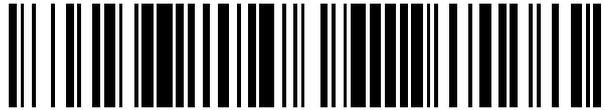


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 688 737**

21 Número de solicitud: 201730655

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

SOLICITUD DE PATENTE

A1

22 Fecha de presentación:

04.05.2017

43 Fecha de publicación de la solicitud:

06.11.2018

71 Solicitantes:

**UNIVERSIDAD DEL PAÍS VASCO / EUSKAL
HERRIKO UNIBERTSITATEA (60.0%)**

Barrio Sarriena, s/n

48940 Leioa (Bizkaia) ES y

ADMINISTRACIÓN GENERAL DE LA

COMUNIDAD AUTÓNOMA DE EUSKADI (40.0%)

72 Inventor/es:

VANDENBROECK, Koen ;

ALLOZA MORAL, Iraide;

FREIJO GUERRERO, María Del Mar y

VEGA MANRIQUE, Reyes

74 Agente/Representante:

ARIAS SANZ, Juan

54 Título: **MÉTODO PARA DIAGNOSTICAR PLACA ATEROESCLERÓTICA INESTABLE**

57 Resumen:

Método para diagnosticar placa aterosclerótica inestable.

La presente invención se relaciona con un método de diagnóstico de placa aterosclerótica inestable, con un método para determinar la probabilidad de un sujeto de sufrir una enfermedad cerebrovascular y con método para seleccionar un paciente susceptible de ser tratado con una terapia para la eliminación o estabilización de la placa aterosclerótica carotidea basados en determinar el nivel de expresión de al menos un gen seleccionado del grupo de genes mostrados en la Tabla 1. La invención también se relaciona con un kit comprende reactivos adecuados para la determinación de los niveles de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 y, opcionalmente, reactivos para la determinación de los niveles de expresión de uno o más genes de mantenimiento y con los usos de dicho kit.

ES 2 688 737 A1

DESCRIPCIÓN

MÉTODO PARA DIAGNOSTICAR PLACA ATEROESCLERÓTICA INESTABLE

5 CAMPO TÉCNICO

La presente invención se relaciona con un método de diagnóstico de placa aterosclerótica inestable y con un método para determinar la probabilidad de un sujeto de sufrir una enfermedad cerebrovascular.

10

ANTECEDENTES

La enfermedad cerebrovascular es una de las principales causantes de muerte y de minusvalías en las sociedades desarrolladas. El riesgo de sufrir un accidente cerebrovascular (ACV) se dobla cada década a partir de los 55 años. Si bien puede ocurrir a cualquier edad, tres cuartas parte de los ACV se dan en personas mayores de 65 años. Teniendo en cuenta el envejecimiento progresivo de la población europea, esta enfermedad será un problema sanitario de primer orden en un futuro cercano. Concretamente, la Organización Mundial de la Salud (OMS) ya considera el ictus, un tipo de ACV, como uno de los problemas sanitarios más graves que padece Europa.

20

La estenosis causada por la formación de placa aterosclerótica en la arteria carótida es uno de los principales factores para desarrollar un enfermedad cerebrovascular. La placa puede llegar a volverse inestable y provocar un accidente cerebrovascular. El proceso por el cual la placa se vuelve inestable no es un proceso aleatorio y existe evidencia de que es un proceso biológico, aunque se desconoce el mecanismo preciso por el cual tiene lugar la desestabilización de la placa aterosclerótica. La inestabilidad de la placa ateromatosa de la carótida es por tanto un factor de riesgo para el desarrollo de un ACV, si bien no siempre va asociada a la presencia de síntomas diagnosticables. Las placas ateroscleróticas pueden definirse como estables o asintomáticas, cuando están constituidas por un núcleo lipídico pequeño y una placa fibrosa estable; o bien sintomáticas, cuando están constituidas por un núcleo lipídico grande y una placa fibrosa fina con tendencia a desprenderse. Una placa inestable se relaciona con un mayor riesgo de desarrollar un ACV.

30

Debido a la complejidad de los procesos que desencadenan un ACV, es conocido en el estado de la técnica la dificultad para encontrar biomarcadores con un verdadero valor

35

diagnóstico (Jickling et al, *Stroke* 46, 915–920, 2015) y no existen test basados en biomarcadores que ayuden a la identificación de placas estables o inestables ni a evaluar el riesgo de desarrollar un ACV (Hermus et al, *Atherosclerosis* 213, 21–29, 2010).

- 5 Por tanto, hay una necesidad de identificar biomarcadores con valor diagnóstico que permitan identificar pacientes con placas ateroscleróticas inestables y por tanto con un mayor riesgo de sufrir una enfermedad cerebrovascular.

BREVE DESCRIPCIÓN DE LA INVENCIÓN

10

En un primer aspecto la invención se relaciona con un método para diagnosticar placas ateroscleróticas inestables en un sujeto que comprende:

(i) determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en una muestra biológica de dicho sujeto, y

15

(ii) comparar el nivel de expresión obtenido en (i) con un valor de referencia, en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 1-33 de la Tabla 1 y su expresión está incrementada respecto al valor de referencia y/o en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 34-93 de la Tabla 1 y su expresión se encuentra reducida respecto al valor de referencia,

20

entonces el sujeto es diagnosticado con placa aterosclerótica inestable.

En un segundo aspecto la invención se relaciona con un método de pronóstico para determinar la probabilidad de un sujeto de sufrir una enfermedad cerebrovascular que comprende

25

(i) determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en una muestra biológica de dicho sujeto, y

(ii) comparar el nivel de expresión obtenido en (i) con un valor de referencia,

en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 1-33 de la Tabla 1 y su expresión está incrementada respecto al valor de referencia y/o

30

en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 34-93 de la Tabla 1 y su expresión se encuentra reducida respecto al valor de referencia, entonces el sujeto tiene una alta probabilidad de sufrir una enfermedad cerebrovascular.

En un tercer aspecto, la invención se relaciona con un método para seleccionar un paciente susceptible de ser tratado con una terapia para la eliminación o estabilización de la placa aterosclerótica carotídea, que comprende:

35

(i) determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en una muestra biológica de dicho sujeto, y

(ii) comparar el nivel de expresión obtenido en (i) con un valor de referencia

5 en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 1-33 y su expresión está incrementada respecto al valor de referencia y/o

en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 34-93 y su expresión se encuentra reducida respecto al valor de referencia, entonces el sujeto es susceptible de ser tratado con una terapia para la eliminación o estabilización de la placa aterosclerótica carotídea.

10

En un cuarto aspecto, la invención se relaciona con un kit que comprende reactivos adecuados para la determinación de los niveles de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 y, opcionalmente, reactivos para la determinación de los niveles de expresión de uno o más genes de mantenimiento

15

En un quinto aspecto, la invención se relaciona con el uso del kit de la invención para diagnosticar placas ateroscleróticas inestables en un sujeto, para determinar la probabilidad de un sujeto de sufrir una enfermedad cerebrovascular o para seleccionar un paciente susceptible de ser tratado con una terapia para la eliminación o estabilización de la placa aterosclerótica carotídea.

20

BREVE DESCRIPCIÓN DE LAS FIGURAS

Figura 1. Análisis de control de calidad. (A) Cuantificación de células positivas para MYH11 versus el número total de células. Los datos son las medias \pm SD. *p <0.05. (B) Detección por Western Blot de MYH11 y PECAM1 con actina como control en cultivos de CML en células Jurkat y U937.

25

Figura 2. Estudio comparativo entre genes subexpresados y sobreexpresados usando el método ClusterProfiler. Análisis comparativo basado en KEGG.

30

Figura 3. Análisis comparativo basado en Gene Ontology. GeneRatio corresponde al número de genes de una categoría específica. El término enriquecimiento está representado por puntos grises (gris claro indica alto enriquecimiento y gris oscuro indica bajo enriquecimiento).

35

Figura 4. Análisis de agrupación de redes mediante MCODE. Ilustración de la agrupación identificada por el algoritmo de agrupación de redes MCODE que muestra la interacción entre señalización de ALK1, señalización de BMP y señalización de TGF-beta, SMAD9, BMP2, ID1 e ID4. Los círculos son genes y los romboides son dominios de proteínas. Las líneas gris claro representan interacciones genéticas. Los círculos gris claro indican genes regulados por disminución en S frente a A y los círculos gris oscuro indican genes regulados por incremento en S frente a A. Las líneas gris oscuro representan rutas. Las líneas grises indican asociaciones de rutas de genes.

10 DESCRIPCIÓN DETALLADA DE LA INVENCION

Los autores de la presente invención han identificado 93 genes cuya expresión se ve modificada al realizar una comparación entre muestras obtenidas de pacientes diagnosticados con placas carotídeas sintomáticas con respecto a pacientes con placas asintomáticas (Ejemplo 3). Las alteraciones en el patrón de expresión de al menos uno de dichos biomarcadores con respecto a un valor de referencia se utilizan para diagnosticar, a partir de una muestra biológica de un sujeto, la estabilidad o inestabilidad de una placa aterosclerótica.

20 Método para diagnosticar placas ateroscleróticas inestables

En un primer aspecto, la invención se relaciona con un método para diagnosticar placas ateroscleróticas inestables en un sujeto que comprende:

- (i) determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en una muestra biológica de dicho sujeto, y
- (ii) comparar el nivel de expresión obtenido en (i) con un valor de referencia en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 1-33 de la Tabla 1 y su expresión está incrementada respecto al valor de referencia y/o en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 34-93 de la Tabla 1 y su expresión se encuentra reducida respecto al valor de referencia, entonces el sujeto es diagnosticado con placa aterosclerótica inestable.

“Método de diagnóstico” o “diagnosticar”, tal como se usa en el presente documento, se refiere tanto al proceso de intentar determinar y/o identificar una posible enfermedad en un sujeto, es decir el procedimiento de diagnóstico, como a la opinión a la que se llega mediante este proceso, es decir la opinión de diagnóstico. Como tal, también puede

considerarse como un intento de clasificación del estado de un individuo en categorías distintas y separadas que permiten que se tomen decisiones médicas sobre el tratamiento y el pronóstico. Tal como entenderán los expertos en la técnica, el diagnóstico de placa aterosclerótica inestable, aunque se prefiere que sea así, no es necesario que sea correcto para el 100% de los sujetos que van a diagnosticarse o evaluarse. Sin embargo, el término requiere que pueda identificarse a una parte estadísticamente significativa de los sujetos que padecen placa aterosclerótica inestable. Si un sujeto es estadísticamente significativo puede determinarse sin más por el experto en la técnica usando diversas herramientas de evaluación estadística bien conocidas, por ejemplo, determinación de intervalos de confianza, determinación del valor de p, prueba de la t de Student, prueba de Mann-Whitney, etc. Se encuentran detalles en Dowdy y Wearden, *Statistics for Research*, John Wiley & Sons, Nueva York 1983. Intervalos de confianza preferidos son de al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90% o al menos el 95%. Los valores de p son, preferiblemente, de 0,05, 0,01, 0,005 o inferior.

15

El término “ateroesclerosis”, como se usa en el presente documento, se refiere a una patología caracterizada por el depósito y la infiltración de sustancias lipídicas en las paredes de las arterias de tamaño medio y grueso. Las células de la pared arterial interpretan este depósito como una invasión y activan monocitos circulantes del sistema inmunitario, que penetran en la pared arterial, se convierten en macrófagos y comienzan a fagocitar partículas LDL, generando un procedimiento inflamatorio. La inflamación a su vez provoca la multiplicación y migración de las células del músculo liso de la pared, que causa un estrechamiento gradual del diámetro arterial. El engrosamiento específico constituye una placa ateromatosa. Entre las enfermedades que cursan con aterosclerosis caracterizadas por el estrechamiento de las arterias debido a placas ateromatosas, y en consecuencia la obstrucción del flujo sanguíneo o isquemia, dependiendo de la arteria del órgano implicado, se encuentran: cardiopatía isquémica (el máximo representante del mismo es infarto agudo de miocardio, en el corazón), enfermedad cerebrovascular (en forma de apoplejía o trombosis cerebral o hemorragia cerebral, en el sistema nervioso central), claudicación intermitente (en su forma más grave, isquemia arterial de las extremidades inferiores), disfunción eréctil, colitis isquémica (un área de inflamación, irritación e hinchazón, provocado por la interferencia con el flujo sanguíneo en el colon, en las arterias de los intestinos) y aneurisma aórtico.

35 “Placa aterosclerótica”, también conocido en la técnica como “placa ateromatosa”, tal y como se usa en la presente invención se refiere a la acumulación de sustancias lipídicas en

las paredes de las arterias según la definición de “ateroesclerosis” hasta constituir una placa formada por lípidos, macrófagos y células de músculo liso. La placa aterosclerótica puede ser estable o inestable.

5 El término "placa estable", también conocido en la técnica como “placa aterosclerótica estable” o “placa asintomática”, se ha usado para describir las placas ateroscleróticas que no son particularmente susceptibles a la rotura. Se caracterizan en general por tener una capa fibrosa gruesa sobre un núcleo de lípidos pequeño.

10 El término "placa aterosclerótica inestable", también conocido en la técnica como “placa inestable” o “placa sintomática”, se ha usado para describir las placas ateroscleróticas que son particularmente susceptibles a la rotura. Se caracterizan en general como las que tienen una capa fibrosa fina inflamada sobre un núcleo de lípidos muy grandes. Está constituida por una mezcla de glóbulos blancos (principalmente macrófagos), células de músculo liso
15 (CML) y lípidos (incluidos colesterol) en la pared de una arteria. Son particularmente inestables y tienen tendencia a producir problemas repentinos tales como un ataque al corazón o apoplejía. Características que definen una placa inestable incluyen: actividad inflamatoria, capa fibrosa fina, núcleo lipídico grande, ulceración visible, intraplaca, hemorragia y la rotura.

20 En una realización particular de los métodos de la invención la placa aterosclerótica inestable es carotídea. El riego sanguíneo a la arteria carótida comienza en el arco de la aorta, la arteria carótida se divide en la arteria carótida interna (que suministra sangre al cerebro) y la arteria carótida externa. La placa aterosclerótica normalmente se acumula en
25 esa división de la carótida, lo que provoca el estrechamiento de estenosis.

El término "sujeto" o "individuo" o "paciente", quiere decir cualquier sujeto, en particular un sujeto mamífero, para el que el diagnóstico, pronóstico, o tratamiento se desea. Sujetos mamíferos incluyen seres humanos, animales domésticos, animales de granja, y de
30 zoológicos, de deportes, o animales domésticos tal como perros, gatos, cobayas, conejos, ratas, ratones, caballos, ganado vacuno, vacas, y así sucesivamente. En un modo de realización preferente de la invención, el sujeto es un mamífero. En un modo de realización más preferente de la invención, el sujeto es un ser humano. En otra realización particular el sujeto es un sujeto que potencialmente padece una enfermedad asociada a una alteración
35 en las placas ateroscleróticas.

En un modo de realización particular, el sujeto cuyo diagnóstico se va a determinar de acuerdo con el primer procedimiento de la invención ya ha padecido una complicación cerebrovascular de manera previa a su diagnóstico de placa aterosclerótica inestable. En un modo de realización particular, dicha complicación cerebrovascular se selecciona del grupo que consiste en accidente isquémico transitorio (AIT), apoplejía, y amaurosis fugaz.

De acuerdo a los métodos de la invención el nivel de expresión se determina en una muestra biológica del sujeto.

10 Por “muestra biológica”, tal y como se usa en la presente invención, se refiere a material biológico aislado de un sujeto. La muestra biológica contiene cualquier material adecuado para detectar el nivel de expresión de un gen y puede ser un material que comprende material genético del sujeto. La muestra biológica puede comprender material celular y/o no celular del sujeto, preferentemente material no celular. En una realización particular, la muestra comprende material genético, por ejemplo, ADN, ADN genómico (ADNg), ADN complementario (ADNc), ARN, ARNm, etc., del sujeto en estudio. En un modo de realización particular, el material genético es ARN. La muestra se puede aislar a partir de cualquier tejido o fluido biológico, tal como, por ejemplo, sangre, saliva, plasma, suero, orina, líquido cefalorraquídeo (CSF), heces, hisopos nasales, bucales o buco-faríngeos, un espécimen, un espécimen obtenido a partir de una biopsia, y una muestra de tejido incrustada en parafina. Los procedimientos para aislar muestras son bien conocidos por los expertos en la técnica.

Antes de analizar la muestra, será frecuentemente deseable realizar una o más operaciones de preparación de dicha muestra para separar la molécula que se va a determinar de otras moléculas que se encuentran en la muestra. En una realización particular, las moléculas son ácidos nucleicos, ADN y/o ARN. Estas operaciones de preparación de las muestras incluyen manipulaciones tales como: concentración, suspensión, extracción de material intracelular (por ejemplo, ácidos nucleicos de muestras de tejido/célula entera y similares), amplificación de ácidos nucleicos, fragmentación, transcripción, etiquetado y/o extensión reacciones. Estos procedimientos son bien conocidos por un experto en la materia. También existen disponibles kits comerciales para purificación de ARNm incluyendo, sin limitación, miRNeasy Mini kit de Qiagen, miARN kits de aislamiento de Life Technologies, mirPremier el kit de aislamiento de microARN de Sigma-Aldrich y High Pure miARN kit de aislamiento de Roche. En una realización particular la integridad del ARN se analizó usando RNA 6000 Nano Chips en el bionalizador Agilent 2100 (Agilent Technologies, Palo Alto, CA, EE.UU.).

En una realización particular, la muestra biológica es una muestra que contiene células de una placa aterosclerótica, más particularmente placa aterosclerótica carotídea. Un experto en la materia conoce técnicas para detección de placas ateroscleróticas, tales como el cálculo del índice tobillo-brazo, la arteriografía por resonancia magnética, la ecografía doble y la hemodinámica de localización o la arteriografía de contraste.

La obtención de la placa de ateroma puede realizarse a modo de ejemplo mediante endarterectomía carotídea. Una vez obtenida la placa, las células de una placa aterosclerótica pueden aislarse empleando diversas técnicas conocidas tales como explantes celulares, digestión enzimática combinada con cultivo en medios específicos para cada tipo celular y/o mediante técnicas de separación. A modo de ejemplo ilustrativo, no limitativo, la separación puede realizarse mediante sedimentación isopícnica, elutriación centrífuga, separación inmunomagnética tales como el uso de esferas de poliestireno con propiedades paramagnéticas (Dynabeads ©), partículas superparamagnéticas (microesferas MACS ©); citometría de flujo, clasificador de células activado por fluorescencia (FACS).

En una realización particular, las células extraídas de la placa aterosclerótica son células endoteliales. En otra realización particular las células extraídas de la placa aterosclerótica son macrófagos. En una realización más preferida las células extraídas de la placa aterosclerótica son células de músculo liso (CML).

Las células de músculo liso de una placa aterosclerótica se pueden obtener por cualquier procedimiento conocido por un experto en la materia. En una realización particular se procesan muestras recién obtenidas de tejido de arteria carótida que posteriormente se digieren con enzimas que permitan la separación de los distintos tipos celulares, a modo de ejemplo ilustrativo no limitativo colagenasa tipo I. Con posterioridad se cultivan en placas con medios específicos para el cultivo de las células de interés, concretamente, de músculo liso, a modo de ejemplo ilustrativo no limitativo M231.

En una realización particular, las CML son positivas para el marcador específico MYH11 y/o negativas para el marcador específico PECAM1. En una realización más particular, las CML son positivas para MYH11 y negativas para PECAM1.

“MYH11”, tal y como se usa en la presente invención, se refiere a la proteína Myosina-11 que en humanos tiene la secuencia mostrada en la base de datos Uniprot con número de acceso P35749 (fecha 15/03/2017).

5 “PECAM1” o CD31, se refiere a la molécula de adhesión de células endoteliales de plaquetas. La secuencia en humanos corresponde a la mostrada en la base de datos Uniprot con número de acceso P16284 (fecha 12/04/2017).

10 Son conocidas diversas técnicas para determinar si una célula es positiva o negativa para un marcador específico.

La presencia/ausencia de un marcador en una célula puede determinarse, por ejemplo, por medio de citometría de flujo usando métodos y aparatos convencionales. Por ejemplo, puede emplearse un citómetro de flujo BD LSR II (BD Biosciences Corp., Franklin Lakes, NJ, US) con anticuerpos comercialmente disponibles y siguiendo protocolos conocidos en la técnica. Por tanto, pueden seleccionarse células que emiten una señal para un marcador de superficie celular específico más intensa que el ruido de fondo. La señal de fondo se define como la intensidad de señal producida por un anticuerpo no específico del mismo isotipo que el anticuerpo específico usado para detectar cada marcador de superficie en el análisis de FACS convencional. Para considerar un marcador positivo, la señal específica observada debe ser un 20%, preferiblemente, un 30%, 40%, 50%, 60%, 70%, 80%, 90%, 500%, 1000%, 5000%, 10000% o superior más intensa que la señal de fondo usando métodos y aparatos convencionales (por ejemplo usando un citómetro de flujo FACSCalibur (BD Biosciences Corp., Franklin Lakes, NJ, US) y anticuerpos comercialmente disponibles). En caso contrario, la célula se considera negativa para dicho marcador.

En una primera etapa, el método de diagnóstico de la invención comprende determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en una muestra biológica de dicho sujeto.

