR-MS patients and HCs, 100 circRNAs were found to be differentially expressed($FC>|1.5|$ and a $p\text{-value}<0.05$), with 47 of them upregulated and 53 downregulated (Figure 53A).

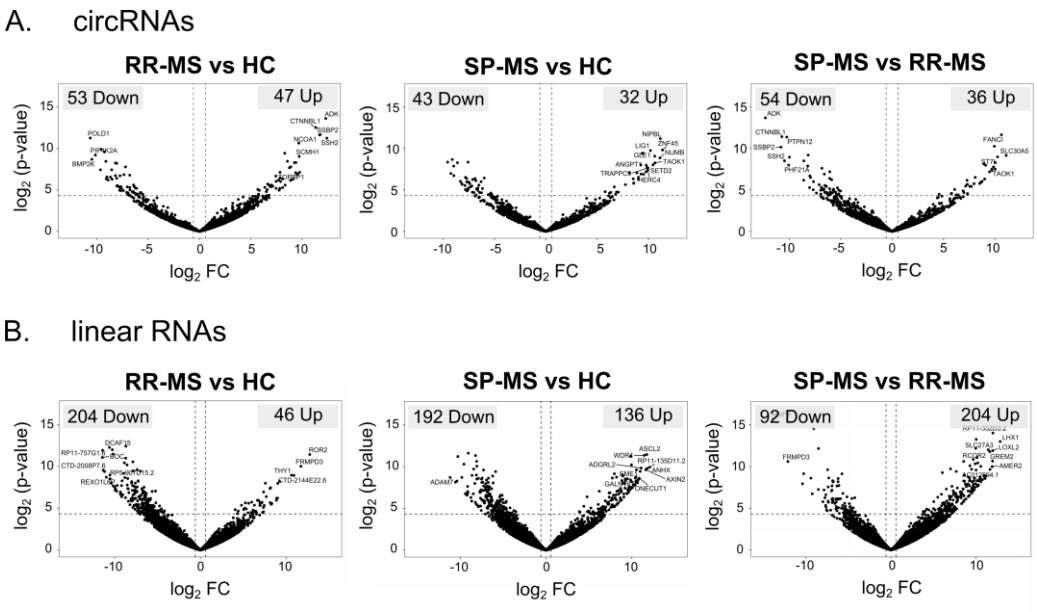

Figure 53. Volcano plots showing the differentially expressed circRNAs (A) and linear RNAs (B) in the different contrasts. Transcripts detected in at least one pool for each of the group have been depicted. The number of downregulated and upregulated transcripts (FC≥|1.5| and p-value<0.05) have been indicated in the upper-left and upper-rigth corner of the volcano plots respectively.

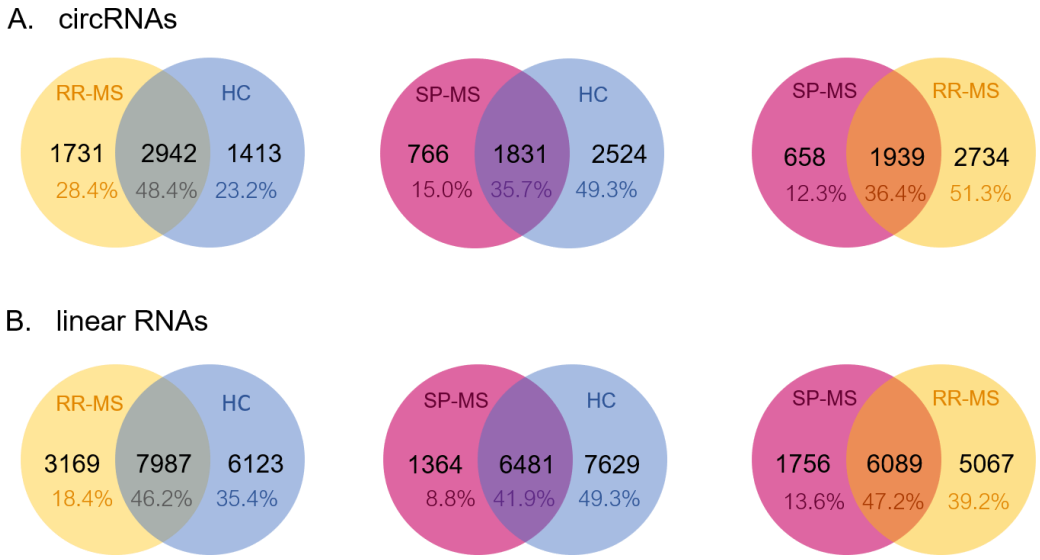

Figure 54. Venn diagrams showing the circRNAs and linear RNAs that have been detected in the different groups.

The comparison between SP-MS patients and HC reported slightly different results. Out of the 5,121 *bona fide* circRNAs that were detected in at least one of the two groups, only 1,831 circRNAs were found in common between SP-MS patients and HCs accounting for only the 35.7% of circRNAs. Among those, 32 circRNAs were found to be upregulated in SP-MS patients and 43 downregulated (Figure 53A). More interestingly, in this comparison, the biggest proportion of circRNAs, accounting for the 49.3% (2,524 circRNAs), were only detected in HCs. Besides, the SP-MS specific profile included 766 circRNAs (15%) (Figure 54A).

Results drawn for these two comparisons indicate that there is a different profile of circRNAs in plasma derived EVs from MS patients when compared to controls. In order to asses whether EV transcriptome profiles from SP-MS and RR-MS patients are also different or not, we also compared these two groups. Among the 5,331 *bona fide* circRNAs present in at least one of the two MS types, 1,939 were common (36.4%) to both, while 2,734 were exclusive to those diagnosed from RR-MS and 658 circRNAs were only found in SP-MS patients (Figure 54A). Apart from the MS type-specific circRNAs, the expression levels of some of the common circRNAs could also distinguish between the two MS types, in fact 36 circRNAs were found significantly upregulated in SP-MS patients, and 54 downregulated (Figure 53A).

The linear transcriptome is also changed in EVs between different MS types and controls.

In the same way that the circular transcriptome is different when compared EVs from RR-MS, SPMS and HC individuals, the linear transcriptome was also found to be different.

From the 18,372 linear RNAsdetected by Kallisto, 17,279 are present in at least one of the 8 pools that comprise the RR-MS and HC groups. Out of them, the 7,987 linear RNAs (46.2%) were detected in both groups, and 250 showed a differentially expressed pattern. Interestingly, most of the DE linear RNAs were downregulated (204 linear RNAs, 81.6%) and only 46 were found to be upregulated in RR-MS patients (Figure 53B). Accordingly, we also found that there are more linear RNAs exclusively detected in HC (6,123 linear RNAs) than exclusively detected in RR-MS patients (3,169 linear RNAs) (Figure 54B).

In the case of SP-MS patients, when compared to HCs, we found that their profiles are more different. The number of common linear RNAs is of 6,481, among which 136 transcripts are found to be upregulated in SP-MS patients and 192 downregulated (Figure 53B). Interestingly, the proportion of downregulated linear RNAs is not as high as it was for RR-

MS (58.5% linear RNAs downregulated between SP-MS and HC compared to the 81.6% for RR-MS vs HC). The biggest difference was found among the transcripts that are present in one of the groups and absent in the other since the 49.3% of the linear RNAs included in the comparison were only detected in HC EVs (7,629 transcripts) whereas only the 8.8% is SP-MS specific (1,364 transcripts)(Figure 54 B).

When SP-MS and RR-MS are compared, similar to what it was observed for the circular transcriptome, only 6,089 linear transcripts (47.1%) are found in samples from both MS types. Besides, 5,067 linear RNAs are unique to RR-MS patients (39.2%) and the 1,756 linear transcripts (13.6%) can only be found in EVs from SP-MS patients (Figure 54B). Out of the common transcripts 204 resulted to be upregulated and 92 downregulated in SP-MS patients (Figure 53B).

Discussion

The global EV transcriptome profile in MS was characterized back in 2017 by Selmaj et al. where they used RNASeq to study the expression of several classes of transcripts including small RNAs and with a particular interest in miRNAs. This study revealed that exosomes have a distinct RNA profile in RR-MS patients and suggested that miRNAs might represent a biomarker for RR-MS relapses[319]. Nevertheless, they focused the research on linear RNAs and did not study the circular transcriptome. Additionaly, their target EV type are mainly exosomes, which are usually smaller than the fraction of of EV isolated with our serial centrifugation protocol.

In this study, we present a genome wide characterization of the transcriptome in EVs, including for the first time the circular transcriptome. In our case, we did not make a particular library preparation to include small RNAs, but we still could detect some pre-miRNAs and snoRNAs (mainly ranging from 60 to 120 bps and from 80 to 200 bps respectively). In spite of the technical differences between studies, it could be confimed that protein-coding genes comprise the most prevalent type of transcripts in EVs, while miRNAs and in our case pre-miRNAs are quite scarce. In contrast, we have found that circRNAs are abundant in EVs, which has also been reported by other studies[87,132], and represent the second most abundant transcript in EVs. Moreover, it is important to metion that circRNA abundance is probably underestimated due to the fact that only reads spanning the BSJ are unique to circRNAs and thus, only those are accounted as circRNA mapping reads. However, there must be many other reads mapping to the rest of the circRNA that have been assigned

to some other transcript derived from the same gene, which in most of the cases corresponds to a protein coding RNA potentially leading to an overestimation of this type of transcripts.

Additionally, Li et al. had already reported that circRNAs were enriched in exosomes compared to the producer cells in cell culture[132]. When making these comparisons for plasma derived EVs, it is important to take into account that different cell types may export their EVs to the bloodstreams and thus, plasma EVs comprise a complex mixture of EVs from different origins. Nevertheless, a recently published study analyzing 101 normal plasma RNA-Seq results has concluded that only the 0.2% of plasma EVs were derived from tissues, with 99.8% of them generated from hematopoietic cells and the 45% particularly from leukocytes[324]. Thus, we can assume that a big proportion of the analyzed EVs are from leukocyte origin and we will discuss the differences between EV and leukocyte transcriptomic profiles as if they were the main producer cells although conclusions drawn from this comparison should be taken with caution.

In our study, circRNAs account for the 4.2% of the total reads in EVs while they only represent the 0.2% in leukocytes, which is a difference worth noting. This result is the first evidence pointing that far from resembling the cytoplasmic content in leukocytes, EVs are actively loaded with a proportionally high number of circRNAs. Further investigation of the function of these circRNAs could help to understand if they are ment to carry some particular information from one cell to other or if they are incorporated into EVs as a mechanism of clearance[87]. In line with the first hypothesis indicating that circRNAs are not incorporated into EVs at random, it is interesting to highlight that a global upregulation of circRNAs has been described for MS leukocytes, which is not observed in EVs from the same type of MS patients. This suggests that the high expression of circRNAs may be playing an important role in the pathological context of MS leukocytes, but that there is no need of exporting them into EVs. For instance, and extra accumulation of circRNAs in MS leukocytes could be implicated in the regulation of the aberrant immune response by interacting with the intracellular sensor PKR. However, the extensive export of these circRNAs into EVs would reduce the intracellular pool of circRNAs which could compromise the regulation of the immune response. Therefore, we suggest that the load of circRNAs from leukocytes to EVs is actively regulated in order to maintain the intracellular circRNA pool.

Apart from the difference in circRNA abundance between EVs and leukocytes there is also a difference in terms of the circRNA size. Based on the results there is no physical limit for incorporating large circRNAs into EVs, however, there is a bias to incorporating smaller

circRNAs which again indicated that circRNAs are selectively released in EVs. A previous study by Preuβer et al. also reported that size could be an important determinant for selective vesicle export of circRNAs mainly considering that the available packaging volume within a EV might be limiting to accommodate not only the circRNAs but also the proteins associated with them[133].

Beyond characterizing the general transcriptome profile of EVs in comparison to that from leukocytes, we have also characterized the differences between EVs from healthy donors and MS patients. The release of a different number of EVs or with a different cargo in pathological conditions have been proposed to contribute to the development of diseases. In fact, EVs play a role in various stages of the immune response and have been related to inflammatory and autoimmune diseases [315,317]. In the particular case of MS, both changes in their concentration and on their cargo have been proposed as biomarkers of the disease although further research is needed[318,319]. In this subchapter, and in agreement with what Selmaj et al. have previously reported[319] we have also found differences in the RNA cargo of EVs between MS patients and healthy controls as well as between RR-MS and SP-MS patients. Both circular and linear transcripts differentially expressed between MS patients and healthy controls as well as transcripts that are unique to the diseased conditions could potentially be used as non-invasive biomarkers of the disease. However, they should first be validated in a bigger cohort of EV samples.

Interestingly, some of the circRNAs that have been proposed as biomarkers in leukocytes have also been detected in some of the EV samples. However, only circNEIL3 has been found to be also deregulated in this type of sample. Whereas it was upregulated in leukocytes, it is 7.86 times downregulated in RR-MS patients' EVs when compared to controls' EVs ($p=0.008$). This result is in agreement with the idea of an actively regulated release of circRNAs that in this case is favouring the accumulation of circNEIL3 in leukocytes in the pathological condition. It is also worth noting that in leukocytes, a very small proportion of the circRNA profile was changed between RR-MS and SP-MS patients whereas this difference is more notable between EV transcriptomic profiles indicating that the EV transcriptomic profile could complement the information obtained from leukocyte circRNAs. Lastly, it is also interesting to note that the second pool of EV samples corresponding to male RR-MS patients, has a higher level of expression of circular and linear RNAs compared to the rest of the RR-MS pools. This pool, also corresponds with patients that debuted with the disease at an earlier age, and do also accumulate a higher degree of disability, this is, they are patients with a more aggressive MS course. In light of these results, it would be interesting to further investigate the EV transcriptome profile, in order

to asses if it is affected by the severity phenotype, which could also lead to promising prognosis biomarkers.

To the best of our knowledge, this study is the first to report the presence of 6,575 circRNAs in plasma derived EVs from MS and HC individuals. Based on the results drawn from the comparison of the transcriptomic profile between EVs and leukocytes, we suggest that the loading of circRNAs into EVs is regulated in order to maintain an equilibrium of the intracellular circRNA pool. Thus, the selective release of circRNAs into EVs could be implicated in the regulation of the physiopathology of the disease, although further research is needed.

Moreover, we report a number of linear and circular RNAs that are differentially enriched between MS and HC controls as well as between RR-MS and SP-MS patients which should be further studied in order to evaluate their role as biomarkers in MS.

CHAPTER 2

Approximation to the function of MS-related circRNAs

General introduction

CircRNAs have been found to be conserved across species[74,79] and endogenously produced in many different tissues[80,82]. Moreover, their expression has been found to be changed in different developmental stages, cellular processes and diseases suggesting their importance[26,97]. Nevertheless, the molecular and cellular roles of the vast majority of circRNAs are still unknown[103,106].

Their cellular localization can be used as an indicator of their potential function: circRNAs retained in the nucleus, such as ciRNAs and EIcircRNAs, are thought to participate in regulation of the transcription, whereas exonic circRNAs are mostly cytoplasmic where they may participate in post-transcriptional regulation by interacting with different RNAs and proteins or as templates for translation[103,105,137].

A growing number of studies have associated different circRNAs to pathological processes and diseases. In many cases, the molecular function that makes them be involved in the pathology remains unknown. But in several other diseases, a circRNA mechanism of action has been proposed. The sponging of a miRNA, the sequestering or interaction with particular proteins, or the peptide synthesized by a circRNA have been identified as the causative factors of different diseases[25,26,150,325–327].

In this chapter we will first try to unravel the molecular functions that the circRNAs involved in MS may be playing by a *in silico* screening of the most interesting circRNA functions. And secondly, we will perform a first approach in order to understand how circRNAs can cross-talk with the immune response.

2.1 Functional screening of circRNAs implicated in Multiple sclerosis

Introduction

Through all the first chapter we have characterized the circRNA profile in different MS type patients and prognostics as well as in different samples types (leukocytes, PBMCs and EVs) revealing a number of circRNAs that may be implicated in the disease. We have also assessed the biomarker potential of some of them, however we have very few clues of their role in the disease.

In order to unravel some of the functions that these circRNAs may be playing in the disease, in this subchapter we will screen the most interesting circRNAs reported from chapter one looking for some of the features that have been linked to the best characterized circRNA molecular functions.

MiRNA sponging was the first function described for a circRNA when Hansen et al. found that ciRS-7 contains more than 70 conserved binding sites for miR-7 and that it effectively sequesters them mitigating the repression of the corresponding mRNA targets[90]. Later, some other additional examples have been reported[137,138]. However, the common feature

to those miRNA sponges is a considerable number of miRNA binding sites which only happens to limited number of circRNAs.

The second most studied cytoplasmic circRNA function is their protein coding potential. Legnini et al. reported the first case of the endogenous human circZNF609 which was found to code for a protein in myoblasts[328]. But again, very few circRNAs have been firmly confirmed to be translated endogenously[144,145]. Moreover, and in contrast with the miRNA sponge function, there is still not a consensus of a feature or a group of features for translatability. Several different criteria have been used for assessing the coding potential of a circRNA, such as the presence of ribosome footprints, mass spectrometry-based detection of peptides translated across the BSJ or presence of N6-methyladenosine modifications that have been described to promote translation of circRNAs[89,144,329], with contradictory results.

Lastly, and although it is still not a widely studied function of circRNAs, in light of the recently published study by Liu et al.[118] and its importance in immunity, we wanted to investigate the potential of our circRNAs of interest to interact with the double-stranded RNA-activated protein kinase (PKR). It has been described that the secondary structure formed by some circRNAs is the feature that enables their interaction with PKR in order to inhibit its activation.

Taking all of this into account, in this subchapter we will assess the number of miRNA binding sites and the secondary structure of the circRNAs that we have detected by RNASeq in leukocytes and EVs and particularly of the most interesting circRNA candidates reported from chapter one.

Materials and Methods

Selection of the most interesting MS related circRNAs

In order to perform a wide screening of the functions for circRNAs implicated in MS, we decided to make a selection of: 1) the 22,835 *bona fide* circRNAs detected by RNASeq in leukocyte samples on the experiment presented in subchapter 1.2 and 2)the dataset of 6,575 *bona fide* circRNAs generated in the RNA-Seq experiment performed in EV samples in subchapter 1.4 as a representative selection of the circRNA populations in cells and EVs.

On the other hand, we also studied the potential functions of the circRNAs reported as the most promising candidates in each of the experiments shown in Table 9.

Table 9. Table gathering the most interesting circRNAs in MS reported from chapter one. CircRNAs whose differential expression has been validated have been included and both their ID with which they have been identified in the corresponding subchapter and genomic position are shown.

Experiment (Subchapter)	circRNA candidate ID	Genomic location (GRCh37)
Microarray based circRNA profiling of leukocytes from RR-MS patients and Healthy controls. (Subchapter 1.1)	hsa_circ_0005402	chr15:60648117-60674640
	hsa_circ_0003452_2	chr15:60653150-60678285
RNASeq based transcriptome profiling of leukocytes from different MS types patients and Healthy controls. (Subchapter 1.2)	circATP8B4	chr15:50294349-50311173
	circAGFG1	chr2:228356262-228389631
	circPADI4	chr1:17668437-17668897
	circAFF2	chrX:147743428-147744289
	circABCA13	chr7:48541721-48542148
	circRELL1	chr4:37633006-37640126
	circNEIL3	chr4:178274461-178274882
RNASeq based transcriptome profiling of PBMCs from patients with IgM oligoclonal band characterization. (Subchapter 1.3)	hsa_circ_0000478	chr13:42439871-42442613
	hsa_circ_0116639	chr22:41568502-41569788

Identifying circRNAs with potential to be miRNA sponges

Dudekula et al. defined circRNAs with more than 20 miRNA binding sites (BS) as "super-sponges" for miRNAs. Thus, to determine the potential of circRNAs to be miRNA sponges, were retrieved from CircInteractome[139] the miRNA-circRNA interactions for circRNAs with more than 20 miRNA binding sites. CircInteractome employes the TargetScan algorithm[330] to predict miRNAs that target circRNA by surveying for 7-mer or 8-mer complementarity to the seed region as well as the 3' end of each miRNA. Based on this dataset we assessed the number of miRNA binding sites for each of the selected circRNAs by intersecting both datases using R in RStudio. To evaluate whether the observed distribution and the expected one were different, we used a Chi square test, performed in R in RStudio.

CircRNA structure determination

The sequence of an RNA molecule is usually not sufficient for a reliable structure prediction. However, its combination with experimental structure data obtained from methods based

in the study of structure-specific chemical modifications such as the SHAPE-MaP[331] method used by Liu et al.[118], or DMS-MaP[332] used by Fischer et al.[119], can result in an accurate assessment of the RNA folding status and RNA structure. SHAPE uses a small hydroxyl-selective electrophilic reagents to probe the reactivity of the RNA ribose 2′-OH group. Only those which are not paired and are flexible will react. In the case of DMS, the dimethylsulfate can directly donate a methyl group to the A and C residues that are not paired. Then both SHAPE and DMS take advantage of the Mutational profiling (Map) which identifies modified residues by using reverse transcriptase to misread them. Fischer et al. have recently found that the *in silico* calculated length normalized minimum thermodynamic free energy (-ΔG/nt) of an RNA molecule has a very good correlation with the structures determined by DMS-Map, therefore we followed the same methodology for the structure determination of circRNAs.

CircRNA sequence was obtained from CircInteractome. The sequence was folded *in silico* using the RNAfold function from Vienna package 2.0. to calculate the minimum thermodynamic free energy and later normalized with the spliced length of the circRNA to calculate the -ΔG/nt value. Based on these value circRNAs were classified as highly structured (-ΔG/nt>0.25) or poorly structured (-ΔG/nt<0.2).

Results

miRNA sponging function is not common among MS-related circRNAs

MiRNA sponging function has been one of the most studied circRNA functions, however, in order to be able to efficiently sequester miRNAs, circRNAs require of a high number of binding sites (BS) for the target miRNA. On this basis, we downloaded from CircInteractome a list of 3,051 circRNAs containing at least 20 BS for a single miRNA.

First, we briefly characterized this set of circRNAs containing more than 20 miRNA BS. The miRNA for which more BS are found within the circRNA sequence will be called “Top miRNA”. However, it is worth noting that most of the circRNAs apart from the BS for the *Top miRNA* do also contain binding sites for other miRNAs. Interestingly, among these circRNAs only 1,428 different parental genes are found, indicating that many circRNAs included in the list are generated from the same gene and may also have overlapping sequences and miRNA binding sites. Moreover, among the 3,051 circRNAs many of them have binding sites for the same miRNAs. In fact, among the list of the *Top miRNAs* only

93 different miRNAs are found. It stands out the fact for 2,093 circRNAs (68.6%) the *Top miRNA* is miR-548c-3p and there are other 517 circRNAs containing BS for miRNA-548c-3p among the four most frequent miRNA BS. This suggests that sequences complementary to the seed sequence of miR-548c-3p ((C)UAAAAA) are very common at least among circRNAs.

The median length of these set of circRNAs (43,826 bps) is significantly higher than the length of the general population of circRNAs included in circBase (8,683 bps)($p<0.0001$)(Figure 55) which could suggest that shorter circRNAs do not accumulate more than 20 miRNA BS. Remarkably, the mean length of circRNAs detected in leukocytes (476bp) or EVs (386bp) is even shorter than the length of circRNAs in circBase ($p<0.0001$ for both comparisons) and consequently also significantly shorter than those circRNAs with more than 20 miRNA BS ($p<0.0001$ for both comparisons) (Figure 55).

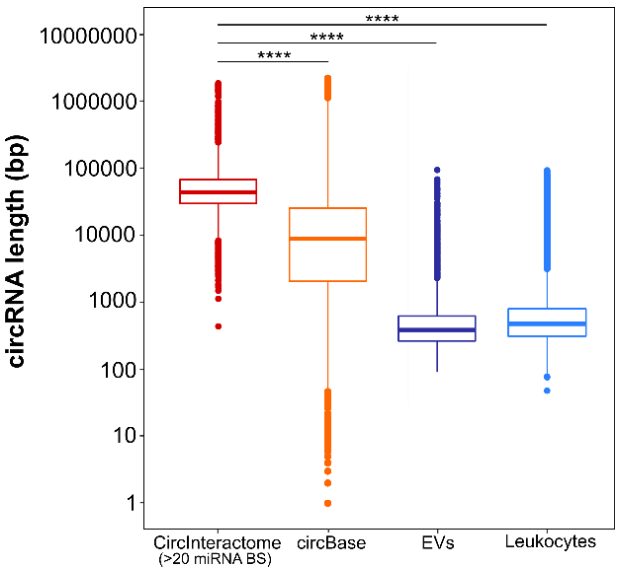

Figure 55. Size distribution of the set of circRNAs with more than 20 miRNA binding sites (BS) retrieved from CircInteractome in comparison with circRNAs included in circBase and circRNAs detected in EVs and leukocytes. Significant differences with respect to the set of circRNAs retrieved from CircInteractome as assessed by T-test are indicated. Length difference between circRNAs in circBase and circRNAs detected in EVs and leukocytes are also significant ($p<0.0001$) although they have not been depicted for clarity. ****; $p<0.00001$.

From these results it can be inferred that there might be a direct relation between circRNA length and number of miRNA binding sites. Thus, next we wanted to assess whether there is a correlation between these two variables among the 3,051 circRNAs retrieved from CircInteractome. Pearson correlation analysis confirmed a strong linear relationship ($R=0.93$ and $p=2.2\cdot10^{-16}$) indicating that the number of miRNA binding sites proportionally increases

with the size of the circRNA (Figure 56A). However, we also observe that there are a few circRNAs whose number of miRNA binding sites is higher than the expected for their size. In order to better observe this phenomenon, we calculated the number of Top miRNA binding sites/bp.

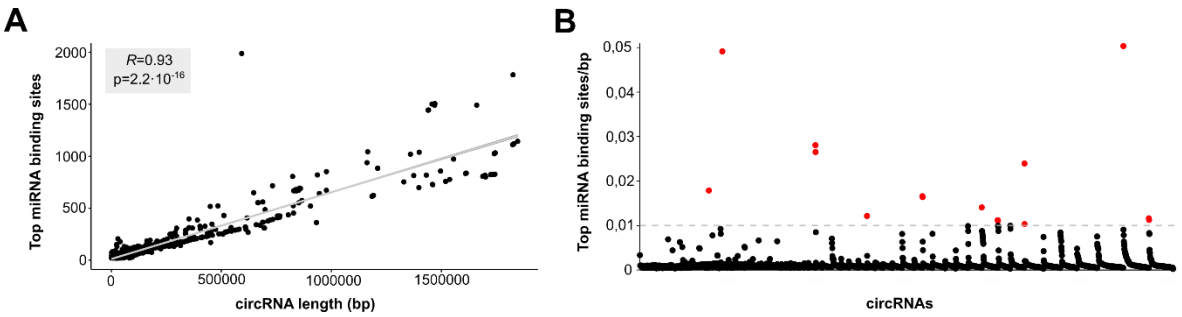

Figure 56. The number of miRNA binding sites (BS) generally increases with the length of the circRNA. A) Pearson correlation analysis between circRNA length (bp) and the maximum number of BS for a single miRNA. B) Ratio between the maximum number of miRNA BS for a single miRNA and circRNA length. CircRNAs with a ratio higher than 0.01 are highlighted in red.

We took as a reference the CiRS-7 whose miRNA sponging function has been extensively demonstrated and within its 437bp of length it presents a total of 22 BS for miR-7. Therefore, the ratio for CiRS-7 is of 0.05. Based on this data, we decided to stablish a threshold of Top miRNA BS/bp>0.01 in order to identify the circRNAs that present at least 1 miRNA binding site for every 100bps. Results reported that only 15 circRNAs (0.49%) meet this criterium (Figure 56B).

Next, we wanted to determine the number of circRNAs with more than 20 miRNA BS within the dataset of circRNAs detected in leukocytes and EVs respectively. To do so, we compared the list of circRNAs with more than 20 miRNA BS retrieved from CircInteractome with the circRNAs detected by RNASeq in each of the sample types. It is important to take into account that this method will automatically exclude those circRNAs that have been newly detected and therefore are not included in CircInteractome, reducing the total number of circRNAs tested to 17,077 circRNAs in leukocytes and 5,017 circRNAs in EVs. Out of the 17,077 circRNAs detected in leukocytes, we found a total of 200 circRNAs with more than 20 miRNA BS which corresponded to the 1.17% of the circRNAs in leukocytes. However, based on the equation inferred from the circRNA length/ number of top miRNA BS correlation for the circRNAs in CircInteractome (N Top miRNA BS = 0.00016·circRNA length+11,513), we would have expected

322 circRNAs (1.88%) with more than 20 BS for a given circRNA, which is significantly different from which we have observed ($p<0.001$). Similar results are found for the dataset of circRNAs detected in EVs where there are 36 circRNAs with more than 20 miRNA BS accounting for the 0.71% of the circRNAs whereas we would have expected twice as many circRNAs, this is, 70 circRNAs with more than 20 miRNA BS (1.39%) ($p<0.001$).

We also checked whether the circRNAs containing more than 20 miRNA BS were overrepresented among the differentially expressed circRNAs in leukocytes or EVs respectively, however we did not find any difference when compared to the proportion observed in the whole dataset. None of the circRNAs with more than 20 miRNA BS has a Top miRNA BS/bp>0.01. Moreover, among the few circRNAs with more than 20 miRNA BS, most of them have miR-548c-3p as the Top miRNA, further reinforcing that even those are miRNA BS accumulated by chance.

Finally, we focused in the study of the 11 most relevant MS related circRNAs (Table 9). In this case, we excluded circABCA13 from the study since it is a ciRNA that stays in the nucleus and thus, even if it presents miRNA BS, in will not function as an sponge. All the remaining 10 exonic circRNAs contained a scarce number of miRNA BS. Circ_0005402 contains binding sites for 25 different miRNAs, however most of them have a single BS within the sequence except for hsa-miR-1248 and hsa-miR-766 that have two BS respectively. Circ_0003452_2 contains a single BS for 18 miRNAs. However, it is worth mentioning that due to the overlap in the sequence between circ_0005402 and circ_0003452_2, 14 of these miRNAs are bound by both of them. Among the six exonic circRNA candidates drawn from the RNASeq study in leukocytes from MS patients and healthy controls, four of them are included in CircInteractome (except for circATP8B4 and circPADI4) and all of them contain single BS for different miRNAs or a maximum of two BS for the same miRNA. Among them circAFF2 is the one containing more miRNA BS with a single BS for 28 miRNAs and two BS for 9 miRNAs. CircAFF2 exceeds the length of the other three circRNAs (circRNA lengths ranging from 434 to 527 bps) by more than 300 bps (861bps), which may explain the higher number of miRNA BS found in its sequence. Taking the four circRNAs together, they accumulate 4 BS for miR-579, , 3 BS for miR-1299, 3 BS for miR-647, , 3 BS for miR-942, 3 BS for miR-637, 2 BS for miR-938, 2 BS for miR-1290, 2 BS for miR-485-3p, 2 BS for miR-548p, 2 BS for miR-587 and 2 BS for miR-892. Similarly, circ_0000478 and circ_0116639, circRNA drawn from subchapter 1.3 contained 13 and 19 single miRNA BS respectively with only one miRNA BS in common for miR-1238.

Highly structured circRNAs are more frequent in leukocytes than in EVs.

