

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 334 084**

21 Número de solicitud: 200703415

51 Int. Cl.:
C12Q 1/68 (2006.01)

12

PATENTE DE INVENCION

B1

22 Fecha de presentación: **21.12.2007**

43 Fecha de publicación de la solicitud: **04.03.2010**

Fecha de la concesión: **12.01.2011**

45 Fecha de anuncio de la concesión: **24.01.2011**

45 Fecha de publicación del folleto de la patente:
24.01.2011

73 Titular/es:
Instituto Científico y Tecnológico de Navarra, S.A.
Avda. Pío XII, 53
31008 Pamplona, Navarra, ES
PROGENIKA BIOPHARMA, S.A.

72 Inventor/es: **Villoslada Díaz, Pablo;**
Martínez Martínez, Antonio;
Simón Buela, Laureano;
Palacios Urtasun, Ricardo y
Sepulcre Bernad, Jorge

74 Agente: **Arias Sanz, Juan**

54 Título: **Marcadores genéticos para el pronóstico de la esclerosis múltiple.**

57 Resumen:

Marcadores genéticos para el pronóstico de la esclerosis múltiple.

La presente invención se relaciona con una serie de genes cuya expresión se ve alterada en sujetos que sufren esclerosis múltiple con respecto a sujetos sanos o en sujetos que sufren esclerosis múltiple de buen pronóstico con respecto a sujetos que sufren esclerosis múltiple de mal pronóstico. A partir de un conjunto inicial de genes que mostraban dicha expresión diferencial, se ha validado un subconjunto formado por 13 genes y dos variables clínicas que permite predecir la evolución de un paciente con una alta fiabilidad. A partir de dichos valores de expresión, la invención proporciona métodos para pronosticar la evolución de un paciente diagnosticado de esclerosis múltiple a partir de tablas de probabilidad condicionada entre los niveles de expresión de un determinado gen o grupo de genes y las probabilidades de que el paciente tenga un buen o mal pronóstico de la enfermedad.

ES 2 334 084 B1

Aviso: Se puede realizar consulta prevista por el art. 37.3.8 LP.

DESCRIPCIÓN

Marcadores genéticos para el pronóstico de la esclerosis múltiple.

5 **Campo técnico de la invención**

La invención se relaciona con un método para el pronóstico de la esclerosis múltiple y, más concretamente, con un método para predecir la evolución clínica de pacientes diagnosticados con esclerosis múltiple mediante el análisis de los niveles de expresión de una serie de genes.

10 **Antecedentes de la invención**

La esclerosis múltiple (EM) tiene una prevalencia de alrededor de 70 casos cada 100.000 habitantes en España y en el mundo occidental es la causa más común de discapacidad neurológica crónica en adultos jóvenes tras los accidentes de tráfico. Aproximadamente, un 70% de casos debuta entre los 20 y 40 años de edad, con un pico en la edad de inicio alrededor de los 25 o 30 años, por lo que resulta fácilmente comprensible el gran impacto que ocasiona sobre la vida profesional, familiar y social de los afectados, así como el enorme gasto económico y social que genera, similar al de la enfermedad de Alzheimer.

La esclerosis múltiple es una enfermedad heterogénea en su presentación y evolución en la que el 80-85% de los pacientes presentan un curso clínico que evoluciona a brotes autolimitados que a medida que se repiten van ocasionando un déficit residual funcional (forma remitente-recurrente; EM-RR). Tras 10-15 años de evolución el 50% de ellos pasarán a tener un curso secundariamente progresivo de incremento de la discapacidad no relacionado con los brotes (forma secundariamente progresiva; EM-SP), y tras 25 años el porcentaje alcanza al 90% de los pacientes. En un 10% de los casos el curso es progresivo desde el inicio (forma primariamente progresiva; EM-PP). Un 10 a 20% de los pacientes se mantendrán sin secuelas importantes 15 años después del inicio de la enfermedad (formas benignas) y en un 1-3% de los casos, sin embargo, los pacientes evolucionarán acumulando una gran discapacidad en pocos meses tras el debut de la enfermedad (formas agresivas o fulminantes).

El interferón beta (Betaferon[®], RebifV, Avonex[®]) y el acetato de glatiramero o copolímero (Copaxone[®]) son los primeros medicamentos que han mostrado efectos beneficiosos en la forma remitente recurrente de la esclerosis múltiple. Estos medicamentos disminuyen la formación de placas y el número de brotes en un tercio comparado con pacientes sin tratamiento. Recientemente se ha incluido el tratamiento con Natalizumab (Tysabri[®]) a la terapia de la esclerosis múltiple, teniendo una eficacia mayor que los previos (previenen de brotes aproximadamente 60%), aunque con potenciales efectos secundarios graves. Sin embargo, la respuesta individual al tratamiento es impredecible y varía desde la excelencia a una total ineficacia. La falta de test biológicos que predigan la actividad y agresividad de la enfermedad impiden prescribir el mejor tratamiento a cada paciente y obliga a realizar un tratamiento preventivo y de por vida con el consiguiente coste en calidad de vida y económico. El poder contar con un test predictivo de agresividad y actividad de la enfermedad permitiría realizar un tratamiento personalizado.

Métodos estándar de diagnóstico de la esclerosis múltiple incluyen determinación de los niveles de IgG en LCR, imagen del cerebro mediante resonancia magnética y de la médula espinal y exclusión de otras enfermedades autoinmunes mediante determinaciones en suero. Aunque recientemente se ha estudiado la utilidad de dichos test en predecir el curso de la esclerosis múltiple, su capacidad predictiva es muy limitada y en la práctica clínica se emplean con fines diagnósticos pero no pronósticos ni para decidir o monitorizar la terapia. Por tanto, sigue siendo problemático el llegar a un diagnóstico y a un pronóstico adecuado.

Los diferentes estudios de historia natural de la esclerosis múltiple han permitido identificar algunas variables clínicas asociadas con la evolución de la esclerosis múltiple. Los factores que mejor predicen un curso relativamente benigno son pertenecer al género femenino, inicio de la enfermedad a temprana edad (menos de 30 años), ataques poco frecuentes, un patrón remitente-recurrente y levedad de la enfermedad en los estudios mediante resonancia magnética del sistema nervioso central. Por el contrario, los factores que predicen un curso más agresivo son el sexo varón e inicio tardío (por estar asociado con las formas progresivas), la recidiva del segundo brote en el primer año tras el primer brote, acumular discapacidad de forma precoz, el inicio clínico con síntomas motores o de coordinación y la persistencia de secuela tras el primer brote, y especialmente el alcanzar ciertos niveles de discapacidad (escala *Kurtzke Expanded Disability Status* EDSS: 3.0, 4.0, 6.0) en edades tempranas.

Hasta la fecha se han intentado desarrollar métodos de diagnóstico y de pronóstico basados en la detección de autoanticuerpos en suero (Bielekova, B. y, Martin R. *Brain*. 2004 Jul;127(Pt 7):1463-78.; Berger, T. y Reindl, M., 2006, *Disease Markers*, 22:207-212). Por ejemplo, Berger, T. *et al.* (*New England J. Medicine*, 2003, 146:181-197) han descrito que la presencia de anticuerpos anti-mielina (anti-MOG) son capaces de predecir el riesgo de primera recaída en pacientes que sufre síndrome clínicamente aislado, sugestivo de esclerosis múltiple. Sin embargo estudios bien diseñados posteriores no han podido confirmar su utilidad predictora (Kuhie J. *et al.*, *N Engl J Med*. 2007, 356:371-8.; Pelayo R. *et al.* *N. Engl. J. Med.* 2007, 356:426-8).

Sin embargo, hasta la fecha no se ha identificado ningún anticuerpo que cumpla los requisitos de biomarcador diagnóstico o pronóstico de esclerosis múltiple, como tampoco la determinación simultánea de varios autoanticuerpos es de utilidad, así como las dificultades y alto coste que supone.

Por otro lado, los estudios de expresión génica (transcriptoma) mediante chips de ADN permiten tener una visión global de los genes que están participando en un proceso, por lo que este tipo de análisis podría convertirse en una valiosa herramienta clínica para el diagnóstico o el pronóstico de MS.

5 Hasta la fecha se han descrito diferencias en los niveles de expresión de diversos genes en la esclerosis múltiple, que podrían llegar a ser candidatos a biomarcadores de esclerosis múltiple (revisados en Goertsches, R. *et al.*, 2006, *Current Pharmaceutical Design*, 12:3761-3779). En estos estudios se compararon los patrones de expresión entre pacientes con esclerosis múltiple y controles, pero no se llegó a estudiar de forma específica la asociación entre dichos patrones y el curso evolutivo y pronóstico de la enfermedad.

10 Ramanathan *et al.* (*J. Neuroimmunol.*, 2001, 116:213-219) han descrito 34 genes diferencialmente expresados en pacientes de EM-RR en comparación con sujetos sanos, la mayoría de ellos relacionados con procesos inflamatorios e inmunes.

15 Bompreszi *et al.* (*Human Molecular Genetics*, 2003, 12:2191-2199) identificaron una serie de genes cuyos niveles de expresión en PBMCs permite distinguir pacientes de EM-RR y de EM-SP y de voluntarios sanos. Mediante este abordaje, se pudieron identificar más de mil genes que permitían distinguir muestras de esclerosis múltiple de controles. Los genes fuertemente dominantes incluyeron HSP70 y la CDC28 proteína quinasa CDC28 (CKS) 2, que combinados con la histona H I de la familia (HIF) 2 y el PAFAH1B1, respectivamente, permitían una buena discriminación entre esclerosis múltiple y controles. Estos pares tenían valor de predicción del 80% para clasificar una muestra independiente en la clase correcta. También se observó una correlación entre los pares más discriminatorios de genes y vías biológicas relevantes de la esclerosis múltiple. Tales moléculas, que se encontraban altamente expresadas en esclerosis múltiple incluían CD27, el receptor de TNF, el locus alfa del receptor de células T y su cadena asociada ZAP70, y la proteína del dedo de zinc (ZNF) 148. Además, el receptor IL-7 (IL-7R), que se requiere para el desarrollo de células T y B, estaba también fuertemente sobre-expresado. Los genes reprimidos en esclerosis múltiple fueron HSP70 y CKS2, ambos implicados en la regulación de la apoptosis. Se ha sugerido previamente que HSP70 puede ser un autoantígeno en esclerosis múltiple, pero puede también estar implicado en la degradación del mRNA en la vía ubiquitina-proteasoma. La activación de los procesos de remodelamiento de la matriz extracelular era evidente por la sobre expresión de la metaloproteinasa de la matriz (MMP) - 19 y la represión de un inhibidor del TIMP 1.

30 Los patrones de expresión para esclerosis múltiple y la patofisiología de la enfermedad se han analizado en varios estudios. Iglesias *et al.* (*J. Neuroimmunol.*, 2004 150:163-77) identificaron un sistema de 553 genes diferencialmente expresados en EM-RR comparado con los controles sanos, 87 de los cuales fueron altamente significativos. Entre los genes diferencialmente expresados se pudieron identificar algunos implicados en la activación y coestimulación de células T, entre los cuales se encontraban varios genes de respuesta a interferón, tales como IL-12, CD40, antígeno citotóxico 4 (CTLA4), receptores de la célula T, inmunoglobulinas, el receptor de IL-6, el receptor de IL-8, e integrinas, por ejemplo VLA4 y VLA6, así como distintos genes de la vía de E2F (E2F2, E2F3, CDC25A, CDK2), la timopoyetina (TMPO), y PRIM1. La importancia de la ruta de E2F en esclerosis múltiple fue validado en encefalitis autoinmune experimental (EAE). Los ratones de E2F 1-deficientes manifestaron solamente un suave curso de la enfermedad de EAE.

45 Los patrones génicos para la actividad de la enfermedad esclerosis múltiple también se han estudiado. En la solicitud de patente internacional WO03081201 se identificó un patrón de 1109 genes en PBMCs a partir de 26 pacientes de esclerosis múltiple comparados con voluntarios sanos con independencia del estado de la activación de la enfermedad. El patrón fue validado con el método LOOCV, que dio en la clasificación solamente dos errores, probando que los patrones observados representan un fenómeno biológico verdadero. En estos genes se incluían aquellos relacionados con la activación y expansión, estímulos inflamatorios de la célula T (citoquinas e integrinas), epitope que se separa, y apoptosis. Comparación de los perfiles de la expresión en PBMCs de pacientes de esclerosis múltiple en brote reveló un patrón de 721 genes. Protease L, CTSL1 y las proteínas MCP1 y MCP2, estaban sobre expresados durante el brote. Por el contrario, varios genes relacionados con apoptosis como ciclina G1 y las caspasas (CASP) 2, 8 y 10 estaban reprimidas.

55 En WO03023056 se describen métodos para el diagnóstico de y/o la susceptibilidad a la esclerosis múltiple mediante la determinación de variaciones en los niveles de expresión de 25 genes.

De manera individual, un gen (CX3CR1) que se ha identificado en análisis de expresión en subpoblaciones de células de T se ha propuesto como marcador para la actividad de la enfermedad. CX3CR1 se encuentra reprimido en pacientes EM-RR y EM-PP comparado con voluntarios sanos. Este hallazgo se ha validado por PCR en tiempo real y por citometría de flujo en cohortes independientes. Las células NK son las responsables del fenotipo, mientras que la expresión de CX3CR1 no está alterada en células citotóxicas CD8 en pacientes de esclerosis múltiple respecto a controles sanos.

65 En US2004/0091915 se describe un método para predecir la tasa de supervivencia de pacientes diagnosticados de esclerosis múltiple mediante la detección de una delección en el gen CCR5.

En WO2005054810 se describe un método para predecir la tasa de supervivencia de pacientes diagnosticados de esclerosis múltiple mediante la detección de una delección en el gen CD24.

ES 2 334 084 B1

En WO03001212 se describe un método para el diagnóstico de esclerosis múltiple basado en detectar en una muestra aislada del sujeto la ausencia de la proteína wt-SARG-1 o del ARNm que lo codifica.

5 En US20050064483 se describe un método para la monitorización de la respuesta de un enfermo de esclerosis múltiple al tratamiento con interferón-beta o con acetato de glatiramer mediante la detección de variaciones en la expresión de al menos 4 genes seleccionados de un grupo de 34 genes.

10 En US20050089919 se describe un método para la detección de le esclerosis múltiple que comprende detectar variaciones en la expresión de al menos un gen seleccionado de una serie de 31 genes.

10 En US20050164253 se describe un método para la detección de la esclerosis múltiple que comprende detectar variaciones en la expresión de al menos un gen seleccionado del grupo de RIPK2, NFKBIE, TNFAIP3, DAXX, TNFSF10, BAG1, TOP1, ADPRT, CREB1, MYC, BAG4, RBBP4, GZMA, BCL2 y E2F5.

15 En US20060115826 se describe un método para el diagnóstico de la esclerosis múltiple que comprende detectar variaciones en la expresión de al menos dos genes seleccionados de un conjunto de 107 genes asociados con procesos inflamatorios.

20 En WO02079218 se describe un método para el diagnóstico de la esclerosis múltiple que comprende detectar variaciones en la expresión de un panel de genes seleccionados porque muestran variaciones en su nivel de expresión en un modelo animal de encefalitis autoinmune experimental. En este estudio, el análisis de la expresión de los distintos genes se llevó a cabo mediante un chip de ADN humano en el que se encontraban representados unos 14000 genes distintos.

25 En WO03081201 se describe un método para el diagnóstico de la esclerosis múltiple basado en detectar variaciones en la expresión de un panel de genes representados en un chip de ADN humano que contenía 12625 genes humanos.

30 En WO03095618 se describen métodos para el diagnóstico de esclerosis múltiple, para el diagnóstico diferencial de esclerosis múltiple con respecto a la esclerosis lateral amiotrófica, para predecir la respuesta de un sujeto diagnosticado de esclerosis múltiple a un tratamiento con Avonex mediante la detección de variaciones de la expresión de una serie de genes implicados en distintas vías de señalización.

35 Sin embargo, todos los métodos descritos hasta la fecha se dirigen a detectar diferencias entre enfermos que se sospecha presentan esclerosis múltiple y sujetos controles, por lo que tienen una utilidad fundamentalmente diagnóstica, pero no permiten predecir la evolución de la enfermedad en pacientes que ya han sido diagnosticados de esclerosis múltiple. Por tanto, existe una necesidad de métodos que permitan predecir la evolución de la enfermedad en pacientes ya diagnosticados.

40 Compendio de la invención

En un primer aspecto, la invención se relaciona con un método *in vitro* para determinar el pronóstico clínico de un paciente afecto de esclerosis múltiple que comprende

- 45 (a) comparar
- (i) el valor correspondiente a la expresión de un gen seleccionado del grupo de KLHDC5, CASP2, EMID1, PRO1073, BTBD7, MGC2518, WDR20bis, NEK4, SYLT2, DOCK10, TTC10, PTPRC y CTLA4 con una tabla de probabilidades condicionadas entre intervalos de valores modales de expresión de dichos genes y valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y/o
 - 50 (ii) el valor de una variable clínica seleccionada del grupo de EDSS y MSFC con una tabla de probabilidades condicionadas entre intervalos de valores modales de dichas variables clínicas y valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y
- 55 (b) asignar una probabilidad de mal y buen pronóstico correspondiente a la probabilidad asociada al intervalo en el que se encuentra el valor de la expresión o de la variable clínica.

60 En otro aspecto, la invención se relaciona con un método *in vitro* para determinar el pronóstico clínico de un paciente afecto de esclerosis que comprende

- (a) comparar
- 65 (i) los valores correspondientes a la expresión de al menos dos genes seleccionados del grupo de KLHDC5, CASP2, EMID1, PRO1073, BTBD7, MGC2518, WDR20bis, NEK4, SYLT2, DOCK10, TTC10, PTPRC y CTLA4 con una tabla de probabilidades condicionadas entre intervalos de valores modales de expresión de dichos genes y valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y/o

ES 2 334 084 B1

(ii) los valores de las variables clínicas EDSS y MSFC con una tabla de probabilidades condicionadas entre intervalos de valores modales de dichas variables clínicas y valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y

- 5 (b) asignar una probabilidad de mal pronóstico correspondiente a la probabilidad condicionada de mal pronóstico asociada a los intervalos de valores modales en los que se encuentran los valores de expresión de cada uno de los genes cuya expresión se ha determinado y/o variables clínicas determinadas y asignar una probabilidad de buen pronóstico correspondiente a la probabilidad condicionada de buen pronóstico asociada a los intervalos de valores modales en los que se encuentra los valores de expresión para cada uno de los genes cuya expresión se ha determinado y/o variables clínicas determinadas.
- 10

En otro aspecto, la invención se relaciona con un método para determinar el pronóstico clínico de un sujeto afecto de esclerosis múltiple, para monitorizar el efecto de la terapia administrada a un sujeto afecto de esclerosis múltiple o para asignar una terapia personalizada a un sujeto afecto de esclerosis que comprende

15

(a) determinar el nivel de expresión de uno o varios genes seleccionados del grupo de genes listados en las posiciones 3, 5, 6, 7, 9, 11, 13, 16, 19, 20, 22, 24, 25, 26, 30, 31, 33, 34, 35, 37, 41 ó 43 de la tabla 3), o de los polipéptidos codificados por dichos genes, en una muestra biológica aislada del paciente y

20

(b) comparar los niveles de expresión de dichos genes o de dichos polipéptidos con un valor de referencia calculado a partir de una o varias muestras obtenidas de un paciente sano,

en donde

25

(i) un aumento de la expresión de los genes en posición 6, 7, 9, 33, 35, 37 ó 43, o de los polipéptidos codificados por dichos genes, o una disminución de la expresión de los genes en posición 3, 5, 11, 13, 16, 19, 22, 24, 25, 26, 30, 31, 34, 41, o de los polipéptidos codificados por dichos genes con respecto al valor de referencia, es indicativo de un mal pronóstico de la esclerosis múltiple en dicho sujeto, de que la terapia es inefectiva o de que el paciente es seleccionado para una terapia agresiva o

30

(ii) un aumento de la expresión de los genes en posiciones 3, 5, 11, 16, 20, 30, o de los polipéptidos codificados por dichos genes, o una disminución de la expresión del gen en posición 43, o del polipéptido codificado por dicho gen con respecto al valor de referencia, es indicativo de un buen pronóstico de la esclerosis múltiple en dicho paciente, de que la terapia es efectiva o de que el paciente es seleccionado para no recibir terapia o para recibir una terapia poco agresiva.

35

En otro aspecto, la invención se relaciona con un método para determinar el pronóstico clínico de un sujeto afecto de esclerosis múltiple, para monitorizar el efecto de la terapia administrada a un sujeto afecto de esclerosis múltiple o para asignar una terapia personalizada a un sujeto afecto de esclerosis que comprende

40

(a) determinar el nivel de expresión de uno o varios genes seleccionados del grupo de genes listados en las posiciones 1 a 21 de la tabla 5, o de los polipéptidos codificados por dichos genes, en una muestra biológica aislada del paciente y

45

(b) comparar los niveles de expresión de dichos genes o de dichos polipéptidos con un valor de referencia calculado a partir de una o varias muestras obtenidas de pacientes de buen pronóstico y con un valor de referencia calculado a partir de una o varias muestras obtenidas de pacientes de mal pronóstico

50

en donde

(i) un aumento de la expresión de los genes en posición 1, 2, 3, 4, 5, 8, 9, 10, 14, 19, 20 ó 21 o de los polipéptidos codificados por dichos genes con respecto a un valor de referencia obtenido de una o varias muestras de pacientes diagnosticados de esclerosis múltiple de mal pronóstico es indicativo de un buen pronóstico de la esclerosis múltiple en dicho sujeto, de que la terapia es efectiva o de que el paciente es seleccionado para no recibir una terapia agresiva y

55

(ii) un aumento de la expresión de los genes en posiciones 6, 7, 11, 12, 13, 15, 16, 17 ó 18 o de los polipéptidos codificados por dichos genes con respecto a un valor de referencia obtenido de una o varias muestras de pacientes diagnosticados de esclerosis múltiple de buen pronóstico es indicativo de un mal pronóstico de la esclerosis múltiple en dicho paciente, de que la terapia no es efectiva o de que el paciente es seleccionado para recibir terapia o para recibir una terapia poco agresiva.

60

65

ES 2 334 084 B1

En otro aspecto, la invención se relaciona con un método para determinar el pronóstico clínico de un sujeto afecto de esclerosis múltiple, para monitorizar el efecto de la terapia administrada a un sujeto afecto de esclerosis múltiple o para asignar una terapia personalizada a un sujeto afecto de esclerosis que comprende

- 5 (a) determinar el nivel de expresión de uno o varios genes seleccionados de la tabla 6), o de los polipéptidos codificados por dichos genes, en una muestra aislada del paciente y
- (b) comparar los niveles de expresión de dichos genes con un valor de referencia calculado a partir de una o varias muestras obtenidas de un paciente sano

10 en donde un aumento de la expresión de los genes en posición 4, 8, 11, 13, 15, 18, 19, 20, 21, 24, 25, 28, 30 ó 32, o de los polipéptidos codificados por dichos genes, o una disminución de los genes en posición 1, 2, 3, 5, 6, 7, 9, 10, 12, 14, 16, 17, 22, 23, 26, 27, 29 ó 31, o de los polipéptidos codificados por dichos genes, con respecto al valor de referencia es indicativo de un mal pronóstico de la esclerosis múltiple, de que la terapia no es efectiva o de que de que el paciente es seleccionado para una terapia agresiva.