30

Num.	Identificador del gen	Símbolo	GRUPO
1	ENSG00000165507	C10orf10	3
2	ENSG00000152049	KCNE4	1
3	ENSG00000117069	ST6GALNAC5	3
4	ENSG00000135424	ITGA7	1,2,5,6
5	ENSG00000128285	MCHR1	6

ES 2 688 737 A1

Num.	Identificador del gen	Símbolo	GRUPO
6	ENSG00000174564	IL20RB	4
7	ENSG00000163815	CLEC3B	1,2
8	ENSG00000075643	MOCOS	7
9	ENSG00000074410	CA12	2
10	ENSG00000174807	CD248	5
11	ENSG00000133454	MYO18B	1
12	ENSG00000179542	SLITRK4	4
13	ENSG00000161570	CCL5	4
14	ENSG00000148948	LRRC4C	4
15	ENSG00000117289	TXNIP	1
16	ENSG00000123689	G0S2	3
17	ENSG00000162706	CADM3	6
18	ENSG00000120088	CRHR1	1,3
19	ENSG00000184557	SOCS3	1,2,4
20	ENSG00000176438	SYNE3	1
21	ENSG00000161958	FGF11	6
22	ENSG00000234745	HLA-B	3,4,6
23	ENSG00000164236	ANKRD33B	7
24	ENSG00000170624	SGCD	1
25	ENSG00000259649	RP11-351M8.1	4
26	ENSG00000172201	ID4	1,2,3,4
27	ENSG00000119938	PPP1R3C	6
28	ENSG00000114268	PFKFB4	4
29	ENSG00000110076	NRXN2	6
30	ENSG00000136378	ADAMTS7	1,2,4
31	ENSG00000152256	PDK1	6
32	ENSG00000112837	TBX18	1,4,6
33	ENSG00000162733	DDR2	1,2,4,5
34	ENSG00000179348	GATA2	1,2,4,5
35	ENSG00000120693	SMAD9	1,4
36	ENSG00000164736	SOX17	1,3,4,5
37	ENSG00000035664	DAPK2	5,6
38	ENSG00000144730	IL17RD	4
39	ENSG00000133216	EPHB2	1,4,6
40	ENSG00000166960	CCDC178	7
41	ENSG00000120279	MYCT1	4
42	ENSG00000121039	RDH10	1
43	ENSG00000183578	TNFAIP8L3	2
44	ENSG00000145681	HAPLN1	3
45	ENSG00000158125	XDH	1,3,4,6

ES 2 688 737 A1

Num.	Identificador del gen	Símbolo	GRUPO
46	ENSG00000139174	PRICKLE1	1,4
47	ENSG00000182368	FAM27A	4
48	ENSG00000118137	APOA1	3,4,5,6
49	ENSG00000169684	CHRNA5	3
50	ENSG00000167644	C19orf33	4
51	ENSG00000116106	EPHA4	1,4,5,6
52	ENSG00000125968	ID1	1,2,3,4,5,6
53	ENSG00000123572	NRK	1
54	ENSG00000121361	KCNJ8	1,3,6
55	ENSG00000091129	NRCAM	1,3,4,5
56	ENSG00000175746	C15orf54	7
57	ENSG00000172602	RND1	4,6
58	ENSG00000095752	IL11	2,6
59	ENSG00000175866	BAIAP2	2,4,6
60	ENSG00000127329	PTPRB	1
61	ENSG00000144366	GULP1	4
62	ENSG00000143140	GJA5	1
63	ENSG00000176399	DMRTA1	6
64	ENSG00000168497	SDPR	1
65	ENSG0000010319	SEMA3G	1,4,5,6
66	ENSG00000254656	RTL1	7
67	ENSG00000178445	GLDC	6
68	ENSG00000162496	DHRS3	1,4
69	ENSG00000102554	KLF5	1,4
70	ENSG00000172159	FRMD3	3
71	ENSG00000164283	ESM1	1,6
72	ENSG00000235109	ZSCAN31	4
73	ENSG00000132321	IQCA1	7
74	ENSG00000008394	MGST1	2,6
75	ENSG00000164530	PI16	1,4
76	ENSG00000150722	PPP1R1C	2
77	ENSG00000116741	RGS2	3,4
78	ENSG00000180660	MAB21L1	7
79	ENSG00000155011	DKK2	4
80	ENSG00000126950	TMEM35	7
81	ENSG00000179104	TMTC2	2
82	ENSG00000118777	ABCG2	3,6
83	ENSG00000125848	FLRT3	1,4,5,6
84	ENSG00000108375	RNF43	2,3,4,6
85	ENSG00000125845	BMP2	1,2,3,4,5,6

Num.	Identificador del gen	Símbolo	GRUPO
86	ENSG00000138759	FRAS1	1
87	ENSG00000108556	CHRNE	6
88	ENSG00000163273	NPPC	1,2,4,6
89	ENSG00000162595	DIRAS3	4
90	ENSG00000109511	ANXA10	4
91	ENSG00000154645	CHODL	2,4
92	ENSG00000138772	ANXA3	1,3,5
93	ENSG00000087494	PTHLH	1,2,3,4,6

5 Tabla 1. Listado de genes que muestran una expresión génica diferencial según método descrito, con la categoría correspondiente indicada en la descripción. El identificador del gen hace referencia al nombre en la base de datos Ensembl.

10 El término “nivel de expresión” de un gen, tal como aquí se usa, se refiere a la cantidad medible de producto del gen en una muestra del sujeto, en el que el producto del gen puede ser un producto de la transcripción o un producto de la traducción de dicho gen. En consecuencia, el nivel de expresión puede corresponder a un ácido nucleico del gen (tal como ARNm o ADNc) o un polipéptido codificado por dicho gen. El nivel de expresión se deriva de la muestra de un sujeto y/o de una muestra o muestras de referencia, y puede ser detectado *de novo* o corresponder a una determinación anterior.

15 Es posible determinar los niveles de expresión de los marcadores mediante la determinación de los niveles de expresión de las proteínas codificadas por los genes, ya que si la expresión de los genes aumenta, cabe esperar que se produzca un aumento en la cantidad de la proteína correspondiente, y si la expresión de los genes disminuye, cabe esperar que se produzca una disminución de la cantidad de la proteína correspondiente.

20 Por tanto, en una realización particular, el nivel expresión de un gen puede llevarse a cabo mediante la cuantificación del nivel de proteína expresada por dicho gen.

25 El nivel de una proteína puede ser cuantificado mediante cualquier método convencional que permita detectar y cuantificar dicha proteína en una muestra de un sujeto. A modo ilustrativo, no limitativo, los niveles de dicha proteína pueden cuantificarse, por ejemplo, mediante el empleo de anticuerpos con capacidad de unirse a la proteína de interés (o a fragmentos de la misma que contengan un determinante antigénico) y la posterior

cuantificación de los complejos formados. Los anticuerpos que se emplean en estos ensayos pueden estar marcados o no. Ejemplos ilustrativos, no limitativos, de marcadores que se pueden utilizar incluyen isótopos radiactivos, enzimas, fluoróforos, reactivos quimioluminiscentes, sustratos enzimáticos o cofactores, inhibidores enzimáticos, partículas, colorantes, etc. Existe una amplia variedad de ensayos conocidos que se pueden utilizar en la presente invención, que utilizan anticuerpos no marcados (anticuerpo primario) y anticuerpos marcados (anticuerpo secundario); entre estas técnicas se incluyen, a modo ilustrativo, no limitativo, Western-blot o transferencia Western, ELISA (ensayo inmunoabsorbente ligado a enzima), RIA (radioinmunoensayo), EIA competitivo (inmunoensayo enzimático competitivo), DAS-ELISA (ELISA sándwich con doble anticuerpo), técnicas inmunocitoquímicas e inmunohistoquímicas, técnicas basadas en el empleo de biochips o microarrays de proteínas que incluyan anticuerpos específicos o ensayos basados en precipitación coloidal en formatos tales como dipsticks. Otras maneras para detectar y cuantificar una proteína, incluyen técnicas de cromatografía de afinidad, ensayos de unión a ligando, etc. Cuando se usa un método inmunológico, se puede usar cualquier anticuerpo o reactivo que se sabe se une a la proteína de interés con alta afinidad para detectar la cantidad de la misma. Sin embargo, se prefiere el uso de un anticuerpo, por ejemplo, sueros policlonales, sobrenadantes de hibridomas o anticuerpos monoclonales, fragmentos de anticuerpos, Fv, Fab, Fab' y F(ab')₂, scFv, diacuerpos, triacuerpos, tetracuerpos y anticuerpos humanizados.

En una realización particular, la cuantificación de los niveles de proteína se realiza mediante western blot, ELISA o mediante un array de proteínas.

En otra realización particular, la cuantificación del nivel de expresión de un gen se lleva a cabo a partir del ARNm que codifica dicho gen o, alternativamente, a partir del ADN complementario (ADNc) de dicho ARNm. Por tanto, en una realización particular, la cuantificación del nivel de expresión de un gen comprende la cuantificación del ARNm que contiene la secuencia que codifica dicho gen, o un fragmento de dicho ARNm, o la cuantificación del ADNc de dicho ARNm, o un fragmento de dicho ADNc, o sus mezclas.

Para ello, la muestra biológica puede ser tratada para disgregar de forma física o mecánica la estructura del tejido o la célula, liberando los componentes intracelulares en una solución acuosa u orgánica para preparar los ácidos nucleicos. Los ácidos nucleicos se extraen mediante procedimientos conocidos para el experto en la materia y disponibles

comercialmente (Sambroock, J., et al., "Molecular cloning: a Laboratory Manual", 3rd ed., Cold Spring Harbor Laboratory Press, N.Y., Vol. 1-3.)

5 Prácticamente cualquier método convencional puede ser utilizado dentro del marco de la presente invención para detectar y cuantificar el nivel de un gen. A modo ilustrativo, no limitativo, el nivel de ARNm codificado por dicho gen puede ser cuantificado mediante el empleo de métodos convencionales, por ejemplo, métodos que comprenden la amplificación del ARNm y la cuantificación del producto de la amplificación de dicho ARNm, tales como electroforesis y tinción, o alternativamente, mediante Southern blot y empleo de sondas apropiadas, northern blot y el empleo de sondas específicas del ARNm que codifica un gen o de su ADNc correspondiente, mapeo con la nucleasa S1, RT-LCR, hibridación, microarrays, etc., preferentemente, mediante PCR cuantitativa a tiempo real usando un marcador apropiado. Análogamente, los niveles del ADNc correspondiente a dicho ARNm que codifica el gen también pueden ser cuantificados mediante el empleo de técnicas convencionales; en este caso, el método de la invención incluye una etapa de síntesis del correspondiente ADNc mediante transcripción inversa (RT) del ARNm correspondiente seguida de amplificación y cuantificación del producto de la amplificación de dicho ADNc. Métodos convencionales de cuantificar los niveles de expresión pueden encontrarse, por ejemplo, en Sambrook y cols., 2001. (citado *ad supra*).

20 Para normalizar los valores de expresión de ARNm entre las diferentes muestras, es posible comparar los niveles de expresión del ARNm de interés en las muestras de prueba con la expresión de un ARN control. Un "ARN control" como se usa en el presente documento, se refiere a un ARN cuyos niveles de expresión no cambian o cambian solo en cantidades limitadas. Preferiblemente, el ARN control es ARNm derivado de genes de mantenimiento y que codifica proteínas que se expresan constitutivamente y llevan a cabo funciones celulares esenciales. Los genes de mantenimiento preferidos para su uso en la presente invención incluyen la proteína ribosómica de 18S, β -2-microglobulina, ubiquitina, ciclofilina, GAPDH, PSMB4, tubulina y β -actina.

30 La determinación de los niveles de ARN, en particular los niveles de ARNm, se puede llevar a cabo por cualquier procedimiento conocido en el estado de la técnica tal como qPCR, northern blot, dot blot de ARN, TaqMan, procedimientos basados en análisis en serie de la expresión génica (SAGE) incluyendo variantes tales como LongSAGE y SuperSAGE, micromatrices. La determinación de los niveles de ARNm también se puede llevar a cabo por hibridación *in situ* con fluorescencia (FISH). La detección se puede llevar a cabo en

muestras individuales o en micromatrices de tejido. En un modo de realización particular, los niveles de expresión de ARN en una muestra de un sujeto que se va a diagnosticar con placa aterosclerótica inestable se determinan por PCR en tiempo real (RT-PCR).

5 La PCR en tiempo real (también conocida como PCR cuantitativa, PCR cuantitativa en tiempo real o RTQ-PCR) es un método de cuantificación y amplificación simultáneas de ADN (Expert Rev. Mol, Diagn. 2005 (2): 209-19). El ADN se amplifica específicamente mediante la reacción en cadena de la polimerasa. Después de cada ronda de amplificación, el ADN se cuantifica. Los métodos comunes de cuantificación incluyen el uso de tintes
10 fluorescentes que se intercalan con ADN de doble hebra y oligonucleótidos de ADN modificados (llamados sondas) que son fluorescentes cuando se hibridan con un ADN

En otra realización particular, la determinación de los niveles de expresión se determinan mediante secuenciación de ARN o ARNseq.

15

En una realización particular, la cuantificación de los niveles de expresión de un gen se realiza mediante una reacción en cadena de la polimerasa (PCR) cuantitativa o un array de ADN o ARN o mediante técnicas de hibridación de nucleótidos a partir del ARNm o de su ADNc correspondiente.

20

ARN-seq, como se usa en la presente memoria, se refiere a una tecnología de secuenciación profunda que proporciona un nivel de medición más preciso de las transcripciones. Esta tecnología implica hacer una colección de fragmentos de ADNc que están flanqueados por secuencias constantes específicas (conocidas como adaptadores) que son necesarias para la secuenciación. Los adaptadores contienen varios elementos
25 funcionales diferentes que son necesarios para la secuenciación y pueden contener uno o más elementos opcionales. En general, una población de ARN (total o fraccionado, tal como poli (A) +) se convierte en una biblioteca de fragmentos de ADNc con adaptadores unidos a uno o ambos extremos. Cada molécula, con o sin amplificación, es entonces secuenciada de una manera de alto rendimiento para obtener secuencias cortas de un extremo (secuenciación de extremo único) o ambos extremos (secuenciación de extremo de par). Las lecturas son típicamente de 30-400 pb, dependiendo de la tecnología de secuenciación de ADN utilizada. Cualquier tecnología de secuenciación de alto rendimiento puede usarse para RNA-Seq, por ejemplo los sistemas HiSeq2500 (Illumina) y Roche 454 Life Science.

30 Después de la secuenciación, las lecturas resultantes se alinean con un genoma de

35

referencia o transcritos de referencia, o se ensamblan de novo sin la secuencia genómica para producir un mapa de transcripción de escala genómica que consiste en la estructura transcripcional y/o nivel de expresión de cada gen. En una realización preferida, la cuantificación se realiza empleando el instrumento Hiseq2500 a partir de 2 µg de ARN de una biblioteca de ARN total.

Una vez obtenidos los valores del nivel de expresión se puede determinar un valor de tasa de cambio (FC) que se puede obtener mediante el método de DeSeq2 y el valor de p ajustado (Pajust), que se puede calcular usando el método de la tasa de falso descubrimiento (FDR).

En una realización particular, el método comprende determinar el nivel de expresión de al menos 1 gen, al menos 2 genes, al menos 3 genes, al menos 4 genes, al menos 5 genes, al menos 6 genes, al menos 7 genes, al menos 8 genes, al menos 9 genes, al menos 10 genes, al menos 11 genes, al menos 12 genes, al menos 13 genes, al menos 14 genes, al menos 15 genes, al menos 16 genes, al menos 17 genes, al menos 18 genes, al menos 19 genes, al menos 20 genes, al menos 21 genes, al menos 22 genes, al menos 23 genes, al menos 24 genes, al menos 25 genes, al menos 26 genes, al menos 27 genes, al menos 28 genes, al menos 29 genes, al menos 30 genes, al menos 31 genes, al menos 32 genes, al menos 33 genes, al menos 34 genes, al menos 35 genes, al menos 36 genes, al menos 37 genes, al menos 38 genes, al menos 39 genes, al menos 40 genes, al menos 41 genes, al menos 42 genes, al menos 43 genes, al menos 44 genes, al menos 45 genes, al menos 46 genes, al menos 47 genes, al menos 48 genes, al menos 49 genes, al menos 50 genes, al menos 51 genes, al menos 52 genes, al menos 53 genes, al menos 54 genes, al menos 55 genes, al menos 56 genes, al menos 57 genes, al menos 58 genes, al menos 59 genes, al menos 60 genes, al menos 61 genes, al menos 62 genes, al menos 63 genes, al menos 64 genes, al menos 65 genes, al menos 66 genes, al menos 67 genes, al menos 68 genes, al menos 69 genes, al menos 70 genes, al menos 71 genes, al menos 72 genes, al menos 73 genes, al menos 74 genes, al menos 75 genes, al menos 76 genes, al menos 77 genes, al menos 78 genes, al menos 79 genes, al menos 80 genes, al menos 81 genes, al menos 82 genes, al menos 83 genes, al menos 84 genes, al menos 85 genes, al menos 86 genes, al menos 87 genes, al menos 88 genes, al menos 89 genes, al menos 90 genes, al menos 91 genes, al menos 92 genes o de los 93 genes de los contenidos en la Tabla 1.

La presente invención se refiere a cualquier combinación de genes 1-93 de la Tabla 1.

En una realización particular el método para diagnosticar placas ateroscleróticas inestables comprende determinar el nivel de expresión de todos los genes de la Tabla 1.

5 En otra realización particular, el método para diagnosticar placas ateroscleróticas inestables comprende determinar el nivel de expresión de al menos 1 gen del grupo 1, al menos 1 gen del grupo 2, al menos 1 gen del grupo 3, al menos 1 gen del grupo 4, al menos 1 gen del grupo 5 y/o al menos 1 gen del grupo 6.

10 Dentro del grupo 1 se encuentran aquellos genes implicados en el desarrollo del sistema circulatorio, del sistema vascular o del desarrollo de tejidos. Concretamente se encuentran los genes ADAMTS7, ANXA3, BMP2, CLEC3B, CRHR1, DDR2, DHRS3, EPHA4, EPHB2, ESM1, FLRT3, FRAS1 GATA2, ID1, ID4, ITGA7, KCNJ8, KLF5, MYO18B, NPPC, NRCAM, NRK, PI16, PRICKLE1, PTHLH, PTPRB, RDH10, SEMA3G, SGCD, SMAD9, SOCS3, SOX17, TBX18, XDH, GJA5, TXNIP, SYNE3, KCNE4 y SDPR.

15 Dentro del grupo 2 se encuentran aquellos genes implicados en la mineralización del hueso y regulación de la diferenciación de osteoclastos, osteoblastos y condrocitos. Concretamente se encuentran los genes ADAMTS7, BAIAP2, BMP2, CA12, CLEC3B, CHODL, DDR2, GATA2, ID1, ID4, IL11, ITGA7, MGST1, NPPC, PPP1R1C, PTHLH, RNF43, 20 SOCS3, TMTC2 y TNFAIP8L3.

Dentro del grupo 3 se encuentran aquellos genes implicados en procesos patológicos. Concretamente se encuentran los genes ABCG2, ANXA3, APOA1, BMP2, C10orf10, CHRNA5, CRHR1, G0S2, HAPLN1, HLA-B, ID1, ID4, KCNJ8, NRCAM, PTHLH, RGS2, 25 RNF43, SOX17, ST6GALNAC5, XDH y FRMD3.

Dentro del grupo 4 se encuentran aquellos genes implicados en diferenciación celular, transducción de señales, diferenciación de adipocitos y supervivencia. Concretamente se encuentran los genes ADAMTS7, APOA1, BAIAP2, BMP2, CHODL, DDR2, DHRS3, DKK2, 30 EPHA4, EPHB2, FLRT3, GATA2, HLA-B, ID1, ID4, IL17RD, KLF5, LRRC4C, NPPC, NRCAM, PI16, PRICKLE1, PTHLH, RGS2, RND1, RNF43, SEMA3G, SLITRK4, SMAD9, SOCS3, SOX17, TBX18, XDH, FAM27A, RP11-351M8.1, ZSCAN31, ANXA10, C19orf33, CCL5, DIRAS3, GULP1, IL20RB, MYCT1 y PFKFB4.

Dentro del grupo 5 se encuentran aquellos genes implicados en migración celular. Concretamente se encuentran los genes BMP2, NRCAM, CD248, ANXA3, DAPK2, DDR2, GATA2, SOX17, ID1, APOA1, FLRT3, SEMA3G, ITGA7 y EPHA4.

5 Dentro del grupo 6 se encuentran aquellos genes implicados en unión de proteínas. Concretamente se encuentran los genes ABCG2, APOA1, BAIAP2, BMP2, CADM3, DAPK2, DMRTA1, EPHA4, EPHB2, ESM1, FGF11, FLRT3, GLDC, HLA-B, ID1, IL11, ITGA7, KCNJ8, MCHR1, MGST1, NPPC, NRXN2, PDK1, PTHLH, RND1, RNF43, SEMA3G, TBX18, XDH, PPP1R3C y CHRNE.

10

En otra realización preferida, el método para diagnosticar placas ateroscleróticas inestables según la invención comprende determinar el nivel de expresión de todos los genes del grupo 1.

15 En otra realización particular, el método para diagnosticar placas ateroscleróticas inestables según la invención comprende determinar el nivel de expresión de todos los genes del grupo 2.

20 En otra realización particular, el método para diagnosticar placas ateroscleróticas inestables según la invención comprende determinar el nivel de expresión de todos los genes del grupo 3.

25 En otra realización particular, el método para diagnosticar placas ateroscleróticas inestables según la invención comprende determinar el nivel de expresión de todos los genes del grupo 4.

30 En otra realización particular, el método para diagnosticar placas ateroscleróticas inestables según la invención comprende determinar el nivel de expresión de todos los genes del grupo 5.

En otra realización particular, el método para diagnosticar placas ateroscleróticas inestables según la invención comprende determinar el nivel de expresión de todos los genes del grupo 6.

35 En otra realización particular, el método para diagnosticar placas ateroscleróticas inestables según la invención comprende determinar el nivel de expresión de todos los genes del grupo

1, de todos los genes del grupo 2, de todos los genes del grupo 3, de todos los genes del grupo 4, de todos los genes del grupo 5 y de todos los genes del grupo 6.

5 En una segunda etapa el método para diagnosticar placas ateroscleróticas de la invención comprende comparar el nivel de expresión obtenido en la etapa (i) con un valor de referencia correspondiente.