Two recent publications have drawn attention to circRNA structure as a determinant feature that defines whether they can bind PKR and regulate the immune response[118], or whether they can be targeted by G3BP1 and UPF1 to promote their structure mediated decay[119]. Based on these evidences, and with the aim of investigating their potential to function as regulators of the immune response, we got interested in assessing the structure of the circRNAs detected in our study both in leukocytes and EVs.

These two publications experimentally determined the circRNA structure by SHAPE-MaP or DMS-MaP sequencing. However, Fischer et al. additionally calculated the length normalized minimum thermodynamic free energy (-ΔG/nt) for 96,123 exonic circRNAs from circBase and reported that it is able to predict the experimentally determined structures by DMS-MaP with high reliability. Liu et al. classified circRNAs in those containing dsRNA regions and those without dsRNA regions, whereas Fischer et al. created three categories based on the -ΔG/nt: highly structured (HS) circRNAs (-ΔG/nt>0.25), poorly structured (PS) circRNAs (-ΔG/nt<0.20), and circRNAs whose structure can not be predicted with a good reliability (ND) (0.20<-ΔG/nt<0.25).

Therefore, we first wanted to assess whether there was a correspondence between these two different classifications. To do so, we calculated the -ΔG/nt for the circRNAs containing dsRNA regions in the study by Liu et al. Out of the 21 dsRNA region containing circRNAs, 17 have a -ΔG/nt> than 0.25 indicating that the 80.9% of the dsRNA containing circRNAs could also be predicted by calculating -ΔG/nt (Figure 57).

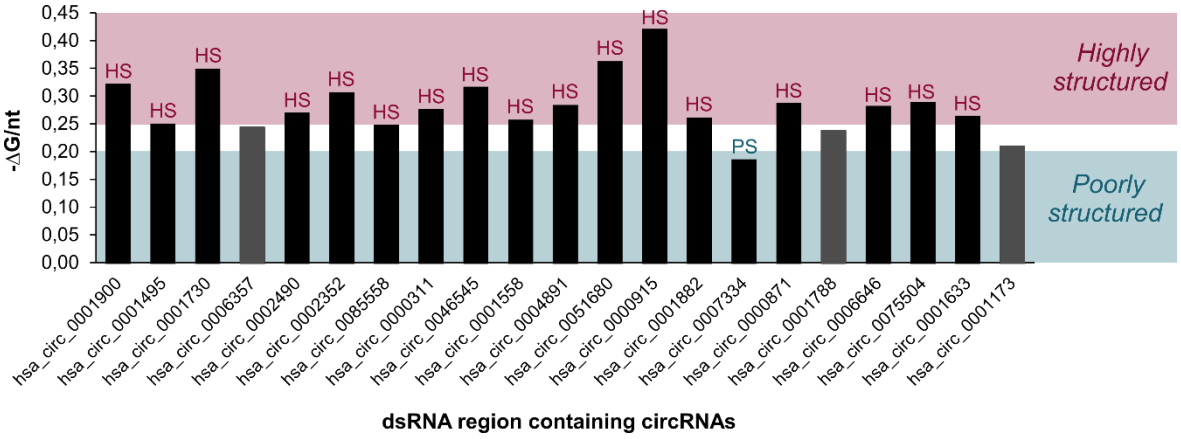

Figure 57. Length normalized Gibbs energy classifies as highly structured the 80.9% of the dsRNA region containing circRNAs.

In light of these results we decided to make an *in silico* prediction of the circRNAs detected in leukocyte and EV samples. To do so, we obtained the list of 96,123 circRNAs for which the -ΔG/nt has been calculated in Fischer's study. Out of these 96,123 circRNAs, 58.7% circRNAs (56,465 circRNAs), were highly structured circRNAs, whereas the 13.7% (13,136 circRNAs) represented poorly structured circRNAs. The structure for the remaining 26.522 circRNAs (27.6%) could not be determined based on their -ΔG/nt. When the 96,123 circRNAs from Fischer's study were compared to the dataset of circRNAs from leukocytes (22,835 circRNAs) and EVs (6,575 circRNAs) we could gather the -ΔG/nt value for 14,014 and 4,186 circRNAs respectively.

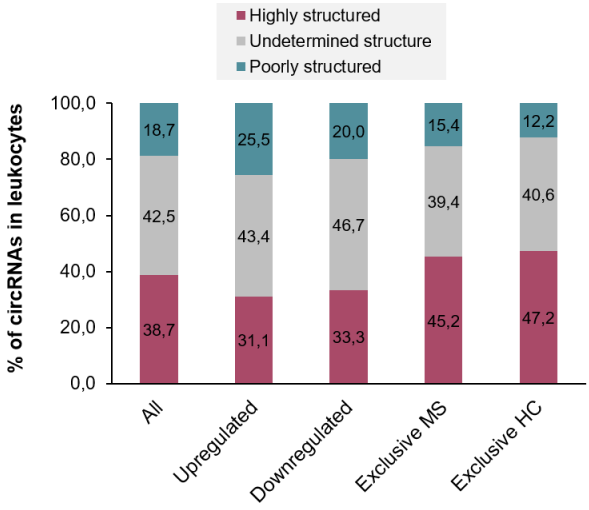

Figure 58. Structure prediction for circRNAs detected in leukocytes based on their -ΔG/nt. Three different structure categories are defined: undetermined structure, highly structured and poorly structured.

Out of the 14,014 leukocyte circRNAs, 5,957 (42.5%) have an -ΔG/nt between 0.2 and 0.25 and thus, we could not predict their structure. From the remaining circRNAs, 5.430 (38.7%) were highly structured and 2.627 (18.7%) poorly structured (Figure 58). Interestingly, this structure distribution is different from the general distribution described by Fischer et al ($p<0.001$). When the structure prediction was limited to the circRNAs reported to be upregulated in MS patients' leukocytes with respect to controls, we found that for 141 circRNAs (43.4%) the structure was indetermined, whereas 101 (31.1%) where highly structured and 83 (25.5%) poorly structured circRNAs. Among the downregulated circRNAs the structure of 7 circRNAs (46.6%) could not be predicted, 5 circRNAs (33.3%) were highly structured and 3(20%) poorly structured (Figure 58). Very similar structure distributions are found between downregulated and upregulated and apparently there is not any bias when compared to the global structure distribution in leukocytes. We also predicted the structure for circRNAs exclusively found in MS patients (1307 HS, 1138 ND, 446PS) as well

as for those exclusive of healthy controls (701HS, 604ND, 181PS)(Figure 58). Remarkably, among the group-specific circRNAs the number of poorly structured circRNAs was reduced and more circRNAs were classified as higly structured circRNAs, But there is no difference between MS or HC exclusive circRNAs.

For the 4,186 circRNAs in EVs, we performed the same analysis. In general, we can observe that the distribution is different from that described by Fischer et al ($p<0.001$). The proportion of PS circRNAs is increased among the DE circRNAs when compared to the global distribution, particularly among the DE circRNAs for the contrast between SP-MS and HC. At the same time, the number of circRNAs whose structure is undetermined is decreased (Figure 59). Finally, we also predicted the structure for those circRNAs that are exclusive for RR-MS patients (228HS, 355NS, 199PS), SP-MS patients (69HS, 101ND, 53PS) or HCs (235HS, 264 ND, 115PS). These results indicate that there is a small difference in the proportion of HS circRNAs that seem to be more frequent in EVs from HC than in EVs from patients (Figure 59).

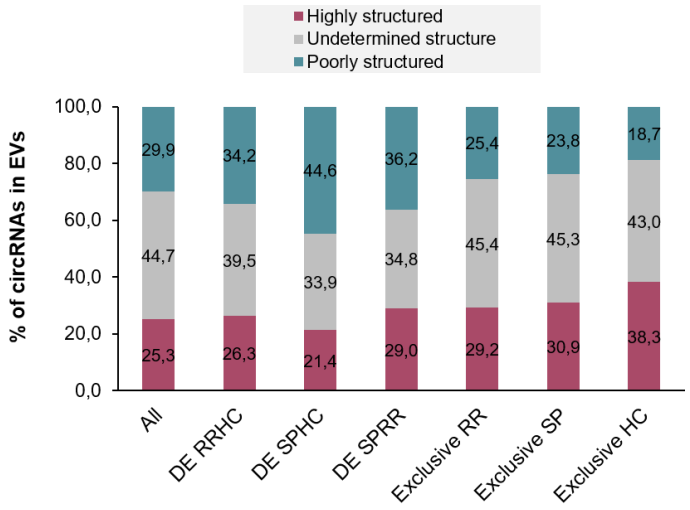

Figure 59. Structure prediction for circRNAs detected in EVs based on their -ΔG/nt. Three different structure categories are defined: undetermined structure, highly structured and poorly structured.

A general comparison between the global structure distribution between leukocytes and EVs reports interesting differences showing that highly structured circRNAs tend to be more retained in the cell (38.7% HS circRNAs in leukocytes versus 25.3% in EVs) than poorly structured circRNAs which are more frequent in EVs (18.7% PS circRNAs in leukocytes versus 29.9% in EVs). This difference is even higher if we only take into account those circRNAs exclusively detected either in one or the other compartment (Figure 60).

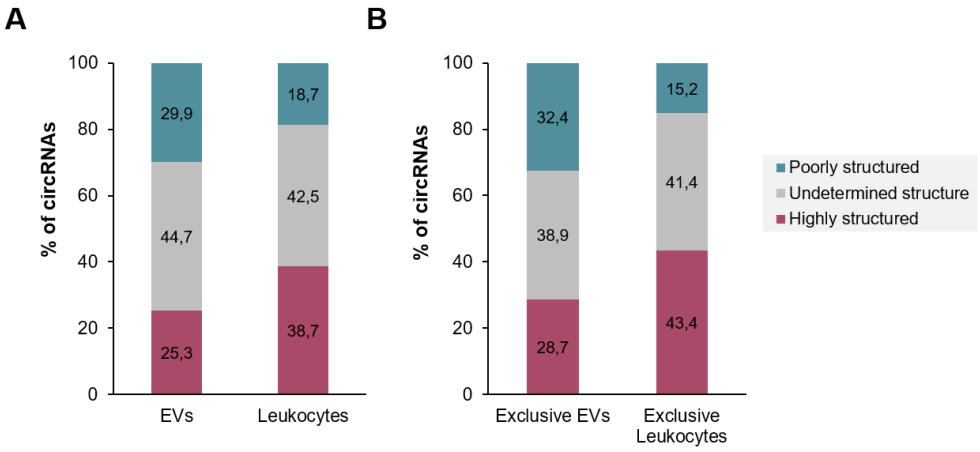

Figure 60. Structure distribution comparison between leukocytes and EVs. Three different structure categories are defined: undetermined structure, highly structured and poorly structured. A) Global distribution of all the circRNAs detected. B) Structure distribution of the circRNAs exclusively detected in each of the sample types.

Finally, we also calculated the -ΔG/nt for the 10 most relevant MS related exonic circRNAs (again excluding circABCA13) and found that 6 out of the 10 circRNAs were highly structured (Table 10).

Table 10. Table showing the length normalized minimum thermodynamic free energy-based structure prediction for the 11 most relevant MS related circRNAs reported from chapter 1.

Position	CircbaseID	GeneName	Length	-ΔG	-ΔG/nt	Structure
chr15:60648117-60674640	hsa_circ_0005402	ANXA2	480	-129,9	0,270	**HS**
chr15:60653150-60678285	hsa_circ_003452_2	ANXA2	357	-100,8	0,282	**HS**
chr15:50294349-50311173	hsa_circ_0141241	ATP8B4	227	-50,7	0,223	ND
chr2:228356262-228389631	hsa_circ_0058514	AGFG1	527	-122,3	0,232	**HS**
chr1:17668437-17668897	NA	PADI4	283	-105,6	0,373	**HS**
chrX:147743428-147744289	hsa_circ_0001947	AFF2	861	-193,1	0,224	ND
chr4:37633006-37640126	hsa_circ_0001400	RELL1	434	-136,4	0,314	**HS**
chr4:178274461-178274882	hsa_circ_0001459	NEIL3	421	-66,4	0,157	PS
chr13:42439871-42442613	hsa_circ_0000478	KIAA0564	345	-80,5	0,233	**HS**
chr22:41568502-41569788	hsa_circ_0116639	EP300	327	-54,7	0,167	PS

Abbreviations: ΔG, minimum thermodynamic free energy or Gibbs free energy. HS, highly structured. ND, undetermined structure. PS, poorly structured.

Discussion

MiRNAs have been the most studied non-coding RNAs, and it is widely known the importance of their regulatory roles in many physiological and pathological processes. Thus, when a circRNA was for the first time described to act as a miRNA sponge[90], many

researchers in the field got interested in further analysing the miRNA sponging potential of circRNAs. Since then some other circRNA acting as miRNA sponges have been reported[137,138], however increasing evidence indicates that it is not a common function among circRNAs but that only a limited number of circRNAs could function as miRNA sponges[105,137].

It is important to consider that mRNAs also compete for binding miRNAs, thus the relative abundances of miRNA, circRNA and mRNA and the number of binding sites play a key role. Thus, only highly abundant circRNAs containing many competing miRNA binding sites are likely to exert this function.

In this subchapter, we have first characterized an online available dataset of 3,051 circRNAs with more than 20 miRNA binding sites for a single miRNA as the most promising candidates to function as miRNA sponges. We have found that these circRNAs are on average the largest circRNAs that have been described and included in CircBase. In line with it, we could confirm that there is a strong relation between the circRNA length and the number of miRNA binding sites. Most of the circRNAs with a high number of circRNAs may have accumulated them by chance. A few circRNAs containing no more miRNA binding sites than expected by chance have shown to have miRNA sponging properties[333,334]. However, circRNAs detected in leukocytes and EVs are generally much shorter, and thus, the number of miRNA binding sites that they accumulate by chance is generally very low. In order to be able to act as miRNA sponges, and taking into account their size, they should be densely covered by miRNA binding sites for a given miRNA as it has been described for ciRS-7 which contains 22 miRNA binding sites along its 437bps [90,92]. Nevertheless, none of the circRNAs detected in leukocytes or EVs presents a minimum number of 1 binding site per 100 bps, which results in very few circRNAs with enough miRNA binding sites in order to act as miRNA sponges.

CircRNAs have also been described to be able to regulate the immune response by binding and inhibiting the activation of the well-known PKR. However, this function is exclusive for circRNAs containing structures such as the dsRNA regions required for the binding to PKR[118]. For circRNA structure determination different *in silico* tools and algorithms are available. However, all the algorithms are sequence based and sequence information alone is generally not sufficient for a good prediction of RNA structure[332], on average, only the 50-70% of the base pairs are correctly predicted[335]. The combination of sequence information with experimental data with methods like SHAPE-MaP[331] or DMS-MaP[332] allow to obtain an accurate approximation of RNA structures. In this study, we have taken

advantage of the *in silico* calculation of the length normalized minimum thermodynamic free energy (-ΔG/nt) of an RNA molecule which Fischer et al. have recently shown to have a very good correlation with the structures determined by DMS-Map. Nevertheless, it should be taken into account as a limitation that the -ΔG/nt based structure predictions could not perfectly meet the real RNA structure. In addition, and although we have found an 80.9% correlation, circRNAs classified as highly structured based on the -ΔG/nt could be slightly different from what Liu et al. have described as dsRNA region containing circRNAs.

Liu et al. reported that the 76% of the circRNAs they analysed contained dsRNA regions. Based on the predictions made by Fischer, most of the circRNAs are also highly structured. They represent the 58.7% of the total number of circRNAs in the dataset, which could seem to be far from the 76% described by Liu et al. However, if we calculate the percentage out of the number of circRNAs for which a structure could be predicted (this is, HS and PS circRNAs) the 81.1% of the circRNAs are HS. Nevertheless, in the leukocyte and EV datasets the number of circRNAs for which the structure could not be predicted (circRNAs with -ΔG/nt between 0.2 and 0.25 values) is increased at the same time that the proportion of highly structured circRNAs is reduced.

In leukocytes, approximately the 40% of the circRNAs are highly structured and poorly structured circRNAs do not exceed the 20%. However, this proportion is more balanced in EVs were each of these groups are close to the 30%. In subchapter 1.4. we already postulated that EV circRNAs far for being a representation of the cellular content could be selectively loaded into EVs resulting in different relative abundances and sizes. The structural analysis performed in this subchapter further supports the hypothesis of a regulated export of circRNAs that may be favouring the accumulation of highly structured circRNAs within leukocytes by loading a smaller HS circRNA/PS circRNA proportion in EVs.

In general, the circRNA structure distribution is almost not changed between MS patients and healthy controls. The only remarkable difference is observed in patient derived EVs were less highly structured circRNAs are found in comparison with EVs from healthy controls. Nevertheless, this observation does not correlate with a bigger proportion of highly structured circRNAs in MS patients when compared to healthy controls.

To sum up, from chapter one we have reported 10 exonic circRNAs that seem to have an implication in the pathology of MS, and in this subchapter, we wanted to investigate some of the molecular functions that these circRNAs may be exerting in order to better understand their role. Based on previous literature, we decided to investigate the miRNA sponging function and the PKR inhibitory function. We found that although none of the 10

circRNAs seem to have potential to extensively sequester miRNAs, 6 of them are highly structured. Therefore, based on the structure prediction we hypothesize that circ_0005402, circ_003452_2, circAGFG1, circPADI4, circRELL1 and circ_0000478 may potentially be implicated in the regulation of the immune response in MS by interacting with PKR. This possibility will be further investigated for some of the HS circRNAs in subchapter 2.2.

2.2

Investigating the cross-talk between circRNAs and PKR and its implication in the immune response

This project was partially developed during a three month stay in the ncRNA lab headed by Dr. Hansen at the Department of Molecular Biology and Genetics in Aarhus University (Denmark).

Introduction

The innate immune system is the first line of defense against pathogens such as bacteria and viruses. To do so, it relies in a set of several membrane and intracellular receptors that recognize different pathogen structures and patterns[336]. The dsRNA-activated protein kinase or PKR is one of these intracellular receptors responsible for recognizing long (>33bp) double stranded RNAs (dsRNAs) (usually viral RNAs) in the cytoplasm. In case of recognizing any of these dsRNAs, it responds by inhibiting protein synthesis in an attempt to limit virus spread. This response needs to be fast in order to be effective, and thus, PKR needs to be readily activable in the cytoplasm in case there is an infection. However, it must be maintained in an inactive state in uninfected cells to prevent inappropriate autoimmune reactions[153].

A recent study by Liu et al. has reported that endogenous circRNAs form intramolecularly imperfect RNA duplexes of about 16-26bp that allow them to bind to PKR stronger than

any of the previously described substrates and blocking its activation. Thus, they highlight the role of circRNAs in maintaining PKR in an inactive state. Interestingly, in this study they also described that upon a viral infection (simulated by the stimulation with poly(I:C) that mimics viral genetic material) the oligoadenylate synthetase (OAS) is activated and produces 2'-5'-linked oligoadenylates(2-5A). These 2-5As ultimately activate the RNase L, that dimerizes and leads to a rapid degradation of circRNAs. PKR is then released and activated to aid in the immune response[118]

In this line, they further found an increased RNAse L activity in patients with the autoimmune disease systemic lupus erythematosus (SLE) which correlated with a global reduction of circRNAs and an enhanced PKR activation. Later on, other studies have reported a similar reduction of circRNAs in other autoimmune diseases such as psoriasis[291]. Interestingly, they also showed that overexpression of circular RNAs was able to alleviate the aberrant PKR activation cascade although it needs further investigation[118].

Regarding MS, and in contrast with what has been described for SLE and psoriasis, we have found that there is a global upregulation of circRNAs in patients' leukocytes when compared to healthy controls. However, it is worth noting that patients' samples were drawn in remission phase which is not the moment when the immune response is more active. However, when we studied the expression of the seven circRNA candidates drawn from the RNASeq analysis, we observed that half of the patients had a global downregulation of the seven circRNAs in the relapse phase when compared to the remission. In light of these results, we hypothesize that circRNAs may also be playing a role in the regulation of the innate immune response in MS by interacting with PKR. In particular, we hypothesize that some factor may induce a circRNA degradation that may lead to the activation of PKR and its downstream cascade similarly to what have been described for SLE or psoriasis, which could contribute to the relapse phenotype. Nevertheless, after a relapse, during the remission phase circRNA expression may be boosted as an attempt of increasing the overall circRNA pool and particularly of the dsRNA region containing circRNAs so that they bind, inhibit and take PKR back to its inactive state (Figure 61).

In this subchapter, we first wanted to stablish a cell model to investigate the interplay between circRNAs and PKR and its effect on immunity. To do so, we have studied the effect of poly(I:C) stimulation on HeLa cells at different timepoints by analyzing the changes at the RNA level and in the phosphorylation and activation of PKR. Moreover, we have also tested the potential of curcumin (as an anti-inflammatory compound) and of the

overexpression of circRNAs (with and without dsRNA regions) to prevent the activation of the PKR cascade. Lastly, we will also study the phosphorylation of PKR in PBMC samples from MS patients in relapse and remission phases.

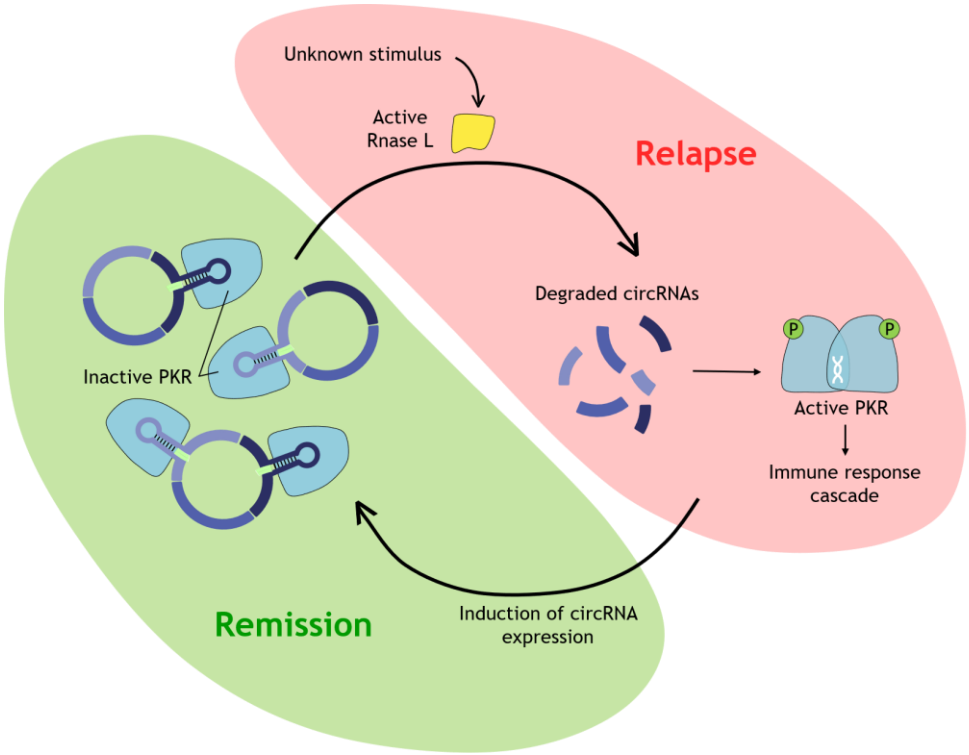

Figure 61. Hypothesis of the cross-talk between circRNAs and PKR in relapsing-remitting multiple sclerosis.

Materials and Methods

Selection of circRNAs for the study of interaction with PKR

CircRNA structural conformation has been reported as a key feature for their interaction with PKR[118]. Liu et al. stated that circRNAs containing regions that tend to form duplexes resulting in regions with double stranded structure (dsRNA containing circRNAs) bind to the PKR and act as inhibitors, whereas those without such region do not have any effect on PKR activation.

In order to replicate these results and to further investigate this interaction, we selected three different circRNAs based on their structure. Structured circRNAs were defined as those with a length normalized minimum thermodynamic free energy (-ΔG/nt) (calculated as previously described in subchapter 2.1) higher than 0.25[119] and containing dsRNA regions determined by SHAPE-MaP[118]. On the other hand, unstructured circRNAs were

defined as those with a -ΔG/nt<0.2 and without any dsRNA region determined by SHAPE-MaP.

Among the structured circRNAs we selected circCAMSAP1 and circRELL1 for being highly expressed circRNAs. Moreover, circRELL1 has been found to be differentially expressed between MS patients and HC both by microarray (subchapter 1.1) and RNASeq (subchapter 1.2). Besides, circSMARCA5 was selected as an example of unstructured circRNA also based on its high expression.

Plasmids for circRNA overexpression

Briefly, all the plasmids used for circRNA overexpression in this study shared a common backbone of pcDNA3 with artificially inverted elements that flank the insert to favour its circularization. The plasmid for circRELL1 overexpression was constructed as described below and plasmids for circCAMSAP1 and circSMARCA5 overexpression were kindly provided by the group headed by Dr. Hansen (Department of Molecular Biology and Genetics, Aarhus University) and the group headed by Dr. Purrello (Department of Biomedical and Biotechnological Sciences-Section of Biology and Genetics, University of Catania) respectively.

Cloning of circRELL1 in a pcDNA3 plasmid

The insert for the plasmid was generated by PCR with subsequent restriction digest and ligation into pcDNA3 (Invitrogen). Primers are listed in Appendix Table 1. Briefly, a 594bp fragment of the human genome, comprising exons 4, 5 and 6 of RELL1 (the three exons that circularize generating circRELL1), 70 bp that flank exons 4 and 6 and which contain the endogenous splicing sites and 10 additional base-pairs that include restriction sites for XhoI and NheI was amplified by PCR (Figure 62). Then, both the 594bp fragment and a modified pcDNA3 vector containing artificial inverted repeats were double digested by XhoI and NheI and later ligated. The correct cloning of RELL1 exons into the plasmid were checked by digestion test and sequencing after transformation and growth of the competent bacteria. Midipreps were performed in order to produce enough amounts of the circRNA plasmids for the transfection experiments.

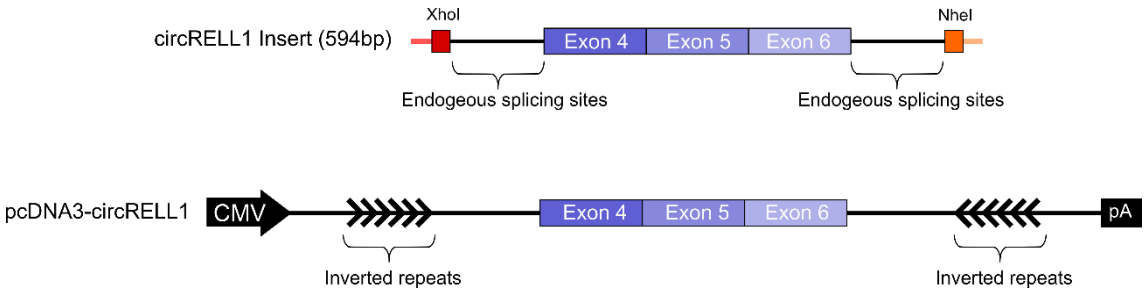

Figure 62. Schematic representation of the insert generated by PCR and the final expression vector for circRNA. It comprises the CMV promoter, exons 4 to 6 known to circularize, the flanking regions containing the endogenous splicing sites, the artificially introduced inverted elements depicted by divergent arrows, and the bovine growth hormone polyadenylation signal.

CircRELL1 probe design and northern blot

To check whether the constructed circRELL1 plasmid was able to produce a single circularized product, cells were transfected and RNA was extracted with TriZol, and subjected to a Northern blot for the detection of the circularized form. A 60nt probe was designed overlapping the exon 5 (See Appendix Table 1).

Northern blotting was performed as described by Hansen in[337]. Briefly, 10 µl RNA (1µg/µl) and 20 µl Northern loading buffer (58.7% formamide, 6.5% formaldehyde, ethidium bromide, 1.18% MOPS, and bromophenol blue) were mixed and denatured at 65°C for 5 min. The RNA was separated by electrophoresis on a 1.2% agarose gel containing 3% formaldehyde and 1× MOPS at 75 V for 3h. After electrophoresis, the gel was briefly washed in water and the EtBr stained ribosomal RNA bands were visualized by UV light. Then, the gel was washed again in 10xSSC buffer and the RNA was transferred from the gel to a Hybond N+ membrane (GE Healthcare) overnight (O/N) in 10× SSC. After transfer, the membrane was shortly washed in 5xSSC and UV cross-linked in order to fix the RNA to the membrane. It was after pre-hybridized in Church buffer (160 mM NaH_2PO_4, 340mM Na_2HPO, 4,7% SDS, 1 mM EDTA, and 0.5% BSA, pH 7.5) for 1 h at 55°C. The probe was radioactively labelled by the action of T4 polynucleotide kinase (10U/ul, ThermoScientific) which catalysed the incorporation of phosphates from the isotope [γ-^{32}P] ATP to it 5' end. Once the probe was prepared the membrane was probed at 55^{o}C O/N.

The next day, the membrane was washed twice in 2× SSC, 0.1% SDS at 50^{o}C for 10 min and exposed on a plate for bioimaging analyser (Fujifilm) and visualized on a Typhoonn FLA 9500 (GE Healthcare).

Cell culture, circRNA plasmid transfection and poly(I:C) stimulation.

HeLa and HEK293T cells (ATCC) were used for all experiments. They were cultured in DMEM with D-Glucose, L-Glutamine and Pyruvate (ThermoFisher) supplemented with 10% Fetal Bovine Serum (FBS) (ThermoFisher) and 0.2% penicillin/streptomycin and kept at 37ºC in a 5% CO_2 cell culture incubator.

Once the culture had reached an 80-90% of confluency, transient transfections were carried out for circRNA overexpression using Lipofectamine 2000 Reagent (ThermoFisher) as transfection reagent according to the manufacturer's protocols. An empty pcDNA3 plasmid was used as control plasmid. A final concentration of 1µg of plasmid/ml of medium was maintained for 16 hours until transfection was stopped by changing the medium. Cells were cultured in fresh medium for 6 hours prior to stimulation in order to allow them to recover and start expressing the transfected circRNAs. For curcumin treatment, curcumin was first dissolved in DMSO and added to the fresh medium to a final concentration of 5µM or 20µM, the same DMSO concentrations were added as control. After the recovery time, poly(I:C)(1µg/ml, Sigma) was transfected into cells for 6 hours using Lipofectamine 2000 Reagent (ThermoFisher) according to the manufacturer's protocols. For unstimulated conditions poly(I:C) was replaced for water. Then, cells were harvested for total RNA isolation. A schematic representation of the culture protocol is presented in Figure 63.