20 En otro aspecto, la invención se relaciona con un método para determinar el pronóstico clínico de un sujeto afecto de esclerosis múltiple, para monitorizar el efecto de la terapia administrada a un sujeto afecto de esclerosis múltiple o para asignar una terapia personalizada a un sujeto afecto de esclerosis que comprende

- (a) determinar el nivel de expresión de uno o varios genes seleccionados de la tabla 7, o de los polipéptidos codificados por dichos genes, en una muestra aislada del paciente y
- 25 (b) comparar los niveles de expresión de dichos genes con un valor de referencia calculado a partir de una o varias muestras obtenidas de un paciente sano

30 en donde un aumento de la expresión de los genes en posición 2, 5, 6, 7, 8 y 10, o de los polipéptidos codificados por dichos genes, o una disminución de la expresión de los genes en posición 1, 3, 4 ó 9, o de los polipéptidos codificados por dichos genes, con respecto al valor de referencia es indicativo de un buen pronóstico de la esclerosis múltiple o de que la terapia administrada es efectiva o de que el paciente es seleccionado para no recibir terapia o para recibir una terapia poco agresiva.

35 En otro aspecto, la invención se relaciona con un método para diagnosticar la esclerosis múltiple en un sujeto que comprende

- (a) determinar el nivel de expresión de uno o varios genes seleccionados del grupo de genes indicado en la tabla 8, o de los polipéptidos codificados por dichos genes, en una muestra aislada del sujeto y
- 40 (b) comparar los niveles de expresión de dichos genes con un valor de referencia calculado a partir de una o varias muestras obtenidas de un paciente sano

45 en donde una disminución de la expresión de los genes en posición 1, 2, 6, 10, 15 ó 16, o de los polipéptidos codificados por dichos genes, o un aumento en la expresión de los genes en posición 3, 4, 5, 7, 8, 9, 11, 12, 13 ó 14, o de los polipéptidos codificados por dichos genes, con respecto al valor de referencia es indicativo de que el sujeto sufre de esclerosis múltiple.

50 En otro aspecto, la invención se relaciona con un kit que comprende un conjunto de sondas en donde dicho conjunto comprende una sonda específica para cada uno de los genes indicados en al menos una tabla seleccionada del grupo de tablas 3, 5-8 y 11.

55 En otro aspecto, la invención se relaciona con el uso de un kit de la invención para la determinación del pronóstico de un enfermo diagnosticado de esclerosis múltiple, para determinar la eficacia de un tratamiento para la esclerosis múltiple o para diagnosticar la esclerosis múltiple en un paciente.

Breve descripción de las figuras

60 Figura 1: Distribución de GO de los 45 genes que presentaban diferencias significativas entre las tres clases.

Figura 2: Análisis de cluster de las muestras.

Figura 3: Análisis de cluster de los genes.

65 Figura 4: Análisis de cluster tanto de muestras como de genes obtenidos tras las comparaciones con $p < 0,001$ de las clases (A) buen y mal pronóstico, (B) control y mal pronóstico, (C) control y buen pronóstico, (D) control y esclerosis múltiple y la comparación con $p < 0,005$ de las tres clases (E).

ES 2 334 084 B1

Figura 5: Representación gráfica en las 60 muestras estudiadas del comportamiento de los 25 genes que presentaron diferencias significativas ($p < 0,01$) entre las tres clases.

Figura 6: Red Bayesiana y matriz de confusión de la validación del clasificador utilizando las variables clínicas EDSS (*Kurtzke Expanded Disability Status*) y MSFC (*Multiple Sclerosis Functional Composite*) y aquellos genes que presentaban diferencias significativas ($p < 0,01$) en los niveles de expresión de las tres clases.

Figura 7: Representación gráfica en las 40 muestras de pacientes del comportamiento de los 13 genes que presentaron diferencias significativas ($p < 0,05$) entre buen y mal pronóstico.

Figura 8: Red Bayesiana y matriz de confusión de la validación del clasificador utilizando las variables clínicas (EDSS y MSFC) y aquellos genes que presentaban diferencias significativas ($p < 0,05$) en los niveles de expresión entre buen y mal pronóstico.

15 Descripción detallada de la invención

Los autores de la presente invención, usando micromatrices de ADN, han identificado una serie de genes que se expresan de forma diferencial en pacientes diagnosticados de esclerosis múltiple en los que la enfermedad es de buen pronóstico con respecto a pacientes en los que la enfermedad muestra un mal pronóstico o a sujetos control.

Asimismo, los autores de la invención han identificado una serie de genes cuya expresión se modifica en pacientes diagnosticados de esclerosis múltiple en los que la enfermedad es de mal pronóstico. A partir de un conjunto inicial de genes identificados, se validó un subconjunto de 13 genes mediante PCR en tiempo real cuyas variaciones de expresión permitían predecir el tipo de pronóstico de los pacientes de forma significativa ($p < 0,05$).

Así, en un primer aspecto, la invención se relaciona con un método *in vitro* (en adelante el primer método de la invención) para determinar el pronóstico clínico de un paciente afecto de esclerosis múltiple que comprende

- (a) comparar
 - (i) el valor correspondiente a la expresión de un gen seleccionado del grupo de KLHDC5, CASP2, EMID1, PRO1073, BTBD7, MGC2518, WDR20bis, NEK4, SYLT2, DOCK10, TTC10, PTPRC y CTLA4 con una tabla de probabilidades condicionadas entre intervalos de valores modales de expresión de dichos genes y valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y/o
 - (ii) el valor de una variable clínica seleccionada del grupo de EDSS y MSFC con una tabla de probabilidades condicionadas entre intervalos de valores modales de dichas variables clínicas y valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y
- (b) asignar una probabilidad de mal y buen pronóstico correspondiente a la probabilidad asociada al intervalo en el que se encuentra el valor de la expresión o de la variable clínica.

Adicionalmente, a partir de los niveles de expresión de estos 13 genes y tomando dos variables clínicas (ESSS y MSFC), se obtuvo un clasificador que permitía predecir la evolución de la enfermedad con una precisión de un 95%. Así, en otro aspecto, la invención se relaciona con un método *in vitro* (en adelante el primer método de la invención) para determinar el pronóstico clínico de un paciente afecto de esclerosis que comprende

- (a) comparar
 - (i) los valores correspondientes a la expresión de al menos dos genes seleccionados del grupo de KLHDC5, CASP2, EMID1, PRO1073, BTBD7, MGC2518, WDR20bis, NEK4, SYLT2, DOCK10, TTC10, PTPRC y CTLA4 con una tabla de probabilidades condicionadas entre intervalos de valores modales de expresión de dichos genes y valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y/o
 - (ii) los valores de las variables clínicas EDSS y MSFC con una tabla de probabilidades condicionadas entre intervalos de valores modales de dichas variables clínicas y valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y
- (b) asignar una probabilidad de mal pronóstico correspondiente a la probabilidad condicionada de mal pronóstico asociada a los intervalos de valores modales en los que se encuentran los valores de expresión de cada uno de los genes cuya expresión se ha determinado y/o variables clínicas determinadas y asignar una probabilidad de buen pronóstico correspondiente a la probabilidad condicionada de buen pronóstico asociada a los intervalos de valores modales en los que se encuentra los valores de expresión para cada uno de los genes cuya expresión se ha determinado y/o variables clínicas determinadas.

Por “determinación del pronóstico clínico” según la presente invención, se entiende el emitir una opinión sobre el estado futuro del paciente (discapacidad clínica (física y cognitiva)) tras un determinado número de años (ej. 2, 5, 10 años a partir del momento de la opinión). El pronóstico clínico se puede realizar tanto en pacientes recién diagnosticados o tras el primer brote, así como en cualquier momento del curso de su enfermedad. El estado del paciente se puede definir en base a los síntomas de la esclerosis múltiple incluyen disminución de la capacidad para controlar movimientos pequeños, disminución en el periodo de atención, disminución de la coordinación, disminución de la capacidad de discernir, disminución de la memoria, depresión, dificultad para hablar o entender el lenguaje, mareos, visión doble, molestia en los ojos, dolor facial, fatiga, pérdida del equilibrio, problemas con el movimiento, lentamente progresivo y comenzando en las piernas, atrofia muscular, espasmos musculares (especialmente en las piernas), espasticidad muscular (espasmo incontrolable de grupos de músculos), entumecimiento o sensación anormal en cualquier área, dolor en brazos y piernas, parálisis de uno o ambos brazos o piernas, mala pronunciación, hormigueo, temblor en uno o ambos brazos o piernas, movimientos incontrolables y rápidos de los ojos, frecuencia urinaria aumentada, dificultad para orinar, urgencia urinaria, incontinencia urinaria, vértigo, pérdida de visión, anomalías para caminar/marcha y debilidad en uno o ambos brazos o piernas.

Por “valor modal” se entiende, en el contexto de la presente invención, un valor de la variable (en este caso, de los niveles de expresión) que particiona el intervalo de valores de dicha variable en dos o más subintervalos. Métodos adecuados para la determinación de dicho valor se han descrito en Dougherty, J. *et al.*, (Proc. of the 12th International Conference on Machine Learning; 1995. p. 194-202) y por Liu H. *et al.* (Data Mining and Knowledge Discovery, 2002, 6:393-423), cuyo contenido se incorpora en la presente solicitud en su totalidad. En una forma preferida de realización, para obtener dicho valor modal se realiza una discretización de la variable mediante un algoritmo de aprendizaje supervisado (computacional) de los umbrales de discretización más informativos con respecto a una variable de referencia (en la presente invención, el diagnóstico). Para el cálculo de los intervalos de discretización, si se parte de la secuencia de valores ordenada $A_i = \{v_1, v_2, \dots, v_n\}$, se evalúa la ganancia de información con respecto a una variable de referencia para todas las particiones posibles ($n - 1$). La partición que mayor ganancia de información tenga es la que se usa para comparar con el resto de atributos. Las decisiones de los nodos serán $[A[x] < v_i]$ y $[A[x] \geq v_i]$.

Posibles algoritmos de discretización adecuados para su utilización en la presente invención incluyen el árbol de decisión, el algoritmo de igual frecuencia y el algoritmo de igual distancia. En una forma preferida de realización, el algoritmo de discretización es el árbol de decisión.

Por “tablas de probabilidades condicionadas” se entiende, en el contexto de la presente invención, una tabla en la que se representan los posibles valores modales de expresión de un determinado gen o variable clínica, y en la que cada uno de los valores modales se correlaciona con una determinada probabilidad de que la enfermedad del paciente siga un pronóstico positivo o negativo. En una forma preferida de realización, las tablas de probabilidades condicionadas entre los valores modales de expresión de cada uno de los genes y los valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y entre los valores modales de cada una de las variables clínicas y los valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico son las indicada en la tabla 14.

Por “gen KLHDC5” se entiende el gen que codifica la proteína “kelch domain containing 5” cuya variante humana se encuentra descrita en la base de datos GenEMBL bajo el número de acceso BC108669.

Por “gen CASP2” se entiende el gen que codifica el precursor de la caspasa 2. La forma humana de dicho gen se encuentra descrita en la base de datos GenEMBL bajo los números de acceso U13021 y U13022.

Por “gen EMID1” se entiende el gen que codifica el precursor de la proteína “EMI domain-containing protein 1”. La forma humana de dicho gen se encuentra descrita en la base de datos GenEMBL bajo el número de acceso AJ416090.

Por “gen PRO1073” se entiende el gen descrito en la base de datos GenEMBL bajo el número de acceso AF113016.

Por “gen BTBD7” se entiende el gen que codifica la “BTB/POZ domain-containing protein 7”. La forma humana de dicho gen se encuentra descrita en la base de datos GenEMBL bajo el número de acceso BX538231.

Por “gen MGC25181” se entiende el gen que codifica la proteína hipotética MGC25181. La forma humana de dicho gen se encuentra descrita en la base de datos GenEMBL bajo el número de acceso AC114730.

Por “gen WDR20bis” se entiende el gen que codifica la “WD repeat-containing protein 20”. La forma humana de dicho gen se encuentra descrita en la base de datos GenEMBL bajo el número de acceso BC028387.

Por “gen NEK4” se entiende el gen que codifica la “quinase de serina/treonina Nek4”. La forma humana de dicho gen se encuentra descrita en la base de datos GenEMBL bajo el número de acceso L20321.

Por “gen SYLT2” se entiende el gen que codifica la “Synaptotagmin-like protein 2”. La forma humana de dicho gen se encuentra descrita en la base de datos GenEMBL bajo el número de acceso AK000170.

Por “gen DOCK10” se entiende el gen que codifica la “dedicator of cytokinesis protein 10”. La forma humana de dicho gen se encuentra descrita en la base de datos GenEMBL bajo el número de acceso AB014594.

ES 2 334 084 B1

Por “gen TTC10” se entiende el gen que codifica la “Tetratricopeptide repeat protein 10”. La forma humana de dicho gen se encuentra descrita en la base de datos GenEMBL bajo el número de acceso U20362.

5 Por “gen PTPRC” se entiende el gen que codifica el receptor tipo C de la proteína tirosina fosfatasa. La forma humana de dicho gen se encuentra descrita en la base de datos GenEMBL bajo el número de acceso BC017863.

Por “gen CTLA4” se entiende el gen que codifica el precursor de “Cytotoxic T-lymphocyte protein 4”. La forma humana de dicho gen se encuentra descrita en la base de datos GenEMBL bajo el número de acceso AF414120.

10 Por “EDSS” se entiende la escala expandida de estado de discapacidad de Kurtzke o “Kurtzke Expanded Disability Status Scale”, tal y como se define en Kurtzke, J.F. (Neurology, 1983, 33:1444-1452).

15 Por “MSFC” se entiende el compuesto funcional de la esclerosis múltiple o “Multiple Sclerosis Functional Composite”, tal y como se define en Fischer, J. S. *et al.* (National MS Society Clinical Outcomes Assessment Task Force. Mult. Scler. 1999, 5:244-250).

La determinación de los valores de expresión de un ácido nucleico se efectúa mediante la medida relativa de los niveles de expresión de un gen de interés frente a los niveles de expresión de un ácido nucleico de referencia. Dicha medidas pueden llevarse a cabo por cualquier método conocido para el experto en la materia, tales como los recogidos en Sambrook, J. *et al.* (Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989)) y Ausubel *et al.* (Current Protocols in Molecular Biology, eds. Ausubel *et al.*, John Wiley & Sons (1992)). Procedimientos típicos para la detección del polinucleótido resultante de la transcripción de un gen de interés incluyen la extracción del ARN de una célula o muestra de tejido, hibridación de dicha muestra con una sonda marcada, esto es, con un fragmento de ácido nucleico de secuencia complementaria a la molécula de ácido nucleico que se desea detectar, y detección de la sonda (por ejemplo, mediante Northern blotting). La invención contempla también la detección de los niveles de expresión de un determinado gen mediante el uso de cebadores en una reacción en cadena de la polimerasa (PCR), tales como PCR ancla, RACE PCR, reacción en cadena de ligasa (LCR). En una forma preferida de realización, la determinación de los valores modales de expresión se lleva a cabo mediante PCR en tiempo real.

30 Estos métodos incluyen las etapas de recoger una muestra de células de un sujeto, aislar los ARNm de dichas muestras, convertir los ARNm presentes en la muestra a ADN complementario (ADNc), poner en contacto la preparación de ADNc con uno o varios cebadores que hibriden específicamente con el gen diana en condiciones adecuadas para la hibridación y amplificación de dicho ácido nucleico seguido de la detección de la presencia de un producto de amplificación. Métodos alternativos de amplificación incluyen la replicación de secuencia autosostenida (Guatelli, J.C. *et al.*, (1990) Proc. Natl. Acad. Sci. USA 87:1874-1878), sistema de amplificación transcripcional Kwok, D.Y. *et al.*, (1989) Proc. Natl. Acad. Sci. USA 86:1173-1177), replicasa Q-Beta (Lizardi, P.M. *et al.* (1988) BioTechnology 6:1197) o cualquier otro método de amplificación de ácidos nucleicos conocido, seguido por la detección de las moléculas amplificadas usando técnicas de sobra conocidas para el experto en la materia. Estos métodos de detección son particularmente adecuados para la detección de ácidos nucleicos cuando dichas moléculas están presentes un número muy reducido de copias.

45 En otras formas de realización, se pueden utilizar los genes *per se* como marcadores de la esclerosis múltiple. Por ejemplo, el aumento de la expresión de un determinado gen puede ser debido a la duplicación del gen correspondiente, de forma que la duplicación pueda usarse como diagnóstico de la enfermedad. La detección del número de copias de un gen diana puede llevarse a cabo usando métodos de sobra conocidos para el experto en la materia. Típicamente, la determinación del número de copias de un determinado gen se lleva a cabo mediante Southern blot en el que el ADN total de una célula o de una muestra de tejido es extraída, hibridada con una sonda marcada y dicha sonda detectada. El marcaje de la sonda puede ser mediante un compuesto fluorescente, mediante un enzima o un cofactor enzimático. Otros métodos habituales para la detección/cuantificación de ADN incluyen secuenciación directa, cromatografía en columna y PCR cuantitativa usando protocolos estándar.

55 La determinación de los niveles de expresión de un gen se puede llevar a cabo en cualquier muestra biológica de un sujeto, incluyendo distintos tipos de fluidos biológicos, tales como sangre, suero, plasma, líquido cefalorraquídeo, líquido peritoneal, heces, orina y saliva, así como muestras de tejidos. Las muestras de fluidos biológicos pueden ser obtenidas por cualquier método convencional al igual que las muestras de tejidos; a modo ilustrativo dichas muestras de tejidos pueden ser muestras de biopsias obtenidas por resección quirúrgica.

60 El segundo método de la invención contempla la determinación simultánea de los valores de expresión de un mayor número de genes. Así, el segundo método de la invención puede incluir la determinación de los valores de expresión de al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10 y al menos 11 genes.

65 En una forma preferida de realización, el segundo método de la invención requiere de la determinación de los valores de expresión del gen KLHDC5 y de la variable clínica EDSS. En otra forma preferida de realización, el segundo método de la invención requiere determinar adicionalmente el valor de expresión del gen CASP2. En otra forma preferida, se determina adicionalmente el valor de expresión del gen EMID1. En una forma aún más preferida, se determina adicionalmente el valor de la variable clínica MSFC. En una forma de realización aún más prefe-

rida, el método comprende, adicionalmente, la determinación del valor de expresión del gen PRO1073. En otra forma preferida de realización, el segundo método de la invención incluye la determinación adicional del valor de expresión del gen BTBD7. En una forma de realización aún más preferida, el método implica la determinación adicional el valor de expresión del gen MGC2518. En otra forma de realización, el método invención implica la determinación adicional del valor de expresión del gen WDR20bis. En una forma de realización aún más preferida, el método de la invención implica la determinación adicional del valor de expresión del gen NEK4. En otra forma de realización, el método de la invención implica la determinación adicional del valor de expresión del gen DOCK10. En otra forma preferida de realización, el método de la invención implica la determinación adicional del valor de expresión del gen TTC10. En una forma de realización aún más preferida, el método de la invención implica la determinación adicional del valor de expresión del gen PTPRC. En otra forma preferida de realización, el método de la invención implica la determinación adicional del valor de expresión del gen CTLA4.

Adicionalmente, los inventores han puesto de manifiesto la existencia de distintos genes que se expresan de forma diferencial en pacientes diagnosticados con esclerosis múltiple de mal pronóstico con respecto a pacientes diagnosticados con esclerosis múltiple de buen pronóstico y con respecto a sujetos control, lo que permite el desarrollo de métodos pronósticos para predecir el desarrollo de la enfermedad. Adicionalmente, los autores de la presente invención han puesto de manifiesto la existencia de genes que se expresan diferencialmente en sujetos diagnosticados de esclerosis múltiple con respecto a pacientes sanos, lo que permite el uso de dichos genes con propósitos diagnósticos.

Así, en otro aspecto, la invención se relaciona con un método para determinar el pronóstico clínico de un sujeto afecto de esclerosis múltiple, para monitorizar el efecto de la terapia administrada a un sujeto afecto de esclerosis múltiple o para asignar una terapia personalizada a un sujeto afecto de esclerosis que comprende

(a) determinar el nivel de expresión de uno o varios genes seleccionados del grupo de genes listados en las posiciones 3, 5, 6, 7, 9, 11, 13, 16, 19, 20, 22, 24, 25, 26, 30, 31, 33, 34, 35, 37, 41 ó 43 de la tabla 3, o de los polipéptidos codificados por dichos genes, en una muestra biológica aislada del paciente y

(b) comparar los niveles de expresión de dichos genes o de dichos polipéptidos con un valor de referencia

en donde

(i) un aumento de la expresión de los genes en posición 6, 7, 9, 33, 35, 37 ó 43 o de los polipéptidos codificados por dichos genes o una disminución de la expresión de los genes en posición 3, 5, 11, 13, 16, 19, 22, 24, 25, 26, 30, 31, 34, 41 o de los polipéptidos codificados por dichos genes es indicativo de un mal pronóstico de la esclerosis múltiple en dicho sujeto, de que la terapia es inefectiva o de que el paciente es seleccionado para una terapia agresiva o

(ii) un aumento de la expresión de los genes en posiciones 3, 5, 11, 16, 20, 30 o de los polipéptidos codificados por dichos genes o una disminución de la expresión del gen en posición 43 o del polipéptido codificado por dicho gen es indicativo de un buen pronóstico de la esclerosis múltiple en dicho paciente, de que la terapia es efectiva o de que el paciente es seleccionado para no recibir terapia o para recibir una terapia poco agresiva.

En otro aspecto, la invención se relaciona con un método para determinar el pronóstico clínico de un sujeto afecto de esclerosis múltiple, para monitorizar el efecto de la terapia administrada a un sujeto afecto de esclerosis múltiple o para asignar una terapia personalizada a un sujeto afecto de esclerosis que comprende

(a) determinar el nivel de expresión de uno o varios genes seleccionados del grupo de genes listados en las posiciones 1 a 21 de la tabla 5, o de los polipéptidos codificados por dichos genes, en una muestra biológica aislada del paciente y

(b) comparar los niveles de expresión de dichos genes o de dichos polipéptidos con un valor de referencia calculado a partir de una o varias muestras obtenidas de pacientes de buen pronóstico y con un valor de referencia calculado a partir de una o varias muestras obtenidas de pacientes de mal pronóstico

en donde

(i) un aumento de la expresión de los genes en posición 1, 2, 3, 4, 5, 8, 9, 10, 14, 19, 20 ó 21 o de los polipéptidos codificados por dichos genes con respecto a un valor de referencia obtenido de una o varias muestras de pacientes diagnosticados de esclerosis múltiple de mal pronóstico es indicativo de un buen pronóstico de la esclerosis múltiple en dicho sujeto, de que la terapia es efectiva o de que el paciente es seleccionado para no recibir una terapia agresiva y

(ii) un aumento de la expresión de los genes en posiciones 6, 7, 11, 12, 13, 15, 16, 17 ó 18 o de los polipéptidos codificados por dichos genes con respecto a un valor de referencia obtenido de una o varias muestras de

ES 2 334 084 B1

pacientes diagnosticados de esclerosis múltiple de buen pronóstico es indicativo de un mal pronóstico de la esclerosis múltiple en dicho paciente, de que la terapia no es efectiva o de que el paciente es seleccionado para recibir terapia o para recibir una terapia poco agresiva.