10 El “valor de referencia” o “nivel de referencia” de acuerdo con cualquiera de los métodos de la invención puede ser un valor absoluto; un valor relativo; un valor que tenga límite superior y/o inferior; un intervalo de valores; un valor promedio; un valor de la mediana, un valor medio, o un valor en comparación con un control particular o un valor de partida. Un valor de referencia se puede basar en un valor de muestra individual, como por ejemplo, un valor obtenido de una muestra del sujeto que se está sometiendo a prueba. El valor de referencia se puede basar en un gran número de muestras, por ejemplo a partir de una población de
15 sujetos de la misma edad, o se puede basar en un conjunto de muestras incluyendo o excluyendo la muestra que se va a someter a prueba. Se tienen en cuenta varias consideraciones cuando se determina el valor de referencia de un gen. Entre dichas consideraciones se encuentran la edad, peso, sexo, afección física general del paciente y similares. Por ejemplo, cantidades iguales de un grupo de al menos 2, al menos 10, al
20 menos 100 a preferentemente de más de 1000 sujetos, preferentemente clasificadas de acuerdo con las consideraciones anteriores, por ejemplo de acuerdo con varias categorías de edad, se toman como el grupo de referencia.

25 En una realización particular, el valor de referencia de un gen corresponde al nivel e expresión de dicho gen en un sujeto sano. En otra realización particular, el valor de referencia de un gen corresponde al nivel de expresión de dicho gen en un sujeto que no padece placas ateroscleróticas inestables. En otra realización particular, el valor de referencia de un gen corresponde al nivel de expresión de dicho gen en una muestra de células de una placa aterosclerótica de un sujeto que tiene placas ateroscleróticas estables.

30 Una vez que el valor de referencia es establecido para el gen de interés correspondiente, el nivel de expresión del gen de interés determinado en una muestra biológica del sujeto se compara con el valor de referencia, y por tanto se puede asignar como “incremento” o “disminución” en el valor de expresión de dicho gen y poder determinar si presenta placas
35 ateroscleróticas inestables.

El término “incremento en el nivel de expresión”, como se usa en el presente documento en relación con el nivel de expresión de un gen, se refiere a que el nivel de expresión es al menos 5 %, al menos un 10 %, al menos 20 %, al menos un 30 %, al menos un 40 %, al menos un 50 %, al menos 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 % o al menos 100 % superior en comparación con el valor de referencia correspondiente.

El término “disminución en el nivel de expresión”, como se usa en el presente documento en relación con el nivel de expresión de un gen, se refiere a que el nivel de expresión disminuye al menos 5 %, al menos un 10 %, al menos 20 %, al menos un 30 %, al menos un 40 %, al menos un 50 %, al menos 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 % o al menos 100 % en comparación con el valor de referencia correspondiente.

De acuerdo al método de diagnóstico de placas ateroscleróticas en un sujeto, si el al menos un gen determinado en la etapa (i) es uno de los genes 1-33 de la Tabla 1 y su expresión se encuentra incrementada respecto al valor de referencia y/o en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 34-93 de la Tabla 1 y su expresión se encuentra reducida respecto al valor de referencia, entonces el sujeto es diagnosticado con placa ateroesclerótica inestable.

Método para determinar la probabilidad de un sujeto de sufrir una enfermedad cerebrovascular

La presencia de placa ateroesclerótica inestable se asocia con un aumento del riesgo de sufrir una enfermedad cerebrovascular (Insull, W. *Am. J. Med.* 122 (1 suppl), S3–S14 (2009)). En consecuencia, los cambios en los niveles de expresión de aquellos genes que se pueden usar como marcadores para determinar la presencia de placa ateroesclerótica inestable permiten también clasificar a sujetos en base al su riesgo de sufrir una enfermedad cerebrovascular.

Por tanto, en un segundo aspecto la invención se refiere a un método de pronóstico para determinar la probabilidad de un sujeto de sufrir una enfermedad cerebrovascular que comprende

(i) determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en una muestra biológica de dicho sujeto, y

(ii) comparar el nivel de expresión obtenido en (i) con un valor de referencia,

en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 1-33 de la Tabla 1 y su expresión está incrementada respecto al valor de referencia y/o

en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 34-93 de la Tabla 1 y su expresión se encuentra reducida respecto al valor de referencia, entonces el

5 sujeto tiene una alta probabilidad de sufrir una enfermedad cerebrovascular.

El término "determinación de probabilidad", o "predecir" como se usa en el presente documento en el contexto de predecir procedimientos de la invención, se refiere a la predicción de un sujeto de padecer una enfermedad cerebrovascular, o se refiere a la

10 predicción de la progresión de la enfermedad cerebrovascular en un sujeto. Como el experto en la materia entenderá, dicha determinación no suele ser correcta para todos (es decir, 100 %) de los pacientes que vayan a ser identificados. Sin embargo, el término requiere que una parte significativa de los sujetos se puedan identificar. El experto en la técnica puede

15 determinar fácilmente si una parte es estadísticamente significativa usando varias herramientas de evaluación estadística bien conocidas, por ejemplo, la determinación de los intervalos de confianza, la determinación de los valores de p, prueba de la t de Student,

prueba Mann-Whitney, etc. los detalles se encuentran en Dowdy and Wearden, Statistics for Research, John Wiley and Sons, New York 1983. Los intervalos preferentes de confianza son de al menos un 90 %, al menos un 95 %, al menos un 97 %, al menos 98 % o al menos

20 un 99 %. Los valores p son preferentemente 0,1, 0,05, 0,01, 0,005 o 0,0001. Más preferentemente, al menos 60 %, al menos un 70 %, al menos un 80 % o al menos el 90 %

de los sujetos de una población de forma adecuada, se pueden identificar por el procedimiento de la presente invención.

25 El término "enfermedad cerebrovascular", como se usa en el presente documento se refiere a un síndrome clínico provocado por la alteración del riego sanguíneo al cerebro, caracterizado por desarrollar rápidamente signos de alteración de la función cerebral global, que duran más de 24 horas o que dan lugar a la muerte. Una enfermedad cerebrovascular

normalmente se conoce como apoplejía, dado que ésta es la enfermedad cerebrovascular

30 más común. Acontecimientos cerebrovasculares de acuerdo con la invención incluyen accidente isquémico transitorio (AIT), apoplejía, arteriopatía carotídea y amaurosis fugaz.

En una realización preferida, la enfermedad cerebrovascular es apoplejía.

El término "accidente isquémico transitorio" o AIT, como se usa en el presente documento,

35 se refiere a un episodio de disfunción neurológica provocada por la isquemia (pérdida del flujo sanguíneo, bien en el cerebro, médula espinal, o de retina) sin infarto agudo (tejido

muerto). AITs tienen la misma causa subyacente que apoplejías: una rotura de flujo sanguíneo cerebral (CBF), y a menudo se denominan como mini-apoplejías o mega-apoplejías.

5 El término “apoplejía”, también conocido como accidente cerebrovascular (ACV), incidente cerebrovascular (CVI), o ataque cerebral, se usa en el presente documento para hacer referencia a la situación en la que falta flujo sanguíneo al cerebro, dando lugar a la muerte celular. Apoplejía isquémica (también conocido como infarto isquémico) es el resultado de la falta del flujo sanguíneo, como consecuencia de una obstrucción dentro de un vaso
10 sanguíneo que suministra el flujo sanguíneo al cerebro. La apoplejía hemorrágica (también conocida como hemorragia intracerebral) es debida a una hemorragia. Como resultado de la apoplejía, el cerebro no funciona correctamente. Los signos y síntomas de una apoplejía pueden incluir: incapacidad para moverse o sentir en un lado del cuerpo, problemas de comprensión u oratoria, mareo, pérdida de la visión entre otros. Los signos y síntomas a
15 menudo aparecen inmediatamente después de que la apoplejía se ha producido. Si los síntomas duran menos de una o dos horas se conoce como accidente isquémico transitorio (AIT).

El término “arteriopatía carotídea”, también conocido como “estenosis de arteria carótida”, o
20 “estenosis carotídea” se refiere a un estrechamiento o constricción de la superficie interna (luz) de la arteria carótida, habitualmente provocada por la aterosclerosis. Los procedimientos para determinar estenosis de arteria carótida son conocidos por el experto en la técnica e incluyen, sin limitación, dúplex Doppler ultrasonidos, arteriografía, tomografía axial computarizada angiografía (CTA), o angiorrsonancia magnética (MRA).

25 El término “amaurosis fugaz”, como se usa en el presente documento, se refiere a una pérdida mono ocular transitoria e indolora de la visión.

La etapa (i) del método de pronóstico de la invención, comprende determinar el nivel de
30 expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en una muestra biológica de dicho sujeto.

En un modo de realización particular la muestra es una muestra que contiene células de placa aterosclerótica. En una realización más preferida las células extraídas de la placa
35 aterosclerótica son células de músculo liso (CML). En una realización aún más preferida, las células de músculo liso son positivas para MYH11 y/o negativas para PECAM1. Los

métodos para obtener e identificar las células de interés se han descrito anteriormente y son igualmente aplicables a este método.

5 En una realización particular el nivel de expresión de un gen se determina mediante técnicas de secuenciación de ARN.

10 En una realización particular, el método de pronóstico de la invención comprende determinar el nivel de expresión de al menos 1 gen del grupo 1, al menos 1 gen del grupo 2, al menos 1 gen del grupo 3, al menos 1 gen del grupo 4, al menos 1 gen del grupo 5 y/o al menos 1 gen del grupo 6.

En otra realización particular, el método de pronóstico de la invención comprende determinar el nivel de expresión de todos los genes del grupo 1.

En otra realización particular, el método de pronóstico de la invención comprende determinar el nivel de expresión de todos los genes del grupo 2.

15 En otra realización particular, el método de pronóstico de la invención comprende determinar el nivel de expresión de todos los genes del grupo 3.

En otra realización particular, el método de pronóstico de la invención comprende determinar el nivel de expresión de todos los genes del grupo 4.

20 En otra realización particular, el método de pronóstico de la invención comprende determinar el nivel de expresión de todos los genes del grupo 5.

En otra realización particular, el método de pronóstico de la invención comprende determinar el nivel de expresión de todos los genes del grupo 6.

25 Los métodos para determinar los niveles de expresión del gen de interés, en particular los niveles de los genes incluidos en la Tabla 1, se han descrito previamente en el contexto del procedimiento de diagnóstico de la invención y son igualmente aplicables en el presente método incluyendo sus limitaciones.

30 En otra realización particular, el método de pronóstico de la invención comprende determinar el nivel de expresión de todos los genes del grupo 1, de todos los genes del grupo 2, de todos los genes del grupo 3, de todos los genes del grupo 4, de todos los genes del grupo 5 y de todos los genes del grupo 6.

En otra realización particular, el método de pronóstico de la invención comprende determinar el nivel de expresión de todos los genes de la Tabla 1.

5 La etapa (ii) del método de pronóstico de la invención, comprende comparar el nivel de expresión obtenido en (i) con un valor de referencia.

10 En una realización particular, el valor de referencia de un gen corresponde al nivel de expresión de dicho gen en una muestra de células de una placa aterosclerótica de un sujeto que tiene placas ateroscleróticas estables.

15 De acuerdo al método de pronóstico de la invención si el nivel de expresión del gen de interés 1-33 aumenta con respecto al valor de referencia, y/o en el que si el nivel de expresión del gen de interés 34-93 disminuye con respecto al valor de referencia a continuación el sujeto muestra una alta probabilidad de sufrir una enfermedad cerebrovascular.

20 “Alta probabilidad”, tal y como se usa en la presente invención, se refiere a la situación en la que el sujeto muestra al menos 5 %, al menos un 10 %, al menos 20 %, al menos un 30 %, al menos un 40 %, al menos un 50 %, al menos 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos 100 % probabilidades de padecer una enfermedad cerebrovascular con el tiempo.

25 “Baja probabilidad”, tal y como se usa en la presente invención, se refiere a la situación en la que el sujeto muestra al menos 5 %, al menos un 10 %, al menos 20 %, al menos un 30 %, al menos un 40 %, al menos un 50 %, al menos 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos 100 % probabilidades de no padecer una enfermedad cerebrovascular con el tiempo.

30 Los términos y limitaciones descritos anteriormente en relación con el método de diagnóstico de la invención son igualmente aplicables a este aspecto.

Método de tratamiento de un sujeto para la eliminación o estabilización de la placa aterosclerótica carótida

En un tercer aspecto la invención se relaciona con un método para la eliminación o estabilización de la placa aterosclerótica carotídea en un sujeto, que comprende:

- (i) determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en una muestra biológica de dicho sujeto, y
- 5 (ii) comparar el nivel de expresión obtenido en (i) con un valor de referencia en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 1-33 y su expresión está incrementada respecto al valor de referencia y/o
- en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 34-93 y su expresión se encuentra reducida respecto al valor de referencia, entonces el sujeto es
- 10 tratado para la eliminación o estabilización de la placa aterosclerótica carotídea.

Alternativamente, la invención se relaciona con un compuesto adecuado para la eliminación o estabilización de la placa aterosclerótica carotídea en un sujeto para su uso en el tratamiento de la placa aterosclerótica carotídea que comprende

- 15 (i) determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en una muestra biológica de dicho sujeto, y
- (ii) comparar el nivel de expresión obtenido en (i) con un valor de referencia en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 1-33 y su expresión está incrementada respecto al valor de referencia y/o
- 20 en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 34-93 y su expresión se encuentra reducida respecto al valor de referencia, entonces el sujeto es tratado para la eliminación o estabilización de la placa aterosclerótica carotídea.

Por “estabilización de la placa aterosclerótica carotídea”, tal y como se usa en la presente

25 invención se refiere a reducir la posibilidad de que la placa aterosclerótica se rompa y/o desprenda.

La etapa (i) del método para la eliminación de la placa aterosclerótica carotídea, comprende determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes

30 mostrados en la Tabla 1 en una muestra biológica de dicho sujeto.

En un modo de realización particular la muestra es una muestra que contiene células de placa aterosclerótica. En una realización más preferida las células extraídas de la placa aterosclerótica son células de músculo liso (CML). En una realización aún más preferida,

35 las células de músculo liso son positivas para MYH11 y negativas para PECAM1. Los

métodos para obtener e identificar las células de interés se han descrito anteriormente y son igualmente aplicables a este método.

5 En una realización particular el nivel de expresión de un gen se determina mediante técnicas de secuenciación de ARN.

10 En una realización particular, el método para la eliminación de la placa aterosclerótica carótida comprende determinar el nivel de expresión de al menos 1 gen del grupo 1, al menos 1 gen del grupo 2, al menos 1 gen del grupo 3, al menos 1 gen del grupo 4, al menos 1 gen del grupo 5 y/o al menos 1 gen del grupo 6.

En otra realización particular, el método para la eliminación de la placa aterosclerótica carótida comprende determinar el nivel de expresión de todos los genes del grupo 1.

En otra realización particular, el método para la eliminación de la placa aterosclerótica carótida comprende determinar el nivel de expresión de todos los genes del grupo 2.

15 En otra realización particular, el método para la eliminación de la placa aterosclerótica carótida comprende determinar el nivel de expresión de todos los genes del grupo 3.

En otra realización particular, el método para la eliminación de la placa aterosclerótica carótida comprende determinar el nivel de expresión de todos los genes del grupo 4.

20 En otra realización particular, el método para la eliminación de la placa aterosclerótica carótida comprende determinar el nivel de expresión de todos los genes del grupo 5.

En otra realización particular, el método para la eliminación de la placa aterosclerótica carótida comprende determinar el nivel de expresión de todos los genes del grupo 6.

25 Los métodos para determinar los niveles de expresión del gen de interés, en particular los niveles de los genes incluidos en la Tabla 1, se han descrito previamente en el contexto del procedimiento de diagnóstico de la invención y son igualmente aplicables en el presente método incluyendo sus limitaciones.

30 En otra realización particular, el método para la eliminación de la placa aterosclerótica carótida comprende determinar el nivel de expresión de todos los genes del grupo 1, de todos los genes del grupo 2, de todos los genes del grupo 3, de todos los genes del grupo 4, de todos los genes del grupo 5 y de todos los genes del grupo 6.

En otra realización particular, el método para la eliminación de la placa aterosclerótica carótida comprende determinar el nivel de expresión de todos los genes de la Tabla 1.

5 En una realización particular, el valor de referencia de un gen corresponde al nivel de expresión de dicho gen en una muestra de células de una placa aterosclerótica de un sujeto que tiene placas ateroscleróticas estables.

10 En una realización particular de manera posterior a la aplicación del tratamiento se monitoriza la respuesta del sujeto al tratamiento terapéutico a través de la detección de los biomarcadores identificados.

En una realización particular, la eliminación de la placa aterosclerótica se realiza mediante endarterectomía carotídea.

15 El término "endarterectomía carotídea" o EAC, como se usa en el presente documento, se refiere a un procedimiento quirúrgico usado para reducir el riesgo de apoplejía, para corregir la estenosis (estrechamiento) en la arteria carótida común o arteria carótida interna. La endarterectomía implica la retirada de material de dentro de una arteria.

20 En una realización particular se utilizan estatinas como tratamiento para estabilizar la placa carotídea.

El término "estatina", tal como aquí se utiliza, se refiere a un inhibidor de la 3-hidroxi-3-metilglutaril-coenzima A (HMG-CoA) reductasa, enzima que cataliza el paso limitante de la biosíntesis del colesterol, e incluye cualquier estatina tanto natural como sintética o semisintética, conocida o de nueva síntesis o de nuevo diseño o moléculas relacionadas; tal como aquí se utiliza, el término "moléculas relacionadas" con las estatinas, se refiere a aquellas moléculas con capacidad hipolipemiente (hipocolesterolémica o hipotrigliceridémica). Ejemplos ilustrativos, no limitativos, de estatinas que pueden ser
30 utilizadas en el contexto de la presente invención incluyen atorvastatina, cerivastatina, fluvastatina, lovastatina, mevastatina, una monacolina (e.g., monacolina M, monacolina J, monacolina N, monacolina L, monacolina X, etc.), pitavastatina (también llamada itavastatina), pravastatina, rosuvastatina, simvastatina, etc., así como combinaciones de dos o más estatinas cualesquiera, por ejemplo, monacolinas, por ejemplo, monacolina M,
35 monacolina J, monacolina N, monacolina L, monacolina X, etc.. o cualquier otra combinación de dos o más estatinas cualesquiera. Algunas estatinas pueden presentarse en forma

cerrada (lactona) o en forma abierta (hidroxiácido). Los hidroxiácidos (forma abierta) pueden prepararse a partir de las lactonas correspondientes por hidrólisis convencional, por ejemplo, con hidróxido sódico en metanol, hidróxido sódico en tetrahidrofurano-agua y similares.

5 En otra realización particular, se usan antiagregantes para estabilizar la placa aterosclerótica.

“Antiagregante”, o “antiagregante plaquetario” tal y como se usa en la presente invención, se refiere a un compuesto cuyo principal efecto es inhibir la agregación de las plaquetas y por lo tanto la formación de trombos o coágulos en el interior de las arterias y venas. Ejemplos
10 ilustrativos, no limitativos de antiagregantes son inhibidores de ciclooxigenasa, tales como ácido acetilsalicílico, Sulfinpirazona, Triflusal, Ditazol, Indobufeno; inhibidores de receptores de ADP tales como ticlopidina, clopidogrel, prasugrel; antagonistas del receptor glicoproteína IIa/IIIb o GPIIb-IIIa, tales como tigramín, Eptifibatide, Tirofiban, Abciximab.

15 Los términos y limitaciones descritos anteriormente en relación con el método de diagnóstico y el método de pronóstico de la invención son igualmente aplicables a este aspecto.

Método para la selección de pacientes para ser sometidos a la eliminación de la placa aterosclerótica carotídea

20 En otro aspecto, la invención se relaciona con un método para seleccionar un paciente susceptible de ser tratado con una terapia para la eliminación o estabilización de la placa aterosclerótica carotídea, que comprende:

(i) determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes
25 mostrados en la Tabla 1 en una muestra biológica de dicho sujeto, y

(ii) comparar el nivel de expresión obtenido en (i) con un valor de referencia
en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 1-33 y su
expresión está incrementada respecto al valor de referencia y/o

en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 34-93 y su
30 expresión se encuentra reducida respecto al valor de referencia, entonces el sujeto es susceptible de ser tratado con una terapia para la eliminación o estabilización de la placa aterosclerótica carotídea.

El término “seleccionar” tal y como se usa aquí, se refiere a la acción de escoger a un sujeto
35 para someterlo a un tratamiento preventivo o curativo de la placa aterosclerótica carotídea y por tanto, poder prevenir una enfermedad cerebrovascular.

En una realización particular, la terapia para la eliminación de la placa aterosclerótica carotídea es endarterectomía carotídea.

- 5 En una realización particular el tratamiento para la estabilización de la placa aterosclerótica carotídea es el tratamiento con estatinas o antiagregantes.

La etapa (i) del método para la selección de un paciente, comprende determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en
10 una muestra biológica de dicho sujeto.

En un modo de realización particular la muestra es una muestra que contiene células de placa aterosclerótica. En una realización más preferida las células extraídas de la placa aterosclerótica son células de músculo liso (CML). En una realización aún más preferida,
15 las células de músculo liso son positivas para MYH11 y negativas para PECAM1. Los métodos para obtener e identificar las células de interés se han descrito anteriormente y son igualmente aplicables a este método.

En una realización particular el nivel de expresión de un gen se determina mediante técnicas
20 de secuenciación de ARN.

En una realización particular, el método para la selección de un paciente comprende determinar el nivel de expresión de al menos 1 gen del grupo 1, al menos 1 gen del grupo 2, al menos 1 gen del grupo 3, al menos 1 gen del grupo 4, al menos 1 gen del grupo 5 y/o al
25 menos 1 gen del grupo 6.

En otra realización particular, el método para la selección de un paciente comprende determinar el nivel de expresión de todos los genes del grupo 1.

En otra realización particular el método para la selección de un paciente comprende determinar el nivel de expresión de todos los genes del grupo 2.