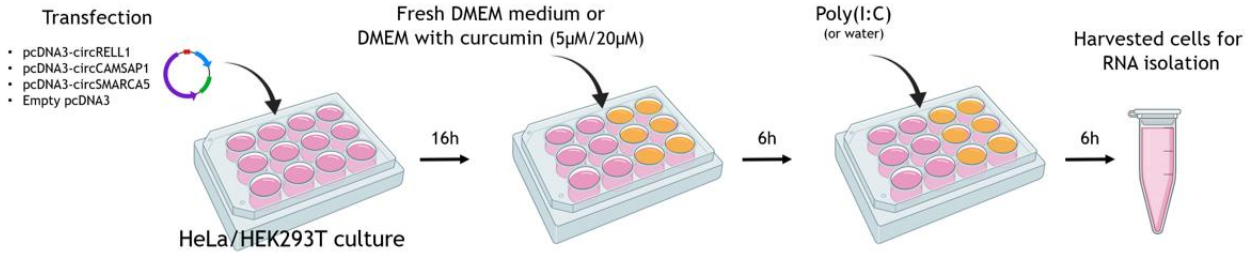

Figure 63. HeLa/HEK293T cell culture protocol for circRNA overexpression and poly(I:C) stimulation. Cells at 80-90% of confluency were transfected with the corresponding plasmid for circRNA overexpression (or control plasmid) for 16h. After that, culture medium was changed to remove transfection reagents and cells were cultured for 6 additional hours either in fresh medium or in medium supplemented with curcumin in order to test its anti-inflammatory effect. Then cells were stimulated with poly(I:C) to mimic a viral infection for 6 hours. Cells were then harvested for subsequent RNA isolation. *Ilustration by Iparraguirre L. Some of the icons have been obtained from BioRender.*

RNA extraction

Two different cell harvesting and RNA extraction protocols were used in order to fulfil the requirements for the different subsequent applications of the RNA samples.

For subsequent RT-PCR analysis cells were harvested and RNA isolated as follows: cells were washed with DPBS 1X, trypsinized, resuspended in DMEM and pelleted (1500rpm x 5 min). They were then lysed with Trizol LS Reagent (ThermoFisher) and RNA was isolated following manufacturers´ protocol. RNA concentration was measured using a NanoDrop ND-1000 spectrophotometer (ThermoFisher).

For RNA-Sequencing, we wanted to ensure the best quality of RNA and a more accurate quantification so we modified the cell harvesting and RNA isolation protocols in consequence. To avoid the use of trypsine, cells were washed with DPBS 1X and scrapped with a cell scraper in QiaZol (Qiagen). RNA was isolated with the miRNeasy Mini Kit (Qiagen) following the manufacturer's instructions. Bioanalyzer was used for RNA quantification and assessment of RNA integrity.

RT-qPCR

As this project was partially developed at the Department of Molecular Biology and Genetics in Aarhus University (MBG), and at my home lab in Biodonostia, two different protocols for RT-qPCR were used.

Regarding the cDNA synthesis, between 600 and 2000ng of total RNA were reverse-transcribed either in two steps at the MBG, by MLV-RT Reverse transcriptase kit (ThermoFisher), or in one step at Biodonostia, using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) following manufacturer's instructions and using random primers. A Verity Thermal cycler (Applied Biosystems) was used in both cases.

For quantitative PCR (qPCR) divergent primers were designed as previously described[259] for the amplification of the circRNAs so that the amplicon spans the BSJ. Primers are listed in Appendix Table 1.

At the MBG, Quantitative PCR was performed using Platinum SYBR Green qPCR SuperMix (Thermofisher) in a LightCycler 480 (Roche Molecular Systems), according to supplier's protocol. Each sample was run in triplicate (starting from 20ng of cDNA in 10 µl total reaction volume). Q-PCR cycling conditions were: 95 ^{o}C for 1 min, 45 cycles of 95 ^{o}C

for 15 sec, 55 ºC for 1 min and 72 ºC for 30s followed by 72 ºC for 5 mins and a dissociation curve.

In Biodonostia, quantitative PCR was performed as described in previous subchapters.

Four different endogenous genes were analysed for the poly(I:C) stimulation experiments, however, EEF1A1, GAPDH and B-actin are not stable after stimulation while 18S amplifies more than 20 Cqs before target genes. Therefore, none of them was considered as a suitable endogenous gene for normalization, and we decided to work with the raw Ct values for each target gene. Due to the lack of the endogenous gene, special care was taken in order to perform the RT-qPCRs starting from the same RNA input for all samples. The relative amount of gene expression and the fold change for each transcript were calculated according to $2^{-\Delta Cq}$ method in Microsoft Excel 2010 in both cases.

RNASeq for quantification of linear and circular RNAs

In order to have a wide picture of the changes at the RNA level induced by the stimulation of poly(I:C) and how these changes are or not affected by the overexpression of different circRNAs we sequenced 16 samples from HeLa cultures performed as previously explained in the *Cell culture, circRNA plasmid transfection and poly(I:C) stimulation section.* Summary information of the samples is found in Figure 64.

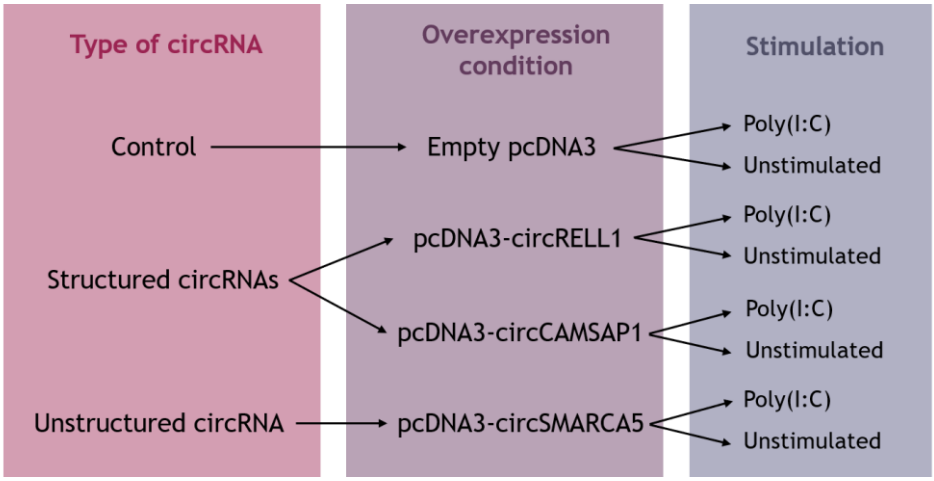

Figure 64. Scheme indicating the different culture conditions for RNA Sequencing. Two replicates of each condition were sequenced.

RNA samples were sequenced in CD Genomics following the same protocol previously explained in subchapters 1.2 and 1.3.

First, quality of the sequencing was checked and then, for circRNA detection reads were mapped to the human genome using STAR [282] or BWA [296] and circRNA prediction was performed by CIRCexplorer2 [111] and CIRI2 [280] as previously described in subchapter 1.2. Only circRNAs supported with at least two reads by both algorithms were considered as *bona fide* circRNAs and used in subsequent analyses. Their expression level was quantified based on back-spliced junctions-spanning reads according to CIRI2. Differential expression analysis between the poly(I:C) stimulated and the unstimulated conditions for each of the different overexpression conditions was performed using DESeq2 package in R-studio. Two different criteria where stablished in order to take a circRNA as differentially expressed: 1) circRNAs present in all the four replicates included in the comparison and with a FC$\geq|2|$, or 2) circRNAs absent in the two replicates of one of the conditions and present in the two replicates of the other condition, this is, circRNAs present in unstimulated samples and absent in poly(I:C) stimulated samples and vice versa.

For the pseudoalignment, identification and quantification of linear transcripts, Kallisto [297] algorithm was used. Differential expression analysis between the poly(I:C) stimulated and the unstimulated conditions for each of the different overexpression conditions was performed using DESeq2 package in R-studio. Several filters were applied in order to stablish the differentially expressed linear transcripts: 1) Linear RNAs detected with more than five reads in each of the four samples included in the comparison (two poly(I:C) stimulated replicates vs two unstimulated replicates). 2) Transcripts with a variation coefficient smaller than the 30% between replicates. 3) Transcripts with a change in the expression between poly(I:C) and unstimulated conditions of at least to times (FC$\geq|2|$), and an adjusted (FDR) p-value<0.05.

Gene ontology analysis

The GOrilla web-based gene ontology tool (http://cbl-gorilla.cs.technion.ac.il/) was used to perform gene enrichment analysis for biological processes (54). For each of the contrasts the list of all the differentially expressed linear RNAs (FC$\geq|2|$ and adjusted p-value<0.05) was subjected to the analysis. Default setting were used for the analysis with a significance cut-off of p-value<0.001 to determine the enriched GO-terms.

Cell-based ELISA for quantification of PKR and p-PKR in HeLa cells

A colorimetric cell-based ELISA kit was used for the qualitative detection of PKR and its phosphorylation (Phospho-Thr446) in response to the treatment with poly (I:C) in plated and fixed HeLa cells.

Briefly, 200,000 cells were seeded onto the bottom of each well in a 96-well plate and cultured as described in the *Cell culture, circRNA plasmid transfection and poly(I:C) stimulation* section. Cells were transfected for the overexpression of circRELL1, circCAMSAP1 or circSMARCA5 including a control condition with no overexpression. Finally, the corresponding cells were stimulated with poly (I:C). In this experiment, we also checked the effect of curcumin. After the corresponding treatments, cells were washed with 1xTBS twice and fixed with 4% formaldehyde for 20 mins. Subsequent steps involved the quenching, permeabilization and blocking of the cells prior to the addition of primary antibodies specific for PKR and pPKR (Thr466). Primary antibodies against GAPDH were also added to some control wells. Cells were incubated with the primary antibody solution for O/N (16 hours) at 4ºC to allow them to bind to their respective epitopes. Next day, cells were again washed and incubated with HRP-conjugated AntiRabbit IgG secondary antibodies for 1.5 hours at room temperature. For GAPDH a HRP-conjugated AntiMouse IgG antibody was used. Finally, substrate was added and the plate was revealed by a chemiluminescent reaction read at 450nm in the Appliskan microplate reader (ThermoFisher Scientific). After recording the chemiluminescent signal, cells were washed and stained with crystal violet (absorbance at 595nm) in order to quantify the final cell density of each well. Data were recorded by the SkanIt Software and after normalized and analysed in Microsoft Excel 2010.

For normalization, the absorbance values for the target proteins (total PKR, p-PKR and GAPDH) (450nm) was divided by the corresponding absorbance value for cell density of each well (595nm). The normalized absorbance (450/595nm) for total PKR was taken as reference (100%) for calculating the percentage of p-PKR. Fold change was calculated by the ratio $\%\text{p-PKR}_{\text{Poly(I:C)}}/\%\text{p-PKR}_{\text{Unstimulated}}$.

Western blot for quantification of PKR an p-PKR in PBMCs from MS patients

The expression of total PKR and p-PKR was determined by Western Blot in PBMCs from 7 RR-MS patients in relapse and 8 RR-MS patients in remission. Additionally, a sample

from a healthy control donor was also included. PBMCs were isolated from blood as previously described in subchapter 1.1, resuspended in DPBS and stored at -80ºC. After thawing protease and phosphatase inhibitors were added at 1% of the final volume.

Briefly, 5µg of protein were denatured in sample buffer (1.25M Tris-Cl, 8% SDS, 40% glycerol, 5% β- mercaptoethanol and 0.04% bromophenol blue) by heat denaturation for 5 min at 95º C. Protein extracts were loaded in 10% sodium dodecyl sulphate polyacrylamide gel to perform the electrophoresis and subsequently electro-transferred to a nitrocellulose membrane (GE healthcare, USA). Before blocking, membranes were stained with the No-Stain Protein Labeling Reagent (ThermoFisher) for total protein normalization following manufacturer's instructions.

Membrane blocking was carried out by adding 5% BSA/tris-buffered saline (TBS)-1% tween (TBS-Tween) after what membranes were incubated in blocking solution overnight at 4º C with anti-PKR (Invitrogen #700286, Monoclonal produced in rabbit), and anti-pPKR (Thr446) (Invitrogen #PA5-99363, Polyclonal produced in rabbit) at 1:750 and 1:1000 respectively. After washing the membrane five times with TBS-1 % Tween-20 for 5 minutes, HRP-conjugated secondary antibody was used in blocking solution and incubated for 1 hour at room temperature (anti-rabbit IgG1 HRP 1:4000 from Cell Signaling). For protein band visualization Novex HRP Chemiluminiscent substrate (Invitrogen) was used and images acquired using iBright FL1000 Imaging Systems (ThermoFisher). Intesity data were analyzed using the iBright Analysis Software.

Results and discussion

The pcDNA3-circRELL1 plasmid produces a single circular product

After constructing the pcDNA3-circRELL1 plasmid, with the aim of confirming that it was useful to overexpress circRELL1, we transfected HeLa cells both with the pcDNA3-circRELL1 plasmid and with an empty pcDNA3 as a control. A northern blot was performed on the RNA isolated from those cells showing that the probe prepared against the exon5 of RELL1 detects an unspecific band in both conditions but it also detects a new band that is unique to pcDNA3-circRELL1 transfected cells. In order to further prove that this product corresponds to the circularized circRELL1, we performed a RT-qPCR using divergent primers to detect its BSJ. RT-PCR results confirmed that we detect more than 4,000 times more circRELL1 in the pcDNA3-circRELL1 transfected condition when compared to control

cells indicating that the band detected in the pcDNA3-circRELL1 conditions corresponds to circRELL1 (Figure 65).

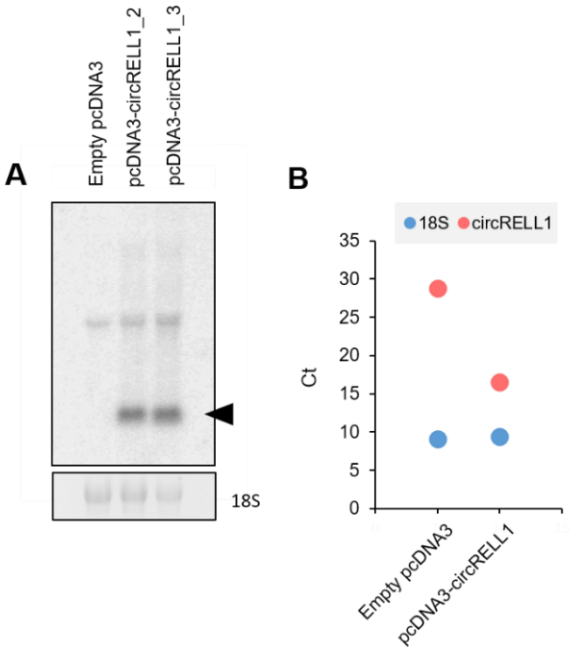

Figure 65. Overexpression of circRELL1 is induced with the transfection of the pcDNA3-circRELL1 plasmid. (A) Northern blot showing a single new band produced in pcDNA3-circRELL1(two different purificated plasmids have been loaded) transfected when compared to the control. (B) RT-qPCR results confirming that there is a notable increase on the levels of circRELL1. 18S is detected both in the northern blot and in the RT-PCR as a reference gene.

CircRNAs are downregulated in Hela cells upon poly(I:C) treatment

CircRNAs are in general stable in cells under normal conditions[121], however they have been reported to be degraded upon viral infection and poly(I:C) treatment, which is used to mimic pathogenic dsRNAs such as viral RNAs. Thus, our first experiments were directed to establish the best culture conditions to investigate the effect of poly(I:C) on the circRNA expression and triggering of immune response pathways, in our hands. To this end, we tested 2 different cell lines (HEK293T and HeLa cells) which were stimulated with a defined poly(I:C) concentration of 1µg/ml and 6 hours of stimulation time.

As the degradation of circRNAs has been linked to the activation of PKR and its cascade, we additionally studied the expression of three genes implicated in the innate immune response pathways and known to be upregulated upon its PKR activation: gene coding for the protein phosphatase-1 regulatory subunit 15 known as GADD34 or PP1R15A[338],

cyclic AMP dependent activating transcription factor 4 (ATF4)[339–341] and the cytokine interferon β1 (IFNB1)[342,343]. The change in expression of circRNAs was inferred from the expression of 4 randomly selected highly expressed circRNAs: circPOLR2A, circCAMSAP1, circHIPK3 and circBMPR2 (See primer sequences in Appendix Table 1).

Results showed that both cells lines respond to poly(I:C) by activating some of the innate immune response cascades. In fact, although ATF4 did not show a big change (FC=1,2 for both cell lines), GADD34 and IFNB1 were upregulated 1,7 and 10,2 times in HEK293T cells and 4,1 and 14,7 times in Hela cells respectively. However, the effect of the stimulation on the circRNA expression was only confirmed for HeLa cells, were all the four circRNAs were downregulated between 1,6 and 3 times. On the contrary, the expression of none of the circRNAs was changed between the unstimulated and stimulated conditions in HEK293T (Figure 66).

These results suggest that at least in HEK293T cells, poly(I:C) could also be able to induce the expression of some immune-response related genes via a mechanism independent of the degradation of circRNAs for the release of PKR. For HeLa cells however, we observe a correlation between circRNA downregulation and the expression of immune-response related genes. Thus, we decided to choose HeLa cells as the study model for further exploring the involvement of circRNAs in the immune response and their cross-talk with PKR.

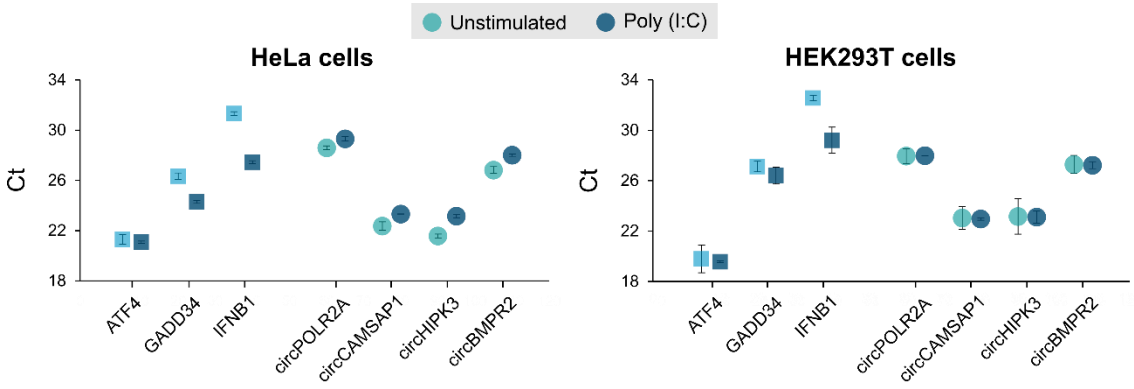

Figure 66. Upon poly(I:C) stimulation, circRNA expression is decreased in HeLa but not in HEK293T cells. The expression of three innate immune related genes (depicted by squares) and four circRNAs (depicted with dots) is represented by their raw Ct values. Note that lower Ct values correspond to higher levels of expression. Error bars represent the standard deviation between replicates.

Once the stimulation is stopped circRNA expression starts to recover.

We have hypothesized that after an immune attack were circRNAs have been downregulated and PKR cascade has been activated, there might be a compensatory mechanism to restore

the normal phenotype. Thus, we wanted to assess whether after the immune attack circRNA levels are restored and PKR activation cascade stopped.

To our knowledge, circRNA expression has not been followed up after the stimulation with poly(I:C). Therefore, for the following experiment we maintained the same stimulation conditions and decided to define two extra time points for the harvesting of cells and analysis of the expression to observe the evolution of the response even after the stimulation was stopped. Expression of the 4 circRNAs and PKR immune response genes was measured by RT-qPCR at 6h, 10h and 24h after transfection.

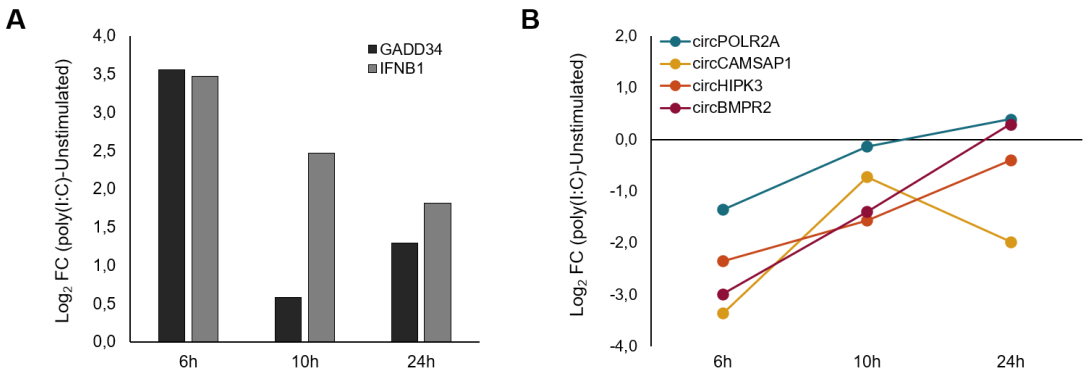

Figure 67. Late response to poly(I:C) shows that circRNA expression starts to recover while immune activation starts to decrease. Expression of two innate immune response genes (A) and of four circRNAs (B) showed as the Log_2 of the fold change between the stimulated and unstimulated conditions at 6, 10 and 24h after poly(I:C) stimulation.

We again confirmed that after 6 hours of stimulation there is an activation of the innate immune response pathways (GADD34 and IFNB1 are upregulated 27,6 and 26,0 times respectively) and a downregulation of circRNAs (Figure 67). Interestingly, after 10 and 24 hours from the stimulation, which correspond with 4 and 18 hours after the stimulation was stopped, the expression of the selected immune genes starts to decrease. In line with these results the downregulation of circRNAs also starts to recover, resulting in a less pronounced downregulation after 10 hours. This trend with time is maintained after 24h for all the circRNAs but for circCAMSAP1 and even a slight upregulation of the circRNAs circPOLR2A and circBMPR2 can be observed at this timepoint (Figure 67). This observation is in agreement with the initial hypothesis of a recovery mechanism and also goes in line with the circRNA upregulated profile that we have observed in MS patients in remission (Figure 36) when compared to healthy controls and with the differences in the expression of the circRNA candidates studied between relapse and remission phases (Figure 41).

Additionally, we also decided to study the early response to poly(I:C) in order to assess how fast was the circRNA expression affected by the stressor agent. Thus, three more timepoints were studied: 30, 60 and 120 mins after the poly(I:C) stimulation.

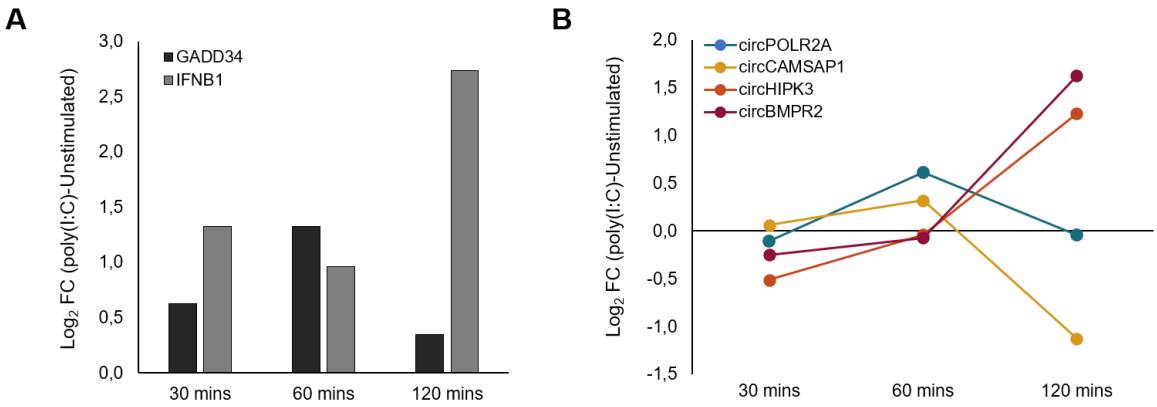

Figure 68. CircRNA expression is only slightly changed at the earliest timepoints after poly(I:C) stimulation. Expression of two innate immune response genes (A) and of four circRNAs (B) showed as the Log_2 of the fold change between the stimulated and unstimulated conditions at 30, 60 and 120 mins after poly(I:C) stimulation.

CircRNA expression was only slightly affected at the earliest timepoints with FC<|1.5|, whereas 120mins after stimulation the expression of three out of the four circRNAs is already changed although they show different trends. Similarly, the increase in the expression of GADD34 and IFNB1 is small at 30 and 60 mins but IFNB1 is already triggered at 120 mins (FC=6,65) although its upregulation is still smaller than the observed after 6h (FC=26,0)(Figure 68). In line of the results we conclude that this timepoints are too short to see a consistent effect and thus decide to keep the 6h timepoint for studying the effect of poly(I:C) stimulation (Figure 67).

Linear RNA degradation takes more time upon poly(I:C) stimulation.

It has been demonstrated that circRNAs bind more strongly than linear RNAs to PKR, however, linear RNAs could also potentially bind and inhibit PKR. Therefore, we also measured the expression of the linear transcripts produced from the genes that give rise to the four circRNAs studied in the previous experiment in order to analyze whether they are also affected by poly(I:C) stimulation and in case they do, whether their changes do also correlated with the changes observed on the expression of immune-related genes.

We observed that after six hours, when the highest circRNA downregulation is found, their linear counterparts are on the contrary, upregulated. Nevertheless, after 10 hours they are all downregulated, and after 24 hours they start to show an upregulation trend (Figure 69).

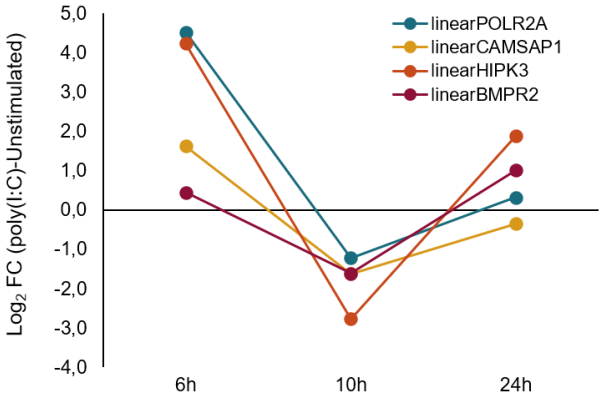

Figure 69. Late response to poly(I:C) shows that linear RNAs are downregulated after 10h and then start to recover. Expression of four circRNAs is shown as the Log_2 of the fold change between the stimulated and unstimulated conditions at 6, 10 and 24h after poly(I:C) stimulation.

We conclude that although linear RNA expression is also changed with poly(I:C) the timing of these changes does not correlate with the timing of the immune response as good as the circRNA expression changes do. Even if these results are quite preliminary and should be further investigated they could go in line with the evidences of a stronger interaction between circRNAs and PKR than between linear RNAs and PKR[118].

Curcumin is not able to prevent neither circRNA downregulation nor PKR phosphorylation upon poly(I:C) stimulation.

Once we have validated that poly(I:C) treatment induces the downregulation of circRNAs and that this hypothetically leads to the release, phosphorylation and activation of PKR resulting in the expression of innate-immune related genes, we wanted to investigate how to prevent these effects. To do so, we first decided to test the effect of curcumin, which has been shown to effectively inhibit the RNase L responsible of degrading circRNAs[344].

Nevertheless, none of the two concentrations we tested (5µM and 20µM) was able to clearly prevent circRNA downregulation. In the case of circPOLR2A, the treatment with 5µM of curcumin slightly reduces the change in its expression upon poly(I:C) stimulation (FC=-2.9 to FC=-2.2). On the contrary, the treatment with 20µM curcumin, far from preventing the circRNA downregulation it increases the changes induced by poly(I:C) stimulation. In line

with these results, although cells grown in curcumin enriched medium seem to have a less pronounced upregulation of GADD34 and IFNB1 when compared to control cells, their expression is still triggered upon poly(I:C) stimulation (Figure 70).

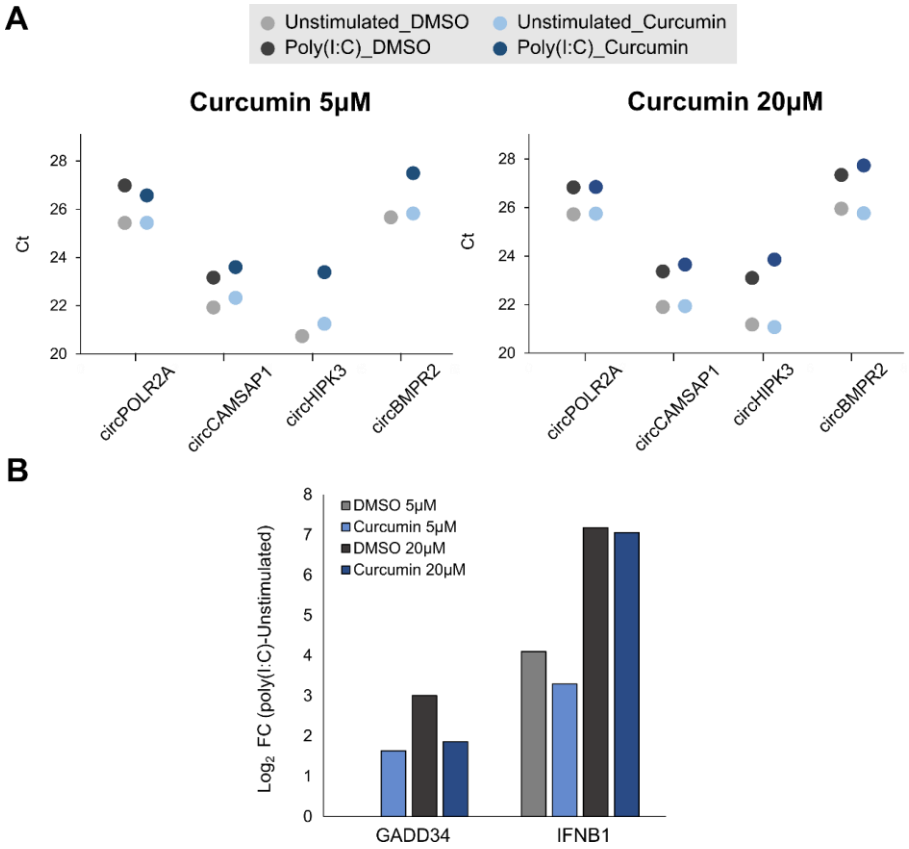

Figure 70. Curcumin is not able neither to stop the innate immune response nor to prevent circRNA downregulation upon poly(I:C). (A) Expression of four circRNAs showed as the raw Ct value obtained by RT-qPCR between the poly(I:C) stimulated and unstimulated cells both with (blue) and without (grey) curcumin treatment. Results for two different curcumin concentrations (5µM and 20µM) are shown. Conditions circHIPK3 and circBMPR2 in 5uM treated cells are missing due to a technical error. (B) The Log2 fold change between stimulated and unstimulated conditions is shown for two innate-immune related genes both with and without curcumin treatment. Results for two different curcumin concentrations (5µM and 20µM) are shown. The expression of GADD34 in the DMSO 5uM condition is missing due to a technical error.

Up to now we have been measuring the expression of some immune genes implicated in the PKR activation cascade as indirect reporters of the PKR activation, however, due to the complexity of the immune response, the expression of those genes could also be influenced by other factors, and thus, we wanted to assess the degree of phosphorylation of PKR as a more direct measure of its activation. To do so, we measured both the total PKR and phosphorylated PKR (p-PKR) concentration by cell-based ELISA and calculated the phosphorylation ratio (p-PKR/total PKR · 100).

Results are not very conclusive since in control cells for which a notable increase of %p-PKR was expected upon poly(I:C) stimulation, the change is very subtle. In the case of the control condition for which 5 µM of DMSO were added the poly(I:C) stimulation even results in a reduction of the %p-PKR. Nevertheless, when the effect of the poly(I:C) is compared between cells treated with curcumin and their respective DMSO controls there is almost no difference (Figure 71).