5

En otro aspecto, la invención se relaciona con un método para identificar el pronóstico clínico de un sujeto afecto de esclerosis múltiple, para monitorizar el efecto de la terapia administrada a un sujeto afecto de esclerosis múltiple o para asignar una terapia personalizada a un sujeto afecto de esclerosis que comprende

10 (a) determinar el nivel de expresión de uno o varios genes seleccionados de la tabla 6 en una muestra aislada del paciente y

(b) comparar los niveles de expresión de dichos genes con un valor de referencia

15 en donde un aumento de la expresión de los genes en posición 4, 8, 11, 13, 15, 18, 19, 20, 21, 24, 25, 28, 30 ó 32 o una disminución de los genes en posición 1, 2, 3, 5, 6, 7, 9, 10, 12, 14, 16, 17, 22, 23, 26, 27, 29 ó 31 con respecto al valor de referencia es indicativo de un mal pronóstico de la esclerosis múltiple, de que la terapia no es efectiva o de que de que el paciente es seleccionado para una terapia agresiva.

20

En otro aspecto, la invención se relaciona con un método para determinar el pronóstico clínico de un sujeto afecto de esclerosis múltiple, para monitorizar el efecto de la terapia administrada a un sujeto afecto de esclerosis múltiple o para asignar una terapia personalizada a un sujeto afecto de esclerosis que comprende

25 (a) determinar el nivel de expresión de uno o varios genes seleccionados de la tabla 7 en una muestra aislada del paciente

(b) comparar los niveles de expresión de dichos genes con un valor de referencia

30 en donde un aumento de la expresión de los genes en posición 2, 5, 6, 7, 8 y 10 o una disminución de la expresión de los genes en posición 1, 3, 4 ó 9 con respecto al valor de referencia es indicativo de un buen pronóstico de la esclerosis múltiple o de que la terapia administrada es efectiva o de que el paciente es seleccionado para no recibir terapia o para recibir una terapia poco agresiva.

35

En otro aspecto, la invención se relaciona con un método para diagnosticar la esclerosis múltiple en un sujeto que comprende

40 (a) determinar el nivel de expresión de uno o varios genes seleccionados del grupo de genes indicado en la tabla 8 en una muestra aislada del sujeto

(b) comparar los niveles de expresión de dichos genes con un valor de referencia

45 en donde una disminución de la expresión de los genes en posición 1, 2, 6, 10, 15 ó 16 o un aumento en la expresión de los genes en posición 3, 4, 5, 7, 8, 9, 11, 12, 13 ó 14 con respecto a la muestra control es indicativo de que el sujeto sufre de esclerosis múltiple.

La definición de “determinación del pronóstico clínico” se ha descrito anteriormente.

50

Por “monitorizar el efecto de la terapia administrada a un sujeto afecto de esclerosis múltiple” se entiende, según la presente invención, el determinar si una terapia tiene alguna incidencia sobre el pronóstico.

55 Por “asignar una terapia personalizada a un sujeto afecto de esclerosis múltiple” se entiende el decidir en base al pronóstico de un individuo el tipo de terapia más adecuada para impedir la aparición de los síntomas anteriormente indicados. En casos de peor pronóstico, se aplica una terapia más agresiva desde el momento en que se detecta dicho peor pronóstico. Así, terapias más agresivas incluyen moduladores inmunitarios para ayudar a controlar el sistema inmunitario, incluyendo interferones (Avonex, Betaseron o Rebif), anticuerpos monoclonales (Tysabri) y acetato de glatirámico (Copaxone) y quimioterapia.

60

Por “valor de referencia” se entiende una medida de la expresión de un determinado gen o polipéptido que puede ser calculado o establecido a partir de una o varias muestras control. Estas pueden proceder de un sujeto sano, de un sujeto con esclerosis múltiple, o de sujetos con buen o mal pronóstico, de acuerdo con el objetivo del método.

65 El experto en la materia apreciará que la determinación de los niveles de expresión de los genes recogidos en las tablas 3, 5, 6, 7 y 8 puede llevarse a cabo usando técnicas conocidas para el experto en la materia.

ES 2 334 084 B1

La determinación de los niveles de expresión de un ácido nucleico relativa a los niveles de un ácido nucleico de referencia puede llevarse a cabo por cualquier método conocido para el experto en la materia, tal y como se ha descrito anteriormente.

5 En otras formas de realización, se pueden utilizar los genes *per se* como marcadores de la esclerosis múltiple en aquellos casos en los que el aumento de la expresión de un determinado gen puede ser debido a la duplicación del gen correspondiente, de forma que la duplicación puede usarse como diagnóstico de la enfermedad. La detección del número de copias de un gen diana puede llevarse a cabo usando los métodos descritos anteriormente.

10 Alternativamente, la invención contempla métodos para determinar el pronóstico clínico de un sujeto afecto de esclerosis múltiple o para monitorizar el efecto de la terapia administrada a un sujeto afecto de esclerosis múltiple o para asignar una terapia a un sujeto afecto de esclerosis múltiple en el que se determina el nivel de expresión de una o varias proteínas codificadas por los genes que se indican en las tablas 1 a 4. En este aspecto, la invención requiere la extracción de una muestra de proteínas de una muestra celular o de tejido seguido de la incubación de dicha muestra con un reactivo marcado capaz de unir específicamente dicha muestra (por ejemplo, un anticuerpo) y detección de dicho reactivo, en donde el marcador que incluye el reactivo se selecciona del grupo de un radioisótopo, un compuesto fluorescente, una enzima o un cofactor enzimático.

15 Métodos típicos de inmunodetección incluyen ELISA, RIA, ensayo inmunoradiométrico, fluoroensayo, ensayos quimioluminiscentes, ensayos bioluminiscentes y ensayos de Western blot.

20 En general, todos los inmunoensayos incluyen una etapa en la que una muestra que se sospecha contiene un antígeno determinado o en la que se desea conocer la concentración de dicho antígeno se pone en contacto con un primer anticuerpo en condiciones adecuadas para la formación de los complejos inmunes. Muestras adecuadas para la determinación incluyen una sección o una biopsia de un tejido, un extracto de un tejido o un fluido biológico. Una vez que se han formado los complejos antígeno anticuerpo, la preparación se somete a uno o varios lavados para eliminar anticuerpos que no se han unido de forma específica.

25 A continuación, se procede a la detección de los complejos inmunes mediante métodos ampliamente conocidos para el experto en la materia, tales como métodos radiactivos, fluorescentes, biológicos o basados en la determinación de una actividad enzimática.

30 Con objeto de aumentar la sensibilidad, es posible el uso de un ligando adicional como por ejemplo un segundo anticuerpo o un ligando acoplado a biotina. En esta situación, es necesaria una etapa adicional de incubación de los complejos ligando anticuerpo obtenidos en la primera etapa con el segundo anticuerpo en condiciones adecuadas para la formación de los complejos inmunes secundarios. Los complejos secundarios son sometidos a un ciclo de lavado para eliminar anticuerpos secundarios que se hayan unido de forma inespecífica y, a continuación, se determina la cantidad de complejo inmune secundario mediante determinación de la señal emitida por el anticuerpo secundario.

35 Métodos adicionales incluyen la detección de los complejos inmunes primarios mediante un procedimiento en dos pasos. En este procedimiento, un ligando secundario (un anticuerpo), que tienen una afinidad de unión por el anticuerpo que se encuentra formando parte de los complejos inmunes se pone en contacto con dichos complejos para formar complejos inmunes secundarios, tal y como se describió anteriormente. Tras una etapa de lavado, los complejos inmunes secundarios se ponen en contacto con un ligando terciario o anticuerpo que se une con alta afinidad al anticuerpo secundario para dar lugar a la formación de los complejos terciarios. El tercer ligando o anticuerpo se encuentra unido a un marcador detectable lo que permite la detección de los complejos inmunes terciarios.

40 Otros métodos de detección incluyen Western blotting, dot blotting, análisis mediante FACS y similares. En una forma de realización, los anticuerpos dirigidos contra los antígenos de la invención se inmovilizan en una superficie que muestra afinidad por las proteínas (por ejemplo poliestireno). A continuación se añade una composición en la que se encuentra el antígeno a detectar. Tras una etapa de lavado para eliminar los complejos unidos de forma inespecífica, el antígeno unido se puede detectar mediante un segundo anticuerpo que se encuentra acoplado a un marcador detectable. Este tipo de ELISA se denomina ELISA sándwich. La detección también puede llevarse a cabo mediante la adición de un segundo anticuerpo y un tercer anticuerpo que tiene afinidad por el segundo anticuerpo y que está unido a un marcador detectable.

45 En otro tipo de ELISA, las muestras que contienen el antígeno se inmovilizan y se detectan mediante un método competitivo en el que la muestra en la que se encuentra el antígeno que se desea detectar se mezcla con anticuerpos marcados frente a dicho antígeno y se añade sobre la superficie en la que se encuentra el anticuerpo inmovilizado. La presencia de antígeno en la muestra impide la unión del anticuerpo al antígeno inmovilizado de forma que la cantidad de anticuerpo que se une al antígeno inmovilizado se encuentra en relación inversa con la cantidad de antígeno en la muestra a analizar.

50 También es posible detectar el antígeno mediante inmunohistoquímica y microscopia confocal en secciones de tejido obtenidas a partir de muestra congeladas, fijadas en formaldehído o embebidas en parafina usando técnicas ampliamente conocidas para el experto en la materia.

ES 2 334 084 B1

La muestra de referencia que se usa para determinar la variación de los niveles de expresión de los genes y proteínas usadas en la presente invención. En una forma de realización, el valor de referencia se obtiene a partir de la señal proporcionada usando una muestra de tejido obtenida de un individuo sano. Preferiblemente, se toman muestras del mismo tejido de varios individuos sanos y se juntan, de forma que la cantidad de ARNm o de polipéptidos en la muestra refleje el valor medio de dichas moléculas en la población.

El método de la presente invención puede ser combinado con otros métodos diagnósticos (ej. Bandas oligoclonales en el LCR, Neuroimagen (RM, OCT), variables clínicas (escalas de discapacidad, tasa de brotes, edad, sexo) o marcadores biológicos: a) genéticos (polimorfismos, haplotipos); b) patrones patológicos en biopsia; c) anticuerpos, etc.

Los métodos de la presente invención son particularmente útiles para establecer el pronóstico en pacientes que han sufrido un único brote de esclerosis múltiple, de un paciente que sufre EM-RR o de un paciente que sufre EM-PP. Este método se aplicaría por tanto una vez durante el trabajo de diagnóstico de la enfermedad. También se podría aplicar a pacientes con la enfermedad ya diagnosticada pero que dado la gran variabilidad de la enfermedad, desconocen si esta estable o si va a progresar, nuevamente con carácter pronóstico y para decidir el tratamiento. Por tanto, la gran mayoría de pacientes realizarían el test al menos una vez, excepto aquellos con la enfermedad muy evolucionada donde la mala evolución ya es obvia y donde no hay posibilidades de elegir entre tratamientos. Un subgrupo de pacientes podrían realizar el test en varias ocasiones si a lo largo de años parece que cambia el curso clínico de la enfermedad y se quiere re-asegurar el pronóstico.

- En el caso de tener un pronóstico favorable, el medico podrá recomendar un seguimiento periódico y valorar si aún así requiere algún tratamiento inmunomodulador, pudiendo elegir el mas conveniente o cómodo para el paciente dado el carácter leve de su enfermedad. También esta información es critica para el paciente porque puede tomar decisiones fundamentales para su vida como casarse, tener hijos, tipo de trabajo, nivel de estrés y riesgos en su vida, seguro médico, seguro de vida, tipo de vivienda, etc.
- En el caso de que el pronóstico sea desfavorable, el medico recomendaría mas enérgicamente realizar tratamientos inmunomoduladores y probablemente emplear ya de inicio los mas eficaces y de segunda línea (por ejemplo el Tysabri) o realizar terapia combinada o quimioterapia. Por otro el paciente puede decir de una forma mas informada los riesgos que puede asumir debido a la terapia más agresiva que se plantea, así como decidir sobre su vida en los aspectos vitales: casarse, tener hijos, tipo de trabajo, nivel de estrés y riesgos en su vida, seguro médico, seguro de vida, tipo de vivienda, etc.

En principio, cualquier muestra aislada de un paciente puede ser usada en los métodos de la invención. Así, la determinación de los niveles de ARNm o polipéptidos puede efectuarse en una biopsia de tejido o en un fluido biológico (suero, saliva, semen, esputo, LCR, lágrimas, mucosidades, sudor, leche y similares). La determinación puede llevarse a cabo en homogenados de tejido o en fracciones más o menos clarificadas de los mismos. En una forma de realización preferida, la determinación de los niveles de ARNm y de los polipéptidos de la invención se lleva a cabo a partir de extractos de células mononucleares obtenidas de la sangre periférica.

En el caso de que se desee determinar simultáneamente los niveles de expresión de varios de los genes identificados en la presente invención, resultan de utilidad composiciones que contienen al menos una copia de una sonda específica para cada uno de los genes indicados en al menos una tabla seleccionada del grupo de las tablas 3, 5-8 y 11.

Así, en otro aspecto, la invención se relaciona con un kit que comprende una sonda específica para cada uno de los genes indicados en al menos una tabla seleccionada del grupo de tablas 3, 5-8 y 11.

Por "kit" se entiende, en el contexto de la presente invención, un producto que contiene los distintos reactivos necesarios para llevar a cabo los métodos de la invención empaquetados para permitir su transporte y almacenamiento. Materiales adecuados para el empaquetado de los componentes del kit incluyen cristal, plástico (polietileno, polipropileno, policarbonato y similares), botellas, viales, papel, sobres y similares. Adicionalmente, los kits de la invención pueden contener instrucciones para el uso simultáneo, secuencial o separado de los distintos componentes que se encuentran en el kit. Dichas instrucciones pueden encontrarse en forma de material impreso o en forma de un soporte electrónico capaz de almacenar instrucciones de forma que puedan ser leídas por un sujeto, tales como medios de almacenamiento electrónicos (discos magnéticos, cintas y similares), medios ópticos (CD-ROM, DVD) y similares. Adicional o alternativamente, los medios pueden contener direcciones de Internet que proporcionen dichas instrucciones.

En una forma preferida de realización, el kit de la invención consiste en una sonda específica para cada uno de los genes indicados en al menos una tabla seleccionada del grupo de tablas 3, 5-8 y 11.

En una forma preferida de realización, los genes que forman parte de la matriz son los genes indicados en la tabla 11 y al menos un gen de referencia.

En otra forma de realización preferida, las sondas o los anticuerpos que forman el kit de la invención se encuentran acoplados a una matriz.

5 En el caso de que se desee determinar simultáneamente los niveles de expresión de varios de los genes identificados en la presente invención, resulta de utilidad la inclusión sondas para todos los genes cuya expresión se desea determinar en una micromatriz de hibridación.

10 Las micromatrices comprenden una pluralidad de ácidos nucleicos distribuidas espacialmente y asociadas de forma estable a un soporte (por ejemplo, un biochip). Los ácidos nucleicos tienen una secuencia complementaria a subsecuencias particulares de los genes cuya expresión se desea detectar, por lo que son capaces de hibridar con dichos ácidos nucleicos. En los métodos de la invención, se pone en contacto una micromatriz que comprende una matriz de ácidos nucleicos con una preparación de ácidos nucleicos aislada del paciente objeto de estudio. La incubación de la micromatriz con la preparación de ácidos nucleicos se lleva a cabo en condiciones adecuadas para la hibridación. Posteriormente, tras la eliminación de los ácidos nucleicos que no han quedado retenidos en el soporte, se detecta el patrón de hibridación 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.

20 La invención contempla una variedad de matrices en cuanto al tipo de sondas como en cuanto al 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 los que el enlace fosfodiéster se ha sustituido por un enlace fosforotioato, metilimino, metilfosfonato, forforamidate, guanidina y similares, ácidos nucleicos en los que la ribosa de los nucleótidos se ha sustituido por otra hexosa, ácidos peptidonucleicos (PNA). La longitud de las sondas puede ser de 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 100 nucleótidos y variar en el rango de 10 a 1000 nucleótidos, preferiblemente en el rango de 15 a 150 nucleótidos, más preferiblemente en el rango de 15 a 100 nucleótidos y pueden ser ácidos nucleicos de cadena sencilla o de cadena doble.

30 La selección de las sondas específicas para los distintos genes diana se lleva a cabo de forma que se unan de forma específica al ácido nucleico diana con una hibridación mínima a genes no relacionados. Sin embargo, existen sondas de 20 nucleótidos que no son únicas para un ARNm determinado. Por tanto, sondas dirigidas a dichas secuencias mostrarán una hibridación cruzada con secuencias idénticas que aparecen en ARNm de genes no relacionados. Por otro lado, existen sondas que no hibridan de forma específica con los genes diana en las condiciones utilizadas (a causa de estructuras secundarias o de interacciones con el sustrato de la matriz). Este tipo de sondas no deben ser incluidas en la matriz. Por lo tanto, el experto en la materia advertirá que las sondas que van a incorporarse en una matriz determinada deberán ser optimizadas antes de su incorporación a la matriz. De forma general, la optimización de las sondas se lleva a cabo generando una matriz que contiene una pluralidad de sondas dirigidas a las distintas regiones de un determinado polinucleótido diana. Esta matriz se pone en contacto en primer lugar con una muestra que contiene el ácido nucleico diana de forma aislada y, en segundo lugar, con una mezcla compleja de ácidos nucleicos. De esta forma se seleccionan sondas que muestran una hibridación altamente específica con el ácido nucleico diana pero no baja o ninguna hibridación con la muestra compleja son seleccionadas para su incorporación a las matrices de la invención. Adicionalmente, es posible incluir en la matriz controles de hibridación para cada una de las sondas que va a ser estudiada. En una forma preferida de realización, 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.

50 En una forma preferida de realización, la matriz contiene una pluralidad de sondas complementarias a subsecuencias del ácido nucleico diana de una longitud constante o de longitud variable en un rango de 5 a 50 nucleótidos. La matriz puede contener todas las sondas específicas de un ARNm determinado de una determinada longitud o puede contener sondas seleccionadas de distintas regiones de un ARNm. Cada sonda se ensaya en paralelo con una sonda con una base cambiada, preferentemente en posición central de la sonda. La matriz se pone en contacto con una muestra que contenga 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 con los correspondientes controles de hibridación. Se seleccionan aquellas sondas en las que se observe una mayor diferencia entre la señal de hibridación con la sonda y su control de hibridación. El proceso de optimización puede incluir una segunda ronda de optimización en el que la matriz de hibridación se hibrida con una muestra que no contiene secuencias complementarias a las sondas de la matriz. Tras la segunda ronda de selección se seleccionarán aquellas sondas que presenten señales de hibridación inferiores a un nivel umbral. De esta forma, se seleccionan sondas que superan ambos controles, es decir, que muestran un nivel mínimo de hibridación inespecífica y un nivel máximo de hibridación específica con el ácido nucleico diana.

65 Las micromatrices de la invención contienen no sólo sondas específicas para los polinucleótidos indicativos de una determinada situación patofisiológica, sino que contienen también una serie de sondas control, que pueden ser de tres tipos: controles de normalización, controles de niveles de expresión y controles de hibridación.

Controles de normalización son oligonucleótidos que son perfectamente complementarios a secuencias de referencia marcadas que se añaden a la preparación de ácidos nucleicos que se desea analizar. Las señales derivadas de los controles de normalización tras la hibridación proporcionan una indicación de las variaciones en las condiciones

de hibridación, intensidad del marcador, eficiencia de la detección y otra serie de factores que puede resultar en una variación de la señal de hibridación entre distintas micromatrices. Preferiblemente, las señales detectadas a partir del resto de sondas de la matriz se dividen por la señal emitida por las sondas control normalizando así las medidas. Prácticamente se puede usar cualquier sonda como control de normalización. Sin embargo, es conocido que la eficacia de la hibridación varía en función de la composición de nucleótidos y la longitud de la sonda. Por tanto, sondas de normalización preferidas son aquellas que representan la longitud media de las sondas presentes en la matriz, aunque pueden ser seleccionadas de forma que recojan un rango de longitudes que reflejen el resto de sondas presentes en la matriz. Las sondas de normalización pueden ser diseñadas de forma que reflejen la composición media de nucleótidos del resto de sondas presentes en la matriz. Preferiblemente, se utiliza un número limitado de sondas de normalización seleccionadas de forma que hibridan adecuadamente, es decir, no presentan estructura secundaria y no muestran similitud de secuencia con ninguna de las sondas de la matriz. Las sondas de normalización pueden localizarse en cualquier posición en la matriz o en múltiple posiciones en la matriz para controlar de forma eficiente variaciones en la eficiencia de hibridación relacionadas con la estructura de la matriz. Preferiblemente, los controles de normalización se localizan en las esquinas de la matriz y/o en el centro de la misma.

Los controles de los niveles de expresión son sondas que hibridan de forma específica 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 con el nivel de expresión del ácido nucleico diana indica si las variaciones en los niveles de expresión son debidas a cambios en los niveles de expresión o son debidas a cambios en la tasa transcripcional global en la célula o en su actividad metabólica general. Así, en el caso de células que presentan deficiencias en un determinado metabolito esencial para la viabilidad celular, se espera que se observe 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 del gen control, es probable que se deba a un aumento de la actividad metabólica de la célula y no a un aumento diferencial en la expresión del gen diana. Se puede emplear cualquier sonda que corresponda a un gen expresado de forma constitutiva tales como genes que codifican para proteínas que ejercen funciones esenciales de las células tales como β -2-microglobulina, ubiquitina, proteína ribosomal 18S, ciclofilina A, receptor de transferina, actina, GAPDH y similares. En una forma preferida de realización, los controles de los niveles de expresión son GAPDH, proteína de activación de la tirosina 3-monooxigenasa/triptófano 5-monooxigenasa (YWHAZ), ubiquitina, beta-actina y β -2-microglobulina.

Los controles de hibridación se pueden incluir tanto para las sondas dirigidas a los genes diana como para las sondas dirigidas al nivel de expresión o a los controles de normalización. Los 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 forma que, aplicando las condiciones adecuadas de hibridación, el gen diana debería hibridar con la sonda específica pero no con el control de hibridación o con una eficiencia reducida. Preferiblemente, los controles de hibridación contienen una o varias posiciones modificadas en el centro de la sonda. Los controles de hibridación proporcionan por tanto una indicación del grado de hibridación inespecífica o de hibridación cruzada a un ácido nucleico en la muestra a una sonda distinta a la que contiene la secuencia exactamente complementaria.

Las matrices de la invención pueden contener también controles de preparación de muestra y de amplificación que son sondas complementarias a subsecuencias de genes controles seleccionados porque normalmente no aparecen en la muestra biológica objeto de 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 a 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 dispone de un conjunto de sondas que muestran la especificidad adecuada y de un conjunto de sondas controles, éstas se disponen en la matriz en una posición conocida de forma que, tras las etapas de hibridación y de detección, sea posible establecer una correlación entre una señal positiva de hibridación y el gen particular a partir de las coordenadas de la matriz en la que se detecte la señal de hibridación positiva.

Las micromatrices pueden ser matrices de alta densidad con miles de oligonucleótidos mediante métodos fotolito-gráficos de síntesis *in situ* (Fodor *et al.*, 1991, Science, 767-773). Este tipo de sondas son habitualmente redundantes, es decir, incluyen varias sondas por cada ARNm que se desea detectar.