30 En otra realización particular, el método para la selección de un paciente comprende determinar el nivel de expresión de todos los genes del grupo 3.

En otra realización particular, el método para la selección de un paciente comprende determinar el nivel de expresión de todos los genes del grupo 4.

En otra realización particular, el método para la selección de un paciente comprende determinar el nivel de expresión de todos los genes del grupo 5.

En otra realización particular, el método para la selección de un paciente comprende determinar el nivel de expresión de todos los genes del grupo 6.

- 5 Los métodos para determinar los niveles de expresión del gen de interés, en particular los niveles de los genes incluidos en la Tabla 1, se han descrito previamente en el contexto del procedimiento de diagnóstico de la invención y son igualmente aplicables en el presente método incluyendo sus limitaciones.

10 En otra realización particular, el método para la selección de un paciente comprende determinar el nivel de expresión de todos los genes del grupo 1, de todos los genes del grupo 2, de todos los genes del grupo 3, de todos los genes del grupo 4, de todos los genes del grupo 5 y de todos los genes del grupo 6.

En otra realización particular, el método para la selección de un paciente comprende determinar el nivel de expresión de todos los genes de la Tabla 1.

15

En una realización particular, el valor de referencia de un gen corresponde al nivel de expresión de dicho gen en una muestra de células de una placa aterosclerótica de un sujeto que tiene placas ateroscleróticas estables.

- 20 Los términos y limitaciones descritos anteriormente en relación con el método de diagnóstico y el método de pronóstico de la invención son igualmente aplicables a este aspecto.

Kits de la invención

- 25 En otro aspecto, la invención se relaciona con un kit que comprende reactivos adecuados para la determinación de los niveles de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 y, opcionalmente, reactivos para la determinación de los niveles de expresión de uno o más genes de mantenimiento.

- 30 En el contexto de la presente invención, "kit" se entiende como un producto que contiene los diferentes reactivos necesarios para llevar a cabo los métodos de la invención empaquetado de forma que permita su transporte y almacenamiento. Los materiales adecuados para el embalaje de los componentes del kit incluyen cristal, plástico (polietileno, polipropileno,

policarbonato y similares), botellas, viales, papel, sobres, y similares. Además, los kits de la invención pueden contener instrucciones para el uso simultáneo, secuencial o separado de los diferentes componentes que están en el kit. Dichas instrucciones pueden estar en forma de material impreso o en forma de un soporte electrónico capaz de almacenar instrucciones de manera que puedan ser leídos por un sujeto, tales como medios electrónicos de almacenamiento (discos magnéticos, cintas y similares), medios ópticos (CD-ROM, DVD) y similares. Adicional o alternativamente, los medios pueden contener direcciones de Internet que proporcionan dichas instrucciones.

5 En una realización particular, el kit comprende reactivos adecuados determinar el nivel de expresión de al menos 1 gen, al menos 2 genes, al menos 3 genes, al menos 4 genes, al menos 5 genes, al menos 6 genes, al menos 7 genes, al menos 8 genes, al menos 9 genes, al menos 10 genes, al menos 11 genes, al menos 12 genes, al menos 13 genes, al menos 14 genes, al menos 15 genes, al menos 16 genes, al menos 17 genes, al menos 18 genes, al menos 19 genes, al menos 20 genes, al menos 21 genes, al menos 22 genes, al menos 23 genes, al menos 24 genes, al menos 25 genes, al menos 26 genes, al menos 27 genes, al menos 28 genes, al menos 29 genes, al menos 30 genes, al menos 31 genes, al menos 32 genes, al menos 33 genes, al menos 34 genes, al menos 35 genes, al menos 36 genes, al menos 37 genes, al menos 38 genes, al menos 39 genes, al menos 40 genes, al menos 41 genes, al menos 42 genes, al menos 43 genes, al menos 44 genes, al menos 45 genes, al menos 46 genes, al menos 47 genes, al menos 48 genes, al menos 49 genes, al menos 50 genes, al menos 51 genes, al menos 52 genes, al menos 53 genes, al menos 54 genes, al menos 55 genes, al menos 56 genes, al menos 57 genes, al menos 58 genes, al menos 59 genes, al menos 60 genes, al menos 61 genes, al menos 62 genes, al menos 63 genes, al menos 64 genes, al menos 65 genes, al menos 66 genes, al menos 67 genes, al menos 68 genes, al menos 69 genes, al menos 70 genes, al menos 71 genes, al menos 72 genes, al menos 73 genes, al menos 74 genes, al menos 75 genes, al menos 76 genes, al menos 77 genes, al menos 78 genes, al menos 79 genes, al menos 80 genes, al menos 81 genes, al menos 82 genes, al menos 83 genes, al menos 84 genes, al menos 85 genes, al menos 86 genes, al menos 87 genes, al menos 88 genes, al menos 89 genes, al menos 90 genes, al menos 91 genes, al menos 92 genes o de los 93 genes de los contenidos en la Tabla 1.

De acuerdo a la presente invención, el kit comprende reactivos específicos para la determinación de cualquier combinación de genes 1-93 de la Tabla 1.

35

En una realización particular el kit de la invención comprende reactivos adecuados para la determinación de los genes 1-93 de la Tabla 1.

5 En otra realización particular, el kit de la invención comprende determinar el nivel de expresión de al menos 1 gen del grupo 1, al menos 1 gen del grupo 2, al menos 1 gen del grupo 3, al menos 1 gen del grupo 4, al menos 1 gen del grupo 5 y/o al menos 1 gen del grupo 6.

10 En una realización particular el kit comprende reactivos adecuados para la determinación del nivel de expresión de al menos un gen seleccionado del grupo 1 que consiste en ADAMTS7, ANXA3, BMP2, CLEC3B, CRHR1, DDR2, DHRS3, EPHA4, EPHB2, ESM1, FLRT3, FRAS1, GATA2, ID1, ID4, ITGA7, KCNJ8, KLF5, MYO18B, NPPC, NRCAM, NRK, PI16, PRICKLE1, PTHLH, PTPRB, RDH10, SEMA3G, SGCD, SMAD9, SOCS3, SOX17, TBX18, XDH, GJA5, TXNIP, SYNE3, KCNE4 y SDPR.

15 En otra realización particular el kit comprende reactivos adecuados para la determinación del nivel de expresión de al menos un gen seleccionado del grupo 2 que consiste en ADAMTS7, BAIAP2, BMP2, CA12, CLEC3B, CHODL, DDR2, GATA2, ID1, ID4, IL11, ITGA7, MGST1, NPPC, PPP1R1C, PTHLH, RNF43, SOCS3, TMTC2 y TNFAIP8L3.

20 En otra realización particular el kit comprende reactivos adecuados para la determinación del nivel de expresión de al menos un gen seleccionado del grupo 3 que consiste en ABCG2, ANXA3, APOA1, BMP2, C10orf10, CHRNA5, CRHR1, G0S2, HAPLN1, HLA-B, ID1, ID4, KCNJ8, NRCAM, PTHLH, RGS2, RNF43, SOX17, ST6GALNAC5, XDH y FRMD3.

25 En otra realización particular el kit comprende reactivos adecuados para la determinación del nivel de expresión de al menos un gen seleccionado del grupo 4 que consiste en ADAMTS7, APOA1, BAIAP2, BMP2, CHODL, DDR2, DHRS3, DKK2, EPHA4, EPHB2, FLRT3, GATA2, HLA-B, ID1, ID4, IL17RD, KLF5, LRRC4C, NPPC, NRCAM, PI16, 30 PRICKLE1, PTHLH, RGS2, RND1, RNF43, SEMA3G, SLITRK4, SMAD9, SOCS3, SOX17, TBX18, XDH, FAM27A, RP11-351M8.1, ZSCAN31, ANXA10, C19orf33, CCL5, DIRAS3, GULP1, IL20RB, MYCT1 y PFKFB4.

35 En otra realización particular el kit comprende reactivos adecuados para la determinación del nivel de expresión de al menos un gen seleccionado del grupo 5 que consiste en BMP2,

NRCAM, CD248, ANXA3, DAPK2, DDR2, GATA2, SOX17, ID1, APOA1, FLRT3, SEMA3G, ITGA7 y EPHA4.

5 En otra realización particular el kit comprende reactivos adecuados para la determinación del nivel de expresión de al menos un gen seleccionado del grupo 6 que consiste en ABCG2, APOA1, BAIAP2, BMP2, CADM3, DAPK2, DMRTA1, EPHA4, EPHB2, ESM1, FGF11, FLRT3, GLDC, HLA-B, ID1, IL11, ITGA7, KCNJ8, MCHR1, MGST1, NPPC, NRXN2, PDK1, PTHLH, RND1, RNF43, SEMA3G, TBX18, XDH, PPP1R3C y CHRNE.

10 En otra realización particular, el kit comprende reactivos para determinar el nivel de expresión de todos los genes del grupo 1, concretamente de ADAMTS7, ANXA3, BMP2, CLEC3B, CRHR1, DDR2, DHRS3, EPHA4, EPHB2, ESM1, FLRT3, FRAS1, GATA2, ID1, ID4, ITGA7, KCNJ8, KLF5, MYO18B, NPPC, NRCAM, NRK, PI16, PRICKLE1, PTHLH, PTPRB, RDH10, SEMA3G, SGCD, SMAD9, SOCS3, SOX17, TBX18, XDH, GJA5, TXNIP,
15 SYNE3, KCNE4 y SDPR.

En otra realización particular, el kit comprende reactivos para determinar el nivel de expresión de todos los genes del grupo 2, concretamente de ADAMTS7, BAIAP2, BMP2, CA12, CLEC3B, CHODL, DDR2, GATA2, ID1, ID4, IL11, ITGA7, MGST1, NPPC, PPP1R1C,
20 PTHLH, RNF43, SOCS3, TMTC2 y TNFAIP8L3.

En otra realización particular, el kit comprende reactivos para determinar el nivel de expresión de todos los genes del grupo 3, concretamente de ABCG2, ANXA3, APOA1, BMP2, C10orf10, CHRNA5, CRHR1, G0S2, HAPLN1, HLA-B, ID1, ID4, KCNJ8, NRCAM,
25 PTHLH, RGS2, RNF43, SOX17, ST6GALNAC5, XDH y FRMD3.

En otra realización particular, el kit comprende reactivos para determinar el nivel de expresión de todos los genes del grupo 4, concretamente de ADAMTS7, APOA1, BAIAP2, BMP2, CHODL, DDR2, DHRS3, DKK2, EPHA4, EPHB2, FLRT3, GATA2, HLA-B, ID1, ID4,
30 IL17RD, KLF5, LRRC4C, NPPC, NRCAM, PI16, PRICKLE1, PTHLH, RGS2, RND1, RNF43, SEMA3G, SLITRK4, SMAD9, SOCS3, SOX17, TBX18, XDH, FAM27A, RP11-351M8.1, ZSCAN31, ANXA10, C19orf33, CCL5, DIRAS3, GULP1, IL20RB, MYCT1 y PFKFB4.

En otra realización particular, el kit comprende reactivos para determinar el nivel de
35 expresión de todos los genes del grupo 5, concretamente de BMP2, NRCAM, CD248, ANXA3, DAPK2, DDR2, GATA2, SOX17, ID1, APOA1, FLRT3, SEMA3G, ITGA7 y EPHA4.

En otra realización particular el kit comprende reactivos adecuados para la determinación del nivel de expresión de todos los genes del grupo 6 , concretamente de ABCG2, APOA1, BAIAP2, BMP2, CADM3, DAPK2, DMRTA1, EPHA4, EPHB2, ESM1, FGF11, FLRT3, GLDC, HLA-B, ID1, IL11, ITGA7, KCNJ8, MCHR1, MGST1, NPPC, NRXN2, PDK1, PTHLH, RND1, RNF43, SEMA3G, TBX18, XDH, PPP1R3C y CHRNE.

En otra realización particular el kit comprende reactivos adecuados para determinar el nivel de expresión de todos los genes del grupo 1, de todos los genes del grupo 2, de todos los genes del grupo 3, de todos los genes del grupo 4, de todos los genes del grupo 5 y de todos los genes del grupo 6.

En una realización particular de los kits de la invención, dichos kits comprenden opcionalmente, reactivos para la determinación de los niveles de expresión de uno o más genes de mantenimiento.

En otra forma de realización, el kit según la presente invención comprende además reactivos adecuados para la determinación de los niveles de expresión de los genes que se expresan específicamente o no se expresan en las poblaciones celulares de músculo liso, concretamente MYH11 y PECAM1.

En una forma de realización preferida, los reactivos adecuados para la determinación de niveles de expresión de uno o más genes comprenden al menos el 10%, al menos el 20%, al menos el 30%, al menos el 40%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90% o al menos el 100% de la cantidad total de los reactivos adecuados para la determinación de los niveles de expresión de genes que forman el kit.

En formas de realización adicionales, los reactivos adecuados para la determinación de los niveles de expresión de uno o más genes comprenden al menos el 55% al menos el 60%, al menos el 65%, al menos el 70%, al menos el 75%, al menos el 80%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98% o al menos el 99% de la cantidad total de reactivos que forman el kit.

Así, en el caso particular de kits que comprenden los reactivos para la determinación de los niveles de expresión de los genes del grupo 1 ADAMTS7, ANXA3, BMP2, CLEC3B, CRHR1, DDR2, DHRS3, EPHA4, EPHB2, ESM1, FLRT3, FRAS1 GATA2, ID1, ID4, ITGA7, KCNJ8,

KLF5, MYO18B, NPPC, NRCAM, NRK, PI16, PRICKLE1, PTHLH, PTPRB, RDH10, SEMA3G, SGCD, SMAD9, SOCS3, SOX17, TBX18, XDH, GJA5, TXNIP, SYNE3, KCNE4 y SDPR, los reactivos específicos para dichos genes (por ejemplo, sondas que son capaces de hibridar en condiciones rigurosas con los genes ADAMTS7, ANXA3, BMP2, CLEC3B, CRHR1, DDR2, DHRS3, EPHA4, EPHB2, ESM1, FLRT3, FRAS1, GATA2, ID1, ID4, ITGA7, KCNJ8, KLF5, MYO18B, NPPC, NRCAM, NRK, PI16, PRICKLE1, PTHLH, PTPRB, RDH10, SEMA3G, SGCD, SMAD9, SOCS3, SOX17, TBX18, XDH, GJA5, TXNIP, SYNE3, KCNE4 y SDPR) comprenden al menos el 10%, al menos el 20%, al menos el 30%, al menos el 40%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90% o al menos el 100% de las sondas presentes en el kit.

La expresión “reactivo que permite determinar el nivel de expresión de un gen” significa un compuesto o conjunto de compuestos que permite determinar el nivel de expresión de un gen, tanto por medio de la determinación del nivel de ARNm como por medio de la determinación del nivel de proteína. Por tanto, los reactivos del primer tipo incluyen sondas capaces de hibridar específicamente con el ARNm codificado por dichos genes. Los reactivos del segundo tipo incluyen compuestos que se unen específicamente a las proteínas codificadas por los genes marcadores y, preferentemente, se incluyen anticuerpos, aunque pueden ser aptámeros específicos.

En una forma de realización particular de los kits de la invención, los reactivos del kit son ácidos nucleicos que son capaces de detectar específicamente el nivel de ARNm de los genes mencionados anteriormente y/o el nivel de proteínas codificadas por uno o más de los genes mencionados anteriormente. Los ácidos nucleicos capaces de hibridar específicamente con los genes antes mencionados pueden ser uno o más pares de oligonucleótidos cebadores para la amplificación específica de fragmentos de los ARNm (o de sus correspondientes ADNc) de dichos genes.

En una forma de realización preferida, el primer componente del kit de la invención comprende una sonda que puede hibridar específicamente con los genes mencionados anteriormente.

El término “que hibrida específicamente”, como se usa aquí, se refiere a las condiciones que permiten la hibridación de dos polinucleótidos en condiciones muy rigurosas o condiciones moderadamente rigurosas.

La “rigurosidad” de las reacciones de hibridación es fácilmente determinable por un experto en la materia, y por lo general es un cálculo empírico dependiente de la longitud de la sonda, la temperatura de lavado, y la concentración de sal. En general, las sondas más largas requieren temperaturas más elevadas para la hibridación adecuada, mientras que sondas más cortas necesitan temperaturas más bajas. La hibridación depende generalmente de la capacidad del ADN desnaturalizado de volver a hibridar cuando están presentes hebras complementarias en un ambiente por debajo de su temperatura de fusión. Cuanto mayor sea el grado de homología deseada entre la sonda y la secuencia hibridable, mayor es la temperatura relativa que se puede usar. Como resultado, se deduce que las temperaturas relativas más altas tenderían a hacer las condiciones de reacción más rigurosas, mientras que las temperaturas más bajas menos. Para obtener detalles y explicaciones adicionales de la rigurosidad de las reacciones de hibridación, véase Ausubel *et al.*, Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995).

“Las condiciones rigurosas” o “condiciones muy rigurosas”, como se define en el presente documento, por lo general: (1) emplean baja fuerza iónica y alta temperatura para el lavado, por ejemplo cloruro de sodio 0,015 M/citrato de sodio 0,0015 M/dodecil sulfato de sodio al 0,1% a 50°C, (2) emplean durante la hibridación un agente de desnaturalización, tal como formamida, por ejemplo, formamida al 50% (v/v) con seroalbúmina bovina al 0,1%/Ficoll al 0.1%/polivinilpirrolidona al 0,1%/tampón fosfato de sodio 50 mM a pH 6,5 con cloruro de sodio 750 mM, citrato sódico 75 mM a 42°C, o (3) emplean formamida al 50%, 5xSSC (NaCl 0,75 M, citrato de sodio 0,075 M), fosfato sódico 50 mM (pH 6,8), pirofosfato de sodio al 0,1%, solución de Denhardt 5x, ADN sonicado de esperma de salmón (50 mg/ml), SDS al 0,1%, y sulfato de dextrano al 10% a 42°C, con lavados a 42°C en 0,2xSSC (cloruro sódico/citrato de sodio) y formamida al 50%, seguido por un lavado de alto rigor que consiste en 0,1xSSC que contiene EDTA a 55°C.

“Las condiciones moderadamente rigurosas” se pueden identificar según lo descrito por Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Nueva York: Cold Spring Harbor Press, 1989, e incluyen el uso de solución de lavado y condiciones de hibridación (por ejemplo, temperatura, fuerza iónica y % de SDS) menos rigurosas que las descritas anteriormente. Un ejemplo de condiciones moderadamente rigurosas es la incubación durante la noche a 37°C en una solución que comprende: formamida al 20%, 5xSSC (NaCl 150 mM, citrato sódico 15 mM), fosfato sódico 50 mM (pH 7,6), solución de Denhardt 5x, sulfato de dextrano al 10%, y 20 mg/ml de ADN desnaturalizado cortado de esperma de salmón, seguido por el lavado de los filtros en 1xSSC a unos 37-50°C. El experto en la

materia reconocerá la forma de ajustar la temperatura, fuerza iónica, etc. según sea necesario para dar cabida a factores como la longitud de la sonda y similares.

5 En el caso de que los niveles de expresión de varios de los genes identificados en la presente invención se determinen al mismo tiempo, es útil incluir las sondas para todos los genes cuya expresión se va a determinar en una hibridación de micromatrices.

10 Las micromatrices comprenden una pluralidad de ácidos nucleicos que están espacialmente distribuidos y establemente asociados a un soporte (por ejemplo, un biochip). Los ácidos nucleicos tienen una secuencia complementaria a subsecuencias particulares de genes cuya expresión se va a detectar, por tanto son capaces de hibridar con dichos ácidos nucleicos. En los métodos de la invención, una micromatriz que comprende un conjunto de ácidos nucleicos se pone en contacto con una preparación de los ácidos nucleicos aislados del paciente objeto del estudio. La incubación de la micromatriz con la preparación de ácidos nucleicos se lleva a cabo en las condiciones adecuadas para la hibridación. Posteriormente, tras la eliminación de los ácidos nucleicos que el soporte no ha retenido, el patrón de hibridación se detecta, lo que proporciona información sobre el perfil genético de la muestra analizada. Aunque las micromatrices son capaces de proporcionar información tanto cualitativa como cuantitativa de los ácidos nucleicos presentes en una muestra, la invención requiere el uso de matrices y metodologías capaces de proporcionar información cuantitativa.

25 La invención contempla una variedad de matrices en relación con el tipo de sondas y en relación con el tipo de soporte utilizado. Las sondas incluidas en las matrices que son capaces de hibridar con los ácidos nucleicos pueden ser ácidos nucleicos o análogos de los mismos que mantienen la capacidad de hibridación, como por ejemplo, ácidos nucleicos en el que el enlace fosfodiéster se ha sustituido con un enlace fosforotioato, metilimina, metilfosfonato, fosforamidato, guanidina y similares, ácidos nucleicos en los que se sustituye la ribosa de los nucleótidos con otra hexosa, ácidos péptido nucleicos (APN). La longitud de las sondas puede ser de 5 a 50 nucleótidos y, preferiblemente, de 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 100 nucleótidos y varían en el intervalo de 10 a 1000 nucleótidos, preferiblemente en el intervalo de 15 a 150 nucleótidos, más preferiblemente en el intervalo de 15 a 100 nucleótidos y pueden ser ácidos nucleicos monocatenarios o bicatenarios. La matriz puede contener todas las sondas específicas de un determinado ARNm de una cierta longitud o puede contener sondas seleccionadas de diferentes regiones de un ARNm. Cada sonda se ensaya en paralelo con una sonda con una base cambiada,

preferiblemente en una zona céntrica de la sonda. La matriz se pone en contacto con una muestra que contiene los ácidos nucleicos con secuencias complementarias a las sondas de la matriz y se determina la señal de hibridación con cada una de las sondas y los controles de hibridación correspondiente. Esas sondas en las que se observa una mayor diferencia entre la señal de hibridación con la sonda y su control de hibridación se seleccionan. El proceso de optimización puede incluir una segunda ronda de optimización en el que se hibrida la matriz de hibridación con una muestra que no contiene secuencias complementarias a las sondas de la matriz. Después de la segunda ronda de selección, las sondas con las señales de hibridación inferior a un umbral se seleccionarán. Por tanto, se seleccionan las sondas que pasan ambos controles, es decir, que muestran un nivel mínimo de hibridación inespecífica y un nivel máximo de hibridación específica con el ácido nucleico diana.