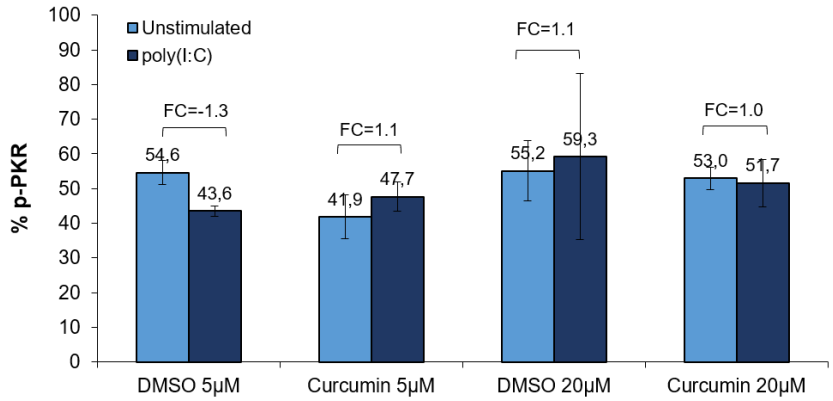

Figure 71. Curcumin has little effect on the PKR phosphorylation upon poly(I:C) stimulation. The % of phosphorylated PKR (out of the total PKR) is shown for the unstimulated cells (light blue) and poly(I:C) stimulated cells (dark blue). Two different curcumin concentrations were tested, and their corresponding controls (DMSO) are also represented. Fold change values between unstimulated and poly(I:C) stimulated cells are shown.

Studies have reported curcumin as a potent anti-inflammatory and antioxidant agent. In fact, some have proposed the therapeutic potential of curcumin against MS [345]. Many different molecular targets of curcumin have been found, from transcriptional factors to enzymes, inflammatory cytokines and protein kinases[345]. Among those, curcumin has also shown evidences of inhibiting RNase L[344]. However, based on the results of our study, we cannot conclude that curcumin could be useful to alleviate the effect of poly(I:C) neither on circRNA expression nor on PKR activation.

In this study, DMSO was added to control cells since it is the solvent were curcumin have been dissolved for its administration. Nevertheless, DMSO itself has recently been shown to possess immunomodulatory effects in the innate and adaptative immunity [346]. The detailed mechanism of the immune modulation by DMSO need still to be elucidated but both effects for immune suppression and immune promotion have been described[346]. These evidences indicate that DMSO may not be an innocuous solvent, and in fact its effects might be interfering with the potential effects of curcumin which could be the reason for our

inconclusive results. Therefore, for future studies, in order to further investigate the effect of curcumin it should be dissolved in a different solvent.

Overexpression of dsRNA containing circRNAs is able to reduce the poly(I:C) induced PKR phosphorylation and the activation of immune-related pathways

In this section, and taking into account that DMSO dissolved curcumin has been shown to be unable to prevent circRNA downregulation and PKR activation, we want to investigate whether the overexpression of circRNAs is able to avoid the activation of PKR and the subsequent activation of the immune response. To do so, we have performed a cell culture experiment with HeLa cells overexpressing different circRNA candidates. As previously introduced, we have selected three different circRNAs based on their structure due to the fact that only circRNAs containing dsRNA regions are predicted to bind and inhibit PKR, whereas those that do not contain these regions have not shown any effect. CircCAMSAP1 and circRELL1 have been selected as examples of dsRNA containing region circRNAs, which will also be called "structured circRNAs" whereas circSMARCA5 is an example of a "unstructured circRNA" or circRNA that does not contain the dsRNA region that enables them to interact with PKR. At the same time an empty pcDNA3 was used as control condition. Each of the conditions was studied with and without poly(I:C) stimulation and in duplicates. On one hand, we have performed the RNA sequencing of the samples to have a wide picture of the changes at the RNA level (not only of a few circRNAs and immune-related genes but for all of them). On the other hand, we have also quantified both the total PKR and p-PKR by ELISA as a measure of the phosphorylation and activation of PKR.

Before sending the samples for RNA-Seq we wanted to validate the correct overexpression of the circRNAs by RT-qPCR. Results indicated that circRELL1, as well as circCAMSAP1 and circSMARCA5 were efficiently overexpressed both in unstimulated and poly(I:C) stimulated conditions (Figure 72).

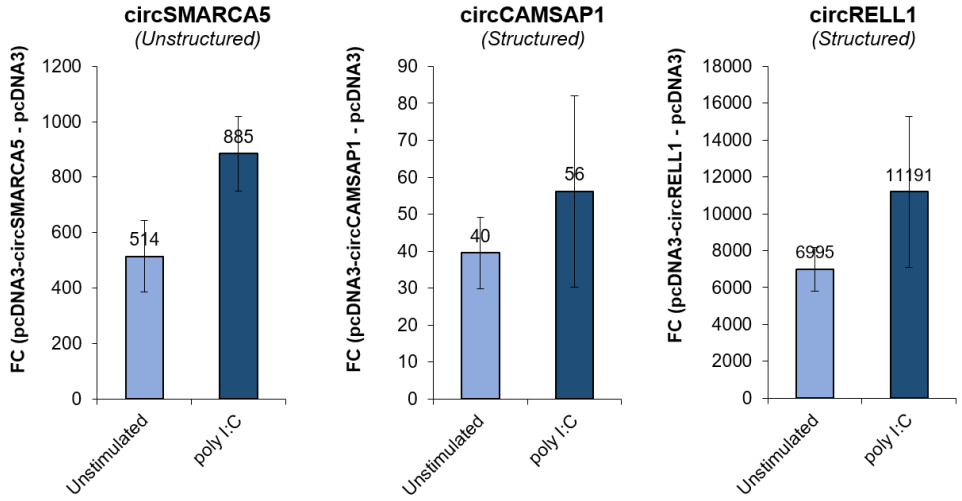

Figure 72. RT-PCR confirms the overexpression of the three circRNAs both in unstimulated and poly(I:C) stimulated conditions. The expression is shown as the fold change between the corresponding circRNA overexpression condition and the control condition in which cells were transfected with and empty pcDNA3. Error bars represent the standard deviation between replicates.

Once we had the RNASeq results, we wanted to double check the correct overexpression of the circRNAs. Interestingly, we found that none of the three overexpressed circRNAs was detected in their respective overexpression conditions by CIRI2. Nevertheless, circExplorer2 confirmed that a high number of circRNA copies were detected in the overexpression conditions when compared to cells transfected with the empty pcDNA3 (Table 11). CircCAMSAP1 and circSMARCA5 are overexpressed a mean of 59 and 362 times respectively. Meanwhile, circRELL1 is expressed 7,035 times more in cells transfected with pcDNA3-circRELL1 when no poly(I:C) was added, however, it could not be detected in the stimulated condition (Table 11).

These results, having previously confirmed the overexpression by RT-qPCR in all the conditions, suggest that circExplorer2 and particularly CIRI2 can show some limitations when quantifying very highly expressed circRNAs. In fact, CIRI2 algorithm has being designed so that it includes an adapted maximum likelihood estimation based on multiple seed matching to identify BSJ reads and to filter false positives derived from repetitive sequences and mapping errors[279]. Due to the fact that the transfection of circRNA plasmids for their overexpression has generated thousands or circRNA copies the maximum defined within the CIRI2 algorithm, could have identified overexpressed circRNAs as false positives reporting a "non-detected" result for all of them.

Table 11. Readcount for circRELL1, circCAMSAP1 and circSMARCA5 detected by circExplorer2 in the different overexpression and stimulation conditions. The number of reads corresponding to the circRNA that has been overexpressed are highlighted in bold.

Overexpression	Stimulation	Replicate	Number of reads by circExplorer2		
			circRELL1	circCAMSAP1	circSMARCA5
Empty pcDNA3	Unstimulated	Replicate 1	1	20	22
		Replicate 2	1	15	32
	poly(I:C)	Replicate 1	2	33	29
		Replicate 2	5	18	18
pcDNA3-circRELL1 *(Structured circRNA)*	Unstimulated	Replicate 1	**6910**	32	18
		Replicate 2	**7161**	35	18
	poly(I:C)	Replicate 1	**ND**	ND	ND
		Replicate 2	**ND**	ND	1
pcDNA3-circCAMSAP1 *(Structured circRNA)*	Unstimulated	Replicate 1	2	**1132**	31
		Replicate 2	ND	**936**	35
	poly(I:C)	Replicate 1	1	**1373**	22
		Replicate 2	1	**1359**	24
pcDNA3-circSMARCA5 *(Unstructured circRNA)*	Unstimulated	Replicate 1	2	43	**7233**
		Replicate 2	5	29	**6564**
	poly(I:C)	Replicate 1	3	27	**11968**
		Replicate 2	3	15	**9040**

With regard to the rest of the circRNAs, and based on what it has been described in literature[118] and in our RT-qPCR results we expected a global downregulation of circRNAs upon poly(I:C) stimulation, in all conditions independent from the different overexpression conditions since they have not been described to interfere with RNase L. As expected, we found that although some circRNAs were upregulated, the majority of circRNAs are downregulated and represent between the 56.4 and the 71.9% of the total number of DE circRNAs for each condition (Figure 73).

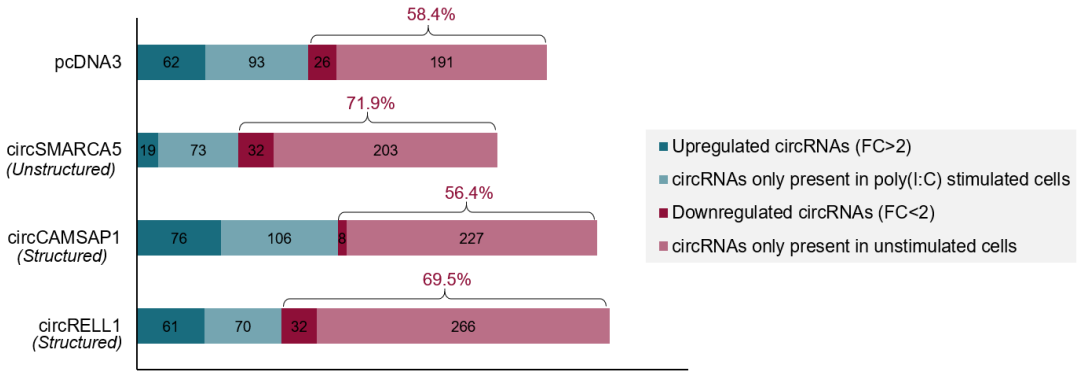

Figure 73. CircRNAs whose expression is changed upon poly(I:C) stimulation in the four different overexpression conditions. Both circRNAs present in the four replicates and differentially expressed and circRNAs that are present only in one of the conditions are represented.

Interestingly, and although the majority of circRNAs are downregulated upon poly(I:C) stimulation, the expression of the overexpressed circRNAs continues to be high, and in fact it is increased (Figure 72). This could allow structured circRNAs to still bind PKR and keep it inhibited upon poly(I:C) stimulation.

To test this hypothesis, we evaluated the degree of PKR phosphorylation in each of the overexpression conditions. ELISA results showed that in unstimulated conditions the percentage of phosphorylated PKR ranges from 32.4 to 45.4%. Interestingly, even in absence of poly(I:C) the overexpression of any of the circRNA candidates seems to promote an increase in the phosphorylation of the PKR when compared to those cells that were transfected with an empty vector (Figure 74). Accordingly, pcDNA3-circRELL1 which is the plasmid producing the highest number of circRNA copies is the one in which the % of p-PKR is more increased with respect to the control condition.

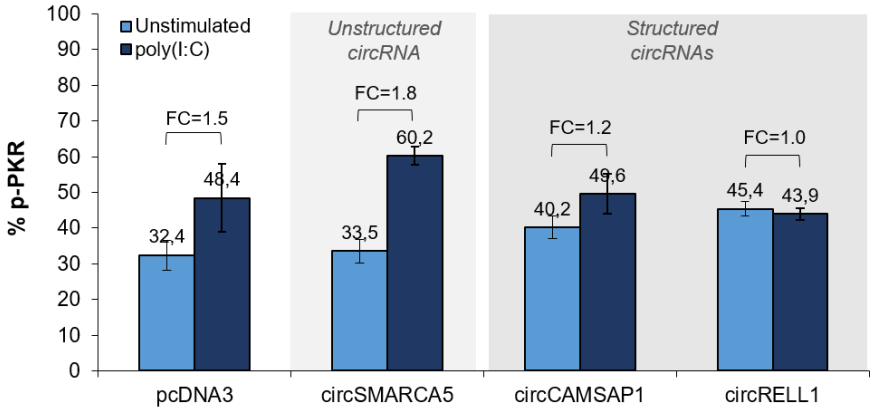

Figure 74. Overexpression of structured circRNAs is able to prevent the increase on the %of p-PKR upon poly(I:C) stimulation. The % of phosphorylated PKR (out of the total PKR) is shown for the unstimulated cells (light blue) and poly(I:C) stimulated cells(dark blue). Three different overexpression conditions were tested and for each of them the fold change values between unstimulated and poly(I:C) stimulated cells are shown. Error bars represent the standard deviation between replicates.

On the other hand, this time, the control condition consisted on cells transfected with the empty vector but without any DMSO and we could confirm that, as expected, the %p-PKR is increased 1,5 fold upon poly(I:C) stimulation reinforcing the idea of DMSO exerting some immunomodulatory effect in the previous experiment. The 1,5 fold increase of the p-PKR form in the empty pcDNA3 condition will be considered as the "normal" response. Compared to this "normal" response, we found that cells overexpressing an unstructured circRNA such as circSMARCA5, have a greater increase of the p-PKR (FC=1,8) upon stimulation. On the contrary, cells overexpressing circRNAs with dsRNA regions such as circCAMSAP1 and

circRELL1 have a less pronounced response to the stimulation. Interestingly, cells overexpressing circRELL1, whose overexpression is higher than for circCAMSAP1, even shows a small decrease of the p-PKR percentage after poly(I:C) stimulation suggesting that the effect could be concentration dependent (Figure 74).

Based on these results we can confirm that although the overexpression of circRNAs can initially promote a small increase on the %p-PKR, the presence of a high number of structured circRNAs seems to be able to at least partially prevent the phosphorylation of PKR upon poly(I:C) stimulation. However, it is worth noting that still more than the 40% of the total PKR is phosphorylated and that a higher number of structured circRNA copies could be needed to reduce this proportion.

Even if part of the total PKR is activated in the four overexpression conditions, the difference between a 60.2% of p-PKR in cells were circSMARCA5 is overexpressed and the 49.6-43.9% of p-PKR in cells were the structured circRNAs circCAMSAP1 and circRELL1 have been overexpressed could be enough to see some difference on the activation of the PKR downstream immune-response pathways. To this aim we studied the changes induced by the poly(I:C) stimulation at the linear transcriptome level.

More than 500 linear transcripts were differentially expressed between the unstimulated and poly(I:C) stimulated cells in each of the four overexpression conditions (Figure 75A). All the genes producing at least one of the DE linear RNAs were selected and subjected to a gene ontology analysis in order to know which are the main functions affected. Unsurprisingly, we found that most of the genes whose expression has been altered are implicated in different immune functions. Results indicate that the DE linear RNA dataset form control cells and from cells in which circSMARCA5 was overexpressed are enriched in genes coding for proteins involved in 20 and 25 different immune related processes such as response to virus or regulation of cytokine-mediated signaling pathways (Figure 75B). Interestingly, and in line with the PKR phosphorylation results, the number of immune-related enriched terms is slightly decreased for cells overexpressing structured circRNAs. In the case of cells overexpressing circCAMSAP1, 17 immune-terms are enriched, and remarkably, circRELL1 overexpression only results in the enrichment of 7 immune-related terms while there are other seven that are related to other processes such as regulation of metabolic processes (Figure 75B).

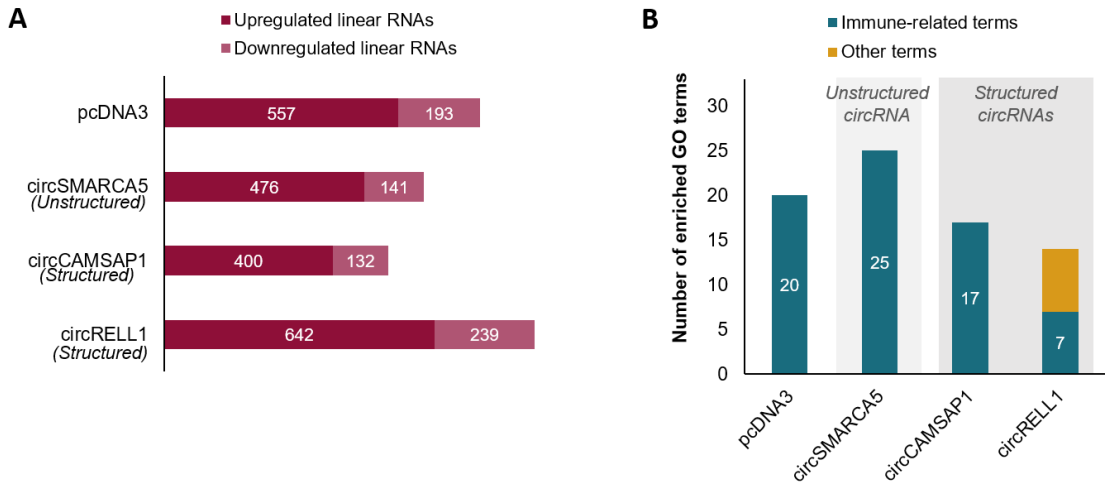

Figure 75. Linear RNAs whose expression is changed upon poly(I:C) stimulation in the four different overexpression conditions are mainly involved in immune-processes. A) Number linear RNAs that are up or downregulated after poly(I:C) stimulation in the four different overexpression conditions. B) Bar plot showing the number of enriched GO-terms for each condition classified as immune related or others.

It is worth highlighting that among the genes whose fold change upon stimulation changes between the different overexpression conditions, we find the gene coding for PKR (EIF2AK2). Interestingly, the overexpression of circSMARCA5 apart from promoting the phosphorylation of the available PKR, it also induces an increase in the expression of more PKR, whereas the overexpression of structured circRNAs, which partially prevented the phosphorylation of PKR, do also experiment a smaller increase of the expression of EIF2AK2 (Figure 76)

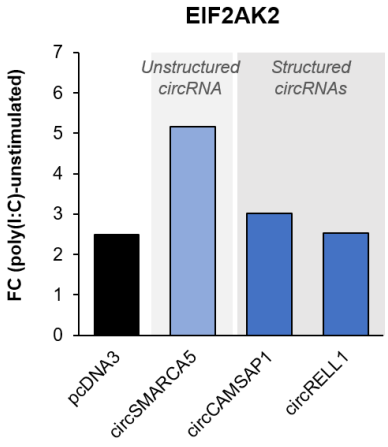

Figure 76. Expression of EIF2AK2 upon poly(I:C) stimulation and in the four different overexpression conditions. Expression is represented as the fold change between the unstimulated and poly(I:C) stimulated conditions. Control cells transfected with the empty pcDNA3 plasmid are coloured in black, unstructured circRNAs are coloured in light blue and structured circRNAs in dark blue.

To sum up, in this section we have found that although the overexpression of circRNAs induces a small increase in the PKR activation, the artificial increase on the pool of dsRNA region containing circRNAs attenuates the activation of PKR and its downstream cascade upon a stimulus such as the poly(I:C). On the contrary, the overexpression of an unstructured circRNA such as circSMARCA5 exerts a synergic effect with poly(I:C) potentiating the immune response. Nevertheless, it should be further investigated whether it is an effect specific for circSMARCA5 or intrinsic to any other unstructured circRNA.

PKR phosphorylation degree is not significantly changed between relapse and remission phases in MS

Dysregulation of immune response plays a critical role in MS, a disease that courses in relapse and remission phases. During the relapse phase the immune response is active, whereas in remission it resolves. In subchapter 1.2 we reported a global upregulation of circRNAs in MS patients in remission which together with the global changes in the circRNA expression reported in other autoimmune diseases[118,291,347], could suggest that the cross-talk between circRNAs and PKR described by Liu et. al for SLE patients may be inherent to other autoimmune diseases.

In HeLa cells we have observed that upon poly(I:C) stimulation, circRNAs are downregulated and PKR phosphorylation is increased followed by and activation of the innate immune response.. Besides, in subchapter 1.2 we observed that for some patients circRNA expression correlates with what we observed in HeLa cells and present a decreased circRNA expression in relapse when compared to remission.

In light of these results and with the hypothesis that the crosstalk mechanism between circRNA and PKR could be one of the mechanisms underlying the disease, we decided to study the phosphorylation state of PKR in patients in relapse and remission phases. The rationale behind this experiment was that during relapse the immune response is active, while it resolves during the remission phase, although we have to keep in mind that the immune mechanisms underlying these processes are much more complex.

Results indicate that there is a big variability between samples both in the total PKR amount and also in the proportion of PKR that is phosphorylated (Figure 77). This variability could be explained by the heterogeneity of the disease itself, and particularly between relapses that could be more or less severe from one to other patient. Contrary to what was expected, the phosphorylation degree of p-PKR in relapse is apparently smaller than in remission, although differences are not statistically significant (Figure 77). In order

to validate this result, more samples should be studied. However, relapse samples are usually drawn at the moment when the patient presents neurologic symptoms and visits the hospital. Thus, we speculate that the circRNA degradation/PKR phosphorylation mechanism might be a very early response that may be happening even before the appearance of the symptoms and that could be the reason why we do not observe the expected increase of p-PKR.

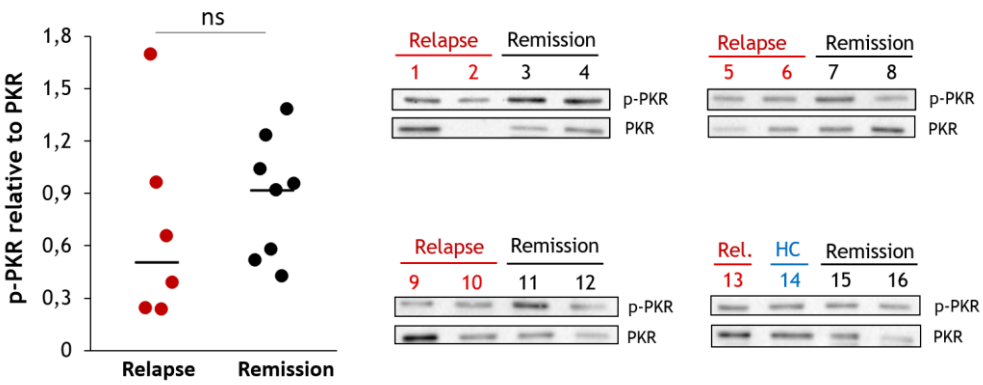

Figure 77. Western blot results showing the phosphorylation levels of PKR in MS patients in relapse and remission phases. P-PKR levels were calculated relative to the intensity of total PKR after total protein normalization. Lines in the dotplot represent the median value of p-PKR.

Conclusion

The research carried out in this subchapter validates several of the observations by Liu et al.[118] reinforcing the evidence on the mechanism describing that circRNAs containing dsRNA regions are able to bind and keep PKR in an inactive state until a stimulus such as a viral infection or the poly(I:C) stimulation activates the RNase L that globally reduces the number of circRNAs resulting in the phosphorylation of PKR.

Additionally, it also studies the recovery phase after the poly(I:C) stimulation, and interestingly reports that after a while, the circRNA expression is again boosted. These results are in line with the findings showing a global upregulation of circRNAs in MS patients in remission when compared to controls. We have additionally studied the phosphorylation level of PKR in PBMCs from patients in relapse and remission phases, but no difference was observed. Yet, a more extensive characterization of the circRNA expression changes and the changes in PKR phosphorylation levels between relapse and remission phases are needed to reach a conclusion. However, these evidences together with the results indicating that the overexpression of dsRNA containing circRNAs could prevent the activation of the PKR cascade can lead to think that circRNAs could be use as therapeutic targets to reduce

the rate of relapses in RR-MS patients although further research is needed. The administration of exogenous circRNAs have reported differing conclusions on their immunogenicity[51,155,156]. Some investigations indicate that cells distinguish between endogenous and exogenous circRNAs based on their modifications or on the RBPs bound to the circRNAs which could be used in order to create circRNAs that could evade the immune response. In this line, different vehicles such as nanoparticles, liposomes or exosomes have been tested for vehiculation of different RNAs as a treatment for MS [348,349], a field that can be incorporated to circRNA research.

General Discussion

The aim of this general discussion is to provide a general overview of the data presented in this thesis. To do so, I will go through the process of constructing this thesis having a comment on the contributions made and the missing gaps for future work. I will also try to discuss from my point of view some aspects of the research that have not been included in previous sections. Besides, I will also gather some of the thoughts that have crossed my mind during these four years as well as those that have emerged from the uncountable discussions with my supervisors and labmates.

First, I will like to start with a comment on biomedical research and translational medicine. Unfortunately, diseases are part of our life from the very beginning, irremediably affecting many unknown people around us, but also our close relatives or even us. The process of growing involves learning how to face the diseases, live with them and accept them. Besides, diseases can also give as a few life lessons. But as biomedical researcher, it also involves many other learning processes. Human diseases are the perfect scenario to learn about the complexity or the human body and all the processes that take place at the same time an interact one with each other. Particularly during this thesis, I have had the opportunity to get continuously surprised by the complexity of the gene expression regulation but also to take a first step in to what I now consider two of the most complex systems: the immune and nervous systems. Working with patient's samples, and gathering all their demographical, clinical and biological data has also make me aware of how unique each person is, and of how important is to gather as many information as possible towards a more personalized and precise medicine.

This thesis has been focused on studying Multiple sclerosis. Multiple sclerosis (MS) is the first most common cause of disability in young adults which courses with very different and unpredictable symptoms ranging from numbness or blurred vision to talking difficulties and movement impairment. Nowadays, thanks to the advances in the diagnosis and the advent of new treatments over the last 20 years the quality of life of patients has been greatly improved. MS itself is at the moment rarely fatal and the life-threatening complications that may arise from the most severe MS phenotypes are also getting less frequent. Even with all these advances and with all the research effort from many years, the MS pathophysiology and cause are still largely unknown. Therefore, nowadays the MS research is focused in the global characterization of the disease from different points of view in order to have a big

picture that could provide researchers and clinicians with clues that could not only treat the disease but ideally cure it and also tools that could help for an earlier diagnostic and a better assessment of the treatment. To this aim, some of the current research lines involve the genetic, transcriptomic, proteomic and metabolomic characterization of patients, as well as the study of the gut microbiota and its communication with the immune system and the central nervous system which has gained attention in the last years.

In this context, and based on the previous expertise of the group, this thesis was aimed to contribute on the characterization of MS at the transcriptome level. Previous works have shown that MS patients present a dysregulated transcriptome both of immune cells and cells from the central nervous system (CNS). Both the immune system and the CNS are closely related to the disease, and thus, the study of any of them could report very interesting insights on the disease. However, the biggest difference between immune and CNS cells is their accessibility. Immune cells can be found circulating in blood, and thus can be accessed by a rudimentary protocol such as the venepuncture. Cells from the CNS, instead are usually only accessible in post mortem brains although a minimal cellular fraction could also be obtained from lumbar puncture. With the double aim of further characterizing the disease but also discovering some potential biomarkers, we decided to focus our efforts in studying the circulating transcriptome of immune cells and plasma derived extracellular vesicles which would results in minimally invasive biomarkers. For the beginning of this project in 2017, several studies have already characterized and reported interesting results from the profiling of the coding transcriptome and part of the non-coding transcriptome, mainly including miRNAs.

At that moment, a newly rediscovered transcript type, circular RNAs (circRNAs), was gaining attention in the RNA field due to its potential to interact and regulate the function of the already well-known miRNAs. At the same time, circRNAs were also opening their path into biomedical researchers due to their implication in many biological processes and diseases as well as for their potential as biomarkers. In this context, we set out the thesis project with two aims: 1) characterising the expression of circRNAs in circulating immune cells in order to integrate these data with the data from mRNAs and miRNAs towards the completion of the whole transcriptome network 2) discovering potential and non-invasive biomarkers for the disease.

In the way towards the completion of this project, we have waged some battles and learned some things about technical and biological aspects of circRNAs.

CircRNAs have turned around the way transcriptome was studied with their circular structure. Their circularity has kept them away from the spotlight for many years, and when they were discovered to be abundant, endogenously produced and conserved among species, they forced the methodologies used for the study of the transcriptome to adapt so they could be studied. On one hand, classical RNASeq preparations used to perform a selection of polyadenylated RNAs consequently removing all non-polyadenylated transcripts as circRNAs[350], had to be avoided. However, even within libraries with no polyA selection, circRNAs are usually expressed at low levels, and in absence of good sequencing depth[351], they could also be missed. To overcome this limitation the treatment of total RNAs with a RNase R has been postulated as the gold standard technique for the circRNA study[94]. RNase R is able to selectively degrade linear RNAs resulting in the selective enrichment of circRNAs. Nevertheless, the RNAse R treatment conditions have not been well stablished resulting in a high variability in RNase treated samples, with linear RNAs that could be resistant to RNase R and circRNAs that could be sensitive under some conditions[79,94]. Similarly, oligodT primers have to be replaced for universal primers in RT-PCRs. On the other hand, apart from ensuring that circRNAs are not removed from the sample to study different experimental and computational tools have been developed in order to actively detect circRNAs and validate their circularity. The main feature of circRNAs is their backspliced junction (BSJ) and thus, the already available techniques for detection of linear RNAs have been adapted for the detection of this BSJ: commercial arrays containing probes that specifically target the BSJ have been designed, convergent primers have been replaced by divergent primers facing away from each other in the genome so that they can only amplify the BSJ in case of a circularization event[352], and bioinformatic algorithms focused on the detection of chimeric reads that do only map on a putative BSJ have been developed [252] as three of the most common examples. Moreover, techniques as the northern blot as well as other electrophoretic techniques have also been used to confirm the circularity of the transcripts[93].

Among all the available techniques we have used microarrays and RNASeq for the screening of circRNAs. On one hand, we performed a first approach using microarrays which have the advantage of not requiring a big bioinformatical background for their analysis, however, the number of circRNAs studied is limited and due to the fact that only one circRNA specific array is commercially available, the freedom to chose different probes, protocols and analysis methods is very limited. To avoid these limitations, we switched to RNASeq and chose to sequence samples of total RNA with rRNA depletion with a high sequencing depth. We chose this protocol instead of the sequencing of RNase R treated samples for two main reasons, not to create any bias with RNase R treatment and to obtain information of both

the linear and circular transcriptome at the same time. When analysing the RNASeq data, many different pipelines have been described in literature, and due to the fact that it is still a nascent field of research, still many other pipelines and algorithms are being developed, but a standard analysis pipeline has not been stablished. In this thesis, we have used two different pipelines for circRNA and linear RNA detection, the first one combines CIRI2 and find_circ for the detection of circRNAs and HTSeq for linear RNAs, whereas the second one combines CIRI2 and CircExplorer2 for circRNAs and uses Kallisto for the detection of linear transcripts. Both pipelines have shown to be able to detect a high number of circRNAs with readcounts ranging from 1 to thousands of reads and the combination of their results has allowed to filter out some of the circRNAs that could be potential false positives[95]. Interestingly, it is worth mentioning that although CIRI2 quantification has been described as one of the most accurate, we have identified a limitation when detecting circRNAs that have been overexpressed that are considered as reads coming from highly repetitive regions and thus they are filtered out. Regarding the linear RNAs, both HTSeq and Kallisto have effectively detected linear transcripts, but it should be mentioned that HTSeq has been used to quantify at the gene level whereas Kallisto has quantified at the transcript level detecting different linear RNAs.