En una forma preferida de realización, las matrices son matrices de baja densidad o LDA que contienen menos de 10000 sondas en cada por centímetro cuadrado. En dichas matrices de baja densidad, las distintas sondas se aplican manualmente con ayuda de una pipeta en distintas localizaciones de un soporte sólido (por ejemplo, una superficie de cristal, una membrana). Los soportes usados para fijar las sondas pueden ser obtenidos a partir de una gran variedad de materiales, incluyendo plástico, cerámica, metales, geles, membranas, cristales y similares. Las micromatrices se pueden obtener usando cualquier metodología conocida para el experto en la materia.

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

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 están marcados con biotina, se pone en contacto la matriz con estreptavidina conjugada con un reactivo fluorescente en condiciones adecuadas para que se produzca la unión entre la biotina y la estreptavidina. Tras la incubación de la micromatriz con el sistema que genera la señal detectable, es necesario efectuar una etapa de lavado para eliminar toda las moléculas que se hayan unido de forma inespecífica a la matriz. Las condiciones de lavado serán determinadas por el experto en la materia usando usando condiciones adecuadas en función del sistema que genera la señal detectable que haya sido utilizado y que son de sobra conocidas por el experto en la materia.

El patrón de hibridación resultante se puede visualizar o detectar de distintas maneras, estando dicha detección determinada por el tipo de sistema utilizado en la micromatriz. Así, la detección del patrón de hibridación puede llevarse a cabo mediante conteo de centelleo, autorradiografía, determinación de una señal fluorescente, determinaciones calorimétricas, detección de una señal luminosa y similares.

Antes de la etapa de detección, es posible tratar las micromatrices con una endonucleasa específica para el ADN de cadena sencilla, de forma que el ADN que se ha unido a la matriz de forma inespecífica se elimina mientras que el ADN de doble cadena resultante de la hibridación de las sondas de la matriz con los ácidos nucleicos de la muestra objeto de estudio permanece inalterada. Endonucleasas adecuadas para este tratamiento incluyen la nucleasa S1, la nucleasa de frijol y similares. En el caso de que el tratamiento con endonucleasa se lleve a cabo en un ensayo en el que al ácido nucleico diana no está marcado con una molécula directamente detectable (por ejemplo, en un ensayo en el que el ácido nucleico diana se encuentra biotinilado), el tratamiento con endonucleasa se llevará a cabo antes de poner en contacto la micromatriz con los otros miembros del sistema que produce la señal detectable.

Tras la hibridación y los eventuales procesos de lavado y tratamiento subsecuentes, el patrón de hibridación se detecta y cuantifica, para lo que la señal correspondiente a cada punto de hibridación en la matriz se compara a un valor de referencia correspondiente a la señal emitida por un número conocido de ácidos nucleicos marcados terminalmente para obtener así un valor absoluto del número de copias de cada ácido nucleico que es hibridado en un punto determinado de la micromatriz.

En el caso de que se desee determinar simultáneamente los niveles de expresión de varios de las proteínas identificadas en la presente invención, resultan de utilidad composiciones que contienen al menos un anticuerpo específico para cada uno de los genes indicados en al menos una tabla seleccionada del grupo de tabla 1 a 4. Para este propósito, resultan de utilidad las matrices de anticuerpos tales como los descritos 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 and Schreiber (2000) *Science* 289:1760-1763; WO 01/40803 y WO 99/51773A1. 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 a antígeno, tales como Fab', Fab, F(ab')₂, anticuerpos de dominio único (single domain antibodies o DABS), Fv, scFv y similares. Las técnicas para preparar dichos anticuerpos son de sobra conocidas para 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 de la matriz se pueden aplicar a alta velocidad, por ejemplo, usando sistemas robóticos disponibles comercialmente (por ejemplo, los producidos por Genetic Microsystems o Biorobotics). El sustrato de la matriz puede ser nitrocelulosa, plástico, cristal o puede ser de un material poroso como por ejemplo, archilamida, agarosa u otro polímero. En otra forma de realización, es posible utilizar células que producen los anticuerpos específicos para detectar las proteínas de la invención mediante su cultivo en filtros en matriz. Tras la inducción de la expresión de los anticuerpos, éstos se inmovilizan en el filtro en la posición de la matriz donde se encontraba la célula productora.

Una matriz de anticuerpos puede ponerse en contacto con una diana marcada y determinar el nivel de unión de la diana a los anticuerpos inmovilizados. Si la diana no se encuentra marcada, se puede utilizar un ensayo tipo sándwich en el que se usa un segundo anticuerpo marcado específico para el polipéptido que se une al polipéptido que forma se encuentra inmovilizado 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 puede producirse por duplicado y utilizarse para comparar los perfiles de unión de dos muestras distintas.

En otro aspecto, la invención se relaciona con el uso de un kit de la invención para la determinación del pronóstico de un enfermo diagnosticado de esclerosis múltiple, para determinar la eficacia de un tratamiento para la esclerosis múltiple o para diagnosticar la esclerosis múltiple en un paciente.

La invención se describe a continuación mediante los siguientes ejemplos que deben ser considerados como ilustrativos y no limitativos del alcance de la invención.

Ejemplos*Materiales y métodos*5 1. *Screening con chips de ADN*

Se reclutaron 6 pacientes de esclerosis múltiple, 3 de ellos diagnosticados como mal pronóstico y 3 como buen pronóstico, y 3 controles sanos sin antecedentes de ninguna enfermedad autoinmune. El pronóstico de los pacientes se determinó mediante datos clínicos asociados con la evolución de la esclerosis múltiple en estudios de historia natural de la esclerosis múltiple como tipo de primer brote, tiempo hasta el segundo brote, número de brotes en los primeros 2 a 5 años y secuelas iniciales (Tabla 1).

15

TABLA 1

Marcadores clínicos de buen y mal pronóstico

20

25

30

35

40

| Marcadores pronósticos clínicos | Valoración + buen pronóstico - mal pronóstico | Referencia bibliográfica |
|--|--|-------------------------------------|
| Clínica de inicio cerebelosa, piramidal o troncoencefálica. | - | 1,2,5,9,6,13,17. |
| Clínica de inicio sensitiva o neuritis óptica. | + | 1,2,9,13 |
| Clínica de inicio polisintomática (afectación de tres o más sistemas funcionales) | - | 1,8,7 |
| Tiempo a 2º brote < 1 año. | - | 1,2,4,9,13 |
| Dos o más brotes en los primeros dos años. | + | 1,2,4,5,13,17. |
| Tiempo mayor o igual a 5 años hasta EDSS de 3 | + | 1,4,13,17. |
| Persistencia de la clínica inicial más de 1 año. | - | 1,9,8 |
| Recuperación buena tras los primeros dos brotes: (valoración al año del brote con EDSS menor o igual a 1,5). | + | 2,7, |

45 **Referencias bibliográficas**

1.- M.A. **Hernández Pérez**. Factores pronósticos en la EM. *Neurología* 2001; 16 [supl 1]: 37-42.

2.- **Scott TF, Schramke CJ, Novero J, Chieffe C**. Short-term prognosis in early relapsing-remitting multiple sclerosis. *Neurology* 2000 Sep 12; 55(5): 689-93.

3.- **Brex PA, Ciccarelli O, O'Riordan JI, Sailer M, Thompson AJ, Miller DH**. A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *N Engl J Med* 2002 Jan 17; 346(3): 158-64.

4.- **Weinshenker BG, Bass B, Rice GP, Noseworthy J, Carriere W, Baskerville J, Ebers GC**. The natural history of multiple sclerosis: a geographically based study. 2. Predictive value of the early clinical course. *Brain* 1989 Dec; 112 (Pt 6):1419-28.

5.- **Weinshenker BG, Rice GP, Noseworthy JH, Carriere W, Baskerville J, Ebers GC**. The natural history of multiple sclerosis: a geographically based study. 3. Multivariate analysis of predictive factors and models of outcome. *Brain* 1991 Apr; 114 (Pt 2):1045-56.

6.- **Miller DH, Hornabrook RW, Purdie G.J**. The natural history of multiple sclerosis: a regional study with some longitudinal data. *Neurol Neurosurg Psychiatry* 1992 May; 55(5):341-6.

7.- **Runmarker B, Andersen O**. Prognostic factors in a multiple sclerosis incidence cohort with twenty-five years of follow-up. *Brain* 1993 Feb; 116 (Pt 1):117-34.

- 8.- **Runmarker B, Andersson C, Oden A, Andersen O.** Prediction of outcome in multiple sclerosis based on multivariate models. *J Neurol* 1994 Oct; 241(10):597-604.
- 9.- **Phadke JG.** Clinical aspects of multiple sclerosis in north-east Scotland with particular reference to its course and prognosis. *Brain* 1990 Dec; 113 (Pt 6):1597-628.
- 10.- **Avasarala JR, Cross AH, Trotter JL.** Oligoclonal band number as a marker for prognosis in multiple sclerosis. *Arch Neurol* 2001 Dec; 58(12):2044-5.
- 11.- **Lin X, Blumhardt LD.** Inflammation and atrophy in multiple sclerosis: MRI associations with disease course. *J Neurol Sci* 2001 Aug 15; 189(1-2):99-104.
- 12.- **Simon JH.** Brain and spinal cord atrophy in multiple sclerosis. *Neuroimaging Clin N Am* 2000 Nov; 10(4):753-70, ix.
- 13.- Multiple sclerosis. McAlpine's. Third edition. Alastair Compston. Churchill Livingstone.
- 14.- **Kappos L, Moeri D, Radue EW, Schoetzau A, Schweikert K, Barkhof F, Miller D, Guttmann CR, Weiner HL, Gasperini C, Filippi M.** Predictive value of gadolinium-enhanced magnetic resonance imaging for relapse rate and changes in disability or impairment in multiple sclerosis: a meta-analysis. Gadolinium MRI Meta-analysis Group. *Lancet* 1999 Mar 20; 353(9157):964-9.
- 15.- **Rovaris M, Filippi M.** Contrast enhancement and the acute lesion in multiple sclerosis. *Neuroimaging Clin N Am* 2000 Nov; 10(4):705-16, viii-ix.
- 16.- **Losseff NA, Miller DH, Kidd D, Thompson AJ.** The predictive value of gadolinium enhancement for long term disability in relapsing-remitting multiple sclerosis-preliminary results. *Mult Scler* 2001 Feb; 7(1):23-5.
- 17.- Esclerosis múltiple, Bases clínicas y patogénicas. Cedric S. **Raine**, Henry F. **McFarland**, Wallace W. **Tourtellotte**. *Edimsa*.

La purificación de RNA total se realizó a partir de sangre periférica utilizando el PAXgene™ Blood RNA Kit de PreAnalytiX. El uso de este kit permite preservar el perfil de expresión de RNA nada más realizar la extracción sanguínea. Durante la purificación de RNA total se realizó un tratamiento con DNasa para eliminar la posible contaminación de DNA. Las muestras se concentraron mediante Speed-vac y se estimó la calidad y cantidad del RNA purificado mediante chequeo de una alícuota en un gel de agarosa y medida espectrofotométrica.

A partir de 6 µg de RNA total se sintetizó cDNA con el Kit SuperScript Choice System de Life Technologies, siguiendo el protocolo del Expression Analysis Technical Manual de Affymetrix. De este cDNA se sintetizó cRNA siguiendo el protocolo del Kit BioArray HighYield RNA Transcript Labeling (T7) de Enzo. El cRNA así sintetizado se purificó con el Kit Clean-up module de Affymetrix, recuperándose en un volumen final de 22 µl de agua. Una vez sintetizado y purificado, el cRNA se fragmentó (15 µg de cada muestra) para preparar las mezclas de hibridación.

Tanto la hibridación como el revelado y escaneo de los chips se realizaron siguiendo los protocolos y equipos oficialmente recomendados por Affymetrix Inc. El chip utilizado fue el HG-U133 Plus 2.0. Los resultados del chip fueron analizados utilizando el software Microarray Suite 5.0 (MAS 5.0; Affymetrix®) y Biometric Research Branch (BRB) Array Tools 3.2.3 (Dr. Richard Simon y Amy Peng Lam).

2. Validación por PCR en tiempo real y construcción del clasificador

Se reclutaron 40 pacientes de esclerosis múltiple, 20 de ellos diagnosticados como mal pronóstico y 20 como buen pronóstico, y 20 controles sanos sin antecedentes de ninguna enfermedad autoinmune empleando los criterios clínicos antes descritos.

La purificación de RNA total se realizó a partir de células mononucleares de sangre periférica. Mediante centrifugación con gradiente de densidad utilizando Ficoll-Paque de Pharmacia Biotech se purificaron los linfocitos y monocitos y se sumergieron inmediatamente en RNAlater RNA Stabilization Reagent de Qiagen para preservar los patrones de expresión génica. El ARN total se purificó utilizando el RNeasy Mini Kit de Qiagen y durante la purificación se eliminaron los restos de ADN mediante un tratamiento con ADNasa utilizando el RNase-Free DNase Set de Qiagen. La síntesis de cDNA a partir del ARN total se realizó utilizando el High-Capacity cDNA Archive Kit de Applied Biosystems.

El análisis de validación de genes se ha realizado utilizando la tecnología de Low Density Arrays (LDAs; Applied Biosystems). Las LDAs contienen 384 pocillos. Los pocillos contienen ensayos TaqMan validados por Applied Biosystems y la distribución de los ensayos es configurable por el usuario. En este proyecto el diseño de las placas elegido es el de 95 genes +1 control analizados por duplicado y con dos muestras estudiadas en cada placa.

Teniendo en cuenta que para las LDAs sólo se pueden seleccionar aquellos ensayos que están inventariados por Applied Biosystems, el proceso de selección de los ensayos más idóneos a partir de los genes que se quieren validar siguió los siguientes criterios:

- 5 - La distancia del probe set de Affymetrix a la sonda de Applied Biosystems será la mínima posible.
- El ensayo no debería detectar ADN genómico.
- 10 - Se seleccionarán un mínimo de cuatro genes constitucionales para el proceso de normalización.

Como primer paso del análisis se descartaron aquellas muestras que presenten una desviación estándar entre replicas del mismo ensayo de PCR mayor a 0,38. Se toma 0,38 como valor límite ya que esto indica que existe una diferencia de 0,75 entre el mínimo y el máximo de las Ct. Como cada Ct equivale a un ciclo de la PCR y en cada ciclo se duplica la cantidad de ADN, la desviación estándar de 0,38 indica que casi hay el doble de cantidad de ADN en un pocillo que en el otro. A continuación y para calcular los valores de expresión de cada gen se aplica la fórmula $2^{\text{exp}} (\text{Ctmin} - \text{Ctsample})$, donde Ctmin es el valor mínimo de Ct de cada gen en todas las muestras y Ctsample es el valor de Ct de ese gen en esa muestra.

20 Tomando los valores de de expresión de los genes constitutivos tras el procesamiento (5 en este caso; GAPDH, YWHAZ, UBC, ACTB, B2M) se calculó el factor de normalización de cada muestra mediante el programa geNorm (<http://medgen.ugent.be/~jvdesomp/genorm/>) que calculará la media geométrica del valor de expresión de un número de genes constitutivos o controles internos. Estos genes controles internos se eligieron de acuerdo a aquellos genes que se observó que en el experimento de análisis de expresión génica en chips de ADN anteriormente realizados menos variaban su expresión entre las condiciones estudiadas. Una vez obtenido el factor de normalización de cada muestra se normalizan los datos de cada gen en cada muestra respecto a este factor de normalización obtenido para dicha muestra del programa geNorm.

30 Para el análisis estadístico los datos normalizados respecto de los genes control se transformaron a escala logarítmica (base 2). Para el cálculo de genes significativos se aplicó un test no paramétrico que dependerá de si se comparan 2 condiciones (U de Mann-Whitney para 2 muestras independientes) o 3 (H de Kruskal Wallis para 3 muestras independientes). En todos los casos los valores de $p < 0.05$ fueron considerados diferencias significativas mientras que los valores de $p < 0.01$ fueron considerados diferencias muy significativas. El análisis estadístico se realizó utilizando el programa SPSS 11.0 (SPSS Inc., Chicago, USA).

35 El clasificador Bayesiano se construyó utilizando el software de análisis bayesiano BayesiaLab 3.2 (Bayesia SA. Laval Cedex, France). Para ello previamente se discretizaron las variables en un máximo de cuatro intervalos usando el algoritmo de discretización "Decisión Tree" teniendo la variable diagnóstico como referencia. El aprendizaje se realizó utilizando el algoritmo "Naive Bayes Aumentado".

40 *Resultados*

1. *Screening con chips de ADN*

45 La Tabla 2 muestra las características demográficas y clínicas de los pacientes de esclerosis múltiple y de los controles sanos utilizados para realizar el screening con chips de ADN.

TABLA 2

Características demográficas y clínicas

| | Controles sanos | Buen pronóstico | Mal pronóstico |
|----------------------------|-----------------|-----------------|----------------|
| 55 N | 3 | 3 | 3 |
| Varón/Mujer | 1/2 | 1/2 | 1/2 |
| 60 Edad (años) | 33,0 ± 2,94 | 38,0 ± 6,83 | 33,8 ± 7,63 |
| Puntuación EDSS | | 0,75 ± 0,50 | 2,86 ± 1,18 |
| Duración enfermedad (años) | | 7,00 ± 0,82 | 1,25 ± 0,50 |
| 65 Brotes en el último año | | 0,25 ± 0,50 | 2,50 ± 0,58 |

ES 2 334 084 B1

Tras normalizar los niveles de expresión utilizando MAS 5.0 se procedió a realizar el filtrado de las sondas (genes) utilizando el BRB Array Tools 3.2.3 según los siguientes criterios:

1. A aquellos genes que presentaban un valor de intensidad menor de 10 se le asignó dicho valor.
2. Un gen era eliminado si menos del 20% de valores de los datos de expresión tenían por lo menos un cambio de 1.5 veces en cualquier dirección del valor de la mediana.
3. Un gen era eliminado si el porcentaje de datos perdidos o filtrados superaba el 50%.
4. Un gen era eliminado si el porcentaje de valores de los datos de expresión ausentes superaba el 70%.

Tras el filtrado cumplieron los criterios 4.705 genes de los 54.675 iniciales. Con estos genes se realizó un test de comparación de clases y se identificaron 45 de ellos que presentaban diferencias significativas entre las tres clases (control, mal y buen pronóstico) con un valor de $p < 0,001$ (Tabla 3).

TABLA 3

Genes que presentaban diferencias significativas entre las tres clases (control, mal y buen pronóstico) con un valor de $p < 0,001$

| | Valor p | Controles Sanos | Mal pronóstico | Buen pronóstico | Sonda | Descripción | Anotación | Símbolo gen | Listas genes |
|---|----------|-----------------|----------------|-----------------|---------------------|---|----------------------|-----------------------|--------------|
| 1 | 2e-07 | 15.9 | 39.7 | 26 | <u>1557278_s_at</u> | CDNA FLJ33199 fis. clone ADRGL2006377 | Info | | |
| 2 | 3e-07 | 20.6 | 10 | 10 | <u>229190_at</u> | CDNA FLJ90295 fis. clone NT2RP2000240. | Info | | |
| 3 | 7.4e-06 | 37.8 | 10.7 | 42.7 | <u>205306_x_at</u> | kynurenine 3-monooxygenase (kynurenine 3-hydroxylase) | Info | KMO | |
| 4 | 1.07e-05 | 59.8 | 36.3 | 35.4 | <u>227541_at</u> | WD repeat domain 20 | Info | WDR20 | |
| 5 | 1.19e-05 | 36.8 | 10.4 | 45.9 | <u>235401_s_at</u> | Fc receptor homolog expressed in B cells | Info | FREB | |
| 6 | 2.07e-05 | 61 | 135.6 | 62.4 | <u>223226_x_at</u> | single stranded DNA binding protein 4 | Info | SSBP4 | |
| 7 | 2.13e-05 | 10.6 | 22.6 | 10 | <u>210436_at</u> | chaperonin | Info | CCT8 | |

ES 2 334 084 B1

| | | | | | | | | | | |
|----|-----------|-----------|-------|-------|-------|--|---|----------------------|---------------------------|--------------------------|
| | | | | | | containing TCP1, subunit 8 (theta) | | | | |
| 5 | 8 | 6.14e-05 | 117.5 | 69.7 | 80.6 | 224945_at | BTB (POZ) domain containing 7 | Info | BTBD7 | |
| | 9 | 9.61e-05 | 10 | 20.3 | 10.7 | 1570043_at | | Info | | |
| 10 | 10 | 0.0001169 | 19.6 | 56.4 | 28.8 | 219805_at | hypothetical protein FLJ22965 | Info | FLJ22965 | |
| | 11 | 0.0002011 | 24.6 | 10.4 | 33.8 | 240394_at | | Info | | |
| 15 | 12 | 0.0002318 | 46 | 10 | 25.9 | 232383_at | transcription factor EC | Info | TFEC | |
| | 13 | 0.0002505 | 31.6 | 12.9 | 32.9 | 221138_s_at | | Info | | |
| | 14 | 0.0002614 | 116.8 | 52.7 | 70.8 | 232914_s_at | synaptotagmin- like 2 | Info | SYTL2 | |
| 20 | 15 | 0.0002713 | 41.4 | 11.8 | 35 | 204634_at | NIMA (never in mitosis gene a)- related kinase 4 | Info | NEK4 | |
| | 16 | 0.0002741 | 127.5 | 62.9 | 133.1 | 201302_at | annexin A4 | Info | ANXA4 | |
| 25 | 17 | 0.0003003 | 82.2 | 35.7 | 38.2 | 216944_s_at | inositol 1,4,5- triphosphate receptor, type 1 | Info | ITPR1 | |
| | 18 | 0.0003018 | 40.5 | 10 | 21.2 | 235412_at | | Info | | |
| 30 | 19 | 0.0003059 | 66 | 40.2 | 61.5 | 203333_at | kinesin- associated protein 3 | Info | KIFAP3 | |
| | 20 | 0.0003135 | 10.8 | 10.5 | 25 | 215151_at | dedicator of cytokinesis 10 | Info | DOCK10 | |
| 35 | 21 | 0.0003224 | 96.8 | 30.2 | 81.1 | 233558_s_at | FLJ12716 protein | Info | FLJ12716 | |
| | 22 | 0.0003732 | 155.2 | 80.9 | 158.1 | 217301_x_at | retinoblastoma binding protein 4 | Info | RBBP4 | |
| 40 | 23 | 0.0003848 | 20.3 | 43.6 | 16.3 | 208050_s_at | caspase 2, apoptosis-related cysteine protease (neural precursor cell expressed, developmentally down-regulated 2) | Info | CASP2 | apoptosis, immunology |
| 45 | 24 | 0.0004095 | 40.1 | 17 | 42.7 | 36920_at | myotubularin 1 | Info | MTM1 | |
| | 25 | 0.0004232 | 26.2 | 10 | 23.7 | 225963_at | KIAA1340 protein | Info | KIAA1340 | |
| 50 | 26 | 0.0004529 | 34.5 | 11.6 | 36.8 | 212310_at | C219-reactive peptide | Info | FLJ39207 | |
| | 27 | 0.0004554 | 45.3 | 10.5 | 34.3 | 235177_at | similar to hepatocellular carcinoma- associated antigen HCA557b | Info | LOC151194 | |
| 55 | 28 | 0.0004832 | 61.7 | 21.2 | 67 | 227268_at | PTD016 protein | Info | LOC51136 | |
| 60 | 29 | 0.000489 | 422 | 205.2 | 370.8 | 208612_at | glucose regulated protein, 58kDa | Info | GRP58 | |
| | 30 | 0.0005587 | 29.2 | 10 | 36.7 | 213659_at | zinc finger protein 75 (D8C6) | Info | ZNF75 | |
| 65 | 31 | 0.0005844 | 147.3 | 88.7 | 152.8 | 217980_s_at | mitochondrial ribosomal protein L16 | Info | MRPL16 | |

ES 2 334 084 B1

| | | | | | | | | | | |
|----|----|-----------|--------|---------|--------|------------------------------|--|----------------------|-------------------------|------------|
| 5 | 32 | 0.0005845 | 68.7 | 21.2 | 45.6 | 205584_at | chromosome X open reading frame 45 | Info | CXorf45 | |
| | 33 | 0.0006039 | 11.1 | 49 | 11.9 | 220366_at | epididymal sperm binding protein 1 | Info | ELSPBP1 | |
| 10 | 34 | 0.0006295 | 140.1 | 75.7 | 134.7 | 201440_at | DEAD (Asp-Glu-Ala-Asp) box polypeptide 23 | Info | DDX23 | |
| | 35 | 0.0006427 | 12 | 24.9 | 10 | 239900_x_at | | Info | | |
| 15 | 36 | 0.0006527 | 53.3 | 28.5 | 63 | 204703_at | tetratricopeptide repeat domain 10 | Info | TTC10 | |
| | 37 | 0.0007041 | 16.7 | 40.1 | 12.1 | 216129_at | ATPase, Class II, type 9A | Info | ATP9A | |
| 20 | 38 | 0.0007168 | 42.2 | 25.1 | 56.5 | 218536_at | MRS2-like, magnesium homeostasis factor (S. cerevisiae) | Info | MRS2L | |
| 25 | 39 | 0.0007406 | 44.9 | 20.8 | 35.3 | 208363_s_at | inositol polyphosphate-4-phosphatase, type I, 107kDa | Info | INPP4A | |
| 30 | 40 | 0.0008726 | 721 | 344.9 | 599.1 | 204588_s_at | solute carrier family 7 (cationic amino acid transporter, y+ system), member 7 | Info | SLC7A7 | immunology |
| 35 | 41 | 0.0008942 | 277 | 140 | 259.8 | 201375_s_at | protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform | Info | PPP2CB | |
| 40 | 42 | 0.0009282 | 31.2 | 63 | 54 | 207681_at | chemokine (C-X-C motif) receptor 3 | Info | CXCR3 | |
| | 43 | 0.0009587 | 53.5 | 98.3 | 41.1 | 220024_s_at | periaxin | Info | PRX | |
| 45 | 44 | 0.0009593 | 8684.9 | 15073.4 | 9366.9 | 1558678_s_at | metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA) | Info | MALAT1 | |
| 50 | 45 | 0.0009911 | 394.4 | 242.1 | 253.3 | 203247_s_at | zinc finger protein 24 (KOX 17) | Info | ZNF24 | |

La distribución de estos genes en Gene Ontology (GO) no era significativamente diferente de la esperada por azar (Figura 1).