La selección de las sondas específicas para los diferentes genes diana se lleva a cabo de tal manera que se unen específicamente al ácido nucleico diana con un mínimo de hibridación con genes no relacionados. Sin embargo, hay sondas de 20 nucleótidos que no son únicas para un ARNm determinado. Por tanto, las sondas dirigidas a dichas secuencias mostrarán una hibridación cruzada con secuencias idénticas que aparecen en el ARNm de genes no relacionados. Además, hay sondas que no hibridan específicamente con los genes diana en las condiciones utilizadas (debido a las estructuras secundarias o de la interacción con el sustrato de la matriz). Este tipo de sonda no se debe incluir en la matriz. Por tanto, el experto en la materia observará que las sondas que se van a incorporar en una matriz determinada se deben optimizar antes de su incorporación a la matriz. La optimización de las sondas se realiza generalmente mediante la generación de una matriz que contiene una pluralidad de sondas dirigidas a las diferentes regiones de un cierto polinucleótido diana. Esta matriz se pone en contacto primero con una muestra que contiene el ácido nucleico diana en una forma aislada y, en segundo lugar, con una mezcla compleja de ácidos nucleicos. Las sondas de hibridación que muestran una hibridación muy específica con el ácido nucleico diana, pero poca o ninguna hibridación con la muestra compleja se seleccionan así para su incorporación a las matrices de la invención. Además, es posible incluir en la matriz controles de hibridación para cada una de las sondas que se va a estudiar. En una forma de realización preferida, los controles de hibridación contienen una posición alterada en la región central de la sonda. En el caso de que se observen altos niveles de hibridación entre la sonda estudiada y su control de hibridación, la sonda no se incluye en la matriz.

Las micromatrices de la invención no sólo contienen sondas específicas para los polinucleótidos que indica una determinada situación patofisiológica, sino que también contiene una serie de sondas control, que puede ser de tres tipos: controles de normalización, controles del nivel de expresión y controles de hibridación.

5

Los controles de normalización son oligonucleótidos que son perfectamente complementarios a las secuencias de referencia marcadas que se añaden a la preparación de los ácidos nucleicos que se va a analizar. Las señales derivadas de los controles de normalización después de la hibridación proporcionan una indicación de las variaciones en las condiciones de hibridación, la intensidad del marcador, eficiencia de la detección y otra serie de factores que pueden dar lugar a una variación de la señal de hibridación entre diferentes micromatrices. Las señales detectadas del resto de las sondas de la matriz preferiblemente se dividen por la señal emitida por las sondas control, normalizando así las medidas. Prácticamente cualquier sonda se puede utilizar como control de normalización. Sin embargo, se sabe que la eficiencia de la hibridación varía en función de la composición de nucleótidos y la longitud de la sonda. Por lo tanto, las sondas de normalización preferidas son las que representan la longitud media de las sondas presentes en la matriz, aunque pueden seleccionarse de forma que incluyan una gama de longitudes que reflejan el resto de las sondas presentes en la matriz. Las sondas de normalización se pueden diseñar de tal forma que reflejen la composición media de los nucleótidos del resto de sondas presentes en la matriz. Un número limitado de sondas de normalización se selecciona preferiblemente de modo que hibriden adecuadamente, es decir, no tienen una estructura secundaria y no presentan similitud de secuencia con cualquiera de las sondas de la matriz se utiliza. Las sondas de normalización pueden ser ubicadas en cualquier posición de la matriz o en varias posiciones de la matriz para un control eficaz de las variaciones en la eficiencia de la hibridación en relación con la estructura de la matriz. Los controles de normalización se localizan preferentemente en las esquinas de la matriz y/o en el centro de la misma.

Los niveles de controles de expresión son sondas que hibridan específicamente con genes que se expresan de forma constitutiva en la muestra que se analiza. Los controles de nivel de expresión están diseñados para controlar el estado fisiológico y la actividad metabólica de la célula. El examen de la covarianza del control del nivel de expresión del ácido nucleico diana indica si las variaciones en los niveles de expresión se deben a cambios en los niveles de expresión o se deben a cambios en la tasa global de transcripción en la célula o en su actividad metabólica general. Así, en el caso de células que tienen deficiencias en un determinado metabolito esencial para la viabilidad celular, se espera la observación de una

disminución tanto en los niveles de expresión del gen diana como en los niveles de expresión del control. Por otro lado, si se observa un aumento en la expresión de la expresión del gen diana y de los genes de control, probablemente es debido a un aumento de la actividad metabólica de la célula y no a un aumento diferencial en la expresión del gen diana. Sondas adecuadas para su uso como control de expresión corresponden a genes expresados constitutivamente, como los genes que codifican las proteínas que ejercen funciones esenciales de la célula, tales como β -2-microglobulina ubiquitina, proteína ribosómica de 18S, ciclofilina A, receptor de transferrina, actina, GAPDH, proteína de activación de tirosina 3-monooxigenasa/triptófano 5-monooxigenasa (YWHAZ) y beta-actina.

Se pueden incluir controles de hibridación tanto para las sondas dirigidas a los genes diana como para las sondas dirigidas al nivel de expresión o a los controles de normalización. Controles de error son sondas de oligonucleótidos idénticas a las sondas dirigidas a los genes diana, pero que contienen mutaciones en uno o varios nucleótidos, es decir, que contienen nucleótidos en ciertas posiciones que no hibridan con el nucleótido correspondiente en el gen diana. Los controles de hibridación se seleccionan de manera que, aplicando las condiciones de hibridación adecuadas, el gen diana debe hibridar con la sonda específica, pero no con el control de hibridación o con una eficacia reducida. Los controles de hibridación contienen preferentemente una o varias posiciones modificadas en el centro de la sonda. Los controles de hibridación por tanto, proporcionan una indicación del grado de hibridación inespecífica o de hibridación cruzada con un ácido nucleico en la muestra a una sonda diferente de la que contiene la secuencia exactamente complementaria.

Las matrices de la invención también pueden contener controles de amplificación y de preparación de muestras que son sondas complementarias a subsecuencias de genes control seleccionados porque normalmente no aparecen en la muestra biológica objeto del estudio, tales como sondas para genes bacterianos. La muestra de ARN se suplementa con una cantidad conocida de un ácido nucleico que hibrida con la sonda control seleccionada. La determinación de la hibridación de dicha sonda indica el grado de recuperación de los ácidos nucleicos durante su preparación, así como una estimación de la alteración causada en los ácidos nucleicos durante el procesamiento de la muestra.

Una vez que se proporcionan un conjunto de sondas que muestra la especificidad adecuadas y un conjunto de sondas control, estas últimos están dispuestas en la matriz en una posición conocida de tal manera que, tras los pasos de hibridación y detección, es posible establecer una correlación entre una señal positiva de hibridación y el gen particular

a partir de las coordenadas de la matriz en las que se detecta la señal positiva de hibridación.

5 Las micromatrices pueden ser matrices de alta densidad, con miles de oligonucleótidos por medio de métodos de síntesis in situ fotolitográficos (Fodor *et al.*, 1991, Science, 767-773). Este tipo de sonda suele ser redundante, es decir, incluyen varias sondas para cada ARNm que se va a detectar. En una forma de realización preferida, las matrices son matrices de baja densidad o LDA que contienen menos de 10.000 sondas por centímetro cuadrado. En
10 dichas matrices de baja densidad, las diferentes sondas se aplican manualmente con la ayuda de una pipeta en diferentes lugares de un soporte sólido (por ejemplo, una superficie de cristal, una membrana). Los soportes utilizados para fijar las sondas se pueden obtener de una gran variedad de materiales, como plástico, cerámica, metales, geles, membranas, cristales, y similares. Las micromatrices se pueden obtener usando cualquier metodología conocida por el experto en la materia.

15

Después de la hibridación, en los casos en que el ácido nucleico no hibridado es capaz de emitir una señal en el paso de detección, es necesaria una etapa de lavado para eliminar dicho ácido nucleico no hibridado. La etapa de lavado se lleva a cabo utilizando métodos y soluciones conocidos por el experto en la materia.

20

En el caso de que el marcaje en el ácido nucleico no sea directamente detectable, es posible conectar la micromatriz que comprende los ácidos nucleicos diana unidos a la matriz con los otros componentes del sistema necesarios para producir la reacción que da lugar a una señal detectable. Por ejemplo, si los ácidos nucleicos diana se marcan con biotina, la
25 matriz se pone en contacto con estreptavidina conjugada con un reactivo fluorescente en condiciones adecuadas para que la unión entre la biotina y la estreptavidina se produzca. Después de la incubación de la micromatriz con el sistema de generación de la señal detectable, es necesario llevar a cabo una etapa de lavado para eliminar todas las moléculas que se han unido de forma no específica a la matriz. Las condiciones de lavado
30 las determinará el experto en la materia, usando condiciones adecuadas según el sistema de generación de la señal detectable y que son bien conocidos por el experto en la materia. El patrón de hibridación resultante se puede ver o detectar de diferentes maneras, dicha detección se determina por el tipo de sistema utilizado en la micromatriz. Por tanto, la detección del patrón de hibridación se puede llevar a cabo por medio de recuento de
35 centelleo, autorradiografía, determinación de una señal fluorescente, determinaciones calorimétricas, detección de una señal luminosa, etc.

Después de la hibridación y los posibles procesos de lavado y tratamiento posteriores, el patrón de hibridación se detecta y cuantifica, por lo que la señal correspondiente a cada punto de la hibridación en la matriz se compara con un valor de referencia correspondiente a la señal emitida por un número conocido de ácidos nucleicos terminalmente marcados con el fin de obtener así un valor absoluto del número de copias de cada ácido nucleico que hibrida en un cierto punto de la micromatriz.

En el caso de que los niveles de expresión de los genes según la presente invención se determinen mediante la medida de los niveles del polipéptido o polipéptidos codificados por dicho gen o genes, los kits según la presente invención comprenden reactivos que son capaces de unirse específicamente a dicho polipéptido o polipéptidos.

Para ello, las matrices de anticuerpos tales como las descritas por De Wildt *et al.* (2000) Nat. Biotechnol. 18:989-994; Lueking *et al.* (1999) Anal. Biochem. 270:103-111; Ge *et al.* (2000) Nucleic Acids Res. 28, e3, I-VII; MacBeath y Schreiber (2000) Science 289:1760-1763, documentos WO 01/40803 y WO 99/51773A1 son útiles. Los anticuerpos de la matriz incluyen cualquier agente inmunológico capaz de unirse a un ligando con alta afinidad, incluyendo IgG, IgM, IgA, IgD e IgE, así como moléculas similares a anticuerpos que tienen un sitio de unión al antígeno, tales como Fab', Fab, F(ab')₂, anticuerpos de dominio único o DABS, Fv, scFv y similares. Las técnicas para preparar dichos anticuerpos las conoce muy el experto en la materia e incluyen los métodos descritos por Ausubel *et al.* (Current Protocols in Molecular Biology, eds. Ausubel *et al.*, John Wiley & Sons (1992).

Los anticuerpos de la matriz se pueden aplicar a gran velocidad, por ejemplo, mediante sistemas robóticos disponibles comercialmente (por ejemplo, los producidos por la Genetics Microsystems o Biorobotics). El sustrato de la matriz puede ser nitrocelulosa, plástico, cristal o puede ser de un material poroso, como por ejemplo, acrilamida, agarosa u otro polímero. En otra forma de realización, es posible utilizar células que producen los anticuerpos específicos para la detección de las proteínas de la invención por medio de su cultivo en filtros matriz. Después de la inducción de la expresión de los anticuerpos, estos últimos se encuentran inmovilizados en el filtro en la posición de la matriz donde se localizó la célula productora. Se puede poner en contacto una matriz de anticuerpos con una diana marcada y se puede determinar el nivel de unión de la diana a los anticuerpos inmovilizados. Si la diana no está marcada, se puede utilizar un ensayo de tipo sándwich en el que se usa un segundo anticuerpo marcado específico para el polipéptido que se une al polipéptido que se

inmoviliza en el soporte. La cuantificación de la cantidad de polipéptido presente en la muestra en cada punto de la matriz se puede almacenar en una base de datos como un perfil de expresión. La matriz de anticuerpos se puede producir por duplicado y se puede utilizar para comparar los perfiles de unión de dos muestras diferentes.

5

En otro aspecto, la invención se relaciona con el uso de un kit de la invención para diagnosticar placas ateroscleróticas inestables en un sujeto, para determinar la probabilidad de un sujeto de sufrir una enfermedad cerebrovascular, para la eliminación de la placa aterosclerótica carotídea o para seleccionar un paciente susceptible de ser tratado con una
10 terapia para la eliminación de la placa aterosclerótica carotídea.

Los términos y limitaciones descritos anteriormente en relación con el método de diagnóstico y el método de pronóstico de la invención son igualmente aplicables a este aspecto.

15 La invención se describe a continuación mediante los siguientes ejemplos, que deben ser considerados como meramente ilustrativos y en ningún caso limitativos del ámbito de la presente invención.

EJEMPLOS

20

Materiales y métodos

Selección de pacientes

25 Se seleccionaron pacientes que se sometieron a endarterectomía carotídea en el Hospital Universitario de Basurto para incluirse en el estudio actual basándose en parámetros clínicos (edad, síntomas, sin otras manifestaciones clínicas). Todos los pacientes se evaluaron antes y después de la operación por un neuropsicólogo especializado. Se completaron pruebas de radiología para evaluar el grado de estenosis con un equipo de
30 ecografía Doppler. Sólo los pacientes que cumplen todos los parámetros requeridos se incluyeron finalmente en el estudio (7A frente a 7S). Se recogió una muestra de tejido carotídeo tras la cirugía y se transportó inmediatamente al laboratorio para el aislamiento celular.

35 *Estrategia experimental para diferenciar las CML aisladas de placas de carótida de pacientes sintomáticos y asintomáticos mediante la tecnología RNAseq.*

El estudio fue diseñado de forma que se obtuviera el poder máximo para detectar expresión génica diferencial (DEG) y expresión de isoformas diferencial (DEI) y que se realizaran análisis complementarios de enriquecimiento funcional y análisis de redes. Los dos factores principales a tener en cuenta en estudios de transcriptómica con RNAseq son el número de réplicas y la profundidad de secuenciación. Por tanto, la secuenciación se realizó con una profundidad para obtener alrededor de 60 millones de lecturas por muestra, lo cual nos permitió alcanzar un poder de detección de expresión diferencial fiable. Secuenciar con esta profundidad permitiría identificar expresión diferencial con máximo poder en 5 réplicas incluso en aquellos genes con niveles de expresión bajos. Además, la variabilidad en análisis de RNAseq se reduce al realizar el estudio en un tipo celular (i.e. CML) comparado con análisis en la placa completa, que es heterogénea y se compone de diferentes tipos celulares como CML, células endoteliales y macrófagos. Por consiguiente, para estudiar el papel que tienen las CML en el desarrollo de la placa de carótida inestable, seleccionamos 14 pacientes (7 sintomáticos y 7 asintomáticos) que fueron seleccionados para una endarterectomía por el comité de médicos expertos del Hospital Universitario de Basurto, quien evaluó los datos clínicos de los pacientes (tabla 3). Las pruebas MRI y dúplex cervical se realizaron en todos los pacientes por el servicio de radiología en el hospital. Los pacientes con estenosis >70% y diagnosticados como sintomáticos (S) y pacientes con estenosis >80% c y sin ningún síntoma (A) fueron incluidos en el estudio. Parámetros adicionales requeridos en este estudio fueron definidos como: (1) edad (70 años \pm 12); (2) ausencia de otras condiciones médicas; (3) ausencia de placa en la contralateral; y (4) ausencia de otras causas de ACV como origen cardioembólico. Durante el periodo de reclutamiento varios pacientes fueron excluidos del estudio porque presentaron por lo menos alguno de los siguientes problemas: (1) pacientes que mostraron signos de demencia según la prueba "minimal state examination" MMSE; (2) pacientes con alteraciones psicológicas debido a fármacos o la intervención cerebral; (3) pacientes que presentaron complicaciones médicas o de cirugía que puedan haber alterado la capacidad intelectual y/o la prueba de neuroimagen y (4) pacientes quienes que no fueron capaces de realizar el test neuropsicológico debido a una dificultad de comunicación (i.e. problema de visión).

Cultivo de células del músculo liso y purificación de ARN

Se procesaron muestras recientes de tejido de arteria carótida procedentes de pacientes sintomáticos y asintomáticos cortándose en fragmentos de tamaño de 2-3 mm, que se digirieron con colagenasa tipo I (Life Technologies, Carlsbad, CA, EE.UU.) para aislar y

cultivar adicionalmente CML. En primer lugar, se realizó una digestión de 3 horas a una concentración de 300 U/ml de colagenasa tipo I. En segundo lugar, se sometió el tejido digerido a digestión durante la noche con colagenasa tipo I 220 U/ml. Se filtró el tejido digerido mediante un filtro Falcon™ Cell Strainer de 100 µm de nailon para retirar el material no digerido y se sembraron las células en placas en medio específico para el crecimiento de CML humanas denominado M231 (Life Technologies, Carlsbad, CA, EE.UU.) complementado con factores de crecimiento de CML tales como IGF-1 20 ng/ml, EGFB 2 ng/ml, EGF 0,5 ng/ml, heparina 5 ng/ml, BSA 0,2 µg/ml y suero bovino fetal al 5% (Sigma, St Louis, MO, EE.UU.). Se seleccionaron células del pase inicial para la extracción de ARN. Se realizó un control de calidad de cultivo de CML puro mediante la identificación de la presencia de marcadores específicos de CML (MYH11) y la ausencia de marcador endotelial (PECAM1). La extracción y purificación de ARN procedente de CML se realizó con el reactivo TRIzol™ según las instrucciones del fabricante. Entonces se purificó el ARN con el kit de aislamiento de ARN total RNeasy (QIAGEN, Valencia, CA, EE.UU.) siguiendo las instrucciones recomendadas por el fabricante. Se analizó la integridad del ARN (RIN>9) usando RNA 6000 Nano Chips en el bioanalizador Agilent 2100 (Agilent Technologies, Palo Alto, CA, EE.UU.).

Inmunotransferencia de tipo Western

Se realizó lisis celular con el tampón de lisis RIPA (Tris HCl 150 mM, NaCl 150 mM, desoxicolato al 0,5%, SDS al 0,1%, NP-40 al 1%) durante 30 min a 4°C seguido por centrifugación a 20.000xg durante 10 min. Se sometieron los lisados celulares a geles de SDS-PAGE al 6% o el 12%. Se transfirieron las proteínas a la membrana de PVDF. Se realizaron inmunodetecciones con los siguientes anticuerpos primarios: anti-cadena pesada de miosina del músculo liso 11 (ab82541, Abcam, Cambridge, RU) y anticuerpo policlonal frente a actina del músculo liso (23081-1-AP, ProteinTech), anticuerpo anti-CD31 (PECAM1-EPR3094 ProteinTech, Chicago, EE.UU.). Se detectaron las proteínas con reactivo de detección de sustrato de HRP mediante quimioluminiscencia para inmunotransferencia de tipo Western Immobilon™ (Millipore, Billerica, MA, EE.UU.) y se visualizaron con el sistema de obtención de imágenes ChemiDoc™XRS (Bio-Rad, Richmond, CA, EE.UU.). Se usó el anticuerpo anti-gliceraldehído-3-fosfato deshidrogenasa (MAB374, Millipore, Billerica, MA, EE.UU.) como control interno para normalización. El análisis de los datos se realizó usando el software Lab™ (Bio-Rad, Richmond, CA, EE.UU.).

Microscopía

Se hicieron crecer las células sobre cubreobjetos de vidrio redondos (Thermo Scientific) seguido por fijación en metanol enfriado en hielo. Después se bloquearon las muestras en PBS/BSA al 3% p/v durante 30 min y se tiñeron con anticuerpo anti-cadena pesada 11 de miosina del músculo liso durante 1 h a temperatura ambiente. Se lavaron las células con PBS y se incubaron con anticuerpo secundario de cabra anti IgG de conejo (H+L) Alexa Fluor[®]488 durante 45 min a temperatura ambiente en la oscuridad y se contratiñó el ADN con 4',6-diamino-2-fenilindol (DAPI). Tras tres lavados, se montaron los cubreobjetos en medio de montaje acuoso Fluoromount[™] (Sigma-Aldrich, St Louis, MO, EE.UU.). Se realizó la adquisición de imágenes con un microscopio de superresolución Leica TCS STED CW SP8 con una lente 40x y registrando secciones ópticas cada 0,3 μ m. Se realizó el análisis de la imagen usando el software ImageJ (Instituto Nacional de Salud, EE.UU.).

Biblioteca de ARN

Se eliminó el ARN ribosómico con el kit de eliminación de ARNr Ribo-Zero. Se realizó la generación de bibliotecas con el kit de preparación de bibliotecas de ARN total monocatenario TruSeq siguiendo las recomendaciones del fabricante. Se comenzó partiendo de 2 μ g de bibliotecas de ARN total (RIN>9), que se secuenciaron usando un instrumento HiSeq2500 (Illumina Inc, San Diego, CA, EE.UU.). Las lecturas de secuenciación se realizaron por extremos apareados con una longitud de lectura de aproximadamente 100 pb realizadas en 14 muestras. La cobertura estimada fue de aproximadamente 60 millones de lecturas por muestra (3 carriles). La generación de bibliotecas y la secuenciación de ARN se realizaron en Sistemas Genómicos S.L. (Valencia, España), siguiendo las instrucciones del fabricante.