For the validation of circRNA expression by PCR, we have generally used RT-qPCR. In the first chapter, we also set up a protocol for studying circRNA expression in cultured PBMCs by ddPCR. This experiment allowed us to compare the sensitivity of both techniques and against what has been described in other studies[126,353], ddPCR does not improve the sensitivity of RT-qPCR for the detection of circRNAs. Regarding the design of divergent primers, we have also progressively improved the design by taking into account different aspects as the exon composition of each circRNA or the presence of other transcripts that may have common exons to ensure the specificity of the primers.

Although it has not been included in this thesis, it is worth mentioning that during its development, a study on circular DNAs has revealed that circular DNAs of chromosomal origin are common in eukaryotic genomes and that they can be transcribed[10]. Transcription events across the junction of circular DNAs would result in a transcript with a junction similar to the BSJ of circRNAs[354]. This evidences still have to be supported by other studies, but the existence of such chimeric transcripts adds a new confounding factor that must be taken into account when studying circRNAs. Thus, previously discussed techniques take particular importance in order to distinguish circRNAs from chimeric transcripts from circular DNA origin.

With regard to other techniques such as the northern blot or the construction of plasmids for the overexpression of circRNAs, they may have not got as much importance as the previously mentioned techniques, in this thesis. But I am sure that they would be very valuable methodologies to the experiments to come.

In terms of the biological aspects of circRNAs learnt during this thesis, I would say that I end this work with more open questions than confirmed responses. We set out this project with the strong believe of circRNAs acting as sponges of miRNAs. However, with the growing evidence that has emerged during the last four years and also supported by the short but revealing functional approximation presented in chapter two, I currently believe that only a few particular circRNAs have the potential to exert this function. The protein coding potential of circRNAs is the second most studied circRNA function. However, it has not been studied in this thesis since this function was very recently described and still there are few and contradictory evidences that make difficult to understand the real relevance of this function. As it has been described for the miRNA sponge function, probably only a few circRNAs may be translatable. However, recent studies have pointed to the extensive translation of circRNAs driven by the m6A modification[144]. The presence of open reading frames and internal ribosome entry sites have been used as features that indicate the translatability of circRNAs. Moreover, detection of circRNAs in RiboSeq libraries of polysome fractions have been proposed as evidences of circRNAs coding for proteins[355]. However, in some cases, circRNAs containing ORFs and IRES where not detected in RiboSeq libraries or polysome fractions[329] and even circRNAs detected in these libraries did not produce peptides that could be detected by mass spectrometry[356]. Those works evidence the importance that this field is taking, in an attempt to elucidate whether circRNA can give rise to proteins, but there is still much more work to characterize the feature that such a circRNA has to fulfil.

During the development of this thesis, a publication by Liu et al. [118] presented for the first time a circRNA function that instead of being performed by a single circRNA, was performed by a whole set of circRNAs acting together. In addition, this function was involved in the regulation of the immune response and described in an autoimmune disease such as systemic lupus erythematosus (SLE). This publication emerged at the time when we have performed the RNASeq on leukocytes from MS patients and HCs and found that circRNAs were globally upregulated in patients. Interestingly, at the same time results from a circRNA profiling study in psoriasis[291], reported a global downregulation of circRNAs. And later on, a study in Sjögren syndrome[347] also reported a global upregulation of circRNAs. Therefore, it was tempting to speculate that these general changes of the whole

circRNA profiles in different autoimmune diseases, are indicative of the implication of circRNAs in the regulation of the aberrant immune response as a group of molecules. The mechanism proposed by Liu et al. [118] indicated that circRNAs form several hairpin structures that allow them to interact with the innate immune sensor PKR and block its activation. The inhibition of PKR activation resulted in a reduction of the aberrant immune response. This mechanism could explain the relation between the global circRNA downregulation found in SLE and in psoriasis, and the aberrant activation of PKR and the immune response. However, based on the mechanism described by Liu et al. the global upregulation of circRNAs observed in MS and Sjögren syndrome should theoretically favour the inhibition of the PKR activation and the immune response, which could be contradictory in the context of an autoimmune disease. Nevertheless, it is known that MS does usually present a relapsing remitting course in which there is a continuous cycle of mechanisms related to the immune attack and the recovery phase. Thus, we suggest that the global increase or decrease of circRNA expression could also be part of the mechanisms that are changed between relapse and remission phases. This hypothesis was in line with the global upregulation observed un leukocytes from patients whose samples were drawn during remission. To further validate this hypothesis, we analysed the expression of the best seven candidates reported from the RNASeq in relapse and remission and observed that again there is a global change in the expression of the seven candidates. However, two different response patterns were identified indicating that whereas some patients presented the expected downregulation of circRNAs in relapse with respect to the remission, some other patients presented the opposite pattern. Due to the small number of samples and circRNAs analysed in this experiment, and in light of these results, we could not confirm or discard the hypothesis of circRNAs regulating the immune response in MS via PKR. Nevertheless, in light of the evidences supporting the implication of circRNAs in different autoimmune diseases, we decided to further investigate this function in a second subchapter. As it has been discussed our results could be promising but further experiments are needed to better characterize the cross-talk between circRNAs and PKR and its implication in the immune regulation.

To sum up, this thesis has contributed to the field by reporting a characterization of the circRNA expression in different sample types from MS patients including leukocytes, PBMCs and EVs. These profiling experiments open a new research line in MS which could on the future report very interesting information on the pathologic processes underlying the disease. Moreover, we have contributed with a first approximation to understand whether the implication of circRNAs in MS could be linked to their potential function on the regulation of the immune response although further research is needed. Lastly, we could

satisfy one of the objectives proposed at the beginning of the thesis, since along the different characterization experiments, we have drawn eleven circRNAs (and three linear RNAs) that have shown potential to be diagnostic or prognostic biomarkers of the disease. Nevertheless, I am aware that this thesis has only put them at the beginning of the path towards the bedside, and that there is still much work left in order to validate their potential in more MS cohorts as well as in other immune diseases.

Conclusions

1. CircRNA expression is altered in RR-MS patients' leukocytes when compared to healthy controls.

2. The leukocyte expression of ANXA2 derived circRNAs circ_0005402 and circ_003452_2 and the expression of their mRNA is able to distinguish RR-MS patients from healthy controls. Their expression is changed with the evolution time and thus, they could also be used to monitor the evolution of the disease.

3. There is a global upregulation of circRNA expression in leukocytes from MS patients compared to controls which is independent from the linear transcriptome expression.

4. The circRNA upregulation pattern in MS is sex specific. Female patients show a clear upregulation pattern for circRNAs whereas this pattern is not found in male patients. Moreover, a sex specific circRNA signature is found pointing to the existence of different molecular pathological mechanisms between both sexes.

5. The expression of six circRNAs (circATP8B4, circAGFG1, circPADI4, circAFF2, circABCA13 and circNEIL3) in leukocytes is identified as a potential diagnostic biomarker for MS. Their expression is also changed between the relapse and remission phases for most of the patients and could be a potential biomarker for relapse.

6. The global transcriptome of PBMCs from patients with positive LS-OCMBs is different from those patients with negative LS-OCMB status.

7. The expression in PBMCs of the circRNAs circ_000478 and circ_0116639 and the linear IRF5 and MTRNR2L8 transcripts could be used as minimally invasive biomarkers to identify patients with highly active disease as an alternative to the current CSF biomarkers.

8. CircRNAs are abundantly detected in plasma derived EVs and represent the second most frequent transcript type in this sample, but not in leukocytes.

9. CircRNA EV cargo tend to be smaller than the whole circRNA pool from leukocytes indicating a regulated selective export of circRNAs.

10. Different transcriptomic profiles are found in EVs from MS patients when compared to healthy controls.

11. MS associated circRNAs do not hold the potential to function as miRNA sponges, but some of them are highly structured, indicating that they could potentially interact with PKR and regulate the immune response.

12. CircRNA expression is decreased in HeLa cells upon poly(I:C) stimulation resulting in the phosphorylation of PKR and upregulation of immune response genes. This phenotype is partially alleviated with the overexpression of highly structured circRNAs, such as the MS-related circRNA circRELL1, indicating that they could have some therapeutical use in controlling aberrant immune responses.

Publications, patents and communications

Research articles during the thesis

- **Iparraguirre L**, Muñoz-Culla M, Prada-Luengo I, Castillo-Triviño T, Olascoaga J, Otaegui D. Circular RNA profiling reveals that circular RNAs from ANXA2 can be used as new biomarkers for multiple sclerosis. Hum Mol Genet. 2017 Sep 15;26(18):3564-3572. doi: 10.1093/hmg/ddx243. PMID: 28651352.

- **Iparraguirre L**, Prada-Luengo I, Regenberg B, Otaegui D. To Be or Not to Be: Circular RNAs or mRNAs From Circular DNAs? Front Genet. 2019 Oct 11;10:940. doi: 10.3389/fgene.2019.00940. PMID: 31681407; PMCID: PMC6797608.

- **Iparraguirre L**, Alberro A, Sepúlveda L, Osorio-Querejeta I, Moles L, Castillo-Triviño T, Hansen TB, Muñoz-Culla M, Otaegui D. RNA-Seq profiling of leukocytes reveals a sex-dependent global circular RNA upregulation in multiple sclerosis and 6 candidate biomarkers. Hum Mol Genet. 2020 Oct 8:ddaa219. doi:10.1093/hmg/ddaa219. Epub ahead of print. PMID: 33030201.

- **Iparraguirre L**, Olaverri D, Blasco T, Sepúlveda L, Castillo-Triviño T, Espiño M, Costa-Frossard L, Prada A, Villar L-M, Otaegui D, Muñoz-Culla M. Whole-transcriptome analysis in peripheral blood mononuclear cells from patients with lipid-specific oligoclonal IgM band characterization reveals two circular RNAs and two linear RNAs as biomarkers of highly active disease. Biomedicines. Accepted the 23 of November 2020

- **Iparraguirre L**, Alberro A, Sepúlveda L, Osorio-Querejeta I, Moles L, Castillo-Triviño T, Hansen TB, Muñoz-Culla M, Otaegui D. CircRNAs join the extracellular vesicle world in multiple sclerosis. In preparation.

Co-authored research articles during the thesis

- Arnaiz E, Sole C, Manterola L**, Iparraguirre L**, Otaegui D, Lawrie CH. CircRNAs and cancer: Biomarkers and master regulators. Semin Cancer Biol. 2019 Oct;58:90-99. doi: 10.1016/j.semcancer.2018.12.002. Epub 2018 Dec 11. PMID: 30550956.

- Osorio-Querejeta I, Carregal-Romero S, Ayerdi-Izquierdo A, Mäger I, A NL, Wood M, Egimendia A, Betanzos M, Alberro A, **Iparraguirre L**, Moles L, Llarena I, Möller M, Goñi-de-Cerio F, Bijelic G, Ramos-Cabrer P, Muñoz-Culla M, Otaegui D. MiR-219a-5p Enriched Extracellular Vesicles Induce OPC Differentiation and EAE Improvement More Efficiently Than Liposomes and Polymeric Nanoparticles.

Pharmaceutics. 2020 Feb 21;12(2):186. doi: 10.3390/pharmaceutics12020186. PMID: 32098213; PMCID: PMC7076664.

- Alberro A, Saenz-Antoñanzas A, Alvarez-Satta M, **Iparraguirre L**, Mateo-Abad M, Berna-Erro A, Garcia-Puga M, Carrasco-Garcia E, Lopez de Munain A, Vergara I, Otaegui D, Matheu A. Transcriptome analysis reveals the association of EGR1, DDX11L1 and miR454 expression with frail individuals. *Under review*

- Moles L, Delgado S, iMSMS, Sepúlveda L, Alberro A, **Iparraguirre L**, Suarez AJ, Romarate L, Arruti M, Castillo-Triviño T, Muñoz-Culla M, Otaegui D. Microbial Dysbiosis and Lack of SCFAs Production on the Gut of Multiple Sclerosis Patients in a Spanish Cohort. *Under review*

Research articles previous to the thesis

- **Iparraguirre L**, Gutierrez-Camino A, Umerez M, Martin-Guerrero I, Astigarraga I, Navajas A, Sastre A, Garcia de Andoin N, Garcia-Orad A. MiR-pharmacogenetics of methotrexate in childhood B-cell acute lymphoblastic leukemia. Pharmacogenet Genomics. 2016 Nov;26(11):517-525. doi: 10.1097/FPC.0000000000000245. PMID: 27649261.

- Lopez-Santillan M, **Iparraguirre L**, Martin-Guerrero I, Gutierrez-Camino A, Garcia-Orad A. Review of pharmacogenetics studies of L-asparaginase hypersensitivity in acute lymphoblastic leukemia points to variants in the GRIA1 gene. Drug Metab Pers Ther. 2017 Mar 1;32(1):1-9. doi: 10.1515/dmpt-2016-0033. PMID: 28259867.

Patent

- Request for grant of a Spanish Patent (P202030782).
  Applicant: Administration of the Autonomous Community of the Basque Country.
  Title: New biomarkers for diagnosis and monitoring of multiple sclerosis

    - Inventors: Leire Iparraguirre, Maider Muñoz-Culla, Tamara Castillo-Triviño and David Otaegui

Scientific communications as first author

- **Leire Iparraguirre**. CircRNAs as multiple sclerosis biomarkers. 1st Neurosciences Annual Meeting 2019. Biodonostia health Research Institute, Donostia-San Sebastain, Spain (08-09 January 2020) **Oral presentation**

- **Leire Iparraguirre**; Lucia Sepulveda; Ainhoa Alberro; Iñaki Osorio Querejeta; Laura Moles; Tamara Castillo Triviño; Maider Muñoz Culla; Thomas B Hansen; David Otaegui. RNA-Seq profiling of circular RNAs in leucocytes of multiple sclerosis patients reveals sex-dependent expression patterns and 6 candidate biomarkers. *EMBL Symposium: The Non-Coding Genome*. Organized by The European Molecular Biology Organization. Heidelberg, Germany (16-19 October 2019). **Poster presentation**

- **Leire Iparraguirre**. Circular RNAs as a new key piece in the transcriptome regulation network for multiple sclerosis. *UPV/EHUko II. Doktorego Jardunaldiak. II Jornadas Doctorales de la UPV/EHU*. Euskal Herriko Unibertsitatea/ Universidad del País Vasco. Bilbo, Spain (3 July 2019). **Oral presentation**

- **Leire Iparraguirre Gil**; Iñigo Prada Luengo; Ainhoa Alberro; Iñaki Osorio Querejeta; Laura Moles; Tamara Castillo Triviño; Maider Muñoz Culla; David Otaegui. ANXA2 circular DNAs are differentially expressed in multiple sclerosis patients. *34th ECTRIMS*. Organized by the European Committee for treatment and research in multiple sclerosis Berlin, Germany. (10-12 October 2018). **Poster presentation.**

- **Leire Iparraguirre**; Maider Muñoz Culla; Iñigo Prada Luengo; Tamara Castillo Triviño; Javier Olascoaga; David Otaegui Bichot. CircularRNA profiling reveals that circularRNAs from ANXA2 could be biomarkers for multiple sclerosis. *Workshop on Non-coding RNA-mediated metabolic regulation in health and Disease.* Organized by the Universidad Internacional de Andalucía. Baeza, Andalucía. (06-08 November 2017). **Poster presentation**

- **Leire Iparraguirre**; Maider Muñoz Culla; Iñigo Prada Luengo; Tamara Castillo Triviño; Javier Olascoaga; David Otaegui Bichot. CircularRNA profiling reveals that circularRNAs from ANXA2 could be biomarkers for multiple sclerosis. *7th Joint ECTRIMS-ACTRIMS meeting*. Organized by the European Comitte Treatment and Research in Multiple Sclerosis. Paris, France. (25-28 October 2017). **Poster presentation.**

References

1. Cobb, M. 60 years ago, Francis Crick changed the logic of biology. *PLoS Biol.* 2017, *15*, 1–8, doi:10.1371/journal.pbio.2003243.
2. National Genome Reseacrh Institute. NIH Available online: https://www.genome.gov/about-genomics/fact-sheets/Deoxyribonucleic-Acid-Fact-Sheet (accessed on Apr 8, 2020).
3. Nelson, D.L.; Cox, M.M. Nucleótidos y ácidos nucleicos. In *Lehninger. Principios de Bioquímica*; Omega, Ed.; 2009; pp. 271–302.
4. Dale, J.W.; Von Schantz, M.; Plant, N. From genes to genomes. In *From genes to genomes*; Wiley-Blackwell: Oxford, 2013; pp. 1–25.
5. Miller, O.J.; Therman, E. The chemistry and packaging of chromosomes. In *Human Chromosomes*; Springer: New York, 2001; pp. 61–74.
6. Annunziato, A. DNA Packaging: Nucleosomes and Chromatin. *Nat. Educ.* 2008, *1*, 26.
7. Nelson, D.L.; Cox, M.M. Genes y cromosomas. In *Lehninger. Principios de Bioquímica*; Omega, Ed.; 2009; pp. 947–974.
8. Shibata, Y.; Kumar, P.; Layer, R.; Willcox, S.; Gagan, J.R.; Griffith, J.D.; Dutta, A. Extrachromosomal MicroDNAs and Chromosomal Microdeletions in Normal Tissues. *Science (80-.).* 2012, *336*, 82–86, doi:10.1126/science.1213307.
9. Dillon, L.W.; Kumar, P.; Shibata, Y.; Wang, Y.H.; Willcox, S.; Griffith, J.D.; Pommier, Y.; Takeda, S.; Dutta, A. Production of Extrachromosomal MicroDNAs Is Linked to Mismatch Repair Pathways and Transcriptional Activity. *Cell Rep.* 2015, *11*, 1749–1759, doi:10.1016/j.celrep.2015.05.020.
10. Møller, H.D.; Mohiyuddin, M.; Prada-Luengo, I.; Sailani, M.R.; Halling, J.F.; Plomgaard, P.; Maretty, L.; Hansen, A.J.; Snyder, M.P.; Pilegaard, H.; et al. Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat. Commun.* 2018, *9*, 1–12, doi:10.1038/s41467-018-03369-8.
11. Jarroux, J.; Morillon, A.; Pinskaya, M. Long Non Coding RNA Biology. In *Advances in experimental medicine and biology*; 2017; Vol. 1008, pp. 1–46 ISBN 978-981-10-5202-6.
12. Craig Venter, J.; Adams, M.D.; Myers, E.W.; Li, P.W.; Mural, R.J.; Sutton, G.G.; Smith, H.O.; Yandell, M.; Evans, C.A.; Holt, R.A.; et al. The sequence of the human genome. *Science (80-.).* 2001, *291*, 1304–1351, doi:10.1126/science.1058040.
13. International Human Genome Sequencing Consortium Initial sequencing and analysis of the human genome. *Nature* 2001, *412*, 565–566, doi:10.1038/35087627.
14. Dunham, I.; Kundaje, A.; Aldred, S.F.; Collins, P.J.; Davis, C.A.; Doyle, F.; Epstein, C.B.; Frietze, S.; Harrow, J.; Kaul, R.; et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012, *489*, 57–74, doi:10.1038/nature11247.
15. Willyard, C. Expanded human gene tally reignites debate. *Nature* 2018, *558*, 334–335, doi:10.1038/d41586-018-05462-w.
16. Salzberg, S.L. Open questions: How many genes do we have? *BMC Biol.* 2018, *16*, 10–12, doi:10.1186/s12915-018-0564-x.

17. Pertea, M.; Shumate, A.; Pertea, G.; Varabyou, A.; Chang, Y.-C.; Madugundu, A.K.; Pandey, A.; Salzberg, S.L. Thousands of large-scale RNA sequencing experiments yield a comprehensive new human gene list and reveal extensive transcriptional noise. *bioRxiv* 2018, 332825, doi:10.1101/332825.
18. Vaquerizas, J.M.; Kummerfeld, S.K.; Teichmann, S.A.; Luscombe, N.M. A census of human transcription factors: Function, expression and evolution. *Nat. Rev. Genet.* 2009, *10*, 252–263, doi:10.1038/nrg2538.
19. Lee, T.I.; Young, R.A. Transcriptional regulation and its misregulation in disease. *Cell* 2013, *152*, 1237–1251, doi:10.1016/j.cell.2013.02.014.
20. Mirabella, A.C.; Foster, B.M.; Bartke, T. Chromatin deregulation in disease. *Chromosoma* 2016, *125*, 75–93, doi:10.1007/s00412-015-0530-0.
21. Boyadjiev, S.; Jabs, E. Online Mendelian Inheritance in Man (OMIM) as a knowledgebase for human developmental disorders. *Clin. Genet.* 2000, *57*, 253–266, doi:10.1034/j.1399-0004.2000.570403.x.
22. Van der Lee, R.; Correard, S.; Wasserman, W.W. Deregulated Regulators: Disease-Causing cis Variants in Transcription Factor Genes. *Trends Genet.* 2020, *36*, 523–539, doi:10.1016/j.tig.2020.04.006.
23. Belluti, S.; Rigillo, G.; Imbriano, C. Transcription Factors in Cancer: When Alternative Splicing Determines Opposite Cell Fates. *Cells* 2020, *9*, 760, doi:10.3390/cells9030760.
24. Sibley, C.R.; Blazquez, L.; Ule, J. Lessons from non-canonical splicing. *Nat. Rev. Genet.* 2016, *17*, 407–421, doi:10.1038/nrg.2016.46.
25. Arnaiz, E.; Sole, C.; Manterola, L.; Iparraguirre, L.; Otaegui, D.; Lawrie, C. CircRNAs and cancer : Biomarkers and master regulators. *Semin. Cancer Biol.* 2018, *58*, 90–99, doi:10.1016/j.semcancer.2018.12.002.
26. Haque, S.; Harries, L.W. Circular RNAs (circRNAs) in Health and Disease. *Genes (Basel).* 2017, *8*, 353, doi:10.3390/genes8120353.
27. Munoz-Culla, M.; Irizar, H.; Otaegui, D. The genetics of multiple sclerosis: review of current and emerging candidates. *Appl. Clin. Genet.* 2013, *6*, 63–73, doi:10.2147/TACG.S29107.
28. Nelson, D.L.; Cox, M.M. Metabolismo del RNA. In *Lehninger. Principios de Bioquímica*; Omega, Ed.; 2009; pp. 1021–1064.
29. Cramer, P. Organization and regulation of gene transcription. *Nature* 2019, *573*, 45–54, doi:10.1038/s41586-019-1517-4.
30. Fuda, N.J.; Ardehali, M.B.; Lis, J.T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* 2009, *461*, 186–192, doi:10.1038/nature08449.
31. Chiarella, A.M.; Lu, D.; Hathaway, N.A. Epigenetic control of a local chromatin landscape. *Int. J. Mol. Sci.* 2020, *21*, 6–8, doi:10.3390/ijms21030943.
32. Alegría-Torres, J.A.; Baccarelli, A.; Bollati, V. Epigenetics and lifestyle. *Epigenomics* 2011, *3*, 267–277, doi:10.2217/epi.11.22.

33. Di Croce, L.; Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* 2013, *20*, 1147–1155, doi:10.1038/nsmb.2669.
34. Nelson, D.L.; Cox, M.M. Regulacion de la expresion genica. In *Lehninger. Principios de Bioquímica*; Omega, Ed.; Figth edit, 2009; pp. 1115–1152.
35. Illingworth, R.S.; Bird, A.P. CpG islands - "A rough guide." *FEBS Lett.* 2009, *583*, 1713–1720, doi:10.1016/j.febslet.2009.04.012.
36. Wilkinson, A.C.; Nakauchi, H.; Göttgens, B. Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. *Cell Syst.* 2017, *5*, 319–331, doi:10.1016/j.cels.2017.07.004.
37. Cavalli, F.M.G. A computational study of transcriptional regulation in eukaryotes on a genomic scale (Doctoral thesis), University of Cambridge, 2010.
38. Jolma, A.; Yin, Y.; Nitta, K.R.; Dave, K.; Popov, A.; Taipale, M.; Enge, M.; Kivioja, T.; Morgunova, E.; Taipale, J. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* 2015, *527*, 384–388, doi:10.1038/nature15518.
39. Kouzine, F.; Wojtowicz, D.; Yamane, A.; Resch, W.; Kieffer-Kwon, K.R.; Bandle, R.; Nelson, S.; Nakahashi, H.; Awasthi, P.; Feigenbaum, L.; et al. Global regulation of promoter melting in naive lymphocytes. *Cell* 2013, *153*, 988, doi:10.1016/j.cell.2013.04.033.
40. Adelman, K.; Lis, J.T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* 2013, *13*, 720–731, doi:10.1038/nrg3293.Promoter-proximal.
41. Tellier, M.; Maudlin, I.; Murphy, S. Transcription and splicing: A two-way street. *WIREs RNA* 2020, 1–25, doi:10.1002/wrna.1593.
42. Sharp, P.A. The discovery of split genes and RNA splicing. *Trends Biochem. Sci.* 2005, *30*, 279–281, doi:10.1016/j.tibs.2005.04.003.
43. Gilbert, W. Why genes in pieces? *Nature* 1978, *271*, 501.
44. Sakharkar, M.K.; Chow, V.T.K.; Kangueane, P. Distributions of exons and introns in the human genome. *In Silico Biol.* 2004, *4*, 387–393.
45. Carrillo Oesterreich, F.; Preibisch, S.; Neugebauer, K.M. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol. Cell* 2010, *40*, 571–581, doi:10.1016/j.molcel.2010.11.004.
46. Wallace, E.W.J.; Beggs, J.D. Extremely fast and incredibly close: Cotranscriptional splicing in budding yeast. *RNA* 2017, *23*, 601–610, doi:10.1261/rna.060830.117.
47. Nojima, T.; Rebelo, K.; Gomes, T.; Grosso, A.R.; Proudfoot, N.J.; Carmo-Fonseca, M. RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol. Cell* 2018, *72*, 369-379.e4, doi:10.1016/j.molcel.2018.09.004.
48. Ji, X.; Zhou, Y.; Pandit, S.; Huang, J.; Li, H.; Lin, C.Y.; Xiao, R.; Burge, C.B.; Fu, X.D. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* 2013, *153*, 855–868, doi:10.1016/j.cell.2013.04.028.

49. Milligan, L.; Sayou, C.; Tuck, A.; Auchynnikava, T.; Reid, J.E.A.; Alexander, R.; de Lima Alves, F.; Allshire, R.; Spanos, C.; Rappsilber, J.; et al. RNA polymerase II stalling at pre-mRNA splice sites is enforced by ubiquitination of the catalytic subunit. *Elife* 2017, *6*, 1–27, doi:10.7554/eLife.27082.
50. Kim, S.; Kim, H.; Fong, N.; Erickson, B.; Bentley, D.L. Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci. U. S. A.* 2011, *108*, 13564–13569, doi:10.1073/pnas.1109475108.
51. Chen, Y.G.; Chen, R.; Ahmad, S.; Verma, R.; Kasturi, S.P.; Amaya, L.; Broughton, J.P.; Kim, J.; Cadena, C.; Pulendran, B.; et al. N6-Methyladenosine Modification Controls Circular RNA Immunity. *Mol. Cell* 2019, *76*, 96-109.e9, doi:10.1016/j.molcel.2019.07.016.
52. Alberts, B.; Johnson, A.; Lewis, J.; Morgan, D.; Raff, M.; Roberts, K.; Walter, P. Control of gene expression. In *Molecular Biology of the Cell*; Garland Sciences, 2015; pp. 369–439.
53. Glažar, P. Expression and possible functions of circular RNAs (Doctoral thesis), The Max Delbrück Center for Molecular Medicine, 2020.
54. Jansen, R.P. mRNA localization: Message on the move. *Nat. Rev. Mol. Cell Biol.* 2001, *2*, 247–256, doi:10.1038/35067016.
55. Valente, L.; Kawahara, Y.; Zinshteyn, B.; Iizasa, H.; Nishikura, K. Postranscriptional gene regulation by an editor: ADAR and its role in RNA editing. In *Posttranscriptional Gene Regulation*; Wu, J., Ed.; Wiley-Blackwell, 2013; pp. 41–69.
56. Berdasco, M.; Esteller, M. Crosstalk Between Non-Coding RNAs and the Epigenome in Development. In *Chromatin Regulation and Dynamics*; Göndör, A., Ed.; Elsevier, 2017; pp. 211–230.
57. Harbers, M.; Carninci, P. Non coding RNA: The mayor output of gene expression. In *Posttranscriptional Gene Regulation*; Wu, J., Ed.; Wiley-Blackwell, 2013; pp. 181–203.
58. Srijyothi, L.; Ponne, S.; Prathama, T.; Ashok, C.; Baluchamy, S. Roles of Non-Coding RNAs in Transcriptional Regulation. In *Transcriptional and Post-transcriptional regulation*; Kais, G., Ed.; IntechOpen, 2018; pp. 55–75.
59. Bartel, D.P. Metazoan MicroRNAs. *Cell* 2018, *173*, 20–51, doi:10.1016/j.cell.2018.03.006.
60. Lee, Y.; Jeon, K.; Lee, J.T.; Kim, S.; Kim, V.N. MicroRNA maturation: Stepwise processing and subcellular localization. *EMBO J.* 2002, *21*, 4663–4670, doi:10.1093/emboj/cdf476.
61. Stavast, C.J.; Erkeland, S.J. The Non-Canonical Aspects of MicroRNAs: Many Roads to Gene Regulation. *Cells* 2019, *8*, 1465, doi:10.3390/cells8111465.
62. Sanger, H.L.; Klotzt, G.; Riesnert, D.; Gross, H.J.; Albrecht, K. Viroids are single stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. U. S. A.* 1976, *73*, 3852–3856.