Con estos 45 genes se realizó un análisis de cluster de las muestras y se observó que se formaban 3 clusters de diagnóstico altamente reproducibles con un valor medio de correlación entre cada cluster de aproximadamente 0,6 (Figura 2).

También con estos 45 genes se realizó un análisis de cluster y se observó que se formaban 4 clusters con un valor medio de correlación de aproximadamente 0,65 (Figura 3 y Tabla 4).

ES 2 334 084 B1

TABLA 4

Listado de genes que conforman los 4 clusters

| Cluster | Sonda | Nombre gen | Símbolo gen |
|--------------------|--|--|-----------------------------|
| Cluster #1 | <u>1557278 s at</u> | CDNA FLJ33199 fis. clone ADRGL2006377 | |
| | <u>207681 at</u> | chemokine (C-X-C motif) receptor 3 | <u>CXCR3</u> |
| | <u>216129 at</u> | ATPase, Class II, type 9A | <u>ATP9A</u> |
| | <u>208050 s at</u> | caspase 2, apoptosis-related cysteine protease (neural precursor cell expressed, developmentally down-regulated 2) | <u>CASP2</u> |
| | <u>220024 s at</u> | penaxin | <u>PRX</u> |
| | <u>239900 x at</u> | | |
| | <u>1559678 s at</u> | metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA) | <u>MALAT1</u> |
| | <u>219805 at</u> | hypothetical protein FLJ22965 | <u>FLJ22965</u> |
| | <u>223226 x at</u> | single stranded DNA binding protein 4 | <u>SSBP4</u> |
| | <u>210436 at</u> | chaperonin containing TCP1, subunit 8 (theta) | <u>CCT8</u> |
| | <u>1570043 at</u> | | |
| | <u>220366 at</u> | epididymal sperm binding protein 1 | <u>ELSPBP1</u> |
| | Cluster #2 | <u>215151 at</u> | dedicator of cytokinesis 10 |
| Cluster #3 | <u>213659 at</u> | zinc finger protein 75 (D8C6) | <u>ZNF75</u> |
| | <u>218536 at</u> | MRS2-like, magnesium homeostasis factor (<i>S. cerevisiae</i>) | <u>MRS2L</u> |
| | <u>204703 at</u> | tetratricopeptide repeat domain 10 | <u>TTC10</u> |
| | <u>240394 at</u> | | |
| | <u>212310 at</u> | C219-reactive peptide | <u>FLJ39207</u> |
| | <u>205306 x at</u> | kynurenine 3-monooxygenase (kynurenine 3-hydroxylase) | <u>KMO</u> |
| | <u>235401 s at</u> | Fc receptor homolog expressed in B cells | <u>FREB</u> |
| | <u>217980 s at</u> | mitochondrial ribosomal protein L16 | <u>MRPL16</u> |
| | <u>227268 at</u> | PTD016 protein | <u>LOC51136</u> |
| | <u>221138 s at</u> | | |
| | <u>201302 at</u> | annexin A4 | <u>ANXA4</u> |
| | <u>36920 at</u> | myotubularin 1 | <u>MTM1</u> |
| | <u>225963 at</u> | KIAA1340 protein | <u>KIAA1340</u> |
| | <u>235177 at</u> | similar to hepatocellular carcinoma-associated antigen HCA557b | <u>LOC151194</u> |
| | <u>217301 x at</u> | retinoblastoma binding protein 4 | <u>RBBP4</u> |
| | <u>201375 s at</u> | protein phosphatase 2 (formerly 2A) catalytic subunit, beta isoform | <u>PPP2CB</u> |
| | <u>233558 s at</u> | FLJ12716 protein | <u>FLJ12716</u> |
| <u>204588 s at</u> | solute carrier family 7 (cationic amino acid transporter, γ - system), member 7 | <u>SLC7A7</u> | |
| <u>205584 at</u> | chromosome X open reading frame 45 | <u>CXorf45</u> | |
| <u>208612 at</u> | glucose regulated protein, 58kDa | <u>GRP58</u> | |
| <u>208363 s at</u> | inositol polyphosphate-4-phosphatase, type I, 107kDa | <u>INPP4A</u> | |
| <u>232383 at</u> | transcription factor EC | <u>TFEC</u> | |
| <u>201440 at</u> | DEAD (Asp-Glu-Ala-Asp) box polypeptide 23 | <u>DDX23</u> | |
| <u>203333 at</u> | kinesin-associated protein 3 | <u>KIFAP3</u> | |
| <u>204634 at</u> | NIMA (never in mitosis gene a)-related kinase 4 | <u>NEK4</u> | |
| Cluster #4 | <u>203247 s at</u> | zinc finger protein 24 (KOX 17) | <u>ZNF24</u> |
| | <u>224945 at</u> | BTB (POZ) domain containing 7 | <u>BTBD7</u> |
| | <u>216944 s at</u> | inositol 1,4,5-triphosphate receptor, type 1 | <u>ITPR1</u> |
| | <u>227541 at</u> | WD repeat domain 20 | <u>WDR20</u> |
| | <u>229190 at</u> | CDNA FLJ90295 fis. clone NT2RP2000240. | |
| | <u>232914 s at</u> | synaptotagmin-like 2 | <u>SYTL2</u> |
| | <u>235412 at</u> | | |

Con el fin de complementar el número de genes que permiten diferenciar entre cada diagnóstico, se realizó un test de comparación de clases con un valor de $p < 0,001$ entre las clases mal y buen pronóstico, control y mal pronóstico, control y buen pronóstico y control y esclerosis múltiple (Tablas 5, 6, 7 y 8) así como un test de comparación de clases entre los tres diagnósticos con un valor de $p < 0,005$ (Tabla 9).

5
10
15
20
25
30
35
40
45
50
55
60
65

| | Valor p | Mal pronóstico | Buen pronóstico | Ratio diferencia | Sonda | Descripción | UG cluster | Símbolo gen | Localización | Lista genes |
|----|-----------|----------------|-----------------|------------------|---------------------|--|------------------|---------------|---------------|--|
| 1 | 1.1e-06 | 10 | 22.6 | 0.442 | <u>210188_at</u> | GA binding protein transcription factor, alpha subunit 60kDa | <u>Hs.78</u> | <u>GABPA</u> | chr21q21.3 | |
| 2 | 4.6e-06 | 10.4 | 45.9 | 0.227 | <u>235401_s_at</u> | Fc receptor homolog expressed in B cells | <u>Hs.266331</u> | <u>FREB</u> | chr1q23.3 | |
| 3 | 8.2e-06 | 10 | 28.9 | 0.346 | <u>1554433_a_at</u> | Zinc finger protein 146 | <u>Hs.44223</u> | <u>ZNF146</u> | chr19q13.1 | |
| 4 | 4.16e-05 | 10.3 | 38.6 | 0.267 | <u>222281_s_at</u> | | | | | |
| 5 | 4.31e-05 | 10 | 39.9 | 0.251 | <u>209602_s_at</u> | GATA binding protein 3 | <u>Hs.169946</u> | <u>GATA3</u> | chr10p15 | gene regulation, immunology, misc, transcription |
| 6 | 4.86e-05 | 28.6 | 10.4 | 2.75 | <u>240990_at</u> | | | | | |
| 7 | 0.0001008 | 24.9 | 10 | 2.49 | <u>239900_x_at</u> | | | | | |
| 8 | 0.0001254 | 10.7 | 42.7 | 0.251 | <u>205306_x_at</u> | kynurenine 3-monooxygenase (kynurenine 3-hydroxylase) | <u>Hs.170129</u> | <u>KMO</u> | chr1q42-q44 | |
| 9 | 0.0001886 | 10.7 | 35.3 | 0.303 | <u>209421_at</u> | mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli) | <u>Hs.440394</u> | <u>MSH2</u> | chr2p22-p21 | DNA_damage, isonc |
| 10 | 0.0001993 | 10.4 | 33.8 | 0.308 | <u>240394_at</u> | | | | | |
| 11 | 0.0002376 | 80.6 | 36.6 | 2.202 | <u>207389_at</u> | glycoprotein Ib (platelet), alpha polypeptide | <u>Hs.1472</u> | <u>GP1BA</u> | chr17pter-p12 | immunology |
| 12 | 0.0002527 | 11.2 | 27 | 0.415 | <u>200606_at</u> | desmoplakin | <u>Hs.349499</u> | <u>DSP</u> | chr6p24 | immunology |
| 13 | 0.0003294 | 25.3 | 10 | 2.53 | <u>219970_at</u> | PDZ domain protein GIPC2 | <u>Hs.13852</u> | <u>GIPC2</u> | chr1p31.1 | |
| 14 | 0.000331 | 10 | 23.9 | 0.418 | <u>217320_at</u> | IgM | <u>Hs.535538</u> | | | |

| | | | | | | | | | | |
|----|-----------|-------|-------|-------|---------------------|--|------------------|-----------------|--------------------|--|
| 15 | 0.0003357 | 32.4 | 10 | 3.24 | <u>228367_at</u> | rheumatoid factor RF-DII, variable heavy chain | <u>Hs.388674</u> | <u>HAK</u> | <u>chr18q21.31</u> | |
| 16 | 0.0003577 | 135.6 | 62.4 | 2.173 | <u>223226_x_at</u> | single stranded DNA binding protein 4 | <u>Hs.324618</u> | <u>SSBP4</u> | <u>chr19p13.1</u> | |
| 17 | 0.0004336 | 52.3 | 10.1 | 5.178 | <u>1560263_at</u> | Hypothetical protein SP192 | <u>Hs.169854</u> | <u>SP192</u> | <u>chr1p34.1</u> | |
| 18 | 0.000437 | 39.5 | 10.4 | 3.798 | <u>242392_at</u> | hypothetical protein MGC35130 | <u>Hs.388746</u> | <u>MGC35130</u> | <u>chr1p31.3</u> | |
| 19 | 0.0004597 | 10 | 26.2 | 0.382 | <u>1563687_a_at</u> | KIAA0826 | <u>Hs.446102</u> | <u>KIAA0826</u> | <u>chr4p12</u> | |
| 20 | 0.0007879 | 12.9 | 32.9 | 0.392 | <u>221138_s_at</u> | | | | | |
| 21 | 0.0009951 | 105.9 | 243.9 | 0.434 | <u>223361_at</u> | Chromosome 6 open reading frame 115 | <u>Hs.238205</u> | <u>C6orf115</u> | <u>chr6q24.1</u> | |

Tabla 5: Genes que presentaban diferencias significativas entre las clases mal y buen pronóstico con un valor de $p < 0,001$.

5
10
15
20
25
30
35
40
45
50
55
60
65

| | Valor P | Contr. sanos | Mal pronóstico | Ratio diferencia | Sonda | Descripción | Anotación | UG cluster | Símbolo gen | Localización | Lista genes |
|----|-----------|--------------|----------------|------------------|---------------------|--|-----------|------------------|-----------------|--------------|-------------|
| 1 | 1.8e-06 | 37.7 | 10 | 3.77 | <u>239431_at</u> | soil-like receptor adaptor molecule 2 | Info | <u>Hs.534007</u> | <u>TICAM2</u> | chr5q23.1 | |
| 2 | 5.7e-06 | 29.2 | 10 | 2.92 | <u>213659_at</u> | zinc finger protein 75 (D8C6) | Info | <u>Hs.131127</u> | <u>ZNF75</u> | chrXq26.3 | |
| 3 | 1.15e-05 | 54.9 | 10 | 5.49 | <u>239842_x_at</u> | | Info | | | | |
| 4 | 2.96e-05 | 14.4 | 54.1 | 0.266 | <u>1559049_a_at</u> | CDNA FLJ30371 fis, clone BRACE2007836 | Info | <u>Hs.154483</u> | | | |
| 5 | 4.08e-05 | 37.4 | 10 | 3.74 | <u>221239_s_at</u> | SH2 domain containing phosphatase anchor protein 1 // SH2 domain containing phosphatase anchor protein 1 | Info | <u>Hs.194976</u> | <u>SPAPI</u> | chr1q21 | |
| 6 | 6.1e-05 | 34.3 | 10 | 3.43 | <u>224163_s_at</u> | DNA methyltransferase 1 associated protein 1 | Info | <u>Hs.8008</u> | <u>DMAPI</u> | chr1p34 | |
| 7 | 0.0001299 | 37.8 | 10.7 | 3.533 | <u>205306_x_at</u> | kynurenine 3-monooxygenase (kynurenine 3-hydroxylase) | Info | <u>Hs.170129</u> | <u>KMO</u> | chr1q42-q44 | |
| 8 | 0.000135 | 19.6 | 56.4 | 0.348 | <u>219805_at</u> | hypothetical protein FLJ22965 | Info | <u>Hs.248572</u> | <u>FLJ22965</u> | chrXq23 | |
| 9 | 0.0002222 | 567.1 | 147.5 | 3.845 | <u>212192_at</u> | potassium channel tetramerisation domain containing 12 | Info | <u>Hs.109438</u> | <u>KCTD12</u> | chr13q22.3 | |
| 10 | 0.0002518 | 10 | 38.6 | 0.259 | <u>1559976_at</u> | CDNA FLJ36082 fis, clone TEST12019998 | Info | <u>Hs.322679</u> | | | |
| 11 | 0.0002543 | 21.1 | 56.1 | 0.376 | <u>1556024_at</u> | SPRY domain-containing SOCS | Info | <u>Hs.7247</u> | <u>SSB3</u> | chr16p13.3 | |

5
10
15
20
25
30
35
40
45
50
55
60
65

| | | | | | | | | | | | |
|----|-----------|-------|-------|-------|--------------------|--|-------------|------------------|-----------------|------------|--|
| 27 | 0.0006781 | 704.4 | 182.4 | 3.862 | <u>213566_at</u> | associated antigen HCA557b ribonuclease, RNase A family, k6 /// ribonuclease, RNase A family, k6 | <u>Info</u> | <u>Hs.23262</u> | <u>RNASE6</u> | chr14q11.2 | |
| 28 | 0.0007095 | 10.9 | 29.7 | 0.367 | <u>239471_at</u> | Leucine rich repeat containing 28 | <u>Info</u> | <u>Hs.390622</u> | <u>LRRC28</u> | chr15q26.3 | |
| 29 | 0.0008529 | 96.8 | 30.2 | 3.205 | <u>233558_s_at</u> | FLJ12716 protein | <u>Info</u> | <u>Hs.443240</u> | <u>FLJ12716</u> | chr4q35.1 | |
| 30 | 0.0008637 | 23.1 | 67.7 | 0.341 | <u>213779_at</u> | EMI domain containing 1 | <u>Info</u> | <u>Hs.289106</u> | <u>EMID1</u> | chr22q12.2 | |
| 31 | 0.0009255 | 25.6 | 10.3 | 2.485 | <u>222412_s_at</u> | signal sequence receptor, gamma (transloc-associated protein gamma) | <u>Info</u> | <u>Hs.28707</u> | <u>SSR3</u> | chr3q25.31 | |
| 32 | 0.0009687 | 10.5 | 29.9 | 0.351 | <u>233538_s_at</u> | | <u>Info</u> | | | | |

Tabla 6: Genes que presentaban diferencias significativas entre las clases control y mal pronóstico con un valor de $p < 0,001$.

ES 2 334 084 B1

TABLA 7

Genes que presentaban diferencias significativas entre las clases control y buen pronóstico con un valor de $p < 0,001$

| | Valor p | Control es sanos | Buen pronósti co | Ratio diferencias | Sonda | Descripción | UG cluster | Símbolo gen | Localizació n | Lista genes |
|----|-----------|---------------------|------------------------|----------------------|------------------|--|------------|----------------|------------------|----------------|
| 1 | 7.55e-05 | 65.6 | 30.3 | 2.165 | 228738_at | hypothetical protein MGC25181 | Hs.511975 | MGC25181 | chr2p25.3 | |
| 2 | 0.0002827 | 11.5 | 37.4 | 0.307 | 240486_at | | | | | |
| 3 | 0.000318 | 49.8 | 11.2 | 4.446 | 227233_at | tetraspan 2 | Hs.234863 | TSPAN-2 | chr1p13.1 | |
| 4 | 0.0004631 | 264.3 | 128.5 | 2.057 | 203231_s_a t | ataxin 1 | Hs.434961 | ATXN1 | chr6p23 | |
| 5 | 0.0005605 | 10.4 | 23.9 | 0.435 | 217320_at | IgM rheumatoid factor RF-D11, variable heavy chain | Hs.535538 | | | |
| 6 | 0.0007329 | 63.8 | 134.2 | 0.475 | 230566_at | hypothetical protein FLJ35801 | Hs.52184 | FLJ35801 | chr22q12.2 | |
| 7 | 0.0008633 | 10.5 | 25.3 | 0.415 | 1553313_s_ at | solute carrier family 5 (inositol transporters), member 3 | Hs.534372 | SLC5A3 | chr21q22.12 | |
| 8 | 0.0009997 | 10 | 24.8 | 0.403 | 1556589_at | CDNA FLJ25645 fis, clone SYN00113 | Hs.368190 | | | |
| 9 | 0.0012821 | 41.4 | 16.7 | 2.479 | 239801_at | Hypothetical LOC400523 | Hs.534916 | | chr16p11.2 | |
| 10 | 0.0017972 | 10.5 | 30.5 | 0.344 | 228559_at | CDNA clone IMAGE:6043059, partial cds | Hs.55028 | | | |

5
10
15
20
25
30
35
40
45
50

| Valor p | Controles sanos | Pacientes EM | Ratio diferencia | Sonda | Descripción | Anotación | UG cluster | Simbolo gen | Localización | Lista genes |
|---------------|-----------------|--------------|------------------|-------------|--|-----------|----------------------------|-----------------------------------|---------------------------|--------------------|
| 1 $p < 1e-07$ | 20.6 | 10 | 2.06 | 229190_at | inositol 1,4,5-triphosphate receptor, type I | Info | | | | |
| 2 2.99e-05 | 82.2 | 36.9 | 2.228 | 216944_s_at | | Info | Hs.149900 | <u>ITPR1</u> | chr3p26-p25 | |
| 3 0.0001192 | 10.8 | 31.5 | 0.343 | 1553491_at | kinase suppressor of Ras-2 | Info | Hs.375836 | <u>KSR2</u> | chr12q24.22-q24.23 | |
| 4 0.0001509 | 32.9 | 72.4 | 0.454 | 228247_at | SLIT-ROBO Rho GTPase activating protein 1 /// Similar to FRG1 protein (FSHD region gene 1 protein) | Info | Hs.446528 /// Hs.450763_04 | <u>SRGAP1</u> /// <u>MGC72104</u> | chr12q14.2 /// chr20q11.1 | |
| 5 0.0002005 | 11.5 | 36.1 | 0.319 | 203683_s_at | vascular endothelial growth factor B | Info | Hs.78781 | <u>VEGFB</u> | chr11q13 | Angiogenesis, misc |
| 6 0.0002331 | 118.2 | 50.2 | 2.355 | 213119_at | solute carrier family 36 (proton/amino acid symporter), member 1 | Info | Hs.409314 | <u>SLC36A1</u> | chr5q33.1 | |
| 7 0.0002652 | 23.8 | 59.5 | 0.4 | 219380_x_at | polymerase (DNA directed), eta | Info | Hs.439153 | <u>POLH</u> | chr6p21.1 | |
| 8 0.0003315 | 177.3 | 458.2 | 0.387 | 214041_x_at | | Info | | | | |
| 9 0.000344 | 15.1 | 63.8 | 0.237 | 210910_s_at | POM (POM121 homolog, rat) and ZP3 fusion | Info | Hs.296380 | <u>POMZP3</u> | chr7q11.23 | |
| 10 0.0004374 | 446.6 | 171.5 | 2.604 | 212033_at | RNA binding motif protein | Info | Hs.197184 | <u>RBM25</u> | chr14q24.3 | |

55
60
65

5
10
15
20
25
30
35
40
45
50
55
60
65

| | | | | | | | | | | | |
|-----------|-----------|-------|------|-------|-------------|---|------|-----------|--------|-----------|------------------|
| 11 | 0.000703 | 13 | 39.3 | 0.331 | 217239_x_at | 25 | Info | | | | |
| 12 | 0.0007383 | 11.3 | 25.4 | 0.445 | 226681_at | | Info | | | | |
| 13 | 0.0008143 | 11.4 | 30.4 | 0.375 | 1563715_at | MRNA; cDNA DKFZp761B0 221 (from clone DKFZp761B0 221) | Info | Hs.541764 | | | |
| 14 | 0.0008227 | 25.6 | 49.4 | 0.518 | 231812_x_at | | Info | | | | |
| 15 | 0.0008812 | 236.4 | 105 | 2.251 | 219242_at | centrosome protein Cep63 | Info | Hs.443301 | Cep63 | chr3q22.1 | |
| 16 | 0.0009018 | 77.9 | 30.7 | 2.537 | 206618_at | interleukin 18 receptor 1 | Info | Hs.159301 | IL18R1 | chr2q12 | Immuno- logia |

Tabla 8: Genes que presentaban diferencias significativas entre las clases control y esclerosis múltiple con un valor de $p < 0,001$.