Análisis de transcriptómica de ARN

El control de calidad de los datos sin procesar se realizó usando la herramienta FastQC. Se mapearon las lecturas por extremos apareados sin procesar frente al genoma humano proporcionado por la base de datos Ensembl (versión GRchr37/hg19) usando el algoritmo tophat2 (Kim, D., et al *Genome Biol* 14, 2013). Se eliminaron las lecturas de calidad insuficiente (puntuación phred <10) usando el software Picard Tools (versión 1.129). En esta etapa, se evaluó la distribución de GC (es decir, la proporción de pb de guanina y citosina a lo largo de las lecturas), que debe tener una distribución deseada de entre el 40-60%. En segundo lugar, se evaluó la distribución de duplicados (indicador de calidad de

secuenciación) para confirmar que nuestra secuenciación contenía una pequeña proporción de duplicados. Se estimaron las predicciones génicas usando el método de Cufflinks (Trapnell, C., et al. *Nat. Biotechnol.* 28, 511–515, 2010) y se calcularon los niveles de expresión usando el software HT Seq (versión 0.6.0, <http://www-huber.embl.de/users/anders/HTSeq/>). Este método emplea lecturas únicas para la estimación de la expresión génica y elimina las lecturas mapeadas múltiples. Se evaluó el análisis de expresión diferencial entre condiciones usando el método DESeq2 (Love, M.I. et al, *Genome Biol* 15, 2014, versión 3.4). Finalmente, seleccionamos genes expresados diferencialmente con un valor de P ajustado mediante FDR (Benjamini, Y., et al *J Roy Stat Soc B* 57, 1995) $< 0,05$ y una tasa de cambio de al menos 1,5. El análisis de DEG entre S y A se realizó usando paquetes estadísticos diseñados por Python y R. Usando el algoritmo DESeq2 (Anders, S. et al *Genome Biol* 11, 2010 que aplica una distribución binominal negativa diferencial para la significación estadística (Love, M.I. et al, *Genome Biol* 15, 2014, versión 3.4) identificamos genes e isoformas expresadas diferencialmente. Consideramos como genes o isoformas expresados diferencialmente aquellos con un valor de FC inferior a -1,5 o superior a 1,5 y con un valor de P (P_{ajust}) corregido mediante FDR $\leq 0,05$ (Benjamini, Y., et al, *J Roy Stat Soc B* 57, 1995) para evitar la identificación de falsos positivos en los datos de expresión diferencial.

20 *Análisis de enriquecimiento funcional y análisis de redes del conjunto de genes*

Se procesaron conjuntos expresados diferencialmente usando ClusterProfiler (Yu, G., et al. *Omi. a J. Integr. Biol.* 16, 284–287 (2012), un paquete de bioconductores, para buscar procesos biológicos implicados en la inestabilidad de la placa. Esta herramienta examina genes en bases de datos específicas (es decir Gene Ontology – GO, Kyoto Encyclopedia of Genes and Genomes – KEGG, DRUG, etc.) para evaluar anotaciones biológicas que surgen como sobrerrepresentadas con respecto al genoma completo. Se generaron redes génicas usando el software GeneMANIA (<http://www.genemania.org/>). GeneMANIA usa interconexiones funcionales entre genes de datos publicados para generar una vista global de interacciones de genes. Se usó la herramienta MCODE (<http://apps.cytoscape.org/apps/mcode>) para identificar agrupaciones altamente interconectadas en una red.

35 *PCR cuantitativa a tiempo real.*

Se utilizó la tecnología SYBRgreen para el análisis de expresión génica de CXCL9, CXCL10, CD5L and BMP2 como control de calidad y para el control de pureza de los cultivos de CML. Se utilizaron ensayos PrimeTime para qPCR (IDT, Leuven, Belgium) y como constitutivos para la normalización se usó GAPDH. Los niveles de expresión génica se detectaron con el equipo ABI7500Fast (Life Technologies, Carlsbad, CA, USA). El análisis se realizó con el método ΔCt y se expresó como $2(-\Delta Ct)$. La significancia se calculó como se ha descrito anteriormente usando el software GraphPad Prism 5.

Ejemplo 1: Enfoque experimental para distinguir CML (células de músculo liso) aisladas de placas carotídeas sintomáticas y asintomáticas mediante tecnología de RNAseq

10

El estudio se diseñó con potencia máxima para detectar DEG (genes expresados diferencialmente) y DEI (isoformas expresadas diferencialmente) así como para realizar análisis de redes y enriquecimiento funcional complementario. Los dos factores principales a tener en cuenta en los estudios de transcriptómica mediante RNAseq son el número de repeticiones y la profundidad de la secuenciación, de modo que se realizó secuenciación de ARN con una profundidad suficiente para alcanzar una potencia de detección fiable de expresión diferencial para obtener aproximadamente 60 millones de lecturas por muestra. La secuenciación a esta profundidad permite identificar expresión diferencial con potencia máxima en 5 repeticiones, incluso en aquellos genes con baja expresión. Permite además reducir la variabilidad en el análisis de RNAseq cuando se estudia un tipo de célula específica (es decir, CML) en comparación con el estudio de toda la placa ateromatosa, que es heterogénea y está compuesta por una proporción variable de células endoteliales, CML y macrófagos. Por consiguiente, para estudiar el papel de las CML en el desarrollo de la placas ateroscleróticas inestables, se seleccionaron 14 pacientes (7 sintomáticos y 7 asintomáticos) que estaban propuestos para endarterectomía carotídea por el comité de expertos médicos del Hospital Universitario de Basurto, que evaluaron los datos clínicos de los pacientes (tabla 2). El servicio de radiología realizó IRM y ecografía Doppler cervical en todos los pacientes en el hospital. Se incluyeron en el estudio los pacientes con >70% de estenosis y diagnosticados de acontecimiento cerebrovascular y los pacientes con >80% estenosis y sin acontecimientos. Otros parámetros requeridos aparte de ser sintomáticos (S) y asintomáticos (A) para la selección de pacientes elegibles que iban a incluirse en este estudio se definieron como: (1) edad (70 años \pm 12); (2) sin otros estados médicos y (3) sin endarterectomía contralateral previa. Durante el periodo de inclusión se excluyeron pacientes del estudio porque presentaban al menos uno de los problemas siguientes: (1) pacientes que mostraban signo de demencia según la prueba del mini-examen del estado mental (MMSE); (2) pacientes con alteraciones psicológicas debidas a fármacos o

35

intervención de cirugía cerebral; (3) pacientes que presentaban complicaciones médicas o quirúrgicas que podían alterar la capacidad intelectual y/o la prueba de neuroimagen y (4) pacientes que no podían realizar la prueba neurofisiológica debido a la dificultad de comunicación (es decir, alteración de la visión).

5

Características de los pacientes	Todos	Sintomáticos	Asintomáticos
Número, n	14	7	7
Años	68	68±8	68±8
Sexo H/M, n	11/3	4/3	7/0
Tratamiento con estatinas	14	7	7
Factores de riesgo (%)			
Diabetes Mellitus	100	3 (43%)	4 (57%)
Dislipidemia	100	7 (100%)	7 (100%)
Hipertensión arterial	100	7 (100%)	7 (100%)
Tabaco	100	4 (57%)	3 (43%)

Tabla 2: Datos clínicos de los pacientes incluidos en el estudio.

10

Ejemplo 2: Las CML aisladas de placa aterosclerótica carotídea expresan biomarcadores típicos de CML

Es conocido en el estado de la técnica que el aislamiento y cultivo de CML vasculares primarias se considera una estrategia aceptable para estudiar el papel de las CML en la aterosclerosis, que pueden agruparse como aislamiento por explante o aislamiento basado en digestión enzimática. La segunda opción generó una mayor cantidad de células aisladas. En particular, para aislar las CML de la placa ateromatosa carotídea se siguió un método enzimático descrito anteriormente con algunas modificaciones (Bonano et al, 2000. *Cytometry* 39, 158-165). Las muestras de tejido de placa de ateroma se seccionaron en fragmentos de 2,3 mm y posteriormente se digirieron con colagenasa I. Primero se realizó una digestión durante 3 horas con 300U/ml de colagenasa I. Después, el tejido digerido se sometió a una segunda digestión durante 16 horas con 200 U/ml de colagenasa I. El tejido digerido es filtrado con filtros de 100 µm para eliminar el material no digerido. Las células son cultivadas en medio específico M231 suplementado con factores de crecimiento. Para determinar la eficacia del aislamiento y la pureza del cultivo de CML primarias, se realizaron

25

dos pruebas de control de calidad. En primer lugar, se detectaron marcadores específicos para células del músculo liso mediante inmunotinción usando anticuerpos específicos para CML (es decir α -MYH11) (Ross, R. *J. Cell Biol.* 50, 172–186 (1971) y células endoteliales (PECAM1- como control negativo), lo que permitió examinar la homogeneidad de los cultivos de CML primarias. También se detectó que en los cultivos de las CML no había contaminación con macrófagos mediante la identificación de marcadores específicos de macrófagos (i.e. CXCL9, CXCL10 y CD5L). En resumen, se sembraron las células en placas de 24 pocillos con cubreobjetos de vidrio y se tiñeron las células con anticuerpos α -MYH11. La adquisición de imágenes se realizó usando un equipo de microscopio Leica TCS STED CW SP8. Todos los cultivos de CML usados en el estudio fueron positivos al 100% para MYH11. La figura 1A muestra la cuantificación de las células positivas para MYH11. En segundo lugar, se analizó la expresión intracelular de proteínas específicas de CML (MYH1 y, PECAM1). El análisis de inmunotransferencia de tipo Western mostró que las CML aisladas expresaban los biomarcadores sometidos a prueba de CML específicas y que eran negativas para la expresión de PECAM1. La figura 1B muestra una ilustración de la detección mediante inmunotransferencia de tipo Western con anticuerpos α -PECAM1 y α -MYH11 en lisados celulares extraídos de cultivos de CML. Para el análisis de control de calidad se usaron cultivos celulares en fase 0-2.

20 *Ejemplo 3: Análisis de DEG y DEI con datos de transcriptómica*

Para evitar posibles cambios en la transcripción debido al número de pases, el ARN usado para el análisis RNAseq se extrajo en todos los casos de células del músculo liso en el pase 0. Se evaluó la integridad del ARN (RIN) para garantizar la buena calidad del ARN. Todas las muestras de ARN mostraron un valor de RIN de entre 9 y 10, lo que las calificaba como muestras adecuadas para el análisis de RNAseq. Se realizó un análisis transcriptómico comparativo entre los dos grupos, A y S, mediante la aplicación de la tecnología de RNAseq para entender el papel de las CML en la sintomatología de la aterosclerosis carotídea. El número de repeticiones por grupo (n=7) se decidió por la disponibilidad máxima de tejido obtenido en un plazo específico teniendo en cuenta el mínimo de repeticiones recomendado para este tipo de estudios. Se secuenciaron las 14 bibliotecas generadas en la plataforma HiSeq2500 de Illumina. Los datos obtenidos con la secuenciación de ARN se convirtieron en niveles cuantitativos de expresión génica con la ayuda de la identificación de genes expresados diferencialmente (DEG) e isoformas expresadas diferencialmente (DEI) a través de la comparación de A y S. Se obtuvieron entre el 85 y el 87 % de lecturas mapeadas de las lecturas generadas totales (tabla 3), lo que estaba dentro del intervalo esperado de

lecturas mapeadas en el genoma humano (70-90%) Risso, D. et al., BMC Bioinformatics 12, (2011).

ID de muestra	Lecturas totales	% de lecturas HQ	% de lecturas de pares correctos	% de lecturas de empalme
1	61449044	86,11	72,15	26,38
2	76876864	86,51	71,27	25,33
3	67990002	85,99	70,8	25,14
4	67892558	86,08	70,74	26,91
5	71976326	86,78	72,31	26,78
6	74497186	82,21	72,82	23,61
7	72479266	87,66	71,85	27,01
8	70654274	84,75	72,09	24,79
9	73679062	87,23	71,45	27,43
10	70845658	86,19	71,21	25,85
11	72278906	86,59	72,24	25,42
12	72592062	86,26	70,91	27,36
13	80682878	85,84	73,04	26,56
14	72040720	86,47	70,94	26,49

5 Tabla 3. Resumen de datos de mapeo de RNAseq. HQ, alta calidad.

Los datos de RNAseq tenían que normalizarse para eliminar cualquier desviación estadística que pudiera alterar el análisis posterior. Las principales desviaciones importantes encontradas fueron la longitud de los genes, el tamaño de la biblioteca y el contenido en GC.

10 Para evitar tales desviaciones se realizó normalización usando el método propuesto de DESeq2 (Hansen, K.D., et al *Biostatistics* 13, 204–216 2012). Se realizó el análisis de componentes principales (PCA) en los perfiles de expresión génica usando el método publicado anteriormente (Stacklies, W., et al, *Bioinforma.* 23, 1164–1167 (2007) para identificar posibles valores atípicos. El PCA obtiene 2 valores atípicos diferenciados (A7 y

15 S1), lo que aumentaría la variabilidad y disminuiría la potencia de nuestros datos estadísticos, y por tanto se eliminaron del análisis posterior. El ajuste de la dispersión ilustra cuánto se desvía la varianza de la media y esto se usó para reducir las estimaciones finales desde las estimaciones en relación con los genes hacia las estimaciones ajustadas, lo que mejora el análisis de la expresión génica. En el análisis de expresión diferencial se

20 identificaron 93 genes dentro del análisis de DEG. En la Tabla 4 puede encontrarse una lista completa de los 93 genes que muestra valores de FC $\geq 1,5$ con un valor de $P \leq 0,05$ (67 con

valor de P (P_{ajust}) corregido por $\leq 0,05$ (Benjamini, Y. *J Roy Stat Soc B* 57, 1995)). Entre los genes incluidos en la Tabla 4, los genes están asociados con: crecimiento y senescencia celular, metabolismo celular, metabolismo de retinol; enfermedad vascular, autofagia, sistema inmunitario, desarrollo / contracción muscular.

5

ID del gen	Símbolo	FC	P_{ajust}
ENSG00000165507	C10orf10	1,94	0,0006
ENSG00000152049	KCNE4	1,87	0,003
ENSG00000117069	ST6GALNAC5	1,83	0,009
ENSG00000135424	ITGA7	1,81	0,005
ENSG00000128285	MCHR1	1,76	0,02
ENSG00000174564	IL20RB	1,75	0,02
ENSG00000163815	CLEC3B	1,75	0,02
ENSG00000075643	MOCOS	1,75	0,001
ENSG00000074410	CA12	1,72	0,0006
ENSG00000174807	CD248	1,71	0,03
ENSG00000133454	MYO18B	1,70	0,04
ENSG00000179542	SLITRK4	1,65	0,06*
ENSG00000161570	CCL5	1,65	0,06*
ENSG00000148948	LRRC4C	1,65	0,06*
ENSG00000117289	TXNIP	1,64	0,04
ENSG00000123689	GOS2	1,63	0,05
ENSG00000162706	CADM3	1,63	0,07*
ENSG00000120088	CRHR1	1,62	0,07*
ENSG00000184557	SOCS3	1,61	0,03
ENSG00000176438	SYNE3	1,60	0,05
ENSG00000161958	FGF11	1,60	0,04
ENSG00000234745	HLA-B	1,59	0,06*
ENSG00000164236	ANKRD33B	1,59	0,09*
ENSG00000170624	SGCD	1,57	0,08*
ENSG00000259649	RP11-351M8,1	1,57	0,09*
ENSG00000172201	ID4	1,55	0,09*
ENSG00000119938	PPP1R3C	1,54	0,09*
ENSG00000114268	PFKFB4	1,54	0,04

ES 2 688 737 A1

ID del gen	Símbolo	FC	P_{ajust}
ENSG00000110076	NRXN2	1,53	0,05
ENSG00000136378	ADAMTS7	1,52	0,02
ENSG00000152256	PDK1	1,52	0,03
ENSG00000112837	TBX18	1,51	0,09*
ENSG00000162733	DDR2	1,50	0,02
ENSG00000179348	GATA2	-1,51	0,09*
ENSG00000120693	SMAD9	-1,52	0,05
ENSG00000164736	SOX17	-1,52	0,09*
ENSG00000035664	DAPK2	-1,53	0,09*
ENSG00000144730	IL17RD	-1,54	0,01
ENSG00000133216	EPHB2	-1,55	0,03
ENSG00000166960	CCDC178	-1,55	0,09*
ENSG00000120279	MYCT1	-1,57	0,05
ENSG00000121039	RDH10	-1,57	0,06*
ENSG00000183578	TNFAIP8L3	-1,58	0,09*
ENSG00000145681	HAPLN1	-1,58	0,09*
ENSG00000158125	XDH	-1,59	0,09*
ENSG00000139174	PRICKLE1	-1,59	0,08*
ENSG00000182368	FAM27A	-1,59	0,09*
ENSG00000118137	APOA1	-1,60	0,07*
ENSG00000169684	CHRNA5	-1,60	0,09*
ENSG00000167644	C19orf33	-1,61	0,06*
ENSG00000116106	EPHA4	-1,61	0,05
ENSG00000125968	ID1	-1,61	0,04
ENSG00000123572	NRK	-1,61	0,09*
ENSG00000121361	KCNJ8	-1,63	0,07*
ENSG00000091129	NRCAM	-1,63	0,07*
ENSG00000175746	C15orf54	-1,64	0,05
ENSG00000172602	RND1	-1,64	0,04
ENSG00000095752	IL11	-1,64	0,07*
ENSG00000175866	BAIAP2	-1,65	0,01
ENSG00000127329	PTPRB	-1,65	0,06*
ENSG00000144366	GULP1	-1,66	0,005

ES 2 688 737 A1

ID del gen	Símbolo	FC	P_{ajust}
ENSG00000143140	GJA5	-1,66	0,04
ENSG00000176399	DMRTA1	-1,68	0,05
ENSG00000168497	SDPR	-1,68	0,02
ENSG00000010319	SEMA3G	-1,69	0,02
ENSG00000254656	RTL1	-1,69	0,03
ENSG00000178445	GLDC	-1,69	0,02
ENSG00000162496	DHRS3	-1,70	0,04
ENSG00000102554	KLF5	-1,70	0,003
ENSG00000172159	FRMD3	-1,70	0,04
ENSG00000164283	ESM1	-1,71	0,03
ENSG00000235109	ZSCAN31	-1,72	0,01
ENSG00000132321	IQCA1	-1,72	0,02
ENSG00000008394	MGST1	-1,72	0,002
ENSG00000164530	PI16	-1,73	0,01
ENSG00000150722	PPP1R1C	-1,73	0,03
ENSG00000116741	RGS2	-1,74	0,02
ENSG00000180660	MAB21L1	-1,76	0,02
ENSG00000155011	DKK2	-1,77	0,01
ENSG00000126950	TMEM35	-1,81	0,02
ENSG00000179104	TMTC2	-1,81	0,004
ENSG00000118777	ABCG2	-1,85	0,01
ENSG00000125848	FLRT3	-1,87	0,007
ENSG00000108375	RNF43	-1,90	0,003
ENSG00000125845	BMP2	-1,90	0,005
ENSG00000138759	FRAS1	-1,90	0,005
ENSG00000108556	CHRNE	-1,91	0,005
ENSG00000163273	NPPC	-1,93	0,003
ENSG00000162595	DIRAS3	-1,93	0,003
ENSG00000109511	ANXA10	-1,93	0,003
ENSG00000154645	CHODL	-2,16	0,0003
ENSG00000138772	ANXA3	-2,23	0,0001
ENSG00000087494	PTHLH	-2,50	1,04E-07

Tabla 4. Expresión génica diferencial entre S (sintomáticos) y A (asintomáticos). El valor de tasa de cambio (FC) se obtiene mediante el método de DeSeq2 y el valor de p ajustado (Pajust) se calculó usando el método de la tasa de falso descubrimiento (FDR). Un valor de Pajust significativo se consideró $P_{ajust} \leq 0,05$. *Genes que muestran una tendencia hacia la significación pero sin alcanzarla ($P_{ajust} \leq 0,09$).

5

La cuantificación exacta de isoformas sigue siendo un reto debido al alto grado de solapamiento entre transcritos. La variabilidad en la precisión de la estimación de la expresión de isoformas a través de las muestras se sesgaría usando paquetes de expresión diferencial tradicionales. Por este motivo, en este caso se utilizó el enfoque bayesiano de Cufflink, que estima la abundancia de transcritos basándose en cuántas lecturas soporta cada transcrito. El análisis de DEI identificó 143 isoformas expresadas diferencialmente entre S y A. La tabla 5 ilustra las isoformas con un valor de FC menor de -1,5 o mayor de 1,5 y con $P_{ajust} \leq 0,05$.