63. Gross, HJ; Domdey, H; Lossow, C; Jank, P; Raba, M; Alberty, H; Sänger, H. Nucleotide sequence and secondary structure of potato spindle tuber viroid. *Nature* 1978, *273*, 203–208.
64. Kos, A; Dijkema, R; Arnberg, AC; Van der Meide, PH; Schellekens, H. The hepatitis delta (delta) virus possesses a circular RNA. *Nature* 1986, *323*, 558–560.
65. Kjems, J.; Garrett, R.A. Novel Splicing Mechanism for the Ribosomal RNA Intron in the Archaebacterium Desulfurococcus mobilis. *Cell* 1988, *54*, 693–703.
66. Grabowski, P.J.; Cech, T.R. The Intervening Sequence of the Ribosomal RNA Precursor Is Converted to a Circular RNA in Isolated Nuclei of Tetrahymena. *Cell* 1981, *23*, 467–476.
67. Nigro, J.M.; Cho, K.R.; Fearon, E.R.; Kern, S.E.; Ruppert, J.M.; Oliner, J.D.; Kinzler, K.W.; Vogelstein, B. Scrambled Exons. *Cell* 1991, *64*, 607–613.
68. Cocquerelle, C.; Daubersies, P.; Majerus, M.; Kerckaert, J.; Bailleul, B. Splicing with inverted order of large introns. *EMBO J.* 1992, *1*, 1095–1098.
69. Zaphiropoulos, P.G. Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene : Correlation with exon skipping. *Proc. Natl. Acad. Sci. U. S. A.* 1996, *93*, 6536–6541.
70. Surono, A; Takeshima, Y; Wibawa, T; Ikezawa, M; Nonaka, I; Matsuo, M. Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. *Hum. Mol. Genet.* 1999, *8*, 493–500.
71. Li, X.F.; Lytton, J. A circularized sodium-calcium exchanger exon 2 transcript. *J. Biol. Chem.* 1999, *274*, 8153–8160, doi:10.1074/jbc.274.12.8153.
72. Capel, B.; Swain, A.; Nicolis, S.; Hacker, A.; Walter, M.; Koopman, P. Circular Transcripts of the Testis-Determining Gene Sry in Adult Mouse Testis. *Cell* 1993, *73*, 1019–1030.
73. Westholm, J.O.; Miura, P.; Graveley, B.R.; Lai, E.C.; Westholm, J.O.; Miura, P.; Olson, S.; Shenker, S.; Joseph, B.; Sanfilippo, P. Genome-wide Analysis of Drosophila Circular RNAs Reveals Their Structural and Sequence Properties and Age-Dependent Neural Accumulation. *CellReports* 2014, *9*, 1966–1980, doi:10.1016/j.celrep.2014.10.062.
74. Wang, P.L.; Bao, Y.; Yee, M.; Barrett, S.P.; Hogan, G.J.; Dinneny, R.; Brown, P.O.; Salzman, J. Circular RNA Is Expressed across the Eukaryotic Tree of Life. *PLoS One* 2014, *9*, doi:10.1371/journal.pone.0090859.
75. Gruner, H.; Cortés-López, M.; Cooper, D.A.; Bauer, M.; Miura, P.; Danan, M.; Schwartz, S.; Edelheit, S.; Sorek, R.; Lu, T.; et al. CircRNA accumulation in the aging mouse brain. *Sci. Rep.* 2016, *6*, 38907, doi:10.1038/srep38907.
76. Cortés-López, M.; Gruner, M.R.; Cooper, D.A.; Gruner, H.N.; Voda, A.I.; van der Linden, A.M.; Miura, P. Global accumulation of circRNAs during aging in Caenorhabditis elegans. *BMC Genomics* 2018, *19*, 1–12, doi:10.1186/s12864-017-4386-y.

77. Venø, M.T.; Hansen, T.B.; Venø, S.T.; Clausen, B.H.; Grebing, M.; Finsen, B.; Holm, I.E.; Kjems, J. Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. *Genome Biol.* 2015, *16*, 245, doi:10.1186/s13059-015-0801-3.

78. Salzman, J.; Gawad, C.; Wang, P.L.; Lacayo, N.; Brown, P.O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* 2012, *7*, doi:10.1371/journal.pone.0030733.

79. Jeck, W.R.; Sorrentino, J.A.; Wang, K.; Slevin, M.K.; Burd, C.E.; Liu, J.; Marzluff, W.F.; Sharpless, N.E. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 2013, *19*, 141–57, doi:10.1261/rna.035667.112.

80. Memczak, S.; Jens, M.; Elefsinioti, A.; Torti, F.; Krueger, J.; Rybak, A.; Maier, L.; Mackowiak, S.D.; Gregersen, L.H.; Munschauer, M.; et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 2013, *495*, 333–338, doi:10.1038/nature11928.

81. Chu, Q.; Bai, P.; Zhu, X.; Zhang, X.; Mao, L.; Zhu, Q.H.; Fan, L.; Ye, C.Y. Characteristics of plant circular RNAs. *Brief. Bioinform.* 2018, *21*, 135–143, doi:10.1093/bib/bby111.

82. Salzman, J.; Chen, R.E.; Olsen, M.N.; Wang, P.L.; Brown, P.O. Cell-Type Specific Features of Circular RNA Expression. *PLoS Genet.* 2013, *9*, e1003777, doi:10.1371/journal.pgen.1003777.

83. Ashwal-Fluss, R.; Meyer, M.; Pamudurti, N.R.; Ivanov, A.; Bartok, O.; Hanan, M.; Evantal, N.; Memczak, S.; Rajewsky, N.; Kadener, S. CircRNA Biogenesis competes with Pre-mRNA splicing. *Mol. Cell* 2014, *56*, 55–66, doi:10.1016/j.molcel.2014.08.019.

84. Ivanov, A.; Memczak, S.; Wyler, E.; Torti, F.; Porath, H.T.; Orejuela, M.R.; Piechotta, M.; Levanon, E.Y.; Landthaler, M.; Dieterich, C.; et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.* 2015, *10*, 170–177, doi:10.1016/j.celrep.2014.12.019.

85. Starke, S.; Jost, I.; Rossbach, O.; Schneider, T.; Schreiner, S.; Hung, L.H.; Bindereif, A. Exon circularization requires canonical splice signals. *Cell Rep.* 2015, *10*, 103–111, doi:10.1016/j.celrep.2014.12.002.

86. Piwecka, M.; Glažar, P.; Hernandez-miranda, L.R.; Memczak, S.; Wolf, S.A.; Rybak-wolf, A.; Filipchyk, A.; Klironomos, F.; Alicia, C.; Jara, C.; et al. Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. 2017, *8526*, 1–14.

87. Lasda, E.; Parker, R. Circular RNAs co-precipitate with extracellular vesicles: A possible mechanism for circrna clearance. *PLoS One* 2016, *11*, 1–11, doi:10.1371/journal.pone.0148407.

88. Memczak, S.; Papavasileiou, P.; Peters, O.; Rajewsky, N. Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. *PLoS One* 2015, *10*, 1–13, doi:10.1371/journal.pone.0141214.

89. Pamudurti, N.R.; Bartok, O.; Jens, M.; Ashwal-Fluss, R.; Stottmeister, C.; Ruhe, L.; Hanan, M.; Wyler, E.; Perez-Hernandez, D.; Ramberger, E.; et al. Translation of CircRNAs. *Mol. Cell* 2017, *66*, 9-21.e7, doi:10.1016/j.molcel.2017.02.021.
90. Hansen, T.B.; Jensen, T.I.; Clausen, B.H.; Bramsen, J.B.; Finsen, B.; Damgaard, C.K.; Kjems, J. Natural RNA circles function as efficient microRNA sponges. *Nature* 2013, *495*, 384–388, doi:10.1038/nature11993.
91. Legnini, I.; Timoteo, G. Di; Rossi, F.; Morlando, M.; Briganti, F.; Sthandier, O.; Fatica, A.; Santini, T.; Andronache, A.; Wade, M.; et al. Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cell* 2017, *66*, 22-37.e9, doi:10.1016/j.molcel.2017.02.017.
92. Guo, J.U.; Agarwal, V.; Guo, H.; Bartel, D.P. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.* 2014, *15*, 409, doi:10.1186/PREACCEPT-1176565312639289.
93. Jeck, W.R.; Sharpless, N.E. Detecting and characterizing circular RNAs. *Nat. Biotechnol.* 2014, *32*, 453–61, doi:10.1038/nbt.2890.
94. Szabo, L.; Salzman, J. Detecting circular RNAs: bioinformatic and experimental challenges. *Nat. Rev. Genet.* 2016, *17*, 679–692, doi:10.1038/nrg.2016.114.
95. Hansen, T.B. Improved circRNA Identification by Combining Prediction Algorithms. *Front. Cell Dev. Biol.* 2018, *6*, 1–9, doi:10.3389/fcell.2018.00020.
96. Haque, S.; Harries, L.W. Circular RNAs (circRNAs) in Health and Disease. *Genes (Basel).* 2017, *8*, 1–17, doi:10.3390/genes8120353.
97. Han, B.; Chao, J.; Yao, H. Circular RNA and its mechanisms in disease: From the bench to the clinic. *Pharmacol. Ther.* 2018, *187*, 31–44, doi:10.1016/j.pharmthera.2018.01.010.
98. Yar Saglam, A.S.; Alp, E.; Onen, H.I. Circular RNAs and Its Biological Functions in Health and Disease. In *Gene Expression and Phenotypic Traits*; Chen, Y.-C., Ed.; IntechOpen, 2020; pp. 1–37.
99. Wang, M.; Yu, F.; Wu, W.; Zhang, Y.; Chang, W.; Ponnusamy, M.; Wang, K.; Li, P. Circular RNAs: A novel type of non-coding RNA and their potential implications in antiviral immunity. *Int. J. Biol. Sci.* 2017, *13*, 1497–1506, doi:10.7150/ijbs.22531.
100. Huang, C.; Liang, D.; Tatomer, D.C.; Wilusz, J.E. A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs. *Genes Dev.* 2018, *32*, 639–644, doi:10.1101/gad.314856.118.
101. Li, Z.; Huang, C.; Bao, C.; Chen, L.; Lin, M.; Wang, X.; Zhong, G.; Yu, B.; Hu, W.; Dai, L.; et al. Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* 2015, *22*, 256–264, doi:10.1038/nsmb.2959.
102. Zhang, Y.; Zhang, X.O.; Chen, T.; Xiang, J.F.; Yin, Q.F.; Xing, Y.H.; Zhu, S.; Yang, L.; Chen, L.L. Circular Intronic Long Noncoding RNAs. *Mol. Cell* 2013, *51*, 792–806, doi:10.1016/j.molcel.2013.08.017.
103. Wilusz, J.E. A 360 view of circular RNAs : From biogenesis to functions. *WIREs RNA* 2018, *9*, 1–17, doi:10.1002/wrna.1478.

104. Meng, X.; Li, X.; Zhang, P.; Wang, J.; Zhou, Y.; Chen, M. Circular RNA: an emerging key player in RNA world. *Brief. Bioinform.* 2016, 1–11, doi:10.1093/bib/bbw045.
105. Patop, I.L.; Wüst, S.; Kadener, S. Past, present, and future of circ RNA s. *EMBO J.* 2019, *38*, 1–13, doi:10.15252/embj.2018100836.
106. Li, X.; Yang, L.; Chen, L. The Biogenesis , Functions , and Challenges of Circular RNAs. *Mol. Cell* 2018, *71*, 428–442, doi:10.1016/j.molcel.2018.06.034.
107. Zhang, Y.; Xue, W.; Li, X.; Zhang, J.; Chen, S.; Zhang, J.L.; Yang, L.; Chen, L.L. The Biogenesis of Nascent Circular RNAs. *Cell Rep.* 2016, *15*, 611–624, doi:10.1016/j.celrep.2016.03.058.
108. Liang, D.; Tatomer, D.C.; Luo, Z.; Wu, H.; Yang, L.; Chen, L.L.; Cherry, S.; Wilusz, J.E. The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. *Mol. Cell* 2017, *68*, 940-954.e3, doi:10.1016/j.molcel.2017.10.034.
109. Liang, D.; Wilusz, J.E. Short intronic repeat sequences facilitate circular RNA production. *Genes Dev.* 2014, *28*, 2233–2247, doi:10.1101/gad.251926.114.
110. Zhang, X.; Wang, H.; Zhang, Y.; Lu, X.; Chen, L.; Yang, L. Complementary Sequence-Mediated Exon Circularization. *Cell* 2014, *159*, 134–147, doi:10.1016/j.cell.2014.09.001.
111. Zhang, X.O.; Dong, R.; Zhang, Y.; Zhang, J.L.; Luo, Z.; Zhang, J.; Chen, L.L.; Yang, L. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* 2016, *26*, 1277–1287, doi:10.1101/gr.202895.115.
112. Gao, Y.; Wang, J.; Zheng, Y.; Zhang, J.; Chen, S.; Zhao, F. Comprehensive identification of internal structure and alternative splicing events in circular RNAs. *Nat. Commun.* 2016, *7*, 1–13, doi:10.1038/ncomms12060.
113. Rybak-Wolf, A.; Stottmeister, C.; Gla??ar, P.; Jens, M.; Pino, N.; Hanan, M.; Behm, M.; Bartok, O.; Ashwal-Fluss, R.; Herzog, M.; et al. Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol. Cell* 2015, *58*, 870–885, doi:10.1016/j.molcel.2015.03.027.
114. Aktaş, T.; Ilik, I.A.; Maticzka, D.; Bhardwaj, V.; Pessoa Rodrigues, C.; Mittler, G.; Manke, T.; Backofen, R.; Akhtar, A. DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* 2017, *544*, 115–119, doi:10.1038/nature21715.
115. Conn, S.J.; Pillman, K.A.; Toubia, J.; Conn, V.M.; Salmanidis, M.; Phillips, C.A.; Roslan, S.; Schreiber, A.W.; Gregory, P.A.; Goodall, G.J. The RNA binding protein quaking regulates formation of circRNAs. *Cell* 2015, *160*, 1125–1134, doi:10.1016/j.cell.2015.02.014.
116. Hansen, T.B.; Wiklund, E.D.; Bramsen, J.B.; Villadsen, S.B.; Statham, A.L.; Clark, S.J.; Kjems, J. miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.* 2011, *30*, 4414–4422, doi:10.1038/emboj.2011.359.

117. Park, O.H.; Ha, H.; Lee, Y.; Boo, S.H.; Kwon, D.H.; Song, H.K.; Kim, Y.K. Endoribonucleolytic Cleavage of m6A-Containing RNAs by RNase P/MRP Complex. *Mol. Cell* 2019, *74*, 494-507.e8, doi:10.1016/j.molcel.2019.02.034.
118. Liu, C.X.; Li, X.; Nan, F.; Jiang, S.; Gao, X.; Guo, S.K.; Xue, W.; Cui, Y.; Dong, K.; Ding, H.; et al. Structure and Degradation of Circular RNAs Regulate PKR Activation in Innate Immunity. *Cell* 2019, *177*, 865-880.e21, doi:10.1016/j.cell.2019.03.046.
119. Fischer, J.W.; Busa, V.F.; Shao, Y.; Leung, A.K.L. Structure-Mediated RNA Decay by UPF1 and G3BP1. *Mol. Cell* 2020, *78*, 70-84.e6, doi:10.1016/j.molcel.2020.01.021.
120. Xiao, M.S.; Ai, Y.; Wilusz, J.E. Biogenesis and Functions of Circular RNAs Come into Focus. *Trends Cell Biol.* 2020, *30*, 226–240, doi:10.1016/j.tcb.2019.12.004.
121. Enuka, Y.; Lauriola, M.; Feldman, M.E.; Sas-Chen, A.; Ulitsky, I.; Yarden, Y. Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. *Nucleic Acids Res.* 2016, *44*, 1370–1383, doi:10.1093/nar/gkv1367.
122. Bachmayr-Heyda, A.; Reiner, A.T.; Auer, K.; Sukhbaatar, N.; Aust, S.; Bachleitner-Hofmann, T.; Mesteri, I.; Grunt, T.W.; Zeillinger, R.; Pils, D. Correlation of circular RNA abundance with proliferation - Exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci. Rep.* 2015, *5*, 8057, doi:10.1038/srep08057.
123. You, X.; Vlatkovic, I.; Babic, A.; Will, T.; Epstein, I.; Tushev, G.; Akbalik, G.; Wang, M.; Glock, C.; Quedenau, C.; et al. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat. Neurosci.* 2015, *18*, 603–610, doi:10.1038/nn.3975.
124. Chen, W.; Schuman, E. Circular RNAs in Brain and Other Tissues : A Functional Enigma. *Trends Neurosci.* 2016, *39*, 597–604, doi:10.1016/j.tins.2016.06.006.
125. Iparraguirre, L.; Muñoz-Culla, M.; Prada-Luengo, I.; Castillo-Triviño, T.; Olascoaga, J.; Otaegui, D. Circular RNA profiling reveals that circular RNAs from ANXA2 can be used as new biomarkers for multiple sclerosis. *Hum. Mol. Genet.* 2017, *26*, 3564–3572, doi:10.1093/hmg/ddx243.
126. Chen, D.; Zhang, L.; Tan, K.; Jing, Q. Application of droplet digital PCR in quantitative detection of the cell-free circulating circRNAs. 2018, *2818*, doi:10.1080/13102818.2017.1398596.
127. Koh, W.; Pan, W.; Gawad, C.; Fan, H.C.; Kerchner, G.A.; Wyss-Coray, T.; Blumenfeld, Y.J.; El-Sayed, Y.Y.; Quake, S.R. Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proc. Natl. Acad. Sci.* 2014, *111*, 7361–7366, doi:10.1073/pnas.1411381111.
128. Bahn, J.H.; Zhang, Q.; Li, F.; Chan, T.-M.; Lin, X.; Yong, K.; Wong, D.T.W.; Xiao, X. The Landscape of MicroRNA, Piwi-Interacting RNA, and Circular RNA in Human Saliva. *Clin. chemestry* 2015, *61*, 221–230, doi:10.1373/clinchem.2014.230433.

129. Dong, W.-W.; Li, H.-M.; Qing, X.-R.; Huang, D.-H.; Li, H.-G.; Griswold, M.D.; Eddy, E.M.; Kimmins, S.; Sassone-Corsi, P.; Papaioannou, M.D.; et al. Identification and characterization of human testis derived circular RNAs and their existence in seminal plasma. *Sci. Rep.* 2016, *6*, 39080, doi:10.1038/srep39080.

130. Lam, W.K.J.; Lo, Y.M.D. Circular RNAs as urinary biomarkers. *Clin. Chem.* 2019, *65*, 1196–1198, doi:10.1373/clinchem.2019.309773.

131. Hulstaert, E.; Morlion, A.; Avila Cobos, F.; Verniers, K.; Nuytens, J.; Vanden Eynde, E.; Yigit, N.; Anckaert, J.; Geerts, A.; Hindryckx, P.; et al. Charting extracellular transcriptomes in The Human Biofluid RNA Atlas. *bioRxiv* 2020, doi:https://doi.org/10.1101/823369.

132. Li, Y.; Zheng, Q.; Bao, C.; Li, S.; Guo, W.; Zhao, J.; Chen, D.; Gu, J.; He, X.; Huang, S. Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res.* 2015, *25*, 981–984, doi:10.1038/cr.2015.82.

133. Preußer, C.; Hung, L.H.; Schneider, T.; Schreiner, S.; Hardt, M.; Moebus, A.; Santoso, S.; Bindereif, A. Selective release of circRNAs in platelet-derived extracellular vesicles. *J. Extracell. Vesicles* 2018, *7*, doi:10.1080/20013078.2018.1424473.

134. Wang, Y.; Liu, J.; Ma, J.; Sun, T.; Zhou, Q.; Wang, W.; Wang, G.; Wu, P.; Wang, H.; Jiang, L.; et al. Exosomal circRNAs: Biogenesis, effect and application in human diseases. *Mol. Cancer* 2019, *18*, 1–10, doi:10.1186/s12943-019-1041-z.

135. Zhang, X.; Wang, H.; Zhang, Y.; Lu, X.; Chen, L.; Yang, L. Complementary Sequence-Mediated Exon Circularization. *Cell* 2014, *159*, 134–147, doi:10.1016/j.cell.2014.09.001.

136. Gualandi, F.; Trabanelli, C.; Rimessi, P.; Calzolari, E.; Toffolatti, L.; Patarnello, T.; Kunz, G.; Muntoni, F.; Ferlini, A. Multiple exon skipping and RNA circularisation contribute to the severe phenotypic expression of exon 5 dystrophin deletion. *J. Med. Genet.* 2003, *40*, 1–6, doi:10.1136/jmg.40.8.e100.

137. Kristensen, L.S.; Andersen, M.S.; Stagsted, L.V.W.; Ebbesen, K.K.; Hansen, T.B.; Kjems, J. The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* 2019, *20*, 675–691, doi:10.1038/s41576-019-0158-7.

138. Bose, R.; Ain, R. Regulation of Transcription by Circular RNAs. In *Circular RNAs, Advances in Experimentar Medicine and Biology*; Xiao, J., Ed.; Springer Nature, 2018; Vol. 1087, pp. 81–94 ISBN 978-981-13-1425-4.

139. Dudekula, D.B.; Panda, A.C.; Grammatikakis, I.; De, S.; Abdelmohsen, K.; Gorospe, M. CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol.* 2016, *13*, 34–42, doi:10.1080/15476286.2015.1128065.

140. Armakola, M.; Higgins, M.J.; Figley, M.D.; Barmada, S.J.; Scarborough, E.A.; Diaz, Z.; Fang, X.; Shorter, J.; Krogan, N.J.; Finkbeiner, S.; et al. Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. *Nat. Genet.* 2012, *44*, 1302–1309, doi:10.1038/ng.2434.

141. Li, X.; Liu, C.X.; Xue, W.; Zhang, Y.; Jiang, S.; Yin, Q.F.; Wei, J.; Yao, R.W.; Yang, L.; Chen, L.L. Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection. *Mol. Cell* 2017, *67*, 214-227.e7, doi:10.1016/j.molcel.2017.05.023.
142. Chen, X.; Han, P.; Zhou, T.; Guo, X.; Song, X.; Li, Y. circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Sci. Rep.* 2016, *6*, 34985, doi:10.1038/srep34985.
143. Abe, N.; Matsumoto, K.; Nishihara, M.; Nakano, Y.; Shibata, A.; Maruyama, H.; Shuto, S.; Matsuda, A.; Yoshida, M.; Ito, Y.; et al. Rolling Circle Translation of Circular RNA in Living Human Cells. *Sci. Rep.* 2015, *5*, 1–9, doi:10.1038/srep16435.
144. Yang, Y.; Fan, X.; Mao, M.; Song, X.; Wu, P.; Zhang, Y.; Jin, Y. Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Res.* 2017, *27*, 626–641, doi:10.1038/cr.2017.31.
145. Schneider, T.; Bindereif, A. Circular RNAs: Coding or noncoding? *Cell Res.* 2017, *27*, 724–725, doi:10.1038/cr.2017.70.
146. Holdt, L.M.; Kohlmaier, A.; Teupser, D. Circular RNAs as therapeutic agents and targets. *Front. Physiol.* 2018, *9*, doi:10.3389/fphys.2018.01262.
147. Zhang, Z.; Yang, T.; Xiao, J. Circular RNAs : Promising Biomarkers for Human Diseases. *EBioMedicine* 2018, *34*, 267–274, doi:10.1016/j.ebiom.2018.07.036.
148. Atkinson, A.J.; Colburn, W.A.; DeGruttola, V.G.; DeMets, D.L.; Downing, G.J.; Hoth, D.F.; Oates, J.A.; Peck, C.C.; Schooley, R.T.; Spilker, B.A.; et al. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 2001, *69*, 89–95, doi:10.1067/mcp.2001.113989.
149. Abu, N.; Jamal, R. Circular RNAs as promising biomarkers: A mini-review. *Front. Physiol.* 2016, *7*, 1–6, doi:10.3389/fphys.2016.00355.
150. Zhou, Z.; Sun, B.; Huang, S.; Zhao, L. Roles of circular RNAs in immune regulation and autoimmune diseases. *Cell Death Dis.* 2019, *10*, doi:10.1038/s41419-019-1744-5.
151. Xia, X.; Tang, X.; Wang, S. Roles of CircRNAs in Autoimmune Diseases. *Front. Immunol.* 2019, *10*, 1–8, doi:10.3389/fimmu.2019.00639.
152. Chen, X.; Yang, T.; Wang, W.; Xi, W.; Zhang, T.; Li, Q.; Yang, A.; Wang, T. Circular RNAs in immune responses and immune diseases. *Theranostics* 2019, *9*, 588–607, doi:10.7150/thno.29678.
153. Wilusz, J.E. Circle the Wagons: Circular RNAs Control Innate Immunity. *Cell* 2019, *177*, 797–799, doi:10.1016/j.cell.2019.04.020.
154. Cadena, C.; Hur, S. Previews Antiviral Immunity and Circular RNA : No End in Sight. *Mol. Cell* 2017, *67*, 163–164, doi:10.1016/j.molcel.2017.07.005.
155. Chen, Y.G.; Kim, M. V.; Chen, X.; Batista, P.J.; Aoyama, S.; Wilusz, J.E.; Iwasaki, A.; Chang, H.Y. Sensing Self and Foreign Circular RNAs by Intron Identity. *Mol. Cell* 2017, *67*, 228-238.e5, doi:10.1016/j.molcel.2017.05.022.
156. Wesselhoeft, R.A.; Kowalski, P.S.; Parker-Hale, F.C.; Huang, Y.; Bisaria, N.; Anderson, D.G. RNA Circularization Diminishes Immunogenicity and Can Extend Translation Duration In Vivo. *Mol. Cell* 2019, *74*, 508-520.e4, doi:10.1016/j.molcel.2019.02.015.

157. Vromman, M.; Vandesompele, J.; Volders, P.-J. Closing the circle: current state and perspectives of circular RNA databases. *Brief. Bioinform.* 2020, bbz175, doi:10.1093/bib/bbz175.
158. Glažar, P.; Papavasileiou, P.; Rajewsky, N. circBase: a database for circular RNAs. *RNA* 2014, *20*, 1666–1670, doi:10.1261/rna.043687.113.overview.
159. Hemmer, B.; Kerschensteiner, M.; Korn, T. Role of the innate and adaptive immune responses in the course of multiple sclerosis. *Lancet Neurol.* 2015, *14*, 406–419, doi:10.1016/S1474-4422(14)70305-9.
160. Pena, J. a.; Lotze, T.E. Pediatric multiple sclerosis: Current concepts and consensus definitions. *Autoimmun. Dis.* 2013, *2013*, 673947, doi:10.1155/2013/673947.
161. Kis, B.; Rumberg, B.; Berlit, P. Clinical characteristics of patients with late-onset multiple sclerosis. *J. Neurol.* 2008, *255*, 697–702, doi:10.1007/s00415-008-0778-x.
162. Tullman, M.J. Overview of the epidemiology, diagnosis, and disease progression associated with multiple sclerosis. *Am. J. Manag. Care* 2013, *19*, S15-20.
163. Harbo, H.F.; Gold, R.; Tintoré, M. Sex and gender issues in multiple sclerosis. *Ther. Adv. Neurol. Disord.* 2013, *6*, 237–248, doi:10.1177/1756285613488434.
164. Gilli, F.; DiSano, K.D.; Pachner, A.R. SeXX Matters in Multiple Sclerosis. *Front. Neurol.* 2020, *11*, 1–20, doi:10.3389/fneur.2020.00616.
165. Bertrams, J.; Kuwert, E.; Liedtke, U. HL-A Antigens and Multiple Sclerosis. *Tissue Antigens* 1972, *2*, 405–408, doi:10.1111/j.1399-0039.1972.tb00060.x.
166. Hollenbach, J.A.; Oksenberg, J.R. The Immunogenetics of Multiple Sclerosis: A Comprehensive Review. *J. Autoimmun.* 2015, *64*, 13–25, doi:10.1016/j.jaut.2015.06.010.
167. Patsopoulos, N.A.; Baranzini, S.E.; Santaniello, A.; Shoostari, P.; Cotsapas, C.; Wong, G.; Beecham, A.H.; James, T.; Replogle, J.; Vlachos, I.S.; et al. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science (80-.).* 2019, *365*, doi:10.1126/science.aav7188.
168. Doshi, A.; B, J.C. Multiple sclerosis , a treatable disease. *Clin. Med. (Northfield. Il).* 2016, *16*, 53–59.
169. Compston, A.; Coles, A. Multiple sclerosis. *Lancet* 2008, *372*, 1502–1517, doi:10.1016/S0140-6736(08)61620-7.
170. Kadowaki, A.; Quintana, F.J. The Gut–CNS Axis in Multiple Sclerosis. *Trends Neurosci.* 2020, 1–13, doi:10.1016/j.tins.2020.06.002.
171. Huynh, J.L.; Casaccia, P. Epigenetic mechanisms in multiple sclerosis: Implications for pathogenesis and treatment. *Lancet Neurol.* 2013, *12*, 195–206, doi:10.1016/S1474-4422(12)70309-5.
172. Chan, V.S.F. Epigenetics in Multiple Sclerosis. In *Epigenetics in Allergy and Autoimmunity*; Springer, 2020; Vol. 1253, pp. 309–374 ISBN 9789811534492.
173. Sospedra, M.; Martin, R. Immunology of multiple sclerosis. *Semin. Neurol.* 2016, *36*, 115–127, doi:10.1007/s11882-007-0043-x.

174. Lee, Y.; Morrison, B.M.; Li, Y.; Lengacher, S.; Farah, M.H.; Hoffman, P.N.; Liu, Y.; Tsingalia, A.; Jin, L.; Zhang, P.W.; et al. Oligodendroglia metabolically support axons and contribute to neurodegeneration. *Nature* 2012, *487*, 443–448, doi:10.1038/nature11314.
175. The National Multiple Sclerosis Society. MS Symptoms. Available online: https://www.nationalmssociety.org/Symptoms-Diagnosis/MS-Symptoms.
176. Lublin, F.D.; Reingold, S.C.; Cohen, J.A.; Cutter, G.R.; Soresen, P.S.; Thompson, A.J.; Wolinsky, J.S.; Balcer, L.J.; Banwell, B.; Barkhof, F.; et al. Defining the clinical course of multiple sclerosis. *Neurology* 2014, *83*, 278–286.
177. Hou, Y.; Jia, Y.; Hou, J. Natural Course of Clinically Isolated Syndrome: A Longitudinal Analysis Using a Markov Model. *Sci. Rep.* 2018, *8*, 1–7, doi:10.1038/s41598-018-29206-y.
178. Cree, B.A.C.; Gourraud, P.A.; Oksenberg, J.R.; Bevan, C.; Crabtree-Hartman, E.; Gelfand, J.M.; Goodin, D.S.; Graves, J.; Green, A.J.; Mowry, E.; et al. Long-term evolution of multiple sclerosis disability in the treatment era. *Ann. Neurol.* 2016, *80*, 499–510, doi:10.1002/ana.24747.
179. Harris, V.K.; Tuddenham, J.F.; Sadiq, S.A. Biomarkers of multiple sclerosis : current findings. *Degener. Neurol. Neuromuscul. Dis.* 2017, *7*, 19–29.
180. Faissner, S.; Gold, R. Progressive multiple sclerosis : latest therapeutic developments and future directions. *Ther. Adv. Neurol. Disord.* 2019, *12*, 1–11, doi:10.1177/https.
181. Filippi, M.; Bar-Or, A.; Piehl, F.; Preziosa, P.; Solari, A.; Vukusic, S.; Rocca, M.A. Multiple sclerosis. *Nat. Rev. Dis. Prim.* 2018, *4*, 1–27, doi:10.1038/s41572-018-0041-4.
182. Polman, C.H.; Reingold, S.C.; Banwell, B.; Clanet, M.; Cohen, J.A.; Filippi, M.; Fujihara, K.; Havrdova, E.; Hutchinson, M.; Kappos, L.; et al. Diagnostic Criteria for Multiple Sclerosis : 2010 Revisions to the McDonald Criteria. *Ann. Neurol.* 2011, *69*, 292–302, doi:10.1002/ana.22366.
183. Thompson, A.J.; Banwell, B.L.; Barkhof, F.; Carroll, W.M.; Coetzee, T.; Comi, G.; Correale, J.; Fazekas, F.; Filippi, M.; Freedman, M.S.; et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 2018, *17*, 162–173, doi:10.1016/S1474-4422(17)30470-2.
184. Link, H.; Huang, Y.M. Oligoclonal bands in multiple sclerosis cerebrospinal fluid: An update on methodology and clinical usefulness. *J. Neuroimmunol.* 2006, *180*, 17–28, doi:10.1016/j.jneuroim.2006.07.006.
185. Villar, L.; García-Barragán, N.; Espiño, M.; Roldán, E.; Sádaba, M.; Gómez-Rial, J.; González-Porqué, P.; Álvarez-Cermeño, J. Influence of oligoclonal IgM specificity in multiple. *Mult. Scler.* 2008, *14*, 183–187.
186. Brownlee, W.J.; Hardy, T.A.; Fazekas, F.; Miller, D.H. Diagnosis of multiple sclerosis: progress and challenges. *Lancet* 2016, *6736*, 292–302, doi:10.1016/S0140-6736(16)30959-X.
187. Klein, S.L. Immune Cells Have Sex and So Should Journal Articles. *Endocrinology* 2012, *153*, 2544–2550, doi:10.1210/en.2011-2120.