ES 2 334 084 B1

TABLA 9

Genes que presentaban diferencias significativas entre las tres clases (control, mal y buen pronóstico) con un valor de $p < 0,005$

| | Valor p | Controles sanos | Mal pronóstico | Buen pronóstico | Sonda | Descripción | Anotación | Símbolo gen | Lista genes | |
|----|---------|-----------------|----------------|-----------------|-------|--------------|--|-------------|-------------|-----------------------|
| 5 | | | | | | | | | | |
| 10 | 1 | 2e-07 | 15.9 | 39.7 | 26 | 1557278_s_at | CDNA FLJ33199 fis, clone ADRGL2006377 | Info | | |
| 15 | 2 | 3e-07 | 20.6 | 10 | 10 | 229190_at | CDNA FLJ90295 fis, clone NT2RP2000240. | Info | | |
| 20 | 3 | 7.4e-06 | 37.8 | 10.7 | 42.7 | 205306_x_at | kynurenine 3-monooxygenase (kynurenine 3-hydroxylase) | Info | KMO | |
| 25 | 4 | 1.07e-05 | 59.8 | 36.3 | 35.4 | 227541_at | WD repeat domain 20 | Info | WDR20 | |
| 30 | 5 | 1.19e-05 | 36.8 | 10.4 | 45.9 | 235401_s_at | Fc receptor homolog expressed in B cells | Info | FREB | |
| 35 | 6 | 2.07e-05 | 61 | 135.6 | 62.4 | 223226_x_at | single stranded DNA binding protein 4 | Info | SSBP4 | |
| 40 | 7 | 2.13e-05 | 10.6 | 22.6 | 10 | 210436_at | chaperonin containing TCP1, subunit 8 (theta) | Info | CCT8 | |
| 45 | 8 | 6.14e-05 | 117.5 | 69.7 | 80.6 | 224945_at | BTB (POZ) domain containing 7 | Info | BTBD7 | |
| 50 | 9 | 9.61e-05 | 10 | 20.3 | 10.7 | 1570043_at | | Info | | |
| 55 | 10 | 0.0001169 | 19.6 | 56.4 | 28.8 | 219805_at | hypothetical protein FLJ22965 | Info | FLJ22965 | |
| 60 | 11 | 0.0002011 | 24.6 | 10.4 | 33.8 | 240394_at | | Info | | |
| 65 | 12 | 0.0002318 | 46 | 10 | 25.9 | 232383_at | transcription factor EC | Info | TFEC | |
| | 13 | 0.0002505 | 31.6 | 12.9 | 32.9 | 221138_s_at | | Info | | |
| | 14 | 0.0002614 | 116.8 | 52.7 | 70.8 | 232914_s_at | synaptotagmin-like 2 | Info | SYTL2 | |
| | 15 | 0.0002713 | 41.4 | 11.8 | 35 | 204634_at | NIMA (never in mitosis gene a)-related kinase 4 | Info | NEK4 | |
| | 16 | 0.0002741 | 127.5 | 62.9 | 133.1 | 201302_at | annexin A4 | Info | ANXA4 | |
| | 17 | 0.0003003 | 82.2 | 35.7 | 38.2 | 216944_s_at | inositol 1,4,5-triphosphate receptor, type 1 | Info | ITPR1 | |
| | 18 | 0.0003018 | 40.5 | 10 | 21.2 | 235412_at | | Info | | |
| | 19 | 0.0003059 | 66 | 40.2 | 61.5 | 203333_at | kinesin-associated protein 3 | Info | KIFAP3 | |
| | 20 | 0.0003135 | 10.8 | 10.5 | 25 | 215151_at | dedicator of cytokinesis 10 | Info | DOCK10 | |
| | 21 | 0.0003224 | 96.8 | 30.2 | 81.1 | 233558_s_at | FLJ12716 protein | Info | FLJ12716 | |
| | 22 | 0.0003732 | 155.2 | 80.9 | 158.1 | 217301_x_at | retinoblastoma binding protein 4 | Info | RBBP4 | |
| 65 | 23 | 0.0003848 | 20.3 | 43.6 | 16.3 | 208050_s_at | caspase 2, apoptosis-related cysteine protease (neural precursor | Info | CASP2 | apoptosis, immunology |

ES 2 334 084 B1

5
10
15
20
25
30
35
40
45
50
55
60
65

| | | | | | | | | | |
|----|-----------|--------|---------|--------|--------------|--|------|-----------|------------|
| | | | | | | cell expressed, developmentally down-regulated 2) | | | |
| 24 | 0.0004095 | 40.1 | 17 | 42.7 | 36920_at | myotubularin I | Info | MTM1 | |
| 25 | 0.0004232 | 26.2 | 10 | 23.7 | 225963_at | KIAA1340 protein | Info | KIAA1340 | |
| 26 | 0.0004529 | 34.5 | 11.6 | 36.8 | 212310_at | C219-reactive peptide | Info | FLJ39207 | |
| 27 | 0.0004554 | 45.3 | 10.5 | 34.3 | 235177_at | similar to hepatocellular carcinoma-associated antigen HCA557b | Info | LOC151194 | |
| 28 | 0.0004832 | 61.7 | 21.2 | 67 | 227268_at | PTD016 protein | Info | LOC51136 | |
| 29 | 0.000489 | 422 | 205.2 | 370.8 | 208612_at | glucose regulated protein, 58kDa | Info | GRP58 | |
| 30 | 0.0005587 | 29.2 | 10 | 36.7 | 213659_at | zinc finger protein 75 (D8C6) | Info | ZNF75 | |
| 31 | 0.0005844 | 147.3 | 88.7 | 152.8 | 217980_s_at | mitochondrial ribosomal protein L16 | Info | MRPL16 | |
| 32 | 0.0005845 | 68.7 | 21.2 | 45.6 | 205584_at | chromosome X open reading frame 45 | Info | CXorf45 | |
| 33 | 0.0006039 | 11.1 | 49 | 11.9 | 220366_at | epididymal sperm binding protein 1 | Info | ELSPBP1 | |
| 34 | 0.0006295 | 140.1 | 75.7 | 134.7 | 201440_at | DEAD (Asp-Glu-Ala-Asp) box polypeptide 23 | Info | DDX23 | |
| 35 | 0.0006427 | 12 | 24.9 | 10 | 239900_x_at | | Info | | |
| 36 | 0.0006527 | 53.3 | 28.5 | 63 | 204703_at | tetratricopeptide repeat domain 10 | Info | TTC10 | |
| 37 | 0.0007041 | 16.7 | 40.1 | 12.1 | 216129_at | ATPase, Class II, type 9A | Info | ATP9A | |
| 38 | 0.0007168 | 42.2 | 25.1 | 56.5 | 218536_at | MRS2-like, magnesium homeostasis factor (S. cerevisiae) | Info | MRS2L | |
| 39 | 0.0007406 | 44.9 | 20.8 | 35.3 | 208363_s_at | inositol polyphosphate-4-phosphatase, type I, 107kDa | Info | INPP4A | |
| 40 | 0.0008726 | 721 | 344.9 | 599.1 | 204588_s_at | solute carrier family 7 (cationic amino acid transporter, y+ system), member 7 | Info | SLC7A7 | immunology |
| 41 | 0.0008942 | 277 | 140 | 259.8 | 201375_s_at | protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform | Info | PPP2CB | |
| 42 | 0.0009282 | 31.2 | 63 | 54 | 207681_at | chemokine (C-X-C motif) receptor 3 | Info | CXCR3 | |
| 43 | 0.0009587 | 53.5 | 98.3 | 41.1 | 220024_s_at | periaxin | Info | PRX | |
| 44 | 0.0009593 | 8684.9 | 15073.4 | 9366.9 | 1558678_s_at | metastasis associated lung | Info | MALAT1 | |

ES 2 334 084 B1

| | | | | | | | | | | |
|----|----|-----------|-------|-------|-------|--|--|------|------------------------|-----------------------|
| | | | | | | adenocarcinoma transcript 1 (non-coding RNA) | | | | |
| 5 | 45 | 0.0009911 | 394.4 | 242.1 | 253.3 | 203247_s_at | zinc finger protein 24 (KOX17) | Info | ZNF24 | |
| | 46 | 0.0010197 | 11.5 | 28.1 | 18 | 244340_x_at | | Info | | |
| 10 | 47 | 0.0010259 | 117.3 | 70 | 147.1 | 213848_at | dual specificity phosphatase 7 | Info | DUSP7 | |
| | 48 | 0.0010309 | 11.3 | 25.1 | 10.1 | 1559441_s_at | cytochrome P450, family 4, subfamily V, polypeptide 2 | Info | CYP4V2 | |
| 15 | 49 | 0.0010448 | 29.4 | 10.7 | 35.3 | 209421_at | mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli) | Info | MSH2 | DNA_damage. tsonc |
| 20 | 50 | 0.0010622 | 19 | 10 | 21.4 | 218884_s_at | hypothetical protein FLJ13220 | Info | FLJ13220 | |
| | 51 | 0.0010626 | 118.2 | 54.7 | 46.1 | 213119_at | solute carrier family 36 (proton/amino acid symporter), member 1 | Info | SLC36A1 | |
| 25 | 52 | 0.0010771 | 136.4 | 67.7 | 139.2 | 228234_at | toll-like receptor adaptor molecule 2 | Info | TICAM2 | |
| 30 | 53 | 0.0010849 | 10.8 | 32.8 | 30.3 | 1553491_at | kinase suppressor of Ras-2 | Info | KSR2 | |
| | 54 | 0.001096 | 180.3 | 89.4 | 149.4 | 218098_at | ADP-ribosylation factor guanine nucleotide-exchange factor 2 (brefeldin A-inhibited) | Info | ARFGEF2 | |
| 35 | 55 | 0.0011018 | 11.5 | 39.9 | 32.7 | 203683_s_at | vascular endothelial growth factor B | Info | VEGFB | angiogenesis. misc |
| 40 | 56 | 0.0011241 | 12.6 | 36.8 | 11.3 | 214997_at | golgi autoantigen, golgin subfamily a, 1 | Info | GOLGA1 | |
| | 57 | 0.0011297 | 33.7 | 15.6 | 46 | 213063_at | nuclear protein UKp68 | Info | FLJ11806 | |
| 45 | 58 | 0.0011697 | 203.2 | 112.4 | 104.3 | 213906_at | v-myb myeloblastosis viral oncogene homolog (avian)-like 1 | Info | MYBL1 | tsonc |
| 50 | 59 | 0.0011949 | 35.2 | 10 | 22.7 | 204113_at | CUG triplet repeat, RNA binding protein 1 | Info | CUGBP1 | |
| | 60 | 0.0012042 | 111.9 | 51.7 | 61.5 | 229510_at | testes development-related NYD-SP21 | Info | NYD-SP21 | |
| 55 | 61 | 0.0012117 | 150.6 | 56.3 | 111.9 | 201816_s_at | glioblastoma amplified sequence | Info | GBAS | |
| 60 | 62 | 0.0012122 | 235.4 | 35.1 | 166.2 | 212956_at | KIAA0882 protein | Info | KIAA0882 | |
| | 63 | 0.0012465 | 32.9 | 75.1 | 69.7 | 228247_at | zinc finger protein 542 /// similar to FRG1 protein (FSHD) | Info | ZNF542 /// MGC72104 | |
| 65 | | | | | | | | | | |

ES 2 334 084 B1

5
10
15
20
25
30
35
40
45
50
55
60
65

| | | | | | | | | | |
|----|-----------|-------|-------|-------|--------------|--|------|----------|---|
| | | | | | | region gene 1 protein) | | | |
| 64 | 0.0012784 | 10.5 | 41 | 10 | 203934_at | kinase insert domain receptor (a type III receptor tyrosine kinase) | Info | KDR | angiogenesis, cell_cycle, cell_signaling, immunology, signal_transduction |
| 65 | 0.0013104 | 445.7 | 221 | 326.1 | 203567_s_at | tripartite motif-containing 38 | Info | TRIM38 | |
| 66 | 0.0013985 | 54.9 | 10 | 34.2 | 239842_x_at | | Info | | |
| 67 | 0.0014599 | 22.6 | 49.9 | 36.6 | 219089_s_at | zinc finger protein 576 | Info | ZNF576 | |
| 68 | 0.0014674 | 62.3 | 24.1 | 47.7 | 238601_at | | Info | | |
| 69 | 0.0014801 | 16.1 | 38.6 | 23.3 | 32540_at | | Info | | |
| 70 | 0.0015065 | 11.1 | 29.3 | 10.3 | 220791_x_at | sodium channel, voltage-gated, type XI, alpha | Info | SCN11A | |
| 71 | 0.0015118 | 38.4 | 12.1 | 50.6 | 212533_at | WEE1 homolog (S. pombe) | Info | WEE1 | immunology |
| 72 | 0.0015299 | 68.4 | 15.1 | 64.8 | 227856_at | hypothetical protein FLJ39370 | Info | FLJ39370 | |
| 73 | 0.0015713 | 126.1 | 57 | 119.8 | 200950_at | actin related protein 2/3 complex, subunit 1A, 41kDa | Info | ARPC1A | |
| 74 | 0.0015729 | 38.4 | 10.3 | 38.6 | 222281_s_at | | Info | | |
| 75 | 0.0015784 | 37.5 | 10 | 28.9 | 1554433_a_at | zinc finger protein 146 | Info | ZNF146 | |
| 76 | 0.0015902 | 33.8 | 11 | 26 | 1553225_s_at | zinc finger protein 75 (D8C6) | Info | ZNF75 | |
| 77 | 0.0016105 | 113.7 | 49 | 97.8 | 211537_x_at | mitogen-activated protein kinase kinase 7 | Info | MAP3K7 | |
| 78 | 0.0016961 | 115.2 | 119.4 | 72.5 | 204046_at | phospholipase C, beta 2 | Info | PLCB2 | |
| 79 | 0.0017071 | 22.5 | 10 | 29.2 | 213132_s_at | malonyl-CoA:acyl carrier protein transacylase, mitochondrial | Info | MT | |
| 80 | 0.0017268 | 10.7 | 34 | 21.3 | 1554106_at | amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 16 | Info | ALS2CR16 | |
| 81 | 0.0017475 | 12 | 38 | 19.1 | 1557292_a_at | mucoylipin 3 | Info | MCOLN3 | |
| 82 | 0.0017591 | 105.4 | 234.7 | 142.1 | 239171_at | | Info | | |
| 83 | 0.0017977 | 33.5 | 61.2 | 28 | 1557961_s_at | | Info | | |
| 84 | 0.0018152 | 15.1 | 72.7 | 55.9 | 210910_s_at | POM (POM121 homolog, rat) and ZP3 fusion | Info | POMZP3 | |
| 85 | 0.0018156 | 13.5 | 10 | 30.2 | 220643_s_at | Fas apoptotic inhibitory molecule | Info | FAIM | |
| 86 | 0.0018321 | 41.1 | 93.6 | 50.1 | 215583_at | KIAA0792 gene product | Info | KIAA0792 | |
| 87 | 0.0018501 | 10 | 33.1 | 16.8 | 204179_at | myoglobin | Info | MB | |
| 88 | 0.0018544 | 27.5 | 10.9 | 28.2 | 1555201_a_at | chromosome 6 | Info | C6orf96 | |

ES 2 334 084 B1

| | | | | | | | | | |
|----|------------|-----------|-------|-------|-------|-----------------------------|---|------|----------------------------------|
| | | | | | | open reading frame 96 | | | |
| 5 | 89 | 0.0018583 | 116.2 | 59.2 | 88.2 | 239243_at | Info | | |
| | 90 | 0.0018584 | 89.7 | 40.4 | 78.7 | 225161_at | mitochondrial elongation factor G1 | Info | EFG1 |
| 10 | 91 | 0.0019076 | 341.6 | 177.2 | 341.6 | 211675_s_at | l-mfa domain-containing protein /// l-mfa domain-containing protein | Info | HIC |
| 15 | 92 | 0.0019228 | 195.8 | 99.8 | 191.4 | 227319_at | chromosome 16 open reading frame 44 | Info | C16orf44 |
| 20 | 93 | 0.0020078 | 28.5 | 24.8 | 50.5 | 213149_at | dihydrolipoamide S-acetyltransferase (E2 component of pyruvate dehydrogenase complex) | Info | DLAT |
| 25 | 94 | 0.0020628 | 226.3 | 89.5 | 169.2 | 209203_s_at | bicaudal D homolog 2 (Drosophila) | Info | BICD2 |
| | 95 | 0.0020776 | 10.7 | 24.1 | 11.6 | 1560204_at | Hypothetical protein LOC284958 | Info | |
| 30 | 96 | 0.0021699 | 34.5 | 84.5 | 45.3 | 202383_at | Jumonji, AT rich interactive domain 1C (RBP2-like) | Info | JARID1C |
| 35 | 97 | 0.002178 | 33.3 | 73.1 | 42.3 | 213681_at | cysteine and histidine rich 1 | Info | CYHR1 |
| | 98 | 0.0022017 | 11.9 | 25.3 | 10 | 219970_at | PDZ domain protein GIPC2 | Info | GIPC2 |
| 40 | 99 | 0.0022244 | 58.4 | 80.6 | 36.6 | 207389_at | glycoprotein Ib (platelet), alpha polypeptide | Info | GP1BA immunology |
| | 100 | 0.0022335 | 205.6 | 94 | 171 | 218715_at | hepatocellular carcinoma-associated antigen 66 | Info | HCA66 |
| 45 | 101 | 0.0022357 | 10 | 23.7 | 18 | 235462_at | | Info | |
| | 102 | 0.0022369 | 23.8 | 61.2 | 57.8 | 219380_x_at | polymerase (DNA directed), eta | Info | POLH |
| 50 | 103 | 0.0022546 | 205.8 | 91.6 | 185.3 | 203922_s_at | cytochrome b-245, beta polypeptide (chronic granulomatous disease) | Info | CYBB immunology |
| 55 | 104 | 0.0022665 | 10.1 | 26.1 | 15.4 | 233246_at | HSPC090 mRNA, partial cds | Info | |
| 60 | 105 | 0.0023625 | 209.6 | 18 | 75.5 | 205321_at | eukaryotic translation initiation factor 2, subunit 3 gamma, 52kDa | Info | EIF2S3 |
| | 106 | 0.0024026 | 34.7 | 69.4 | 42.2 | 243051_at | | Info | |
| 65 | 107 | 0.0024423 | 82.3 | 24.9 | 67.6 | 222646_s_at | ERO1-like (S. cerevisiae) | Info | ERO1L |
| | 108 | 0.0024457 | 13 | 28.6 | 10.4 | 240990_at | | Info | |

ES 2 334 084 B1

5
10
15
20
25
30
35
40
45
50
55
60
65

| | | | | | | | | | |
|-----|-----------|-------|-------|-------|-------------|--|------|-----------|--|
| 109 | 0.0024576 | 183.4 | 78.7 | 161.3 | 201301_s_at | annexin A4 | Info | ANXA4 | |
| 110 | 0.0025048 | 43.3 | 18.2 | 47.8 | 205427_at | zinc finger protein 354A | Info | ZNF354A | |
| 111 | 0.0025413 | 13 | 34.2 | 45.1 | 217239_x_at | | Info | | |
| 112 | 0.0025506 | 49.5 | 15.1 | 37.2 | 207968_s_at | MADS box transcription enhancer factor 2, polypeptide C (myocyte enhancer factor 2C) | Info | MEF2C | |
| 113 | 0.0025715 | 10.1 | 15.3 | 19.6 | 232962_x_at | CDNA FLJ11549 fis, clone HEMBA1002968 | Info | | |
| 114 | 0.0025796 | 31.2 | 13.1 | 26.3 | 204995_at | cyclin-dependent kinase 5, regulatory subunit 1 (p35) | Info | CDK5R1 | |
| 115 | 0.0026228 | 12.7 | 41.9 | 12.9 | 204042_at | WAS protein family, member 3 | Info | WASF3 | |
| 116 | 0.0026255 | 24.9 | 12.7 | 29.6 | 231913_s_at | c6.1A | Info | C6.1A | |
| 117 | 0.0027291 | 57 | 67.7 | 114 | 201614_s_at | RuvB-like 1 (E. coli) | Info | RUVBL1 | |
| 118 | 0.0027564 | 357.6 | 164.7 | 296.4 | 207467_x_at | calpastatin | Info | CAST | |
| 119 | 0.0027577 | 13.8 | 52.3 | 32.4 | 1553647_at | chromodomain protein, Y-like 2 | Info | CDYL2 | |
| 120 | 0.0028454 | 11.4 | 25.8 | 16.7 | 231985_at | flavoprotein oxidoreductase MICAL3 | Info | MICAL3 | |
| 121 | 0.0028521 | 177.3 | 458.5 | 458 | 214041_x_at | | Info | | |
| 122 | 0.0028675 | 115.5 | 73.7 | 128.2 | 203630_s_at | component of oligomeric golgi complex 5 | Info | COG5 | |
| 123 | 0.0029057 | 24.6 | 13.6 | 36.9 | 227751_at | programmed cell death 5 | Info | PDCD5 | |
| 124 | 0.0029409 | 541.3 | 310.4 | 529.5 | 208819_at | RAB8A, member RAS oncogene family | Info | RAB8A | |
| 125 | 0.0029651 | 319.5 | 97.6 | 235.8 | 227260_at | | Info | | |
| 126 | 0.0029731 | 196.8 | 114.1 | 203.9 | 218604_at | integral inner nuclear membrane protein | Info | MAN1 | |
| 127 | 0.0029877 | 14 | 34.9 | 12.7 | 206654_s_at | polymerase (RNA) III (DNA directed) polypeptide G (32kD) | Info | POLR3G | |
| 128 | 0.0029972 | 129.2 | 56.3 | 116.1 | 224876_at | hypothetical protein FLJ37562 | Info | FLJ37562 | |
| 129 | 0.0030549 | 11.4 | 34.4 | 26.9 | 1563715_at | | Info | | |
| 130 | 0.0031053 | 206.5 | 82.2 | 151.2 | 241993_x_at | | Info | | |
| 131 | 0.0031432 | 12 | 22.2 | 10 | 233086_at | chromosome 20 open reading frame 106 | Info | C20orf106 | |
| 132 | 0.0031432 | 92.3 | 88.9 | 42.9 | 226018_at | hypothetical protein Ells1 | Info | Ells1 | |
| 133 | 0.003172 | 29 | 11.1 | 25.4 | 231975_s_at | hypothetical protein FLJ35954 | Info | FLJ35954 | |