15

ID isoformas	ID gen	Símbolo	FC	P_{valor}	$P_{ajust.}$
ENST00000358321	ENSG00000099994	SUSD2	11,61	5,00E-05	0,01
ENST00000260227	ENSG00000137673	MMP7	9,25	5,00E-05	0,01
ENST00000506608	ENSG00000185149	NPY2R	8,39	0,0002	0,04
ENST00000264613	ENSG00000047457	CP	7,90	0,0002	0,04
ENST00000374975	ENSG00000198502	HLA-DRB5	5,72	5,00E-05	0,01
ENST00000477717	ENSG00000117069	ST6GALNAC5	4,84	5,00E-05	0,01
ENST00000075322	ENSG00000136960	ENPP2	4,59	5,00E-05	0,01
ENST00000296130	ENSG00000163815	CLEC3B	4,37	5,00E-05	0,01
ENST00000394887	ENSG00000196616	ADH1B	4,07	5,00E-05	0,01
ENST00000259486	ENSG00000136960	ENPP2	3,88	5,00E-05	0,01
ENST00000013222	ENSG00000241644	INMT	3,84	5,00E-05	0,01
ENST00000393078	ENSG00000196136	SERPINA3	3,57	5,00E-05	0,01
ENST00000368125	ENSG00000162706	CADM3	3,53	0,00015	0,03
ENST00000296506	ENSG00000164106	SCRG1	3,43	5,00E-05	0,01
ENST00000256861	ENSG00000123243	ITIH5	3,42	0,00015	0,03
ENST00000294728	ENSG00000162692	VCAM1	3,38	5,00E-05	0,01
ENST00000274938	ENSG00000146197	SCUBE3	3,30	5,00E-05	0,01
ENST00000167586	ENSG00000186847	KRT14	3,06	5,00E-05	0,01
ENST00000374737	ENSG00000155659	VSIG4	3	0,00015	0,03
ENST00000372330	ENSG00000100985	MMP9	2,82	5,00E-05	0,01
ENST00000447544	ENSG00000135218	CD36	2,65	5,00E-05	0,01
ENST00000298295	ENSG00000165507	C10orf10	2,50	5,00E-05	0,01
ENST00000604125	ENSG00000152049	KCNE4	2,48	5,00E-05	0,01
ENST00000542220	ENSG00000159403	C1R	2,42	5,00E-05	0,01

ES 2 688 737 A1

ENST00000553804	ENSG00000135424	ITGA7	2,38	5,00E-05	0,01
ENST00000367662	ENSG00000116183	PAPPA2	2,34	0,00015	0,03
ENST00000006053	ENSG00000006210	CX3CL1	2,29	5,00E-05	0,01
ENST00000361673	ENSG00000198796	ALPK2	2,20	5,00E-05	0,01
ENST00000437043	ENSG00000164308	ERAP2	2,18	0,0001	0,02
ENST00000339732	ENSG00000131386	GALNT15	2,15	0,0002	0,04
ENST00000284322	ENSG00000154175	ABI3BP	2,12	5,00E-05	0,01
ENST00000261799	ENSG00000113721	PDGFRB	2,10	5,00E-05	0,01
ENST00000369317	ENSG00000117289	TXNIP	2,07	5,00E-05	0,01
ENST00000435422	ENSG00000170624	SGCD	2,06	5,00E-05	0,01
ENST00000345517	ENSG00000163017	ACTG2	2	0,00015	0,03
ENST00000257435	ENSG00000134986	NREP	1,99	5,00E-05	0,01
ENST00000429713	ENSG00000172403	SYNPO2	1,99	5,00E-05	0,01
ENST00000412585	ENSG00000234745	HLA-B	1,98	5,00E-05	0,01
ENST00000307365	ENSG00000168209	DDIT4	1,96	5,00E-05	0,01
ENST00000261326	ENSG00000075643	MOCOS	1,95	5,00E-05	0,01
ENST00000178638	ENSG00000074410	CA12	1,94	5,00E-05	0,01
ENST00000254722	ENSG00000132386	SERPINF1	1,93	5,00E-05	0,01
ENST00000335174	ENSG00000186352	ANKRD37	1,90	0,0002	0,04
ENST00000378700	ENSG00000172201	ID4	1,89	5,00E-05	0,01
ENST00000294507	ENSG00000162511	LAPTM5	1,89	5,00E-05	0,01
ENST00000540010	ENSG00000056558	TRAF1	1,89	5,00E-05	0,01
ENST00000330871	ENSG00000184557	SOCS3	1,88	5,00E-05	0,01
ENST00000344366	ENSG00000074410	CA12	1,87	5,00E-05	0,01
ENST00000238994	ENSG00000119938	PPP1R3C	1,86	5,00E-05	0,01
ENST00000217233	ENSG00000101255	TRIB3	1,84	5,00E-05	0,01
ENST00000327299	ENSG00000162433	AK4	1,83	0,0002	0,04
ENST00000328916	ENSG00000182326	C1S	1,79	0,0001	0,02
ENST00000323076	ENSG00000136167	LCP1	1,76	5,00E-05	0,01
ENST00000157600	ENSG00000071282	LMCD1	1,75	0,0001	0,02
ENST00000297848	ENSG00000187955	COL14A1	1,73	5,00E-05	0,01
ENST00000369663	ENSG00000112837	TBX18	1,73	5,00E-05	0,01
ENST00000367288	ENSG00000163431	LMOD1	1,72	5,00E-05	0,01
ENST00000338981	ENSG00000114374	USP9Y	1,70	5,00E-05	0,01
ENST00000566620	ENSG00000140945	CDH13	1,68	5,00E-05	0,01
ENST00000394308	ENSG00000070669	ASNS	1,65	0,0001	0,02
ENST00000402744	ENSG00000109472	CPE	1,65	5,00E-05	0,01
ENST00000367922	ENSG00000162733	DDR2	1,64	5,00E-05	0,01
ENST00000220244	ENSG00000103888	KIAA1199	1,62	0,0002	0,04
ENST00000380250	ENSG00000073910	FRY	1,58	0,00015	0,03
ENST00000409652	ENSG00000221963	APOL6	1,56	0,00015	0,03
ENST00000396680	ENSG00000128272	ATF4	1,55	0,00015	0,03
ENST00000372764	ENSG00000122861	PLAU	1,55	0,0001	0,02

ES 2 688 737 A1

ENST00000375377	ENSG00000165757	KIAA1462	1,54	0,00015	0,03
ENST00000263734	ENSG00000116016	EPAS1	1,53	5,00E-05	0,01
ENST00000295992	ENSG00000163710	PCOLCE2	1,53	5,00E-05	0,01
ENST00000376588	ENSG00000135069	PSAT1	1,53	0,0002	0,04
ENST00000274276	ENSG00000145623	OSMR	1,50	0,0002	0,04
ENST00000358432	ENSG00000142627	EPHA2	-1,50	5,00E-05	0,01
ENST00000260630	ENSG00000138061	CYP1B1	-1,53	0,00015	0,03
ENST00000229416	ENSG00000001084	GCLC	-1,56	0,0002	0,04
ENST00000367466	ENSG00000116711	PLA2G4A	-1,56	0,0001	0,02
ENST00000005257	ENSG00000006451	RALA	-1,56	0,0001	0,02
ENST00000245185	ENSG00000125148	MT2A	-1,57	0,0001	0,02
ENST00000361441	ENSG00000244509	APOBEC3C	-1,58	0,0001	0,02
ENST00000381434	ENSG00000137033	IL33	-1,62	0,0001	0,02
ENST00000272542	ENSG00000144136	SLC20A1	-1,63	5,00E-05	0,01
ENST00000371697	ENSG00000148677	ANKRD1	-1,64	0,00015	0,03
ENST00000248598	ENSG00000127951	FGL2	-1,67	5,00E-05	0,01
ENST00000315711	ENSG00000091972	CD200	-1,68	0,0002	0,04
ENST00000222543	ENSG00000105825	TFPI2	-1,69	5,00E-05	0,01
ENST00000237305	ENSG00000118515	SGK1	-1,72	5,00E-05	0,01
ENST00000367468	ENSG00000073756	PTGS2	-1,77	5,00E-05	0,01
ENST00000258888	ENSG00000136383	ALPK3	-1,78	5,00E-05	0,01
ENST00000426693	ENSG00000164619	BMPER	-1,81	0,0002	0,04
ENST00000379375	ENSG00000078401	EDN1	-1,86	5,00E-05	0,01
ENST00000276925	ENSG00000147883	CDKN2B	-1,91	5,00E-05	0,01
ENST00000377861	ENSG00000184226	PCDH9	-1,91	5,00E-05	0,01
ENST00000265080	ENSG00000113319	RASGRF2	-1,91	5,00E-05	0,01
ENST00000377103	ENSG00000178726	THBD	-1,93	5,00E-05	0,01
ENST00000376112	ENSG00000125968	ID1	-1,99	5,00E-05	0,01
ENST00000252593	ENSG00000130303	BST2	-2,03	5,00E-05	0,01
ENST00000396210	ENSG00000008394	MGST1	-2,04	5,00E-05	0,01
ENST00000330688	ENSG00000077092	RARB	-2,04	5,00E-05	0,01
ENST00000310613	ENSG00000173597	SULT1B1	-2,09	0,0001	0,02
ENST00000327536	ENSG00000183578	TNFAIP8L3	-2,09	5,00E-05	0,01
ENST00000274341	ENSG00000145681	HAPLN1	-2,10	5,00E-05	0,01
ENST00000304141	ENSG00000168497	SDPR	-2,10	5,00E-05	0,01
ENST00000262259	ENSG00000105497	ZNF175	-2,15	5,00E-05	0,01
ENST00000306773	ENSG00000171246	NPTX1	-2,21	5,00E-05	0,01
ENST00000222482	ENSG00000128510	CPA4	-2,27	5,00E-05	0,01
ENST00000377474	ENSG00000178695	KCTD12	-2,35	5,00E-05	0,01
ENST00000321196	ENSG00000179104	TMTC2	-2,35	5,00E-05	0,01
ENST00000315274	ENSG00000196611	MMP1	-2,40	0,00015	0,03
ENST00000444842	ENSG00000082482	KCNK2	-2,43	5,00E-05	0,01
ENST00000286614	ENSG00000156103	MMP16	-2,43	5,00E-05	0,01

ENST00000376468	ENSG00000120937	NPPB	-2,43	5,00E-05	0,01
ENST00000235382	ENSG00000116741	RGS2	-2,45	5,00E-05	0,01
ENST00000546561	ENSG00000127324	TSPAN8	-2,54	5,00E-05	0,01
ENST00000541194	ENSG00000128567	PODXL	-2,59	5,00E-05	0,01
ENST00000247829	ENSG00000127324	TSPAN8	-2,59	5,00E-05	0,01
ENST00000232892	ENSG00000114771	AADAC	-2,69	5,00E-05	0,01
ENST00000297785	ENSG00000165092	ALDH1A1	-2,72	5,00E-05	0,01
ENST00000591504	ENSG00000166510	CCDC68	-2,74	5,00E-05	0,01
ENST00000307407	ENSG00000169429	IL8	-2,82	5,00E-05	0,01
ENST00000284136	ENSG00000153993	SEMA3D	-2,85	5,00E-05	0,01
ENST00000537928	ENSG00000128567	PODXL	-2,86	0,0002	0,04
ENST00000376223	ENSG00000162496	DHRS3	-2,87	5,00E-05	0,01
ENST00000304195	ENSG00000172159	FRMD3	-3	0,00015	0,03
ENST00000359299	ENSG00000109511	ANXA10	-3,13	5,00E-05	0,01
ENST00000355754	ENSG00000162654	GBP4	-3,13	5,00E-05	0,01
ENST00000372930	ENSG00000126950	TMEM35	-3,22	5,00E-05	0,01
ENST00000296575	ENSG00000164161	HHIP	-3,48	5,00E-05	0,01
ENST00000326245	ENSG00000179914	ITLN1	-3,49	0,0001	0,02
ENST00000299502	ENSG00000197632	SERPINB2	-3,50	5,00E-05	0,01
ENST00000261292	ENSG00000101670	LIPG	-3,65	5,00E-05	0,01
ENST00000237612	ENSG00000118777	ABCG2	-3,81	5,00E-05	0,01
ENST00000397463	ENSG00000150551	LYPD1	-3,98	5,00E-05	0,01
ENST00000381405	ENSG00000164283	ESM1	-4,14	5,00E-05	0,01
ENST00000295461	ENSG00000163293	NIPAL1	-4,22	5,00E-05	0,01
ENST00000261405	ENSG00000110799	VWF	-4,25	5,00E-05	0,01
ENST00000378827	ENSG00000125845	BMP2	-4,27	5,00E-05	0,01
ENST00000263174	ENSG00000099260	PALMD	-4,36	5,00E-05	0,01
ENST00000395201	ENSG00000162595	DIRAS3	-4,39	5,00E-05	0,01
ENST00000271348	ENSG00000143140	GJA5	-4,86	5,00E-05	0,01
ENST00000409852	ENSG00000163273	NPPC	-4,93	5,00E-05	0,01
ENST00000282849	ENSG00000140873	ADAMTS18	-6,48	5,00E-05	0,01
ENST00000297316	ENSG00000164736	SOX17	-6,48	5,00E-05	0,01
ENST00000373674	ENSG00000164530	PI16	-8,28	5,00E-05	0,01

Tabla 5: Análisis de la expresión diferencial de isoformas entre pacientes sintomáticos y asintomáticos con usando el método Cufflinks. El valor FC se calculó por el método DeSeq1 y Padj fue calculado aplicando FDR. El valor Padj significativo considerado fue de

5 $P_{ajust} \leq 0,05$.

Ejemplo 4: Análisis de redes y enriquecimiento funcional

A continuación se realizó un análisis de redes génicas y funcionales, por medio de un análisis de enriquecimiento que permite la identificación de combinaciones de anotaciones significativas asociadas con las listas generadas de DEG y DEI para identificar relaciones funcionales entre las muestras sintomáticas y asintomáticas. El análisis de enriquecimiento basado en bioconductores usando como bases de datos GO (Gene Ontology (ontología génica)), KEGG (Kyoto Encyclopedia of Genes and Genome (enciclopedia de genes y genomas de Kyoto)); reactoma, DO (disease ontology (ontología de enfermedades), OMIM (Online Mendelian Inheritance in Man (herencia mendeliana en el hombre en línea) y el Cincinnati Children's Hospital Medical centre (Centro Médico del Hospital Infantil de Cincinnati) (Consortium, G. O. *Nucleic Acids Res.* 43, D1049–D1056 (2015); Kanehisa, M. & Goto, S. *Nucleic Acids Res.* 28, 27–30 (2000); Fabregat, A. et al. *Nucleic Acids Res.* 44, D481–D487 (2016); Yu, G. et al. *Bioinformatics* 31, 608–609 (2015)., generó una lista de grupos de anotación que son significativamente diferentes entre S y A.. Aunque el análisis basado en bioconductores permitió identificar categorías funcionales comunes para una agrupación de genes, a continuación se comparan estas categorías entre las agrupaciones de genes mediante implementación del algoritmo propuesto por la herramienta ClusterProfiler. Esta herramienta, que se usó para la comparación de agrupaciones de genes (es decir genes regulados por disminución y regulados por incremento en S frente a A), permitió visualizar las categorías biológicas comparadas entre agrupaciones de genes. Los perfiles funcionales identificados basados en KEGG, GO, Reactome y DO se ilustran en la figura 2, la figura 3. La comparación basada en KEGG (figura 2) identificó la ruta de señalización de TGF- β , la guía axonal y las moléculas de adhesión celular (CAM) como las categorías con mayor número de GeneRatio y menor valor de P_{ajust} . Se identificó la actividad de la retinol deshidrogenasa en la comparación basada en GO como la categoría con menor P_{ajust} (figura 3). Además, se realizó el análisis de redes de interacciones funcionales usando varias fuentes de información tales como interacción de proteínas, genómica, rutas, etc. usando el programa denominado GeneMANIA (Gene Multiple Association Network Integration Algorithm) (Montejo, J., et al., *F1000Research* 3, 2014) La figura 4 muestra la red funcional de genes obtenida mediante GeneMANIA incluyendo agrupaciones relacionadas con la señalización de TGF- β , la guía axonal, las moléculas de adhesión celular y la señalización de ALK1.

Posteriormente, se realizó un análisis de redes adicional para identificar una red central a partir de la red funcional generada con GeneMANIA con el método MCODE (Molecular Complex Detección) (Saito, R. et al. *Nat Meth* 9, 1069–1076 2012). MCODE realiza una etapa de agrupación de redes, que reduce los núdulos no conectados en las agrupaciones

de red. Por tanto, la agrupación final identificada por MCODE fue aquella con mayor especificidad y mostró que las rutas de señalización de TGF- β , BMP y ALK1 surgen como un proceso funcional, que distinguen CML aisladas para placas carotídeas sintomáticas o asintomáticas. También es notable la identificación en este análisis de agrupación de redes

5 de 3 genes/isoformas regulados por disminución [*BMP2* (DEG: FC -1,9; P_{ajust} 0,004 / DEI: FC-4,27; P_{ajust} 0,01) / *SMAD9* (DEG: FC -1,51; P_{ajust} 0,05) / *ID1* (DEG: FC -1,61; P_{ajust} 0,03 / DEI: FC-1,99; P_{ajust} 0,01)] y uno regulado por incremento [*ID4* (DEG: FC 155; P_{ajust} 0,09 / DEI: FC 1,89; P_{ajust} 0,01)] en CML asociadas con desestabilización de la placa (figura 4).

REIVINDICACIONES

- 1 - Un método para diagnosticar placas ateroscleróticas inestables en un sujeto que
5 comprende:
- (i) determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en una muestra biológica de dicho sujeto, y
 - (ii) comparar el nivel de expresión obtenido en (i) con un valor de referencia correspondiente,
- 10 en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 1-33 de la Tabla 1 y su expresión está incrementada respecto al valor de referencia y/o en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 34-93 de la Tabla 1 y su expresión se encuentra reducida respecto al valor de referencia, entonces el sujeto es diagnosticado con placa aterosclerótica inestable.
- 15
- 2- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 1 que comprende determinar el nivel de expresión de al menos 1 gen del grupo 1, al menos 1 gen del grupo 2, al menos 1 gen del grupo 3, al menos 1 gen del grupo 4, al menos 1 gen del grupo 5 y/o al menos 1 gen del grupo 6.
- 20
- 3- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 2 que comprende determinar el nivel de expresión de todos los genes del grupo 1.
- 4- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 2 que
25 comprende determinar el nivel de expresión de todos los genes del grupo 2.
- 5- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 2 que comprende determinar el nivel de expresión de todos los genes del grupo 3.
- 30
- 6- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 2 que comprende determinar el nivel de expresión de todos los genes del grupo 4.
- 7- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 2 que
35 comprende determinar el nivel de expresión de todos los genes del grupo 5.

- 8- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 2 que comprende determinar el nivel de expresión de todos los genes del grupo 6.
- 5 9- Método para diagnosticar placas ateroscleróticas inestables según cualquiera de las reivindicaciones 2 a 8 que comprende determinar el nivel de expresión de todos los genes del grupo 1, de todos los genes del grupo 2, de todos los genes del grupo 3, de todos los genes del grupo 4, de todos los genes del grupo 5 y de todos los genes del grupo 6.
- 10 10- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 1 que comprende determinar el nivel de expresión de todos los genes de la Tabla 1.
- 11- Método según cualquiera de las reivindicaciones 1 a 10 en donde la muestra biológica es una muestra que contiene células de una placa aterosclerótica.
- 15 12- Método según cualquiera de las reivindicaciones 1 a 11 en donde las células de una placa aterosclerótica son células de músculo liso.
- 13- Método según la reivindicación 12 en donde las células de músculo liso son positivas para MYH11 y negativas para PECAM1.
- 20 14- Método según cualquiera de las reivindicaciones 1-13 en donde el valor de referencia de un gen de la Tabla 1 corresponde al nivel de expresión de dicho gen en una muestra de células de una placa aterosclerótica de un sujeto que tiene placas ateroscleróticas estables.
- 25 15- Método según cualquiera de las reivindicaciones 1-14 en donde el nivel de expresión se determina mediante técnicas de secuenciación de ARN.
- 16- Un método de pronóstico para determinar la probabilidad de un sujeto de sufrir una enfermedad cerebrovascular que comprende
- 30 (i) determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en una muestra biológica de dicho sujeto, y
(ii) comparar el nivel de expresión obtenido en (i) con un valor de referencia, en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 1-33 de la Tabla 1 y su expresión está incrementada respecto al valor de referencia y/o

en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 34-93 de la Tabla 1 y su expresión se encuentra reducida respecto al valor de referencia, entonces el sujeto tiene una alta probabilidad de sufrir una enfermedad cerebrovascular.

- 5 17- Método según la reivindicación 16 que comprende determinar el nivel de expresión de al menos 1 gen del grupo 1, al menos 1 gen del grupo 2, al menos 1 gen del grupo 3, al menos 1 gen del grupo 4, al menos 1 gen del grupo 5 y/o al menos 1 gen del grupo 6.
- 18- Método según la reivindicación 16 que comprende determinar el nivel de expresión de
10 todos los genes del grupo 1.
- 19- Método según la reivindicación 16 que comprende determinar el nivel de expresión de todos los genes del grupo 2.
- 15 20- Método según la reivindicación 16 que comprende determinar el nivel de expresión de todos los genes del grupo 3.
- 21- Método según la reivindicación 16 que comprende determinar el nivel de expresión de todos los genes del grupo 4.
20
- 22- Método según la reivindicación 16 que comprende determinar el nivel de expresión de todos los genes del grupo 5.
- 23- Método según la reivindicación 16 que comprende determinar el nivel de expresión de
25 todos los genes del grupo 6.
- 24- Método según cualquiera de las reivindicaciones 16 a 23 que comprende determinar el nivel de expresión de todos los genes del grupo 1, de todos los genes del grupo 2, de todos los genes del grupo 3, de todos los genes del grupo 4, de todos los genes del grupo 5 y de
30 todos los genes del grupo 6.
- 25-Método de pronóstico según la reivindicación 16 que comprende determinar el nivel de expresión de todos los genes de la Tabla 1.
- 35 26- Método según cualquiera de las reivindicaciones 16 a 25 en donde la muestra biológica es una muestra que contiene células de una placa aterosclerótica.

27- Método según cualquiera de las reivindicaciones 16 a 26 en donde las células de una placa aterosclerótica son células de músculo liso.

5 28- Método según la reivindicación 27 en donde las células de músculo liso son positivas para MYH11 y negativas para PECAM1.

29- Método según cualquiera de las reivindicaciones 16-28 en donde el valor de referencia de un gen de la Tabla 1 corresponde al nivel de expresión de dicho gen en una muestra de células de una placa aterosclerótica de un sujeto que tiene placas ateroscleróticas estables.

10 30- Método según cualquiera de las reivindicaciones 16-29 en donde el nivel de expresión se determina mediante técnicas de secuenciación de ARN.

15 31- Método según cualquiera de las reivindicaciones 16-30, en donde la enfermedad cerebrovascular es apoplejía.

32- Método para seleccionar un paciente susceptible de ser tratado con una terapia para la eliminación o estabilización de la placa aterosclerótica carotídea, que comprende:

20 (i) determinar el nivel de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 en una muestra biológica de dicho sujeto, y

(ii) comparar el nivel de expresión obtenido en (i) con un valor de referencia

en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 1-33 y su expresión está incrementada respecto al valor de referencia y/o

25 en donde si el al menos un gen determinado en la etapa (i) es uno de los genes 34-93 y su expresión se encuentra reducida respecto al valor de referencia, entonces el sujeto es susceptible de ser tratado para la eliminación o estabilización de la placa aterosclerótica carotídea.