188. Libert, C.; Dejager, L.; Pinheiro, I. The X chromosome in immune functions: When a chromosome makes the difference. *Nat. Rev. Immunol.* 2010, *10*, 594–604, doi:10.1038/nri2815.
189. Marriott, I.; Huet-Hudson, Y.M. Sexual dimorphism in innate immune responses to infectious organisms. *Immunol. Res.* 2006, *34*, 177–192, doi:10.1385/IR:34:3:177.
190. Weinstein, Y.; Ran, S.; Segal, S. Sex-associated differences in the regulation of immune responses controlled by the MHC of the mouse. *J. Immunol.* 1984, *132*, 656–661.
191. Butterworth, M.; McClellan, B.; Allansmith, M. Influence of Sex on Immunoglobulin Levels. *Nat. Publ. Gr.* 1967, *216*, 615–616.
192. Cook, I.F. Sexual dimorphism of humoral immunity with human vaccines. *Vaccine* 2008, *26*, 3551–3555, doi:10.1016/j.vaccine.2008.04.054.
193. Villacres, M.C.; Longmate, J.; Auge, C.; Diamond, D.J. Predominant type 1 CMV-specific memory T-helper response in humans: Evidence for gender differences in cytokine secretion. *Hum. Immunol.* 2004, *65*, 476–485, doi:10.1016/j.humimm.2004.02.021.
194. Wikby, A.; Månsson, I.A.; Johansson, B.; Strindhall, J.; Nilsson, S.E. The immune risk profile is associated with age and gender: Findings from three Swedish population studies of individuals 20-100 years of age. *Biogerontology* 2008, *9*, 299–308, doi:10.1007/s10522-008-9138-6.
195. Belanger, K.M.; Crislip, G.R.; Gillis, E.E.; Abdelbary, M.; Musall, J.B.; Mohamed, R.; Baban, B.; Elmarakby, A.; Brands, M.W.; Sullivan, J.C. Greater T Regulatory Cells in Females Attenuate DOCA-Salt-Induced Increases in Blood Pressure Versus Males. *Hypertension* 2020, 1615–1623, doi:10.1161/HYPERTENSIONAHA.119.14089.
196. Dunn, S.E.; Lee, H.; Pavri, F.R.; Zhang, M.A. Sex-Based Differences in Multiple Sclerosis (Part I): Biology of Disease Incidence. *Curr. Top. Behav. Neurosci.* 2015, 29–56, doi:10.1007/7854.
197. Bergamaschi, R. Prognostic Factors in Multiple Sclerosis. *Int. Rev. Neurobiol.* 2007, *79*, 423–447, doi:10.1016/S0074-7742(07)79019-0.
198. Kalincik, T.; Vivek, V.; Jokubaitis, V.; Lechner-Scott, J.; Trojano, M.; Izquierdo, G.; Lugaresi, A.; Grand'Maison, F.; Hupperts, R.; Oreja-Guevara, C.; et al. Sex as a determinant of relapse incidence and progressive course of multiple sclerosis. *Brain* 2013, *136*, 3609–3617, doi:10.1093/brain/awt281.
199. Pozzilli, C.; Tomassini, V.; Marinelli, F.; Paolillo, A.; Gasperini, C.; Bastianello, S. "Gender gap" in multiple sclerosis: Magnetic resonance imaging evidence. *Eur. J. Neurol.* 2003, *10*, 95–97, doi:10.1046/j.1468-1331.2003.00519.x.
200. Guneykaya, D.; Ivanov, A.; Hernandez, D.P.; Haage, V.; Wojtas, B.; Meyer, N.; Maricos, M.; Jordan, P.; Buonfiglioli, A.; Gielniewski, B.; et al. Transcriptional and Translational Differences of Microglia from Male and Female Brains. *Cell Rep.* 2018, *24*, 2773-2783.e6, doi:10.1016/j.celrep.2018.08.001.

201. Avila, M.; Culberson, J.; Peiris, N. The Role of Sex Hormones in Multiple Sclerosis. *Eur. Neurol.* 2018, *19430*, 93–99, doi:10.1159/000494262.
202. Krysko, K.M.; Graves, J.S.; Dobson, R.; Altintas, A.; Amato, M.P.; Bernard, J.; Bonavita, S.; Bove, R.; Cavalla, P.; Clerico, M.; et al. Sex effects across the lifespan in women with multiple sclerosis. *Ther. Adv. Neurol. Disord.* 2020, *13*, 1–30, doi:10.1177/1756286420936166.
203. Bhatia, A.; Sekhon, H.K.; Kaur, G. Sex hormones and immune dimorphism. *Sci. World J.* 2014, *2014*, 1–9, doi:10.1155/2014/159150.
204. Trigunaite, A.; Dimo, J.; Jørgensen, T.N. Suppressive effects of androgens on the immune system. *Cell. Immunol.* 2015, *294*, 87–94, doi:10.1016/j.cellimm.2015.02.004.
205. Chitnis, T. The role of testosterone in MS risk and course. *Mult. Scler.* 2018, *24*, 36–41, doi:10.1177/1352458517737395.
206. Bove, R.; Musallam, A.; Healy, B.C.; Raghavan, K.; Glanz, B.I.; Bakshi, R.; Weiner, H.; De Jager, P.L.; Miller, K.K.; Chitnis, T. Low testosterone is associated with disability in men with multiple sclerosis. *Mult. Scler. J.* 2014, *20*, 1584–1592, doi:10.1177/1352458514527864.
207. Straub, R.H. The complex role of estrogens in inflammation. *Endocr. Rev.* 2007, *28*, 521–574, doi:10.1210/er.2007-0001.
208. Spence, R.D.; Voskuhl, R.R. Neuroprotective effects of estrogens and androgens in CNS inflammation and neurodegeneration. *Front. Neuroendocrinol.* 2012, *33*, 105–115, doi:10.1016/j.yfrne.2011.12.001.
209. Shah, N.M.; Imami, N.; Johnson, M.R. Progesterone modulation of pregnancy-related immune responses. *Front. Immunol.* 2018, *9*, doi:10.3389/fimmu.2018.01293.
210. De Nicola, A.F.; Coronel, F.; Garay, L.I.; Gargiulo-Monachelli, G.; Gonzalez Deniselle, M.C.; Gonzalez, S.L.; Labombarda, F.; Meyer, M.; Guennoun, R.; Schumacher, M. Therapeutic effects of progesterone in animal models of neurological disorders. *CNS Neurol. disorders drug targets* 2013, *12*, 1205–1218.
211. Kashani, I.R.; Hedayatpour, A.; Pasbakhsh, P.; Kafami, L.; Khallaghi, B.; Malek, F. Progesterone enhanced remyelination in the mouse corpus callosum after cuprizone induced demyelination. *Iran. J. Med. Sci.* 2015, *40*, 507–514.
212. Garay, L.; Deniselle, M.C.G.; Lima, A.; Roig, P.; De Nicola, A.F. Effects of progesterone in the spinal cord of a mouse model of multiple sclerosis. *J. Steroid Biochem. Mol. Biol.* 2007, *107*, 228–237, doi:10.1016/j.jsbmb.2007.03.040.
213. Garay, L.; Deniselle, M.C.G.; Meyer, M.; Costa, J.J.L.; Lima, A.; Roig, P.; DeNicola, A.F. Protective effects of progesterone administration on axonal pathology in mice with experimental autoimmune encephalomyelitis. *Brain Res.* 2009, *1283*, 177–185, doi:10.1016/j.brainres.2009.04.057.
214. Fish, E.N. The X-files in immunity: sex-based differences predispose immune responses. *Nat. Rev. Immunol.* 2008, *8*, 737–744, doi:10.1016/j.cbpc.2008.10.051.

215. Pinheiro, I.; Dejager, L.; Libert, C. X-chromosome-located microRNAs in immunity: Might they explain male/female differences?: The X chromosome-genomic context may affect X-located miRNAs and downstream signaling, thereby contributing to the enhanced immune response of females. *BioEssays* 2011, *33*, 791–802, doi:10.1002/bies.201100047.
216. Carrel, L.; Willard, H.F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nat. Lett.* 2005, *434*, 400–404, doi:10.1038/nature03479.
217. Menon, R.; Di, D.M.; Cordiglieri, C.; Musio, S.; La, M.L.; Milanese, C.; Di Stefano, A.L.; Crabbio, M.; Franciotta, D.; Bergamaschi, R.; et al. Gender-based blood transcriptomes and interactomes in multiple sclerosis: involvement of SP1 dependent gene transcription. *J.Autoimmun.* 2012, *38*, J144–J155.
218. Irizar, H.; Muñoz-Culla, M.; Zuriarrain, O.; Goyenechea, E.; Castillo-Triviño, T.; Prada, A.; Saenz-Cuesta, M.; De Juan, D.; Lopez De Munain, A.; Olascoaga, J.; et al. HLA-DRB1*15:01 and multiple sclerosis: A female association? *Mult. Scler. J.* 2012, *18*, 569–577, doi:10.1177/1352458511426813.
219. Camiña-Tato, M.; Morcillo-Suárez, C.; Bustamante, M.F.; Ortega, I.; Navarro, A.; Muntasell, A.; López-Botet, M.; Sánchez, A.; Carmona, P.; Julià, E.; et al. Gender-Associated Differences of Perforin Polymorphisms in the Susceptibility to Multiple Sclerosis. *J. Immunol.* 2010, *185*, 5392–5404, doi:10.4049/jimmunol.1000102.
220. Irizar, H.; Muñoz-Culla, M.; Sepúlveda, L.; Saenz-Cuesta, M.; Prada, Á.; Castillo-Triviño, T.; Zamora-López, G.; Lopez de Munain, A.; Olascoaga, J.; Otaegui, D. Transcriptomic Profile Reveals Gender-Specific Molecular Mechanisms Driving Multiple Sclerosis Progression. *PLoS One* 2014, *9*, e90482, doi:10.1371/journal.pone.0090482.
221. Achiron, A.; Gurevich, M. Gender effects in relapsing-remitting multiple sclerosis: correlation between clinical variables and gene expression molecular pathways. *J. Neurol. Sci.* 2009, *286*, 47–53.
222. Muñoz-Culla, M.; Irizar, H.; Sáenz-Cuesta, M.; Castillo-Triviño, T.; Osorio-Querejeta, I.; Sepúlveda, L.; López de Munain, A.; Olascoaga, J.; Otaegui, D. SncRNA (microRNA &snoRNA) opposite expression pattern found in multiple sclerosis relapse and remission is sex dependent. *Sci. Rep.* 2016, *6*, 1–10, doi:10.1038/srep20126.
223. Duttagupta, R.; Jiang, R.; Gollub, J.; Getts, R.C.; Jones, K.W. Impact of cellular miRNAs on circulating miRNA biomarker signatures. *PLoS One* 2011, *6*, e20769, doi:10.1371/journal.pone.0020769.
224. Morgan, C.P.; Bale, T.L. Sex differences in microRNA regulation of gene expression: No smoke, just miRs. *Biol. Sex Differ.* 2012, *3*, 1–9, doi:10.1186/2042-6410-3-22.
225. Clayton, J.A.; Collins, F.S. NIH to balance sex in cell and animal studies. *Nature* 2014, *509*, 282–283, doi:10.1038/509282a.

226. Comabella, M.; Montalban, X. Body fluid biomarkers in multiple sclerosis. *Lancet Neurol.* 2014, *13*, 113–126, doi:10.1016/s1474-4422(13)70233-3.
227. Sorensen, P.S.; Sellebjerg, F.; Hartung, H.-P.; Montalban, X.; Comi, G.; Tintoré, M. The apparently milder course of multiple sclerosis: changes in the diagnostic criteria, therapy and natural history. *Brain* 2020, doi:10.1093/brain/awaa145.
228. Teunissen, C.E.; Malekzadeh, A.; Leurs, C.; Bridel, C.; Killestein, J. Body fluid biomarkers for multiple sclerosis—the long road to clinical application. *Nat. Rev. Neurol.* 2015, *11*, 585–596, doi:10.1038/nrneurol.2015.173.
229. Yamout, B.; Sahraian, M.; Bohlega, S.; Al-Jumah, M.; Goueider, R.; Dahdaleh, M.; Inshasi, J.; Hashem, S.; Alsharoqi, I.; Khoury, S.; et al. Consensus Recommendations for the Diagnosis and Treatment of Multiple Sclerosis: 2019 revisions to The MENACTRIMS Guidelines. *Mult. Scler. Relat. Disord.* 2019, *37*, 1–13, doi:10.1016/j.msard.2019.101459.
230. Polivka, J.; Polivka Jr., J.; Krakorova, K.; Peterka, M.; Topolcan, O. Current status of biomarker research in neurology. *EPMA J.* 2016, *7*, 1–13, doi:10.1186/s13167-016-0063-5.
231. Martinez, B.; Peplow, P. V MicroRNAs in blood and cerebrospinal fluid as diagnostic biomarkers of multiple sclerosis and to monitor disease progression. *Neural Regen. Res.* 2020, *15*, 606–619, doi:10.4103/1673-5374.266905.
232. Kuhle, J.; Barro, C.; Disanto, G.; Mathias, A.; Soneson, C.; Bonnier, G.; Yaldizli, Ö.; Regeniter, A.; Derfuss, T.; Canales, M.; et al. Serum neurofilament light chain in early relapsing remitting MS is increased and correlates with CSF levels and with MRI measures of disease severity. *Mult. Scler.* 2016, *22*, 1550–1559, doi:10.1177/1352458515623365.
233. Lennon, V.A.; Wingerchuk, D.M.; Kryzer, T.J.; Pittock, S.J.; Lucchinetti, C.F.; Fujihara, K.; Nakashima, I.; Weinshenker, B.G. A serum autoantibody marker of neuromyelitis optica: Distinction from multiple sclerosis. *Lancet* 2004, *364*, 2106–2112, doi:10.1016/S0140-6736(04)17551-X.
234. Calabresi, P. a; Giovannoni, G.; Confavreux, C.; Galetta, S.L.; Havrdova, E.; Hutchinson, M.; Kappos, L.; Miller, D.H.; O'Conner PW The incidence and significance of anti-natalizumab antibodies. *Neurology* 2007, *69*, 1391–1403, doi:10.1212/01.wnl.0000277457.17420.b5.
235. Bloomgren, G.; Richman, S.; Hotermans, C.; Subramanyam, M.; Goelz, S.; Natarajan, A.; Lee, S.; Plavina, T.; Scanlon, J. V.; Sandrock, A.; et al. Risk of natalizumab-associated progressive multifocal leukoencephalopathy. *N. Engl. J. Med.* 2012, *366*, 1870–1880, doi:10.1056/NEJMoa1107829.
236. Polman, C.H.; Bertolotto, A.; Deisenhammer, F.; Giovannoni, G.; Hartung, H.P.; Hemmer, B.; Killestein, J.; McFarland, H.F.; Oger, J.; Pachner, A.R.; et al. Recommendations for clinical use of data on neutralising antibodies to interferon-beta therapy in multiple sclerosis. *Lancet Neurol.* 2010, *9*, 740–750, doi:10.1016/S1474-4422(10)70103-4.

237. Loebermann, M.; Winkelmann, A.; Hartung, H.P.; Hengel, H.; Reisinger, E.C.; Zettl, U.K. Vaccination against infection in patients with multiple sclerosis. *Nat. Rev. Neurol.* 2012, *8*, 143–151, doi:10.1038/nrneurol.2012.8.

238. Jagot, F.; Davoust, N. Is it worth considering circulating microRNAs in multiple sclerosis? *Front. Immunol.* 2016, *7*, 1–11, doi:10.3389/fimmu.2016.00129.

239. Paul, A.; Comabella, M.; Gandhi, R. Biomarkers in Multiple Sclerosis. *Cold Spring Harb Perspect Med* 2018, *9*, pii: a029058, doi:10.1101/cshperspect.a029058.

240. Muñoz-Culla, M.; Irizar, H.; Castillo-Triviño, T.; Sáenz-Cuesta, M.; Sepúlveda, L.; Lopetegi, I.; De Munain, A.L.; Olascoaga, J.; Baranzini, S.E.; Otaegui, D. Blood miRNA expression pattern is a possible risk marker for natalizumab-associated progressive multifocal leukoencephalopathy in multiple sclerosis patients. *Mult. Scler. J.* 2014, *20*, 1851–1859, doi:10.1177/1352458514534513.

241. Fenoglio, C.; De Riz, M.; Pietroboni, A.M.; Calvi, A.; Serpente, M.; Cioffi, S.M.G.; Arcaro, M.; Oldoni, E.; Scarpini, E.; Galimberti, D. Effect of fingolimod treatment on circulating miR-15b, miR23a and miR-223 levels in patients with multiple sclerosis. *J. Neuroimmunol.* 2016, *299*, 81–83, doi:10.1016/j.jneuroim.2016.08.017.

242. Quintana, E.; Ortega, F.J.; Robles-Cedeño, R.; Villar, M.L.; Buxó, M.; Mercader, J.M.; Alvarez-Cermeño, J.C.; Pueyo, N.; Perkal, H.; Fernández-Real, J.M.; et al. miRNAs in cerebrospinal fluid identify patients with MS and specifically those with lipid-specific oligoclonal IgM bands. *Mult. Scler. J.* 2017, *23*, 1716–1726, doi:10.1177/1352458516684213.

243. Regev, K.; Paul, A.; Healy, B.; von Glenn, F.; Diaz-Cruz, C.; Gholipour, T.; Mazzola, M.A.; Raheja, R.; Nejad, P.; Glanz, B.I.; et al. Comprehensive evaluation of serum microRNAs as biomarkers in multiple sclerosis. *Neurol. Neuroimmunol. neuroinflammation* 2016, *3*, e267, doi:10.1212/NXI.0000000000000267.

244. Kemppinen, A.K.; Kaprio, J.; Palotie, A.; Saarela, J. Systematic review of genome-wide expression studies in multiple sclerosis. *BMJ Open* 2011, *1*, e000053–e000053, doi:10.1136/bmjopen-2011-000053.

245. Nickles, D.; Chen, H.P.; Li, M.M.; Khankhanian, P.; Madireddy, L.; Caillier, S.J.; Santaniello, A.; Cree, B.A.C.; Pelletier, D.; Hauser, S.L.; et al. Blood RNA profiling in a large cohort of multiple sclerosis patients and healthy controls. *Hum. Mol. Genet.* 2013, *22*, 4194–4205, doi:10.1093/hmg/ddt267.

246. Muñoz-Culla, M.; Irizar, H.; Sáenz-Cuesta, M.; Castillo-Triviño, T.; Osorio-Querejeta, I.; Sepúlveda, L.; López De Munain, A.; Olascoaga, J.; Otaegui, D. SncRNA (microRNA &snoRNA) opposite expression pattern found in multiple sclerosis relapse and remission is sex dependent. *Sci. Rep.* 2016, *6*, 1–10, doi:10.1038/srep20126.

247. Otaegui, D.; Baranzini, S.E.; Armañanzas, R.; Calvo, B.; Muñoz-Culla, M.; Khankhanian, P.; Inza, I.; Lozano, J.A.; Castillo-Triviño, T.; Asensio, A.; et al. Differential micro RNA expression in PBMC from multiple sclerosis patients. *PLoS One* 2009, *4*, doi:10.1371/journal.pone.0006309.

248. Irizar, H.; Muñoz-Culla, M.; Saenz-Cuesta, M.; Osorio-Querejeta, I.; Sepulveda, L.; Castillo-Triviño, T.; Prada, Á.; Lopez de Munain, A.; Olascoaga, J.; Otaegui, D. Identification of ncRNAs as potential therapeutic targets in multiple sclerosis through differential ncRNA – mRNA network analysis. *BMC Genomics* 2015, *16*, doi: 10.1186/s12864-015-1396-5.
249. Dolati, S.; Marofi, F.; Babaloo, Z.; Aghebati-Maleki, L.; Roshangar, L.; Ahmadi, M.; Rikhtegar, R.; Yousefi, M. Dysregulated Network of miRNAs Involved in the Pathogenesis of Multiple Sclerosis. *Biomed. Pharmacother.* 2018, *104*, 280–290, doi:10.1016/j.biopha.2018.05.050.
250. Aufiero, S.; Reckman, Y.J.; Pinto, Y.M.; Creemers, E.E. Circular RNAs open a new chapter in cardiovascular biology. *Nat. Rev. Cardiol.* 2019, *16*, 503–514, doi:10.1038/s41569-019-0185-2.
251. Floris, G.; Zhang, L.; Follesa, P.; Sun, T. Regulatory Role of Circular RNAs and Neurological Disorders. *Mol. Neurobiol.* 2016, *54*, 5156–5165, doi:10.1007/s12035-016-0055-4.
252. Zeng, X.; Lin, W.; Guo, M.; Zou, Q. A comprehensive overview and evaluation of circular RNA detection tools. *PloS Comput. Biol.* 2017, *13*, 1–21.
253. Lu, D.; Xu, A.D. Mini Review: Circular RNAs as potential clinical biomarkers for disorders in the central nervous system. *Front. Genet.* 2016, *7*, 1–5, doi:10.3389/fgene.2016.00053.
254. Untergasser, A.; Nijveen, H.; Rao, X.; Bisseling, T.; Geurts, R.; Leunissen, J.A.M. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 2007, *35*, 71–74, doi:10.1093/nar/gkm306.
255. DeLong, E.R.; DeLong, D.M.; Clarke-Pearson, D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988, *44*, 837–845.
256. Reina-San-Martín, B.; Cosson, A.; Minoprio, P. Lymphocyte polyclonal activation: A pitfall for vaccine design against infectious agents. *Parasitol. Today* 2000, *16*, 62–67, doi:10.1016/S0169-4758(99)01591-4.
257. Kulcheski, F.R.; Christoff, A.P.; Margis, R. Circular RNAs are miRNA sponges and can be used as a new class of biomarker. *J. Biotechnol.* 2016, *238*, 42–51, doi:10.1016/j.jbiotec.2016.09.011.
258. Hwang, J.M.; Chang, B.L.; Park, S.S. Leber's hereditary optic neuropathy mutations in Korean patients with multiple sclerosis. *Ophthalmologica 215*, 398–400.
259. Iparraguirre, L.; Muñoz-Culla, M.; Prada-Luengo, I.; Castillo-Triviño, T.; Olascoaga, J.; Otaegui, D. Circular RNA profiling reveals that circular RNAs from ANXA2 can be used as new biomarkers for multiple sclerosis. *Hum. Mol. Genet.* 2017, *26*, 3564–3572, doi:10.1093/hmg/ddx243.
260. Cardamone, G.; Paraboschi, E.M.; Rimoldi, V.; Duga, S.; Soldà, G.; Asselta, R. The Characterization of GSDMB Splicing and Backsplicing Profiles Identifies Novel Isoforms and a Circular RNA That Are Dysregulated in Multiple Sclerosis. *Int. J. Mol. Sci.* 2017, *18*, 1–15, doi:10.3390/ijms18030576.

261. Yamamoto, T.; Kudo, M.; Peng, W.-X.; Takata, H.; Takakura, H.; Teduka, K.; Fujii, T.; Mitamura, K.; Taga, A.; Uchida, E.; et al. Identification of aldolase A as a potential diagnostic biomarker for colorectal cancer based on proteomic analysis using formalin-fixed paraffin-embedded tissue. *Tumor Biol.* 2016, *37*, 13595–13606, doi:10.1007/s13277-016-5275-8.

262. El-Abd, N.; Fawzy, A.; Elbaz, T.; Hamdy, S. Evaluation of annexin A2 and as potential biomarkers for hepatocellular carcinoma. *Tumor Biol.* 2016, *37*, 211–216, doi:10.1007/s13277-015-3524-x.

263. Kling, T.; Ferrarese, R.; Ó hAilín, D.; Johansson, P.; Heiland, D.H.; Dai, F.; Vasilikos, I.; Weyerbrock, A.; Jörnsten, R.; Carro, M.S.; et al. Integrative Modeling Reveals Annexin A2-mediated Epigenetic Control of Mesenchymal Glioblastoma. *EBioMedicine* 2016, *12*, 72–85, doi:10.1016/j.ebiom.2016.08.050.

264. Cañas, F.; Simonin, L.; Couturaud, F.; Renaudineau, Y. Annexin A2 autoantibodies in thrombosis and autoimmune diseases. *Thromb. Res.* 2015, *135*, 226–230, doi:10.1016/j.thromres.2014.11.034.

265. Pianta, A.; Drouin, E.E.; Crowley, J.T.; Arvikar, S.; Strle, K.; Costello, C.E.; Steere, A.C. Annexin A2 is a target of autoimmune T and B cell responses associated with synovial fibroblast proliferation in patients with antibiotic-refractory Lyme arthritis. *Clin. Immunol.* 2015, *160*, 336–341, doi:10.1016/j.clim.2015.07.005.

266. Swisher, J.F.A.; Khatri, U.; Feldman, G.M. Annexin A2 is a soluble mediator of macrophage activation. *J. Leukoc. Biol.* 2007, *82*, 1174–1184, doi:10.1189/jlb.0307154.

267. Chao, P.-Z.; Hsieh, M.-S.; Cheng, C.-W.; Hsu, T.-J.; Lin, Y.-T.; Lai, C.-H.; Liao, C.-C.; Chen, W.-Y.; Leung, T.-K.; Lee, F.-P.; et al. Dendritic cells respond to nasopharygeal carcinoma cells through annexin A2-recognizing DC-SIGN. *Oncotarget* 2015, *6*, 159–170, doi:10.18632/oncotarget.2700.

268. Fang, W.; Fa, Z.-Z.; Xie, Q.; Wang, G.-Z.; Yi, J.; Zhang, C.; Meng, G.-X.; Gu, J.-L.; Liao, W.-Q. Complex Roles of Annexin A2 in Host Blood-Brain Barrier Invasion by *Cryptococcus neoformans*. *CNS Neurosci. Ther.* 2017, doi:10.1111/cns.12673.

269. Lopez-Ramirez, M.A.; Wu, D.; Pryce, G.; Simpson, J.E.; Reijerkerk, A.; King-Robson, J.; Kay, O.; De Vries, H.E.; Hirst, M.C.; Sharrack, B.; et al. MicroRNA-155 negatively affects blood-brain barrier function during neuroinflammation. *FASEB J.* 2014, *28*, 2551–2565, doi:10.1096/fj.13-248880.

270. Zhang, J.; Cheng, Y.; Cui, W.; Li, M.; Li, B.; Guo, L. MicroRNA-155 modulates Th1 and Th17 cell differentiation and is associated with multiple sclerosis and experimental autoimmune encephalomyelitis. *J. Neuroimmunol.* 2014, *266*, 56–63, doi:10.1016/j.jneuroim.2013.09.019.

271. Moore, C.S.; Rao, V.T.S.; Durafourt, B.A.; Bedell, B.J.; Ludwin, S.K.; Bar-Or, A.; Antel, J.P. MiR-155 as a multiple sclerosis-relevant regulator of myeloid cell polarization. *Ann. Neurol.* 2013, *74*, 709–720, doi:10.1002/ana.23967.

272. Hagiwara, K.; Katsuda, T.; Gailhouste, L.; Kosaka, N.; Ochiya, T. Commitment of Annexin A2 in recruitment of microRNAs into extracellular vesicles. *FEBS Lett.* 2015, *589*, 4071–4078, doi:10.1016/j.febslet.2015.11.036.
273. Sand, I.K.; Krieger, S.; Farrell, C.; Miller, A.E. Diagnostic uncertainty during the transition to secondary progressive multiple sclerosis. *Mult. Scler. J.* 2014, *20*, 1654–1657, doi:10.1177/1352458514521517.
274. Acquaviva, M.; Menon, R.; Di Dario, M.; Dalla Costa, G.; Romeo, M.; Sangalli, F.; Colombo, B.; Moiola, L.; Martinelli, V.; Comi, G.; et al. Inferring Multiple Sclerosis Stages from the Blood Transcriptome via Machine Learning. *Cell Reports Med.* 2020, *1*, 100053, doi:10.1016/j.xcrm.2020.100053.
275. Srinivasan, S.; Dario, M. Di; Russo, A.; Menon, R.; Brini, E.; Romeo, M.; Sangalli, F.; Costa, G.D.; Rodegher, M.; Radaelli, M.; et al. Dysregulation of MS risk genes and pathways at distinct stages of disease. *Neurol. Neuroimmunol. NeuroInflammation* 2017, *4*, 1–10, doi:10.1212/NXI.0000000000000337.
276. Irizar, H.; Munoz-Culla, M.; Sepulveda, L.; Saenz-Cuesta, M.; Prada, A.; Castillo-Trivino, T.; Zamora-Lopez, G.; de Munain, A.L.; Olascoaga, J.; Otaegui, D. Transcriptomic profile reveals gender-specific molecular mechanisms driving multiple sclerosis progression. *PLoS One* 2014, *9*, e90482, doi:10.1371/journal.pone.0090482.
277. Achiron, A.; Gurevich, M. Gender effects in relapsing-remitting multiple sclerosis: Correlation between clinical variables and gene expression molecular pathways. *J. Neurol. Sci.* 2009, *286*, 47–53, doi:10.1016/j.jns.2009.06.038.
278. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009, *25*, 1754–1760, doi:10.1093/bioinformatics/btp324.
279. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009, *10*, R25.2-R25.10, doi:10.1186/gb-2009-10-3-r25.
280. Gao, Y.; Zhang, J.; Zhao, F. Circular RNA identification based on multiple seed matching. *Brief. Bioinform.* 2018, *19*, 803–810, doi:10.1093/bib/bbx014.
281. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014, *15*, 1–21, doi:10.1186/s13059-014-0550-8.
282. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 2013, *29*, 15–21, doi:10.1093/bioinformatics/bts635.
283. Anders, S.; Pyl, P.T.; Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015, *31*, 166–169, doi:10.1093/bioinformatics/btu638.
284. Eden, E.; Navon, R.; Steinfeld, I.; Lipson, D.; Yakhini, Z. GOrilla : a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 2009, *7*, 1–7, doi:10.1186/1471-2105-10-48.

285. Iparraguirre, L.; Alberro, A.; Sepúlveda, L.; Osorio-Querejeta, I.; Moles, L.; Castillo-Triviño, T.; Hansen, T.B.; Muñoz-Culla, M.; Otaegui, D. RNA-Seq profiling of leukocytes reveals a sex-dependent global circular RNA upregulation in multiple sclerosis and 6 candidate biomarkers. *Hum. Mol. Genet.* 2020, 1–12, doi:10.1093/hmg/ddaa219.

286. Zurawska, A.; Mycko, M.P.; Selmaj, K.W. Circular RNAs as a novel layer of regulatory mechanism in multiple sclerosis. *J. Neuroimmunol.* 2019, *334*, 576971, doi:10.1016/j.jneuroim.2019.576971.

287. Hansen, T.B.; Venø, M.T.; Damgaard, C.K.; Kjems, J. Comparison of circular RNA prediction tools. *Nucleic Acids Res.* 2016, *44*, e58, doi:10.1093/nar/gkv1458.

288. Chen, L. The biogenesis and emerging roles of circular RNAs. *Nat. Rev. Mol. Cell Biol.* 2016, *17*, 205–211.

289. Palmer, C.; Diehn, M.; Alizadeh, A.A.; Brown, P.O. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics* 2006, *16*, 1–15, doi:10.1186/1471-2164-7-115.