ES 2 334 084 B1

5

10

15

20

25

30

35

40

45

50

55

60

65

| | | | | | | | | | |
|-----|-----------|-------|-------|-------|-------------|---|------|-----------|--|
| 134 | 0.0031828 | 13.1 | 39.5 | 10.4 | 242392_at | hypothetical protein MGC35130 | Info | MGC35130 | |
| 135 | 0.0032055 | 41.4 | 14.7 | 36.4 | 220201_at | membrane associated DNA binding protein | Info | MNAB | |
| 136 | 0.0032743 | 446.6 | 164.6 | 178.8 | 212033_at | RNA binding motif protein 25 | Info | RBM25 | |
| 137 | 0.0033044 | 82.6 | 132.4 | 72 | 215504_x_at | Clone 25061 mRNA sequence | Info | | |
| 138 | 0.003329 | 145.3 | 494.4 | 222 | 224321_at | transmembrane protein with EGF-like and two follistatin-like domains 2 /// transmembrane protein with EGF-like and two follistatin-like domains 2 | Info | TMEFF2 | |
| 139 | 0.0033665 | 71.3 | 20.2 | 42 | 225922_at | KIAA1450 protein | Info | KIAA1450 | |
| 140 | 0.0033774 | 171.7 | 76.7 | 175.5 | 201259_s_at | synaptophysin-like protein | Info | SYPL | |
| 141 | 0.0033795 | 50.5 | 19.5 | 39.1 | 225754_at | adaptor-related protein complex 1, gamma 1 subunit | Info | APIG1 | |
| 142 | 0.0034069 | 11.7 | 40.3 | 27.9 | 243343_at | | Info | | |
| 143 | 0.003445 | 171.9 | 61.3 | 144.7 | 201711_x_at | RAN binding protein 2 | Info | RANBP2 | gene regulation, transcription |
| 144 | 0.0034662 | 26.6 | 15.8 | 34.7 | 228561_at | | Info | | |
| 145 | 0.0035038 | 36.8 | 76 | 52.7 | 1552646_at | interleukin 11 receptor, alpha | Info | IL11RA | immunology |
| 146 | 0.0035692 | 65.1 | 39 | 64.7 | 217043_s_at | mitofusin 1 | Info | MFN1 | |
| 147 | 0.0036899 | 30.6 | 11.5 | 29.9 | 219608_s_at | F-box protein 38 | Info | FBXO38 | |
| 148 | 0.0037154 | 130.1 | 53.4 | 149.4 | 231736_x_at | microsomal glutathione S-transferase 1 | Info | MGST1 | pharmacology |
| 149 | 0.0037166 | 62.4 | 23.3 | 51.9 | 226894_at | | Info | | |
| 150 | 0.0037726 | 175.8 | 88.5 | 133.8 | 222000_at | hypothetical protein LOC339448 | Info | LOC339448 | |
| 151 | 0.0037966 | 325.1 | 103 | 234 | 221841_s_at | Kruppel-like factor 4 (gut) | Info | KLF4 | |
| 152 | 0.0038994 | 60.9 | 31.9 | 60.1 | 223404_s_at | chromosome 1 open reading frame 25 | Info | C1orf25 | |
| 153 | 0.0039171 | 52.7 | 16.8 | 43.6 | 210635_s_at | kelch-like ECT2 interacting protein | Info | KLEIP | |
| 154 | 0.003963 | 33.8 | 34.7 | 14.3 | 222426_at | mitogen-activated protein kinase associated protein 1 | Info | MAPKAP1 | |
| 155 | 0.0039777 | 58.7 | 117.2 | 96.7 | 236346_at | | Info | | |
| 156 | 0.0040561 | 16.1 | 33.1 | 22.1 | 216261_at | integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61) | Info | ITGB3 | cell signaling, immunology, metastasis |
| 157 | 0.0040985 | 13.8 | 28.3 | 19.7 | 241695_s_at | | Info | | |
| 158 | 0.0041185 | 112.2 | 45.9 | 97.4 | 238077_at | potassium channel | Info | KCTD6 | |

ES 2 334 084 B1

| | | | | | | | | | | |
|----|-----|-----------|-------|-------|-------|-------------------------------------|--|------|------------------|--|
| | | | | | | tetramerisation domain containing 6 | | | | |
| 5 | 159 | 0.004191 | 19.2 | 52.1 | 42.3 | 206569_at | interleukin 24 | Info | IL24 | |
| | 160 | 0.0041965 | 74.6 | 43.7 | 77.8 | 225538_at | zinc finger, CCHC domain containing 9 | Info | ZCCHC9 | |
| 10 | 161 | 0.0042477 | 12.4 | 10 | 26.3 | 203650_at | protein C receptor, endothelial (EPCR) | Info | PROCR | |
| | 162 | 0.0042591 | 103.6 | 54 | 88.2 | 222476_at | KIAA1194 | Info | KIAA1194 | |
| 15 | 163 | 0.0042876 | 200 | 162.9 | 311.1 | 221602_s_at | regulator of Fas-induced apoptosis | Info | TOSO | |
| | 164 | 0.0043282 | 74 | 36.8 | 67.6 | 212214_at | optic atrophy 1 (autosomal dominant) | Info | OPA1 | immunology |
| 20 | 165 | 0.0043616 | 40.4 | 12.5 | 58.8 | 235400_at | Fc receptor homolog expressed in B cells | Info | FREB | |
| | 166 | 0.0043619 | 98.2 | 39.5 | 91 | 211256_x_at | butyrophilin, subfamily 2, member A1 | Info | BTN2A1 | |
| 25 | 167 | 0.0044282 | 187.6 | 465.9 | 238 | AFFX-r2-Ec-bioB-5_at | | Info | | |
| | 168 | 0.0044636 | 358.6 | 172.4 | 317.3 | 201386_s_at | DEAH (Asp-Glu-Ala-His) box polypeptide 15 | Info | DHX15 | |
| 30 | 169 | 0.0045185 | 81.4 | 41.7 | 82.4 | 204168_at | microsomal glutathione S-transferase 2 | Info | MGST2 | pharmacology |
| 35 | 170 | 0.0045364 | 156 | 305 | 246.3 | 213041_s_at | ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit | Info | ATP5D | |
| 40 | 171 | 0.0045462 | 37.8 | 76.9 | 35.9 | 222041_at | DPH2-like 1 (S. cerevisiae) /// candidate tumor suppressor in ovarian cancer 2 | Info | DPH2L1 /// OVCA2 | |
| 45 | 172 | 0.0045494 | 58.6 | 33.3 | 30.3 | 204109_s_at | nuclear transcription factor Y, alpha | Info | NFYA | gene regulation, immunology, transcription |
| 50 | 173 | 0.0046367 | 27.3 | 10 | 39.9 | 209602_s_at | GATA binding protein 3 | Info | GATA3 | gene regulation, immunology, misc, transcription |
| | 174 | 0.0046481 | 214.3 | 225.4 | 129.1 | 228768_at | KIAA1961 protein | Info | KIAA1961 | |
| 55 | 175 | 0.0046552 | 21.2 | 25.2 | 37 | 231843_at | DEAD (Asp-Glu-Ala-Asp) box polypeptide 55 | Info | DDX55 | |
| 60 | 176 | 0.0047735 | 32.6 | 118.6 | 74.7 | 217390_x_at | | Info | | |
| | 177 | 0.0047736 | 17.6 | 10 | 22.2 | 240557_at | CDNA FLJ41867 fis, clone OCBBF2005546 | Info | | |
| 65 | 178 | 0.0048014 | 57.9 | 101.3 | 65.3 | 217499_x_at | | Info | | |
| | 179 | 0.0048023 | 300.9 | 170.1 | 280.1 | 220742_s_at | N-glycanase 1 | Info | NGLY1 | |

ES 2 334 084 B1

| | | | | | | | | | | |
|----|------------|-----------|------|------|------|-----------------------------|---|----------------------|-------------------------|------------|
| 5 | 180 | 0.00482 | 76.8 | 36.1 | 63.5 | 207629_s_at | rho/rac guanine nucleotide exchange factor (GEF) 2 | Info | ARHGEE2 | |
| | 181 | 0.0048333 | 10.8 | 14.2 | 29.2 | 238057_at | | Info | | |
| 10 | 182 | 0.0048512 | 77.9 | 28.4 | 33.3 | 206618_at | interleukin 18 receptor 1 | Info | IL18R1 | immunology |
| | 183 | 0.0048879 | 28.3 | 71.6 | 37.2 | 203389_at | kinesin family member 3C | Info | KIF3C | |
| 15 | 184 | 0.0048938 | 41.1 | 94 | 67.4 | 243216_x_at | | Info | | |
| 20 | 185 | 0.0049691 | 45.8 | 70.1 | 36.6 | 208022_s_at | CDC14 cell division cycle 14 homolog B (S. cerevisiae) /// CDC14 cell division cycle 14 homolog B (S. cerevisiae) | Info | CDC14B | |

25 Los resultados de los análisis de cluster tanto de muestras como de genes obtenidos tras estas comparaciones de clases se muestran en la figura 5.

30 (Tabla pasa a página siguiente)

35

40

45

50

55

60

65

ES 2 334 084 B1

2. Validación por PCR en tiempo real y construcción del clasificador

La Tabla 10 recoge el listado de los 95+1 genes y ensayos seleccionados para configurar las LDAs a partir de los resultados obtenidos en el screening con chip de ADN.

TABLA 10

Listado de genes y ensayos seleccionados para configurar LDA

| | Código ensayo | Símbolo gen | Nombre gen |
|----|---------------|-------------|--|
| 15 | Hs00154040_m1 | ANXA4 | annexin A4 |
| | Hs00154242_m1 | CASP2 | caspace 2, apoptosis-related cysteine peptidase (neural precursor cell expressed developmentally down-regulated 2) |
| | Hs00164982_m1 | JAG1 | jagged 1 (Alagille syndrome) |
| | Hs00165656_m1 | ATXN1 | ataxin 1 |
| 20 | Hs00166163_m1 | CYBB | cytochrome b-245, beta polypeptide (chronic granulomatous disease) |
| | Hs00168405_m1 | IL12A | interleukin 12A (natural killer cell stimulatory factor 1, cytotoxic lymphocyte maturation factor 1, p35) |
| | Hs00168433_m1 | ITGA4 | integrin, alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor) |
| | Hs00168469_m1 | ITGB7 | integrin, beta 7 |
| | Hs00169680_m1 | MTM1 | myotubularin 1 |
| 25 | Hs00171041_m1 | CXCR3 | chemokine (C-X-C motif) receptor 3 |
| | Hs00171257_m1 | TGFB1 | transforming growth factor, beta 1 (Camurati-Engelmann disease) |
| | Hs00172915_m1 | RBM6 | RNA binding motif protein 6 |
| | Hs00173149_m1 | ZNF24 | zinc finger protein 24 (KOX 17) |
| | Hs00173196_m1 | ZNF146 | zinc finger protein 146 |
| 30 | Hs00173947_m1 | GP1BA | glycoprotein Ib (platelet), alpha polypeptide |
| | Hs00174086_m1 | IL10 | interleukin 10 |
| | Hs00174122_m1 | IL4 | interleukin 4 |
| | Hs00174128_m1 | TNF | tumor necrosis factor (TNF superfamily, member 2) |
| | Hs00174143_m1 | IFNG | interferon, gamma |
| 35 | Hs00174796_m1 | CD28 | CD28 molecule |
| | Hs00175480_m1 | CTLA4 | cytotoxic T-lymphocyte-associated protein 4 |
| | Hs00175738_m1 | KMO | kynurenine 3-monooxygenase (kynurenine 3-hydroxylase) |
| | Hs00177323_m1 | NEK4 | NIMA (never in mitosis gene a)-related kinase 4 |
| | Hs00179887_m1 | MSH2 | mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli) |
| 40 | Hs00181881_m1 | ITPR1 | inositol 1,4,5-triphosphate receptor, type 1 |
| | Hs00182073_m1 | MX1 | myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse) |
| | Hs00183973_m1 | KIFAP3 | kinesin-associated protein 3 |
| | Hs00189422_m1 | DSP | desmoplakin |
| | Hs00194836_m1 | TSPAN2 | tetraspanin 2 |
| 45 | Hs00197926_m1 | TTC10 | tetrapeptide repeat domain 10 |
| | Hs00203436_m1 | TBX21 | T-box 21 |
| | Hs00208425_m1 | HELZ | helicase with zinc finger |
| | Hs00211612_m1 | LOC51136 | PTD016 protein |
| | Hs00214273_m1 | GIPC2 | GIPC PDZ domain containing family, member 2 |
| 50 | Hs00215231_m1 | MRPL16 | mitochondrial ribosomal protein L16 |
| | Hs00216842_m1 | BTBD7 | BTB (POZ) domain containing 7 |
| | Hs00219525_m1 | DMAP1 | DNA methyltransferase 1 associated protein 1 |
| | Hs00219575_m1 | HLA-DRA | major histocompatibility complex, class II, DR alpha |
| | Hs00221246_m1 | PRX | periaxin |
| | Hs00222575_m1 | FLJ12716 | FLJ12716 protein |
| 55 | Hs00223326_m1 | ELSPBP1 | epididymal sperm binding protein 1 |
| | Hs00227238_m1 | CXorf45 | chromosome X open reading frame 45 |
| | Hs00229156_m1 | FCRL2 | Fc receptor-like 2 |
| | Hs00231122_m1 | GATA3 | GATA binding protein 3 |

ES 2 334 084 B1

| | | | |
|----|---------------|------------------|--|
| | Hs00232612_m1 | TFEC | transcription factor EC |
| | Hs00234829_m1 | STAT1 | signal transducer and activator of transcription 1, 91kDa |
| | Hs00237047_m1 | YWHAZ | tyrosine 3-monooxygenase;tryptophan 5-monooxygenase activation protein, zeta polypeptide |
| 5 | Hs00244467_m1 | INPP4A | inositol polyphosphate-4-phosphatase type I, 107kDa |
| | Hs00252895_m1 | MRS2L | MRS2-like, magnesium homeostasis factor (<i>S. cerevisiae</i>) |
| | Hs00261786_m1 | SSBP4 | single stranded DNA binding protein 4 |
| | Hs00262988_m1 | SYTL2 | synaptotagmin-like 2 |
| | Hs00266139_m1 | CA1 | carbonic anhydrase I |
| 10 | Hs00272857_s1 | SLC5A3 | solute carrier family 5 (inositol transporters), member 3 |
| | Hs00273907_s1 | PRO1073.MALAT1 | PRO1073 protein,metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA) |
| | Hs00288176_s1 | WDR20 | WD repeat domain 20 |
| | Hs00292260_m1 | MGC25181 | hypothetical protein MGC25181 |
| | Hs00294940_m1 | EMID1 | EMI domain containing 1 |
| 15 | Hs00325227_m1 | KLHDC5 | kelch domain containing 5 |
| | Hs00325689_m1 | KIAA1447 | KIAA1447 protein |
| | Hs00364763_m1 | ALPK2 | alpha-kinase 2 |
| | Hs00365634_g1 | PTPRC | protein tyrosine phosphatase, receptor type, C |
| | Hs00366948_m1 | ZNF75 | zinc finger protein 75 (DBC6) |
| 20 | Hs00374418_m1 | SLC7A7 | solute carrier family 7 (cationic amino acid transporter, y+ system), member 7 |
| | Hs00375921_m1 | WDR20bis | WD repeat domain 20 |
| | Hs00377819_m1 | RNASE6 | ribonuclease, RNase A family, k6 |
| | Hs00378993_m1 | KIAA0826 | KIAA0826 |
| | Hs00381019_m1 | UBE2U | ubiquitin-conjugating enzyme E2U (putative) |
| 25 | Hs00388776_m1 | ARHGEF7 | Rho guanine nucleotide exchange factor (GEF) 7 |
| | Hs00391058_m1 | ATP9A | ATPase, Class II, type 9A |
| | Hs00391515_m1 | DOCK10 | dedicator of cytokinesis 10 |
| | Hs00395930_m1 | CHD5 | chromodomain helicase DNA binding protein 5 |
| | Hs00396464_g1 | ABCC13 | ATP-binding cassette, sub-family C (CFTR/MRP), member 13 |
| 30 | Hs00400812_m1 | LRRC28 | leucine rich repeat containing 28 |
| | Hs00402198_m1 | RBM25 | RNA binding motif protein 25 |
| | Hs00409790_m1 | HLA-DQB1 | major histocompatibility complex, class II, DQ beta 1 |
| | Hs00410715_m1 | C6orf115 | chromosome 6 open reading frame 115 |
| | Hs00412706_m1 | KIAA0268.UNQ6077 | C219-reactive peptide_AAAP6077 |
| 35 | Hs00428123_m1 | DDX23 | DEAD (Asp-Glu-Ala-Asp) box polypeptide 23 |
| | Hs00428403_g1 | RBBP4 | retinoblastoma binding protein 4 |
| | Hs00540758_m1 | SSB3 | SPRY domain-containing SOCS box protein SSB-3 |
| | Hs00540818_s1 | KCTD12 | potassium channel tetramerisation domain containing 12 |
| | Hs00541844_m1 | FLJ35801 | hypothetical protein FLJ35801 |
| 40 | Hs00541858_m1 | TNPO1 | transportin 1 |
| | Hs00559595_m1 | ITGB1 | integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12) |
| | Hs00598618_m1 | STAT6 | signal transducer and activator of transcription 6, interleukin-4 induced |
| | Hs00602137_m1 | PPP2CB | protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform |
| | Hs00606481_m1 | SSR3 | signal sequence receptor, gamma (translocon-associated protein gamma) |
| 45 | Hs00607126_m1 | PDIA3 | protein disulfide isomerase family A, member 3 |
| | Hs00607229_mH | CCT8 | chaperonin containing TCP1, subunit 8 (theta) |
| | Hs00697611_m1 | LOC151194 | similar to hepatocellular carcinoma-associated antigen HCA557b |
| | Hs00742415_s1 | CCT8 | chaperonin containing TCP1, subunit 8 (theta) |
| | Hs00745591_s1 | GABPA | GA binding protein transcription factor, alpha subunit 60kDa |
| 50 | Hs00824723_m1 | UBC | ubiquitin C |
| | Hs99999903_m1 | ACTB | actin, beta |
| | Hs99999907_m1 | B2M | beta-2-microglobulin |

55 El análisis estadístico identificó 25 genes que presentaban diferencias significativas ($p < 0,01$) en los niveles de expresión entre las tres clases (Figura 6).

60 Utilizando estos 25 genes y las variables clínicas de EDSS y MSFC al inicio de la enfermedad se construyó un clasificador Bayesiano que mostró una precisión de un 91,66% entre los tres diagnósticos (Figuras 7 y 8). Este clasificador tenía una precisión de un 87,5% al distinguir entre buen y mal pronóstico.

65 Con el fin de aumentar la precisión a la hora de distinguir entre buen y mal pronóstico se construyó un nuevo clasificador utilizando las variables clínicas y únicamente aquellos genes que presentaban diferencias significativas ($p < 0,05$) en los niveles de expresión de ambas clases (Figuras 9, 10 y 11). El número de genes utilizados fue de 13 (Tabla 11) y la precisión que se obtuvo entonces fue de un 95%.

ES 2 334 084 B1

TABLA 11

Genes que diferencian entre buen y mal pronóstico

| 5 | Código ensayo | Símbolo gen | Nombre gen |
|----|---------------|-----------------|---|
| | Hs00216842_m1 | BTBD7 | BTB (POZ) domain containing 7 |
| 10 | Hs00154242_m1 | CASP2 | caspase 2, apoptosis-related cysteine peptidase (neural precursor cell expressed, developmentally down-regulated 2) |
| | Hs00175480_m1 | CTLA4 | cytotoxic T-lymphocyte-associated protein 4 |
| | Hs00391515_m1 | DOCK10 | dedicator of cytokinesis 10 |
| 15 | Hs00294940_m1 | EMID1 | EMI domain containing 1 |
| | Hs00325227_m1 | KLHDC5 | kelch domain containing 5 |
| | Hs00292260_m1 | MGC25181 | hypothetical protein MGC25181 |
| 20 | Hs00177323_m1 | NEK4 | NIMA (never in mitosis gene a)-related kinase 4 |
| | Hs00273907_s1 | PRO1073, MALAT1 | PRO1073 protein, metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA) |
| 25 | Hs00365634_g1 | PTPRC | protein tyrosine phosphatase, receptor type, C |
| | Hs00262988_m1 | SYTL2 | synaptotagmin-like 2 |
| | Hs00197926_m1 | TTC10 | tetratricopeptide repeat domain 10 |
| 30 | Hs00375921_m1 | WDR20bis | WD repeat domain 20 |

35 La Tabla 12 presenta el análisis de la información aportada por cada variable para el establecimiento del pronóstico. El peso representa la cantidad relativa de información que aporta cada variable a la clasificación del pronóstico. La lista de variables está ordenada en orden descendente según la información que aporta cada una. El valor modal *a priori* describe el valor más probable que posee cada variable cuando el pronóstico es desconocido mientras que el valor modal para cada pronóstico describe el valor más probable para ese pronóstico. Todos los valores modales se acompañan de su probabilidad. La variación para cada pronóstico es una medida que indica la diferencia de probabilidad entre el valor modal *a priori* y el valor modal para el pronóstico cuando se conoce el mismo. La fórmula usada para su cálculo es: $-\log_2(P(\text{valor modal para el pronóstico})) + \log_2(P(\text{valor modal para el pronóstico} | \text{valor observado}))$. Los valores subrayados simple indican variaciones positivas (la probabilidad del valor modal para el pronóstico es mayor que la del valor modal *a priori*) mientras que los valores en *cursiva (Italicas)* indican variaciones negativas. Obviamente, no se indica ninguna variación si el valor modal para el pronóstico es diferente del valor modal *a priori*. Entonces, el valor modal para el pronóstico se representa en **negrita**.

50 (Tabla pasa a página siguiente)

55

60

65

| Variable | Peso | Valor modal <i>a priori</i> | Valor modal mal pronóstico | Valor modal buen pronóstico | Variación mal pronóstico | Variación buen pronóstico |
|----------|--------|-----------------------------|----------------------------|-----------------------------|--------------------------|---------------------------|
| klhdc5 | 1,0000 | <=-2,640 | <=-3,145 | 60,00% | <=-2,640 | 65,00% |
| EDSS | 0,8512 | <=2,000 | >2,000 | 85,00% | <=2,000 | 85,00% |
| casp2 | 0,6791 | <=-2,385 | <=-2,385 | 90,00% | <=-1,920 | 50,00% |
| emid1 | 0,6367 | >-3,295 | <=-3,295 | 66,67% | >-3,295 | 91,92% |
| MSFC | 0,5951 | <=0,665 | <=0,665 | 100,00% | <=0,665 | 55,00% |
| pro1073 | 0,5951 | >-2,765 | >-2,765 | 55,00% | >-2,765 | 100,00% |
| btbd7 | 0,5273 | >-2,825 | <=-2,825 | 55,00% | >-2,825 | 95,00% |
| mgc2518 | 0,4406 | >-3,585 | >-3,585 | 65,00% | >-3,585 | 100,00% |
| wdr20bis | 0,4406 | >-2,740 | >-2,740 | 65,00% | >-2,740 | 100,00% |
| nek4 | 0,3978 | <=-2,335 | <=-2,335 | 78,95% | >-2,335 | 70,00% |
| sytl2 | 0,3902 | <=-2,650 | <=-2,650 | 90,00% | >-2,650 | 55,00% |
| dock10 | 0,3691 | >-3,320 | >-3,320 | 70,00% | >-3,320 | 100,00% |
| ttc10 | 0,3066 | >-2,735 | >-2,735 | 60,00% | >-2,735 | 95,00% |
| ptprc | 0,3009 | >-3,210 | >-3,210 | 75,00% | >-3,210 | 100,00% |
| ctla4 | 0,2860 | >-3,720 | <=-3,720 | 61,11% | >-3,720 | 80,00% |

Tabla 12: Peso, valores modales *a priori*, de mal pronóstico y de buen pronóstico para los 13 marcadores seleccionados y para las variables clínicas EDSS y MSFC.