30 33- Método según la reivindicación 32 en donde la terapia para la eliminación de placa aterosclerótica carotídea es endarterectomía carotídea o la terapia para la estabilización de la placa es estatinas o antiagregantes.

34- Método según cualquiera de las reivindicaciones 32 o 33 que comprende determinar el nivel de expresión de al menos 1 gen del grupo 1, al menos 1 gen del grupo 2, al menos 1

gen del grupo 3, al menos 1 gen del grupo 4, al menos 1 gen del grupo 5 y/o al menos 1 gen del grupo 6.

5 35- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 34 que comprende determinar el nivel de expresión de todos los genes del grupo 1.

36- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 34 que comprende determinar el nivel de expresión de todos los genes del grupo 2.

10 37- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 34 que comprende determinar el nivel de expresión de todos los genes del grupo 3.

38- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 34 que comprende determinar el nivel de expresión de todos los genes del grupo 4.

15

39- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 34 que comprende determinar el nivel de expresión de todos los genes del grupo 5.

20 40- Método para diagnosticar placas ateroscleróticas inestables según la reivindicación 34 que comprende determinar el nivel de expresión de todos los genes del grupo 6.

41- Método para diagnosticar placas ateroscleróticas inestables según cualquiera de las reivindicaciones 34 a 40 que comprende determinar el nivel de expresión de todos los genes del grupo 1, de todos los genes del grupo 2, de todos los genes del grupo 3, de todos los genes del grupo 4, de todos los genes del grupo 5 y de todos los genes del grupo 6.

25

42-Método según cualquiera de las reivindicaciones 33 o 34 que comprende determinar el nivel de expresión de todos los genes de la Tabla 1.

30 43-Método según cualquiera de las reivindicaciones 33 a 42 en donde la muestra biológica es una muestra que contiene células de una placa aterosclerótica.

44- Método según la reivindicación 43 en donde las células de una placa aterosclerótica son células de músculo liso.

35

45- Método según la reivindicación 44 en donde las células de músculo liso son positivas para MYH11 y negativas para PECAM1.

5 46- Método según cualquiera de las reivindicaciones 33-45 en donde el valor de referencia de un gen de la Tabla 1 corresponde al nivel de expresión de dicho gen en una muestra de células de una placa aterosclerótica de un sujeto que tiene placas ateroscleróticas estables.

47- Método según cualquiera de las reivindicaciones 33-46 en donde el nivel de expresión se determina mediante técnicas de secuenciación de ARN.

10

48- Un kit que comprende reactivos adecuados para la determinación de los niveles de expresión de al menos 1 gen seleccionado del grupo de genes mostrados en la Tabla 1 y, opcionalmente, reactivos para la determinación de los niveles de expresión de uno o más genes de mantenimiento.

15

49- Kit según la reivindicación 48 que comprende reactivos adecuados para la determinación del nivel de expresión de al menos 1 gen del grupo 1, al menos 1 gen del grupo 2, al menos 1 gen del grupo 3, al menos 1 gen del grupo 4, al menos 1 gen del grupo 5 y/o al menos 1 gen del grupo 6.

20

50- Kit según la reivindicación 49 que comprende reactivos adecuados para la determinación del nivel de expresión de todos los genes del grupo 1.

25 51- Kit según la reivindicación 49 que comprende reactivos adecuados para la determinación del nivel de expresión de todos los genes del grupo 2.

52- Kit según la reivindicación 49 que comprende reactivos adecuados para la determinación del nivel de expresión de todos los genes del grupo 3.

30 53- Kit según la reivindicación 49 que comprende reactivos adecuados para la determinación del nivel de expresión de todos los genes del grupo 4.

54-Kit según la reivindicación 49 que comprende reactivos adecuados para la determinación del nivel de expresión de todos los genes del grupo 5.

35

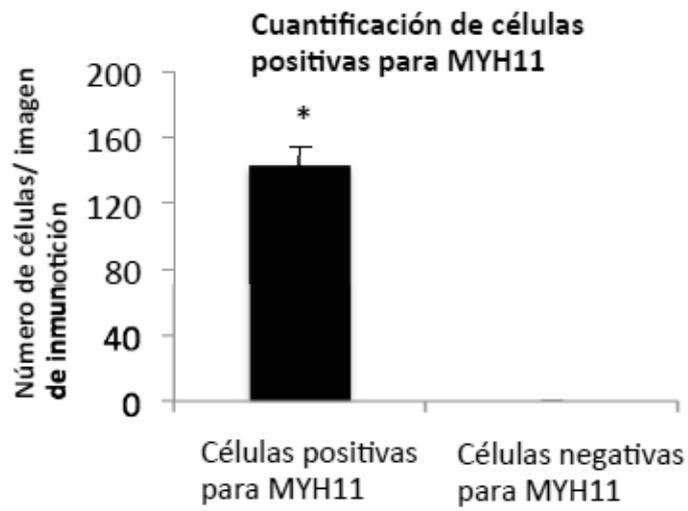
55- Kit según la reivindicación 49 que comprende reactivos adecuados para la determinación del nivel de expresión de todos los genes del grupo 6.

5 56- Kit según cualquiera de las reivindicaciones 49 a 56 que comprende reactivos adecuados para determinar el nivel de expresión de todos los genes del grupo 1, de todos los genes del grupo 2, de todos los genes del grupo 3, de todos los genes del grupo 4, de todos los genes del grupo 5 y de todos los genes del grupo 6.

10 57-Kit según la reivindicación 48 que comprende reactivos adecuados para determinar el nivel de expresión de todos los genes de la Tabla 1.

15 58- Uso de un kit según cualquiera de las reivindicaciones 48 a 57 para diagnosticar placas ateroscleróticas inestables en un sujeto, para determinar la probabilidad de un sujeto de sufrir una enfermedad cerebrovascular o para seleccionar un paciente susceptible de ser tratado con una terapia para la eliminación o estabilización de la placa aterosclerótica carotídea.

A



B

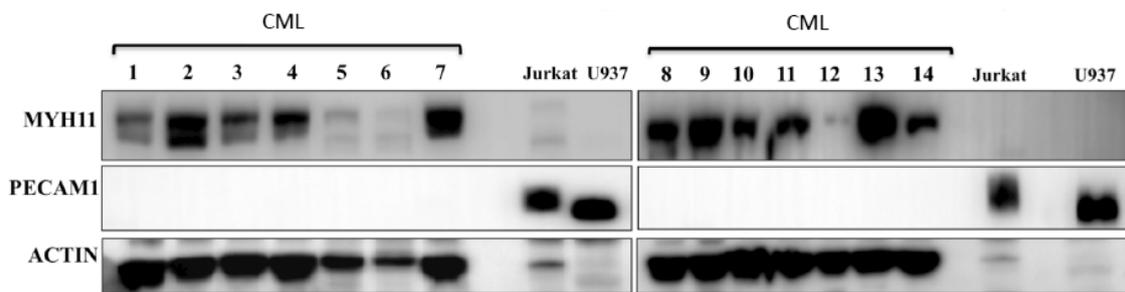


Fig. 1

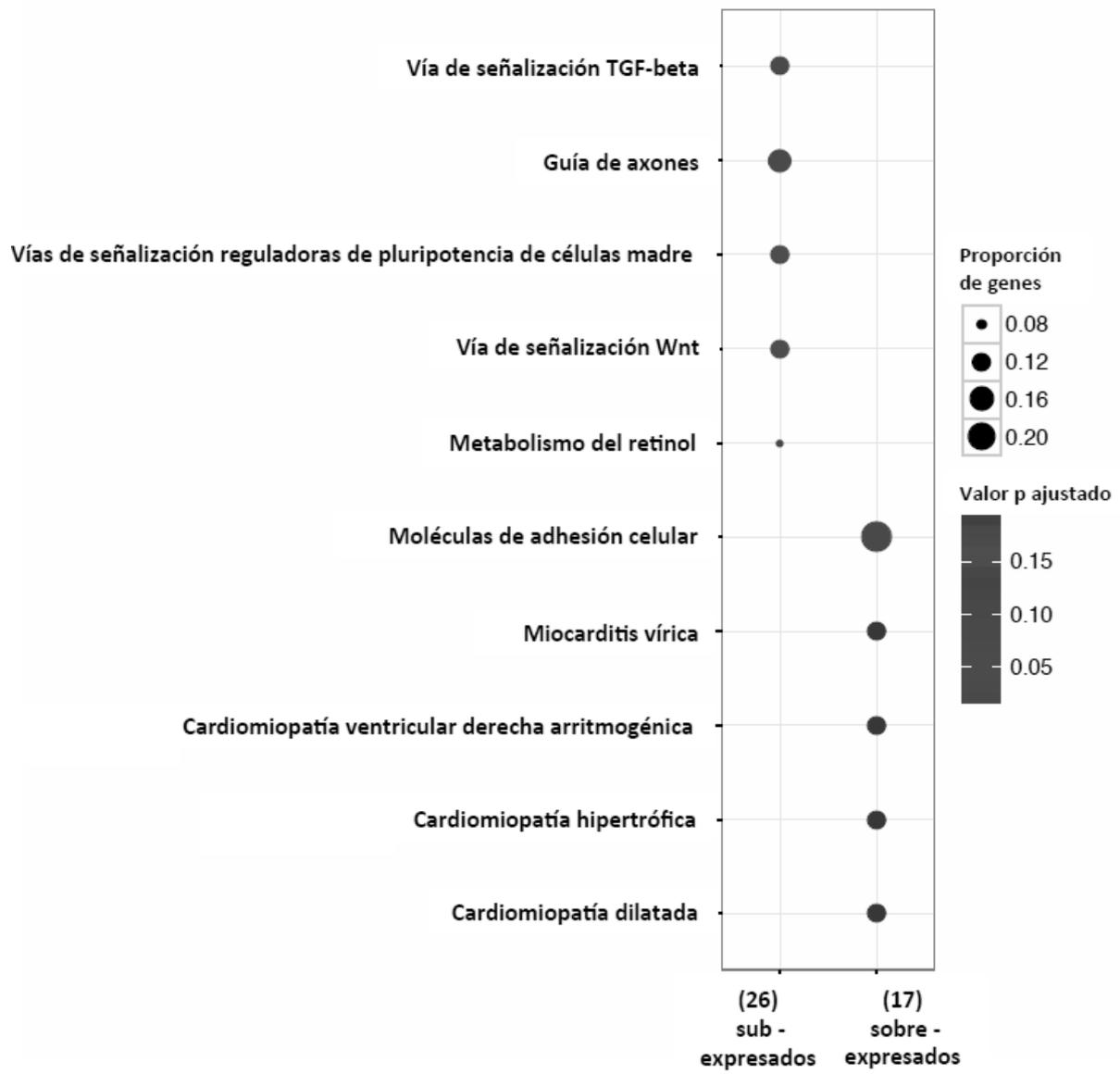


Fig. 2

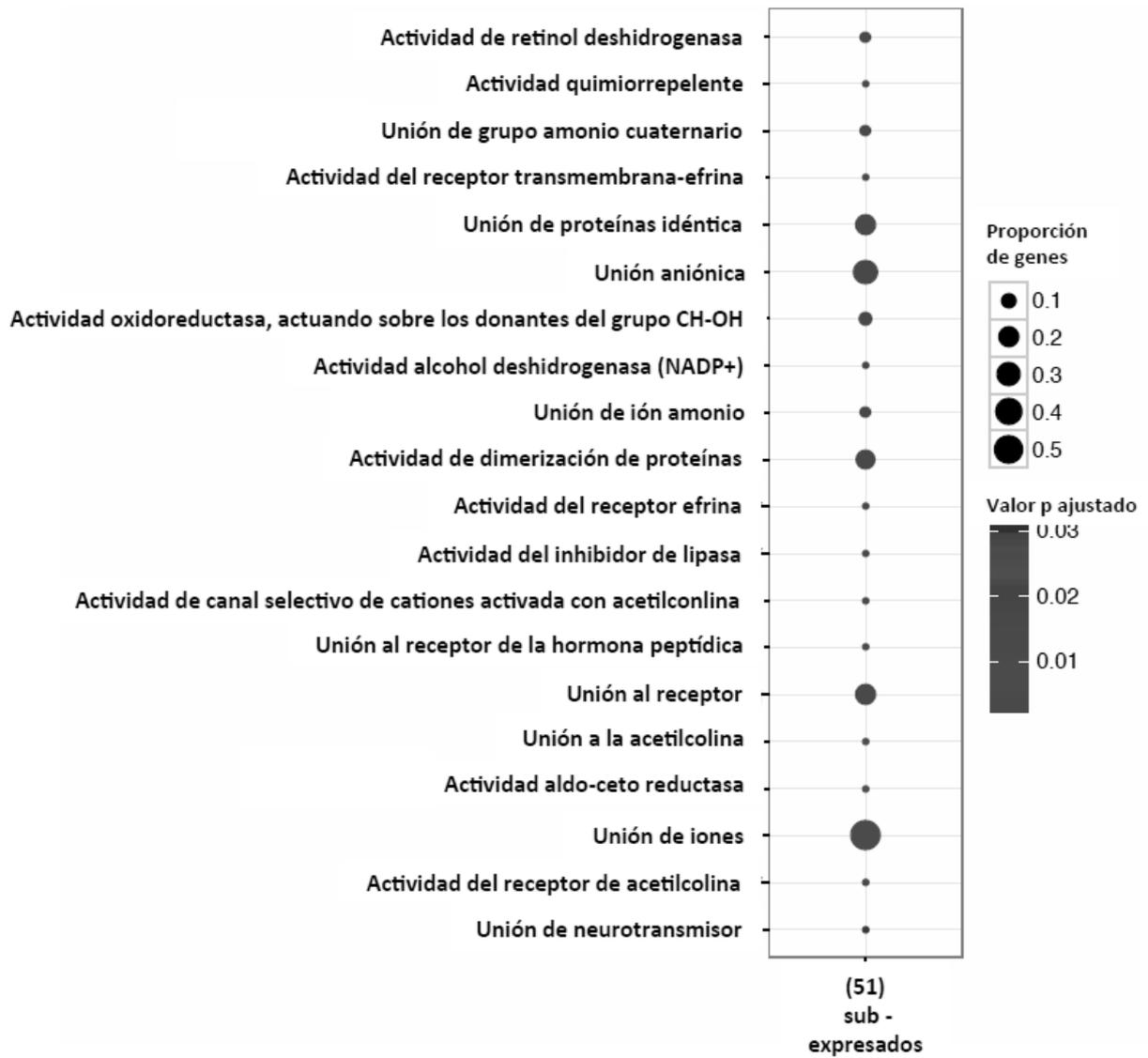


Fig. 3

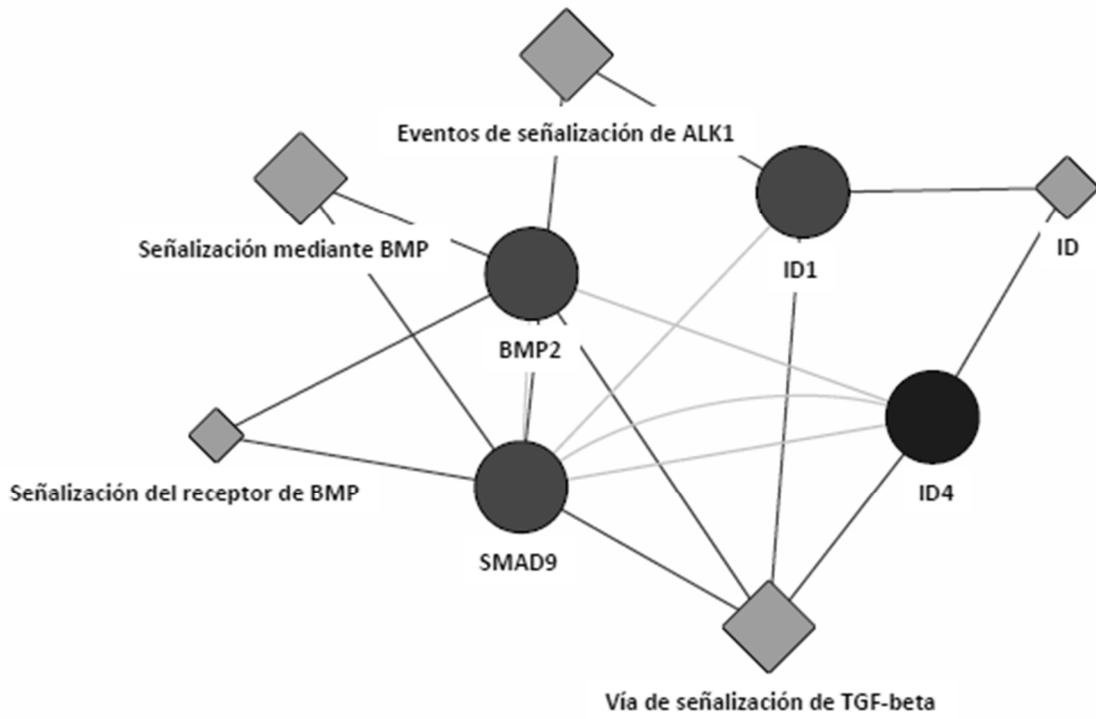


Fig. 4



- ②¹ N.º solicitud: 201730655
②² Fecha de presentación de la solicitud: 04.05.2017
③² Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

⑤¹ Int. Cl.: **C12Q1/68** (2006.01)

DOCUMENTOS RELEVANTES

Categoría	⑤ ⁶ Documentos citados	Reivindicaciones afectadas
X	DENG JUNJIANG et al. Decreased Expression of C10orf10 and Its Prognostic Significance in Human Breast Cancer. PLoS One JUN 17 2014. , 31/05/2014, Vol. 9, Páginas Article No.: e99730, ISSN 1932-6203(print) ISSN 1932-6203(electronic), <DOI: doi:10.1371/journal.pone.0099730>. Materiales y métodos.	48
X	WO 03078658 A2 (NOVARTIS AG et al.) 25/09/2003, Ejemplos 1 y 2; Reivindicaciones.	48
A	PERISIC L et al. Gene expression signatures, pathways and networks in carotid atherosclerosis. Journal of Internal Medicine MAR 2016. , 29/02/2016, Vol. 279, Páginas 293-308, ISSN 0954-6820(print) ISSN 1365-2796(electronic), <DOI: doi:10.1111/joim.12448>. todo el documento.	1, 11-16, 26-33, 43-48, 58
A	FERRONATO SILVIA et al. HDAC9, TWIST1 and FERD3L gene expression in asymptomatic stable and unstable carotid plaques. Inflammation Research APR 2016. , 31/03/2016, Vol. 65, Páginas 261-263, ISSN 1023-3830(print) ISSN 1420-908X(electronic), <DOI: doi:10.1007/s00011-015-0904-z>. todo el documento.	1, 11-16, 26-33, 43-48, 58
A	SIMATS ALBA et al. Neuroinflammatory biomarkers: From stroke diagnosis and prognosis to therapy. Biochimica et Biophysica Acta MAR 2016. , 29/02/2016, Vol. 1862, Páginas 411-424, ISSN 0925-4439(print) ISSN 0006-3002(electronic), <DOI: doi:10.1016/j.bbadis.2015.10.025>. todo el documento.	1, 11-16, 26-33, 43-48, 58
A	JICKLING GLEN C et al. Biomarker Panels in Ischemic Stroke. Stroke MAR 2015. , 28/02/2015, Vol. 46, Páginas 915-920, ISSN 0039-2499(print) ISSN 1524-4628(electronic), <DOI: doi:10.1161/STROKEAHA.114.005604>. Todo el documento.	1, 11-16, 26-33, 43-48, 58
A	SWAMINATHAN BHAIRAVI et al. Autophagic Marker MAP1LC3B Expression Levels Are Associated with Carotid Atherosclerosis Symptomatology. PLoS One DEC 12 2014. , 30/11/2014, Vol. 9, Páginas Article No.: e115176, ISSN 1932-6203(print) ISSN 1932-6203(electronic), <DOI: doi:10.1371/journal.pone.0115176>. <p>todo el documento.</p>	1, 11-16, 26-33, 43-48, 58

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

para todas las reivindicaciones

para las reivindicaciones nº: 1, 11-16, 26-33, 43-48, 58

Fecha de realización del informe
28.11.2017

Examinador
J. L. Vizán Arroyo

Página
1/3



OFICINA ESPAÑOLA
DE PATENTES Y MARCAS
ESPAÑA

- ②¹ N.º solicitud: 201730655
②² Fecha de presentación de la solicitud: 04.05.2017
③² Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

⑤¹ Int. Cl.: **C12Q1/68** (2006.01)

DOCUMENTOS RELEVANTES

Categoría	⑤ ⁶ Documentos citados	Reivindicaciones afectadas
A	PAPASPYRIDONOS MARIANNA et al. Novel candidate genes in unstable areas of human atherosclerotic plaques. Arteriosclerosis Thrombosis and Vascular Biology AUG 2006. , 31/07/2006, Vol. 26, Páginas 1837-1844, ISSN 1079-5642(print) ISSN 1524-4636(electronic), <DOI: doi:10.1161/01.ATV.0000229695.68416.76>. todo el documento.	1, 11-16, 26-33, 43-48, 58
A	US 2008166734 A1 (DENG DAVID XING-FEI) 10/07/2008, todo el documento.	1, 11-16, 26-33, 43-48, 58
A	US 2003129176 A1 (JONES KAREN ANNE et al.) 10/07/2003, todo el documento.	1, 11-16, 26-33, 43-48, 58

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

para todas las reivindicaciones

para las reivindicaciones nº: 1, 11-16, 26-33, 43-48, 58

Fecha de realización del informe
28.11.2017

Examinador
J. L. Vizán Arroyo

Página
2/3

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

C12Q

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, BIOSIS, MEDLINE, EMBASE, INTERNET