290. Woodberry, T.; Bouffler, S.E.; Wilson, A.S.; Buckland, R.L.; Brüstle, A. The Emerging Role of Neutrophil Granulocytes in Multiple Sclerosis. *J. Clin. Med.* 2018, *17*, 10–12, doi:10.3390/jcm7120511.

291. Moldovan, L.; Hansen, T.B.; Venø, M.T.; Line, T.; Okholm, H.; Andersen, T.L.; Hager, H.; Iversen, L.; Kjems, J.; Johansen, C. High-throughput RNA sequencing from paired lesional- and non-lesional skin reveals major alterations in the psoriasis circRNAome. *BMC Genomics* 2019, *12*, 1–17.

292. Zheng, Q.; Bao, C.; Guo, W.; Li, S.; Chen, J.; Chen, B.; Luo, Y.; Lyu, D.; Li, Y.; Shi, G.; et al. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat. Commun.* 2016, *7*, 1–13, doi:10.1038/ncomms11215.

293. Villar, L.M.; Sádaba, M.C.; Roldán, E.; Masjuan, J.; González-Porqué, P.; Villarrubia, N.; Espiño, M.; García-Trujillo, J.A.; Bootello, A.; Álvarez-Cermeño, J.C. Intrathecal synthesis of oligoclonal IgM against myelin lipids predicts an aggressive disease course in MS. *J. Clin. Invest.* 2005, *115*, 187–194, doi:10.1172/JCI22833.

294. Villar, L.M.; González-Porqué, P.; Masjuán, J.; Álvarez-Cermeño, J..; Bootello, A.; Keir, G. A sensitive and reproducible method for the detection of oligoclonal IgM bands. *J. Immunol. Methods* 2001, *258*, 151–155.

295. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, a. D. The Human Genome Browser at UCSC. *Genome Res.* 2002, *12*, 996–1006, doi:10.1101/gr.229102.

296. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows – Wheeler transform. 2009, *25*, 1754–1760, doi:10.1093/bioinformatics/btp324.

297. Bray, N.L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 2016, *34*, 525–528, doi:10.1038/nbt.3519.

298. Mi, H.; Thomas, P. PANTHER Pathway: an ontology-based pathway database coupled with data analysis tools. *Nat. Protoc.* 2019, *14*, 703–721, doi:10.1007/978-1-60761-175-2.

299. Luo, M.; Ma, W.; Sand, Z.; Finlayson, J.; Wang, T.; Diaz, R.; Willis, W.T.; Mandarino, L.J. Von Willebrand factor A domain-containing protein 8 (VWA8) localizes to the matrix side of the inner mitochondrial membrane. *Biochem. Biophys. Res. Commun.* 2020, *521*, 158–163, doi:10.1016/j.bbrc.2019.10.095.

300. Vandenbroeck, K.; Alloza, I.; Swaminathan, B.; Antiguedad, A.; Otaegui, D.; Olascoaga, J.; Barcina, M.G.; de las Heras, V.; Bartolome, M.; Fernandez-Arquero, M.; et al. Validation of IRF5 as multiple sclerosis risk gene: putative role in interferon beta therapy and human herpes virus-6 infection. *Genes Immun* 2011, *12*, 40–45, doi:10.1038/gene.2010.46.

301. Lindén, M.; Khademi, M.; Bomfim, I.L.; Piehl, F.; Jagodic, M.; Kockum, I.; Olsson, T. Multiple sclerosis risk genotypes correlate with an elevated cerebrospinal fluid level of the suggested prognostic marker CXCL13. *Mult. Scler. J.* 2012, *19*, 863–870, doi:10.1177/1352458512463482.

302. Sellebjerg, F.; Christiansen, M.; Garred, P. MBP, anti-MBP and anti-PLP antibodies, and intrathecal complement activation in multiple sclerosis. *Mult. Scler.* 1998, *4*, 127–131, doi:10.1177/135245859800400307.

303. Mead, R.J.; Singhrao, S.K.; Neal, J.W.; Lassmann, H.; Morgan, B.P. The Membrane Attack Complex of Complement Causes Severe Demyelination Associated with Acute Axonal Injury. *J. Immunol.* 2002, *168*, 458–465, doi:10.4049/jimmunol.168.1.458.

304. Ingram, G.; Hakobyan, S.; Hirst, C.L.; Harris, C.L.; Loveless, S.; Mitchell, J.P.; Pickersgill, T.P.; Robertson, N.P.; Morgan, B.P. Systemic complement profiling in multiple sclerosis as a biomarker of disease state. *Mult. Scler. J.* 2012, *18*, 1401–1411, doi:10.1177/1352458512438238.

305. Mandrekar, J.N. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *J. Thorac. Oncol.* 2010, *5*, 1315–1316, doi:10.1097/JTO.0b013e3181ec173d.

306. Iacobaeus, E.; Ryschkewitsch, C.; Gravell, M.; Khademi, M.; Wallstrom, E.; Olsson, T.; Brundin, L.; Major, E. Analysis of cerebrospinal fluid and cerebrospinal fluid cells from patients with multiple sclerosis for detection of JC virus DNA. *Mult. Scler.* 2009, *15*, 28–35, doi:10.1177/1352458508096870.

307. Håkansson, I.; Tisell, A.; Cassel, P.; Blennow, K.; Zetterberg, H.; Lundberg, P.; Dahle, C.; Vrethem, M.; Ernerudh, J. Neurofilament levels , disease activity and brain volume during follow-up in multiple sclerosis. *J. Neuroinflammation* 2018, *15*, 1–10.

308. Khademi, M.; Kockum, I.; Andersson, M.L.; Iacobaeus, E.; Brundin, L.; Sellebjerg, F.; Hillert, J.; Piehl, F.; Olsson, T. Cerebrospinal fluid CXCL13 in multiple sclerosis : a suggestive prognostic marker for the disease course. *Mult. Scler. J.* 2011, *17*, 335–343, doi:10.1177/1352458510389102.

309. Cantó, E.; Tintore, M.; Villar, L.M.; Costa, C.; Nurtdinov, R.; Deisenhammer, F.; Hegen, H.; Khademi, M.; Olsson, T.; Piehl, F.; et al. Chitinase 3-like 1 : prognostic biomarker in clinically isolated syndromes. *Brain* 2015, *138*, 918–931, doi:10.1093/brain/awv017.

310. Vistbakka, J.; Sumelahti, M.; Lehtimäki, T.; Elovaara, I.; Hagman, S. Evaluation of serum miR-191-5p, miR-24-3p, miR-128-3p and miR-376c-3 in multiple sclerosis patients. *Acta Neurol. Scand.* 2018, 1–7, doi:10.1111/ane.12921.

311. Gupta, M.; Martens, K.; Metz, L.M.; Jason, A.P.; Koning, D.; Pfeffer, G. Long noncoding RNAs associated with phenotypic severity in multiple sclerosis. *Mult. Scler. Relat. Disord.* 2019, *36*, 101407, doi:10.1016/j.msard.2019.101407.

312. Iacobaeus, E.; Arrambide, G.; Amato, M.P.; Derfuss, T.; Vukusic, S.; Hemmer, B.; Tintore, M. Aggressive multiple sclerosis (1): Towards a definition of the phenotype. 2020, 1031–1044, doi:10.1177/1352458520925369.

313. Díaz, C.; Zarco, L.A.; Rivera, D.M. Highly active multiple sclerosis : An update. *Mult. Scler. Relat. Disord.* 2019, *30*, 215–224, doi:10.1016/j.msard.2019.01.039.

314. Raposo, G.; Stoorvogel, W. Extracellular vesicles: Exosomes, microvesicles, and friends. *J. Cell Biol.* 2013, *200*, 373–383, doi:10.1083/jcb.201211138.

315. Yáñez-Mó, M.; Siljander, P.R.M.; Andreu, Z.; Zavec, A.B.; Borràs, F.E.; Buzas, E.I.; Buzas, K.; Casal, E.; Cappello, F.; Carvalho, J.; et al. Biological properties of extracellular vesicles and their physiological functions. *J. Extracell. Vesicles* 2015, *4*, 1–60, doi:10.3402/jev.v4.27066.

316. Selmaj, I.; Mycko, M.P.; Raine, C.S.; Selmaj, K.W. The role of exosomes in CNS inflammation and their involvement in multiple sclerosis. *J. Neuroimmunol.* 2017, *306*, 1–10, doi:10.1016/j.jneuroim.2017.02.002.

317. Xu, K.; Liu, Q.; Wu, K.; Liu, L.; Zhao, M.; Yang, H.; Wang, X.; Wang, W. Extracellular vesicles as potential biomarkers and therapeutic approaches in autoimmune diseases. *J. Transl. Med.* 2020, *18*, 1–8, doi:10.1186/s12967-020-02609-0.

318. Sáenz-Cuesta, M.; Osorio-Querejeta, I.; Otaegui, D. Extracellular vesicles in multiple sclerosis: What are they telling us? *Front. Cell. Neurosci.* 2014, *8*, 1–9, doi:10.3389/fncel.2014.00100.

319. Selmaj, I.; Cichalewska, M.; Namiecinska, M.; Galazka, G.; Horzelski, W.; Selmaj, K.W.; Mycko, M.P. Global Exosome Transcriptome Profiling Reveals Biomarkers for Multiple Sclerosis. *Ann. Neurol.* 2017, *81*, 703–717.

320. Sáenz-Cuesta, M.; Alberro, A.; Muñoz-Culla, M.; Osorio-Querejeta, I.; Fernandez-Mercado, M.; Lopetegui, I.; Tainta, M.; Prada, Á.; Castillo-Triviño, T.; Falcón-Pérez, J.M.; et al. The first dose of fingolimod affects circulating extracellular vesicles in multiple sclerosis patients. *Int. J. Mol. Sci.* 2018, *19*, doi:10.3390/ijms19082448.

321. Sáenz-Cuesta, M.; Arbelaiz, A.; Oregi, A.; Irizar, H.; Osorio-Querejeta, I.; Muñoz-Culla, M.; Banales, J.M.; Falcón-Pérez, J.M.; Olascoaga, J.; Otaegui, D. Methods for extracellular vesicles isolation in a hospital setting. *Front. Immunol.* 2015, *6*, doi:10.3389/fimmu.2015.00050.

322. Gardiner, C.; Ferreira, Y.J.; Dragovic, R.A.; Redman, C.W.G.; Sargent, I.L. Extracellular vesicle sizing and enumeration by nanoparticle tracking analysis. *J. Extracell. Vesicles* 2013, *2*, 19671, doi:10.3402/jev.v2i0.19671.
323. Julián, P.; Milon, P.; Agirrezabala, X.; Lasso, G.; Gil, D.; Rodnina, M. V.; Valle, M. The cryo-EM structure of a complete 30S translation initiation complex from Escherichia coli. *PLoS Biol.* 2011, *9*, 1–11, doi:10.1371/journal.pbio.1001095.
324. Li, Y.; He, X.; Li, Q.; Lai, H.; Zhang, H.; Hu, Z.; Li, Y.; Huang, S. EV-origin: Enumerating the tissue-cellular origin of circulating extracellular vesicles using exLR profile. *Comput. Struct. Biotechnol. J.* 2020, *18*, 2851–2859, doi:10.1016/j.csbj.2020.10.002.
325. Aufiero, S.; Reckman, Y.J.; Pinto, Y.M.; Creemers, E.E. Circular RNAs open a new chapter in cardiovascular biology. *Nat. Rev. Cardiol.* 2019, *16*, 503–514, doi:10.1038/s41569-019-0185-2.
326. Shao, Y.; Chen, Y. Roles of Circular RNAs in Neurologic Disease. 2016, *9*, 1–5, doi:10.3389/fnmol.2016.00025.
327. Kristensen, L.S.; Hansen, T.B.; Venø, M.T.; Kjems, J. Circular RNAs in cancer : opportunities and challenges in the fi eld. *Nat. Publ. Gr.* 2017, *37*, 555–565, doi:10.1038/onc.2017.361.
328. Legnini, I.; Timoteo, G. Di; Rossi, F.; Morlando, M.; Briganti, F.; Sthandier, O.; Fatica, A.; Santini, T.; Andronache, A.; Wade, M.; et al. Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cell* 2017, *66*, 22-37.e9, doi:10.1016/j.molcel.2017.02.017.
329. Ragan, C.; Goodall, G.J.; Shirokikh, N.E.; Preiss, T. Insights into the biogenesis and potential functions of exonic circular RNA. *Sci. Rep.* 2019, *9*, 1–18, doi:10.1038/s41598-018-37037-0.
330. Agarwal, V.; Bell, G.W.; Nam, J.W.; Bartel, D.P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 2015, *4*, 1–38, doi:10.7554/eLife.05005.
331. Smola, M.J.; Rice, G.M.; Busan, S.; Siegfried, N.A.; Weeks, K.M. Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* 2015, *10*, 1643–1669, doi:10.1038/nprot.2015.103.
332. Zubradt, M.; Gupta, P.; Persad, S.; Lambowitz, A.M.; Weissman, J.S.; Rouskin, S. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* 2016, *14*, 75–82, doi:10.1038/nmeth.4057.
333. Zheng, Q.; Bao, C.; Guo, W.; Li, S.; Chen, J.; Chen, B.; Luo, Y.; Lyu, D.; Li, Y.; Shi, G.; et al. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat. Commun.* 2016, *7*, doi:10.1038/ncomms11215.
334. Yu, C.Y.; Li, T.C.; Wu, Y.Y.; Yeh, C.H.; Chiang, W.; Chuang, C.Y.; Kuo, H.C. The circular RNA circBIRC6 participates in the molecular circuitry controlling human pluripotency. *Nat. Commun.* 2017, *8*, doi:10.1038/s41467-017-01216-w.

335. Mortimer, S.A.; Weeks, K.M. A Fast-acting reagent for accurate analysis RNA secondary and tertiary structure by SHAPE Chemistry. *J. Am. Chem. Soc.* 2007, *129*, 4144–4145.
336. Mogensen, T.H. Pathogen recognition and inflammatory signaling in innate immune defenses. *Clin. Microbiol. Rev.* 2009, *22*, 240–273, doi:10.1128/CMR.00046-08.
337. Hansen, T.B. Characterization of circular RNA concatemers. In *Methods in Molecular Biology*; Dieterich, C., Papantonis, A., Eds.; Humana Press: New York, 2018; Vol. 1724, pp. 143–157 ISBN 9781493975624.
338. Clavarino, G.; Cláudio, N.; Couderc, T.; Dalet, A.; Judith, D.; Camosseto, V.; Schmidt, E.K.; Wenger, T.; Lecuit, M.; Gatti, E.; et al. Induction of GADD34 is necessary for dsRNA-dependent interferon-β production and participates in the control of Chikungunya virus infection. *PLoS Pathog.* 2012, *8*, doi:10.1371/journal.ppat.1002708.
339. Yuen, K.C.; Xu, B.; Krantz, I.D.; Gerton, J.L. NIPBL Controls RNA Biogenesis to Prevent Activation of the Stress Kinase PKR. *Cell Rep.* 2016, *14*, 93–102, doi:10.1016/j.celrep.2015.12.012.
340. Lee, E.S.; Yoon, C.H.; Kim, Y.S.; Bae, Y.S. The double-strand RNA-dependent protein kinase PKR plays a significant role in a sustained ER stress-induced apoptosis. *FEBS Lett.* 2007, *581*, 4325–4332, doi:10.1016/j.febslet.2007.08.001.
341. Segev, Y.; Barrera, I.; Ounallah-Saad, H.; Wibrand, K.; Sporild, I.; Livne, A.; Rosenberg, T.; David, O.; Mints, M.; Bramham, C.R.; et al. PKR inhibition rescues memory deficit and atf4 overexpression in apoe ε4 human replacement mice. *J. Neurosci.* 2015, *35*, 12986–12993, doi:10.1523/JNEUROSCI.5241-14.2015.
342. Pham, A.M.; Santa Maria, F.G.; Lahiri, T.; Friedman, E.; Marié, I.J.; Levy, D.E. PKR Transduces MDA5-Dependent Signals for Type I IFN Induction. *PLoS Pathog.* 2016, *12*, 1–27, doi:10.1371/journal.ppat.1005489.
343. McAllister, C.S.; Taghavi, N.; Samuel, C.E. Protein kinase PKR amplification of interferon β induction occurs through initiation factor eIF-2α-mediated translational control. *J. Biol. Chem.* 2012, *287*, 36384–36392, doi:10.1074/jbc.M112.390039.
344. Gupta, A.; Rath, P.C. Curcumin, a Natural Antioxidant, Acts as a Noncompetitive Inhibitor of Human RNase L in Presence of Its Cofactor 2-5A In Vitro. *Biomed. Res. Int.* 2014, *2014*, doi:10.1155/2014/817024.
345. Qureshi, M.; Al-Suhaimi, E.A.; Wahid, F.; Shehzad, O.; Shehzad, A. Therapeutic potential of curcumin for multiple sclerosis. *Neurol. Sci.* 2018, *39*, 207–214, doi:10.1007/s10072-017-3149-5.
346. Huang, S.H.; Wu, C.H.; Chen, S.J.; Sytwu, H.K.; Lin, G.J. Immunomodulatory effects and potential clinical applications of dimethyl sulfoxide. *Immunobiology* 2020, *225*, 151906, doi:10.1016/j.imbio.2020.151906.
347. Li, F.; Liu, Z.; Zhang, B.; Jiang, S.; Wang, Q.; Du, L.; Xue, H.; Zhang, Y.; Jin, M.; Zhu, X.; et al. Circular RNA sequencing indicates circ-IQGAP2 and circ-ZC3H6 as noninvasive biomarkers of primary Sjögren's syndrome. *Rheumatology (Oxford).* 2020, *59*, 2603–2615, doi:10.1093/rheumatology/keaa163.

348. Osorio-Querejeta, I.; Carregal-Romero, S.; Ayerdi-Izquierdo, A.; Mäger, I.; Nash, L.A.; Wood, M.; Egimendia, A.; Betanzos, M.; Alberro, A.; Iparraguirre, L.; et al. MiR-219a-5p enriched extracellular vesicles induce OPC differentiation and EAE improvement more efficiently than liposomes and polymeric nanoparticles. *Pharmaceutics* 2020, *12*, doi:10.3390/pharmaceutics12020186.

349. Osorio-Querejeta, I.; Alberro, A.; Muñoz-Culla, M.; Mäger, I.; Otaegui, D. Therapeutic potential of extracellular vesicles for demyelinating diseases; Challenges and opportunities. *Front. Mol. Neurosci.* 2018, *11*, 1–8, doi:10.3389/fnmol.2018.00434.

350. Gao, Y.; Wang, J.; Zhao, F. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.* 2015, *13*, doi: 10.1186/s13059-014-0571-3.

351. Wang, J.; Liu, K.; Liu, Y.; Lv, Q.; Zhang, F.; Wang, H. Evaluating the bias of circRNA predictions from total RNA-Seq data. *Oncotarget* 2017, *8*, 110914–110921, doi:10.18632/oncotarget.22972.

352. Panda, A.C.; Gorospe, M. Detection and Analysis of Circular RNAs by RT-PCR. 2018, *8*, doi:10.21769/BioProtoc.2775.Detection.

353. Taylor, S.C.; Laperriere, G.; Germain, H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: From variable nonsense to publication quality data. *Sci. Rep.* 2017, *7*, 1–8, doi:10.1038/s41598-017-02217-x.

354. Iparraguirre, L.; Prada-Luengo, I.; Regenberg, B.; Otaegui, D. To Be or Not to Be: Circular RNAs or mRNAs From Circular DNAs? *Front. Genet.* 2019, *10*, doi:10.3389/fgene.2019.00940.

355. Shi, Y.; Jia, X.; Xu, J. The new function of circRNA: translation. *Clin. Transl. Oncol.* 2020, *22*, 2162–2169, doi:10.1007/s12094-020-02371-1.

356. Stagsted, L.V.W.; Nielsen, K.M.; Daugaard, I.; Hansen, T.B. Noncoding AUG circRNAs constitute an abundant and conserved subclass of circles. *Life Sci. Alliance* 2019, *2*, 1–16, doi:10.26508/lsa.201900398.

Appendix

Appendix Table 1. List of primers and probes used during the thesis.

Gene	Transcript	Position	Primer	Sequence	Subchapter 1.1	Subchapter 1.2	Subchapter 1.3	Subchapter 2.2
RPPH1	hsa_circ_0000518	chr14:20811404-20811554	Forward	AGGTGAGTTCCCAGAGAACG	X			
			Reverse	GGACATGGGAGTGGAGTGAC				
	hsa_circ_0000517	chr14:20811404-20811492	Forward	AGTTCCCAGAGAACGGGG	X			
			Reverse	CCCGAAGCTCAGGGAGAG				
	hsa_circ_0000519	chr14:20811436-20811534	Forward	GGTCTGAGACTAGGGCCAGA	X			
			Reverse	GGACATGGGAGTGGAGTGA				
	hsa_circ_0000520	⊠hr14:20811436-20811559	Forward	CAGAGGCGGCCCTAACAG	X			
			Reverse	AAGGGACATGGGAGTGGAGT				
	NR_002312	chr14: 20811207-20811844	Forward	CCGGAGCTTGGAACAGACT	X			
			Reverse	AATGGGCGGAGGAGAGTAGT				
RELL1	hsa_circ_0001400	chr4:37633006-37640126	Forward	GAACGGAGAAGCCTGATGTC	X	X		X
			Reverse	GGAGCACTCACCTTTCAGGA				
	NM_001085399	chr4: 37636509-37667003	Forward	CCCCTCCCCTAATCAGTCTC	X	X		X
			Reverse	CATGCCTGTGGTTGACTCAG				
CACNB4	hsa_circ_0056731	chr2:152698416-152717334	Forward	ACGGCACTCTCACCATATCC	X			
			Reverse	TTGCTGGGATTATTTAGGACAGA				
	NM_001005747	chr2: 152732994-152828502	Forward	AAGAAAATCGAAATACCAATAGAAGC	X			
			Reverse	GGCACTATCAATACATTTGGTAAGTC				
RBMS3	hsa_circ_0064644	chr3:29910348-29941246	Forward	TGTTTGATGCATAGGTCCAGAG	X			
			Reverse	GCCTCCCATTCTGGGTATATT				
	NM_001177712	chr3: 29322864-29925717	Forward	GCACAGATGGCTAAGCAACA	X			
			Reverse	CAACACCTCTGCTGACTCCA				
ANXA2	hsa_circ_0005402	chr15:60648117-60674640	Forward	TTTCGGACACATCTGGTGAC	X			
			Reverse*	CCGCTCAGCATCAAAGTTAGT				
	hsa_circ_0003452_2	chr15:60653150-60678285	Forward	CACCTGGAGACGGTGATTTT	X			
			Reverse*	CCGCTCAGCATCAAAGTTAGT				
	NM_001136015	chr15: 60677756-60690185	Forward	TGCCTCCTGTGTACTGTGTCA	X			
			Reverse	CCCAAGCACGTTTAATTTTCA				
SNX19	hsa_circ_0024892	chr11:130749518-130749606	Forward	GGGGTCTAGTCCTGGAGTCAC	X			
			Reverse	GCTCAGCCGGCATTTG				
	NM_014758	chr11: 130747946-130763769	Forward	TTCTGTCGGCATTGCTAGTG	X			
			Reverse	GGTTGATCTCCCGTTCCAG				
EEF1A1	NM_001402	chr6: 74225473-74230741	Forward	TGGTATTGGTACTGTTCCTG	X	X	X	
			Reverse	CTTCACTCAAAGCTTCATGG				
ATP8B4	hsa_circ_0141241	chr15:50294349-50311173	Forward	TGAGCCACATGGTCTCTGTT		X		
			Reverse	TGAGCCACATGGTCTCTGTT				
AGFG1	hsa_circ_0058514	chr2:228356262-228389631	Forward	TCCCCAGTTGTAGGTCGTTC		X		
			Reverse	TAATCCTCGCCTGCATGACT				
PADI4	circPADI4	chr1:17668437-17668897	Forward	CTGTGGTGTTCCAAGACAGC		X		
			Reverse	GGTAGTGAGAGGGCCACTTG				
AFF2	hsa_circ_0001947	chrX:147743428-147744289	Forward	TTTGGGAATCTGTCATTTGGA		X		
			Reverse	TTTGTTTGTTTCACTTGTTTG				
ABCA13	hsa_circ_0001707	chr7:48541721-48542148	Forward	ATTGAGCTCCCCACAGAAAC		X		
			Reverse	CTCGCTGGTGAAATGATTGA				
NEIL3	hsa_circ_0001459	chr4:178274461-178274882	Forward	TGGATATTCTAACAGATTCAGTGCTC		X		
			Reverse	CTTGACTTCTGTATGAGGTTTTCC				
ANKHD1	hsa_circ_0001541	chr5:139819703-139828890	Forward	GCCTGCATGTCGCAGTCTA			X	
			Reverse	CACAGCAGGCTTTCTCCTTC				
METRNL	circMETRNL	chr17:81042814-81043199	Forward	AGGACCACAGGCTTCCAGTA			X	
			Reverse	AGCACCTGTTGGGTACATCC				
ASAP1	hsa_circ_0001824	chr8:131164981-131181313	Forward	CACATGCCACATTTCTTTCAA			X	
			Reverse	TTCTTCATCCTGGGTCTGTTTT				
KIAA0564	hsa_circ_0000478	chr13:42439871-42442613	Forward	GCTCTATCAGCGCTTTGAACT			X	
			Reverse	AATCCGGATGGTCACAGAAG				
ANKRD44	hsa_circ_0008443	⊠hr2:197943426-197954756	Forward	GTCTGGAATTGGTGTGGAATC			X	
			Reverse	TCAAAGCCCGATGATAACAA				
EP300	hsa_circ_0116639	chr22:41568502-41569788	Forward	CCTGAGTAGGGGCAACAAGA			X	
			Reverse	CAGAACATTGGGCCAGAAAT				
RASA3	hsa_circ_0004790	chr13:114806475-114822949	Forward	CATCCAGAAGGAGGACTTGC			X	
			Reverse	GCTACAGGCCTTGCACTCAT				
CREBBP	hsa_circ_0007637	chr16:3900297-3901010	Forward	TGGCCAAGATTTTGGATCAT			X	
			Reverse	AGCAGCATCTGGAACAAGGT				
MCCC1	hsa_circ_0008550	chr3:182788786-182789145	Forward	ACACACCGAGCACATCCAA			X	
			Reverse	AATTCTCCTGGCGTGTTCCT				
BANP	hsa_circ_0040823	chr16:88061088-88071617	Forward	GGACGGTCAGGTCACCTTT			X	
			Reverse	GCAGTAGGAGGACGAGTTGG				
MTRNR2L8	ENST00000536684	chr11: 10529434-10530723	Quantitech	Hs_MTRNR2L8_1_SG			X	
DUSP3	ENST00000590342	chr17: 41845754-41856303	Quantitech	Hs_DUSP3_1_SG			X	
IRF5	ENST00000249375	chr7: 128578271-128590084	Quantitech	Hs_IRF5_1_SG			X	
TENT4B	ENST00000561678	chr16: 50187705-50264314	Quantitech	Hs_PAPD5_1_SG			X	
CCHCR1	ENST00000383527	chr6:31102653-31118444	Quantitech	Hs_CCHCR1_1_SG			X	
VPS28	ENST00000526054	chr8: 145649000-145652366	Quantitech	Hs_VPS28_1_SG			X	
SYMPK	ENST00000599814	chr19: 46322213-46358968	Quantitech	Hs_SYMPK_1_SG			X	
TNFAIP2	ENST00000333007	chr14: 103592664-103603776	Quantitech	Hs_TNFAIP2_1_SG			X	
PPARA	ENST00000262735	chr22: 46546499-46639653	Quantitech	Hs_PPARA_1_SG			X	

Gene	Transcript	Position	Primer	Sequence	Subchapter 1.1	Subchapter 1.2	Subchapter 1.3	Subchapter 2.2
SETD5	ENST00000402198	chr3: 9439403-9519838	Quantitech	Hs_FLJ10707_1_SG			X	
B2M	NM_004048	chr15: 45007698-45010343	Quantitech	Hs_B2M_1_SG			X	
RELL1	Primers for cloning_Circularized exons		Forward	CGAATGCTGATGTCTTAAAGG				X
			Reverse	CTGCTACTCTGTGCCACTGC				
	Primers for cloning_Whole insert		Forward_XhoI	GATCCTCGAGAGCTTTTCTGATGTGATTACT CATCTTATTCATTCTATTTTAATTCTTTAATT TTATTTTTATTCCTTAGCGAATGCTGATGTC TTAAAGG				X
			Reverse_NheI	GCAGTGGCACAGAGTAGCAGGTGAGCCGTG GTTTTGGTGACATTGGGGGCAGAGTGGTGC AGGGTGAGGAGAAGGTACTTGGAGCCTCCC GCTAGCGATC				
GADD34	NM_014330	chr19: 49375649-49379314	Forward	GACCTGTGATCGCTTCTGG				X
			Reverse	CAGGCCAGCTTCTACCAGAG				
IFNB1	NM_002176	chr9: 21077104-21077943	Forward	ACTGCCTCAAGGACAGGATG				X
			Reverse	AGCCAGGAGGTTCTCAACAA				
ATF4	NM_001675	chr22: 39916569-39918688	Forward	CTGACCACGTTGGATGACAC				X
			Reverse	TCTGGAGTGGAGGACAGGAC				
POLR2A	NM_000937	chr17: 7387685-7417933	Forward	GCTTTTTCCTCCCAACTCG				X
			Reverse	TCAGGACTCCGAACTGGACT				
	hsa_circ_0000741	chr17:7402357-7402810	Forward	CTACCTTTCGCTTGAATCTTAG				X
			Reverse	GGTTGAAGATAACAATGTCCCC				
CAMSAP1	NM_015447	chr9: 138702336-138799074	Forward	CACGGAGCTCCCTTACTTGA				X
			Reverse	AGCCAAGATTGAGCCTCAGT				
	hsa_circ_0001900	chr9:138773478-138774924	Forward	CAACGTTCAGTGCCTCGAAA				X
			Reverse	TTTCAGGATGAGGCTGCAGA				
HIPK3	NM_005734	chr11: 33279133-33378569	Forward	TTAGCCATGGCACAAGAAAA				X
			Reverse	GCAGGAGCAAAACTGGACTT				
	hsa_circ_0000284	chr11:33307958-33309057	Forward	TATGTTGGTGGATCCTGTTCGGCA				X
			Reverse	TGGTGGGTAGACCAAGACTTGTGA				
BMPR2	NM_001204	chr2: 203241659-203432474	Forward	CAGGGCAGTTCATTCCAAAT				X
			Reverse	GGGTGCTGACAGGAGGATAA				
	hsa_circ_0003218	chr2:203329531-203332412	Forward	ACCACTCCTCCCTCAATTCA				X
			Reverse	GCCGTTCTTGATTCTGCGAA				
HIPK3	Northern blot probe_Exon2		Probe	ACTCTTGCGCTTCAATCCACATCGCTGGGGG CCTTCTAGGAAATGCAATCTGTTTCGC				X
RELL1	Northern blot probe_Exon5		Probe	TTTATAAAGTGCCACCGCTTGTGCCTACACC GATGACACACATCCCTCTCGACAACACC				X