ES 2 334 084 B1

La Tabla 13 presenta la precisión del clasificador a medida que se aumenta el número de variables que lo componen. Las variables se introducen en orden según la información que aporta cada uno para el establecimiento del pronóstico. Dentro de los pronósticos los valores se presentan como individuos clasificados correctamente (correcto) respecto a los individuos clasificados incorrectamente (incorrecto).

5

TABLA 13

10

Precisión del clasificador a medida que se incorporan genes y variables clínicas

15

| Variable | Precisión | Buen pronóstico (correcto / incorrecto) | Mal pronóstico (correcto / incorrecto) |
|----------|-----------|---|--|
| klhdc5 | 85,00% | 18 / 2 | 16 / 4 |
| + | | | |
| EDSS | 90,00% | 16 / 4 | 20 / 0 |
| + | | | |
| casp2 | 92,50% | 17 / 3 | 20 / 0 |
| + | | | |
| emid1 | 92,50% | 17 / 3 | 20 / 0 |
| + | | | |
| MSFC | 90,00% | 16 / 4 | 20 / 0 |
| + | | | |
| pro1073 | 95,00% | 18 / 2 | 20 / 0 |
| + | | | |
| btbd7 | 95,00% | 19 / 1 | 19 / 1 |
| + | | | |
| mgc2518 | 92,50% | 18 / 2 | 19 / 1 |
| + | | | |
| wdr20bis | 95,00% | 19 / 1 | 19 / 1 |
| + | | | |
| nek4 | 92,50% | 18 / 2 | 19 / 1 |
| + | | | |
| syt12 | 92,50% | 18 / 2 | 19 / 1 |
| + | | | |
| dock10 | 92,50% | 18 / 2 | 19 / 1 |
| + | | | |
| ttc10 | 95,00% | 19 / 1 | 19 / 1 |
| + | | | |
| ptprec | 92,50% | 18 / 2 | 19 / 1 |
| + | | | |
| ctla4 | 95,00% | 19 / 1 | 19 / 1 |

60

65

ES 2 334 084 B1

La Tabla 14 presenta las probabilidades condicionadas al pronóstico de la enfermedad para cada variable utilizadas por el clasificador.

TABLA 14

Probabilidades condicionadas al pronóstico de la enfermedad para cada variable utilizadas por el clasificador

| Variable | Valor modal | Probabilidad <i>a priori</i> | Probabilidad mal pronóstico | Probabilidad buen pronóstico |
|----------|-------------|------------------------------|-----------------------------|------------------------------|
| | <=-3.145 | 35.14% | 60.00% | 10.00% |
| klhdc5 | <=-2.640 | 42.37% | 20.00% | 65.00% |
| | <=-2.475 | 10.06% | 20.00% | 0.00% |
| | >-2.475 | 12.43% | 0.00% | 25.00% |
| EDSS | <=2.000 | 49.80% | 15.00% | 85.00% |
| | >2.000 | 50.20% | 85.00% | 15.00% |
| casp2 | <=-2.385 | 67.63% | 90.00% | 45.00% |
| | <=-1.920 | 24.86% | 0.00% | 50.00% |
| emid1 | >-1.920 | 7.51% | 10.00% | 5.00% |
| | <=-3.295 | 37.54% | 66.67% | 8.08% |
| MSFC | >-3.295 | 62.46% | 33.33% | 91.92% |
| | <=-0.665 | 77.63% | 100.00% | 55.00% |
| pro1073 | >0.665 | 22.37% | 0.00% | 45.00% |
| | <=-2.765 | 22.63% | 45.00% | 0.00% |
| btbd7 | >-2.765 | 77.37% | 55.00% | 100.00% |
| | <=-2.825 | 30.14% | 55.00% | 5.00% |
| mgc2518 | >-2.825 | 69.86% | 45.00% | 95.00% |
| | <=-3.585 | 17.60% | 35.00% | 0.00% |
| wdr20bis | >-3.585 | 82.40% | 65.00% | 100.00% |
| | <=-2.740 | 17.60% | 35.00% | 0.00% |
| nek4 | >-2.740 | 82.40% | 65.00% | 100.00% |
| | <=-2.335 | 54.61% | 78.95% | 30.00% |
| sytl2 | >-2.335 | 45.39% | 21.05% | 70.00% |
| | <=-2.650 | 67.63% | 90.00% | 45.00% |
| dock10 | >-2.650 | 32.37% | 10.00% | 55.00% |
| | <=-3.320 | 15.09% | 30.00% | 0.00% |
| ttc10 | >-3.320 | 84.91% | 70.00% | 100.00% |
| | <=-2.735 | 22.60% | 40.00% | 5.00% |
| ptprc | >-2.735 | 77.40% | 60.00% | 95.00% |
| | <=-3.210 | 12.57% | 25.00% | 0.00% |
| ctla4 | >-3.210 | 87.43% | 75.00% | 100.00% |
| | <=-3.720 | 41.22% | 61.11% | 20.00% |
| | >-3.720 | 58.78% | 38.89% | 80.00% |

REIVINDICACIONES

- 5 1. Un método *in vitro* para determinar el pronóstico clínico de un paciente afecto de esclerosis múltiple que comprende
- (a) comparar
 - 10 (i) el valor correspondiente a la expresión de los genes KLHDC5, CASP2, EMID1, PRO1073, BTBD7, MGC2518, WDR20bis, NEK4, SYLT2, DOCK10, TTC10, PTPRC y CTLA4 con una tabla de probabilidades condicionadas entre intervalos de valores modales de expresión de dichos genes y valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y/o
 - 15 (ii) el valor de una variable clínica seleccionada del grupo de EDSS y MSFC con una tabla de probabilidades condicionadas entre intervalos de valores modales de dichas variables clínicas y valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y
 - (b) asignar una probabilidad de mal y buen pronóstico correspondiente a la probabilidad asociada al intervalo en el que se encuentra el valor de la expresión o de la variable clínica.
- 20
2. Un método *in vitro* para determinar el pronóstico clínico de un paciente afecto de esclerosis que comprende
- (a) comparar
 - 25 (i) los valores correspondientes a la expresión de los genes seleccionados del grupo de KLHDC5, CASP2, EMID1, PRO1073, BTBD7, MGC2518, WDR20bis, NEK4, SYLT2, DOCK10, TTC10, PTPRC y CTLA4 con una tabla de probabilidades condicionadas entre intervalos de valores modales de expresión de dichos genes y valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y/o
 - 30 (ii) los valores de las variables clínicas EDSS y MSFC con una tabla de probabilidades condicionadas entre intervalos de valores modales de dichas variables clínicas y valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y
 - (b) asignar una probabilidad de mal pronóstico correspondiente a la probabilidad condicionada de mal pronóstico asociada a los intervalos de valores modales en los que se encuentran los valores de expresión de cada uno de los genes cuya expresión se ha determinado y/o variables clínicas determinadas y asignar una probabilidad de buen pronóstico correspondiente a la probabilidad condicionada de buen pronóstico asociada a los intervalos de valores modales en los que se encuentra los valores de expresión para cada uno de los genes cuya expresión se ha determinado y/o variables clínicas determinadas.
- 35
- 40
- 45 3. Método según la reivindicación 1 o 2 en donde la determinación de los valores de expresión de los distintos genes se determina mediante PCR en tiempo real.
- 50 4. Método según cualquiera de las reivindicaciones 1 a 3 en donde la tabla de probabilidades condicionadas entre los niveles de expresión de cada uno de los genes y los valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico y entre los valores modales de cada una de las variables clínicas y los valores de probabilidad de que la esclerosis múltiple sea de buen o de mal pronóstico son las indicada en la tabla 14.
- 55 5. Un kit que comprende un conjunto de sondas en donde dicho conjunto comprende una sonda específica para cada uno de los genes indicados en al menos una tabla seleccionada del grupo de tablas 3, 5-8 y 11, en donde las sondas específicas para los genes del grupo de tablas 3, 5-8 y 11 son las sondas mencionadas en dichas tablas.
- 60 6. Un kit según la reivindicación 5, en donde el conjunto de sondas consiste en una sonda específica para cada uno de los genes indicados en al menos una tabla seleccionada del grupo de tablas 3, 5-8 y 11.
7. Un kit según la reivindicación 6 en donde el conjunto de sondas consiste en una sonda específica para cada uno de los genes que se indica en la tabla 11.
8. Un kit según cualquiera de las reivindicaciones 5 a 7 en donde el kit comprende, adicionalmente, al menos una sonda específica para un gen de referencia de expresión constitutiva.
- 65 9. Un kit según la reivindicación 8 en donde el al menos un gen de referencia se selecciona del grupo de GABPA, UBC, beta-actin y beta-microglobulina.

ES 2 334 084 B1

10. Un kit según cualquiera de las reivindicaciones 5 a 9 en donde las sondas se encuentran formando parte de una matriz.

5 11. Un kit según la reivindicación 10 en donde la matriz es una LDA (low-density array).

12. Uso de un kit según cualquiera de las reivindicaciones 5 a 11 para la determinación del pronóstico de un enfermo diagnosticado de esclerosis múltiple, para determinar la eficacia de un tratamiento para la esclerosis múltiple o para diagnosticar la esclerosis múltiple en un paciente.

10

15

20

25

30

35

40

45

50

55

60

65

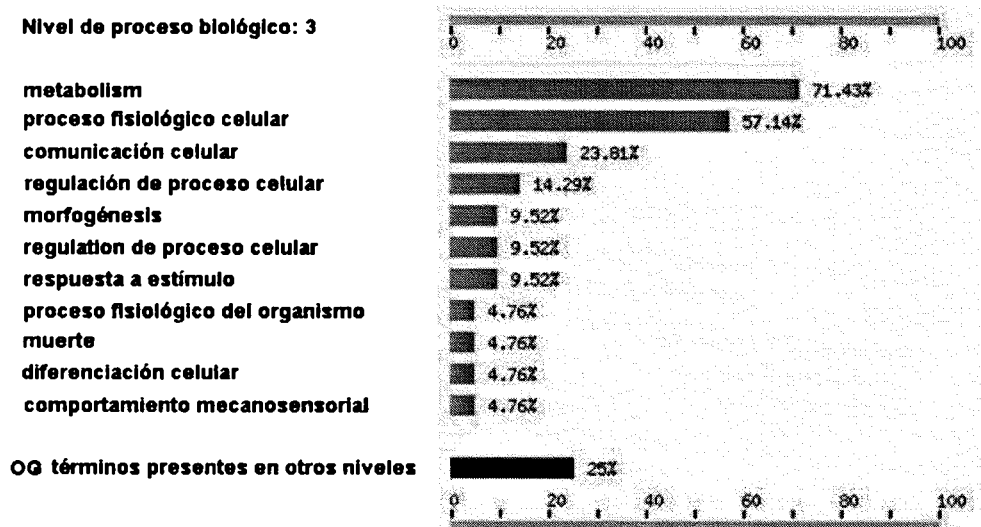


Figura 1

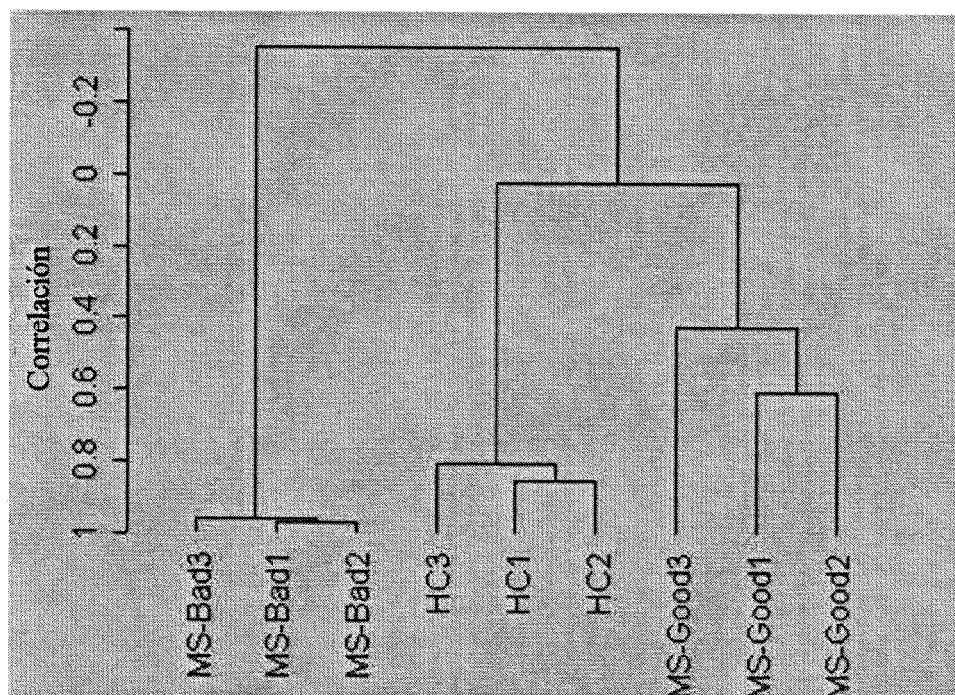


Figura 2

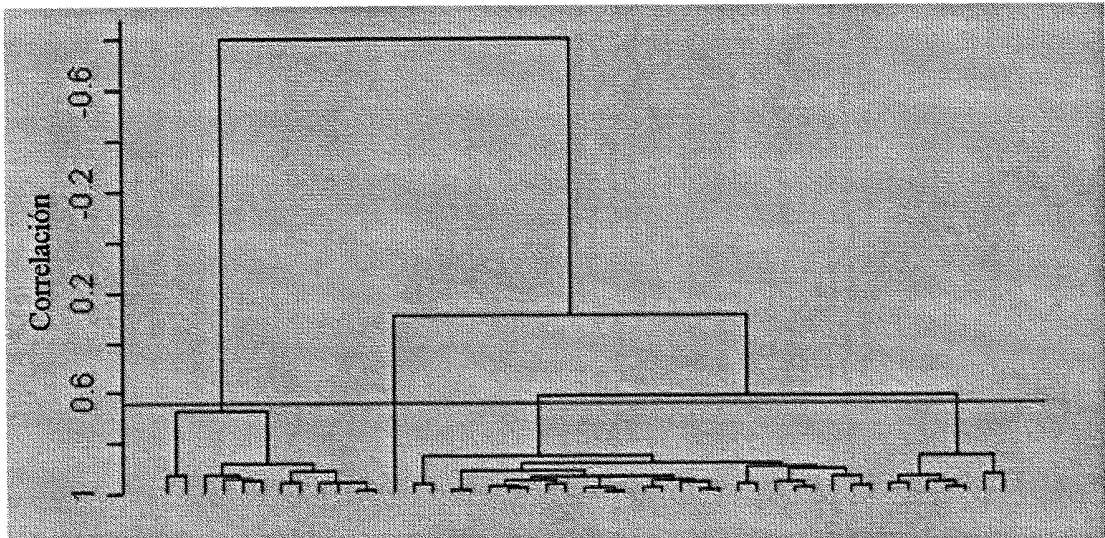


Figura 3

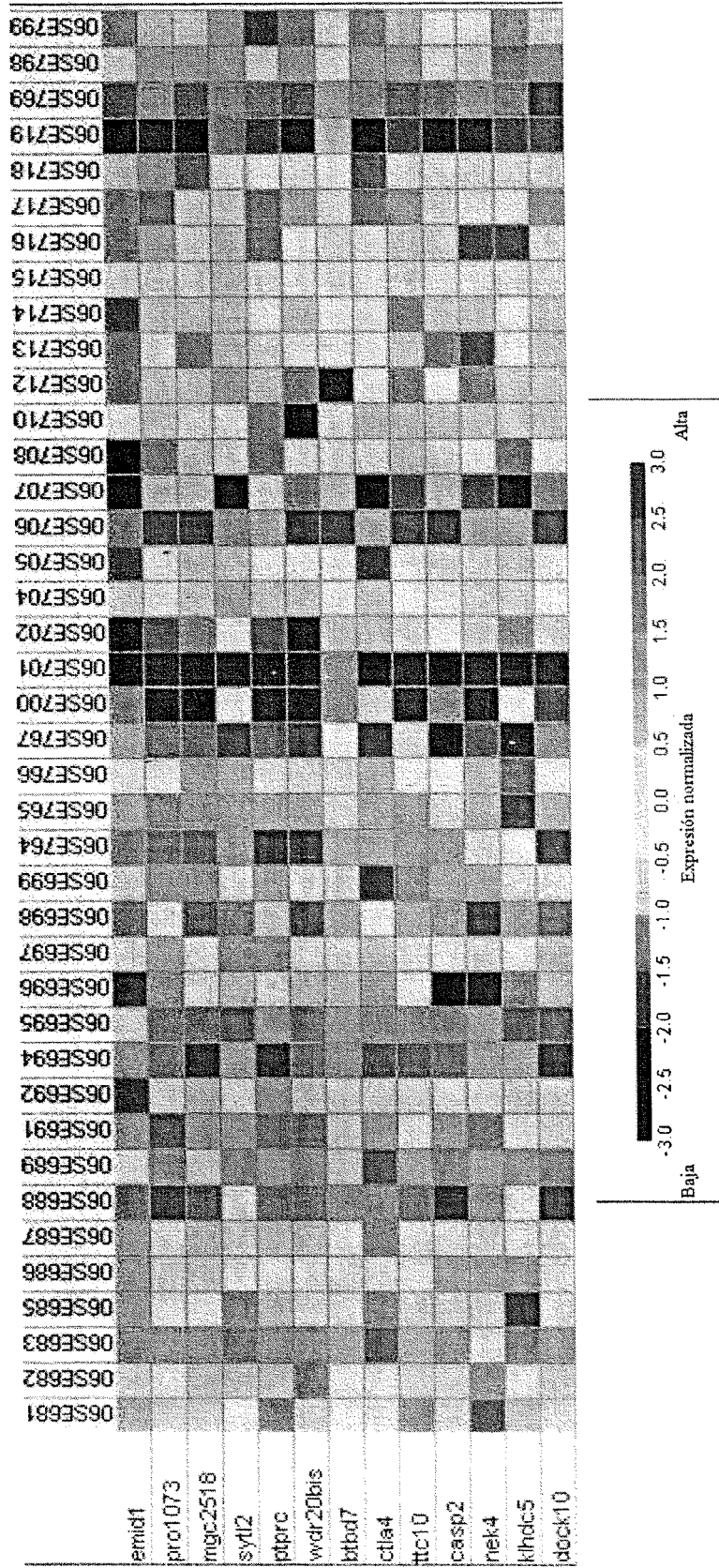


Figura 7

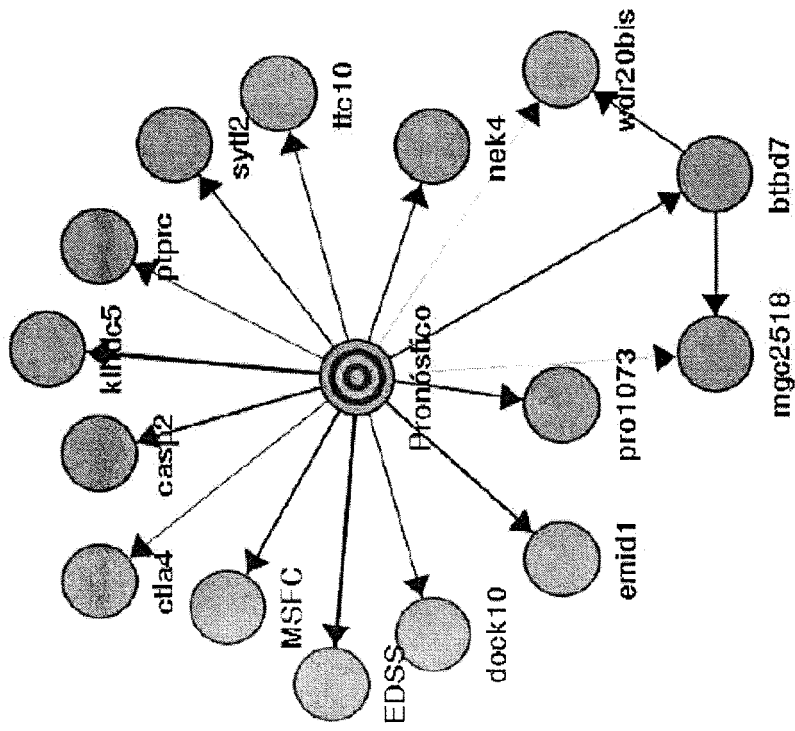
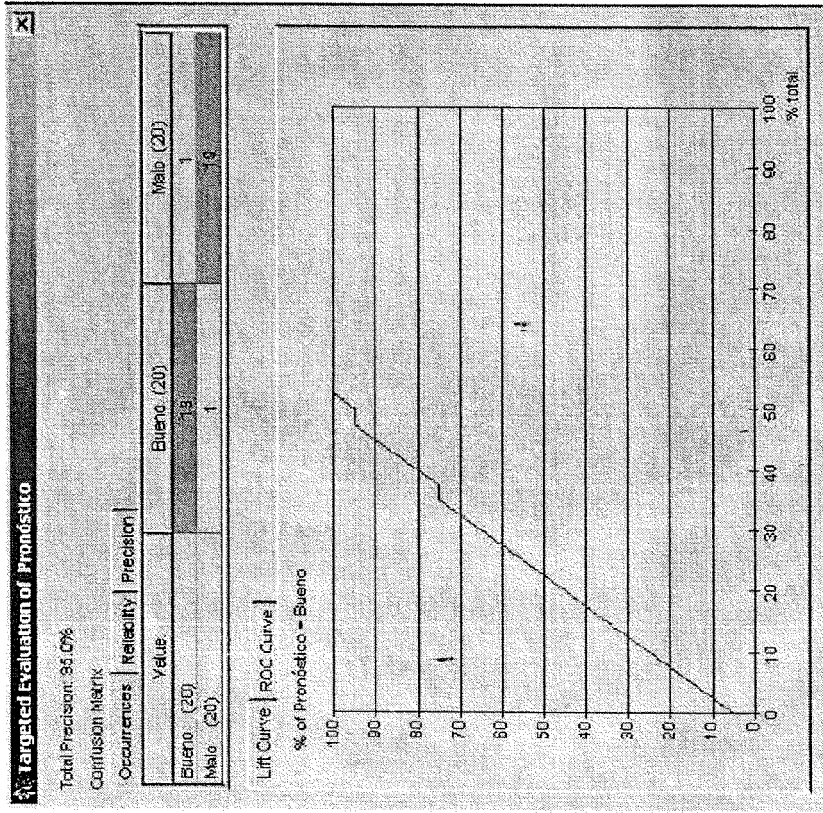


Figura 8



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

① ES 2 334 084

② Nº de solicitud: 200703415

③ Fecha de presentación de la solicitud: 21.12.2007

④ Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤ Int. Cl.: C12Q 1/68 (2006.01)

DOCUMENTOS RELEVANTES

| Categoría | ⑥ Documentos citados | Reivindicaciones afectadas |
|-----------|---|----------------------------|
| A | IGLESIAS AH. Microarray detection of E2F pathway activation and other targets in multiple sclerosis peripheral blood mononuclear cells. Journal of Neuroimmunology. 2004, Vol. 150, páginas 163-177, todo el documento. | 1-12 |
| A | CHRISTOPHER B. Gene microarray analysis of multiple sclerosis lesions. TRENDS in Molecular Medicine. 2003, VOL. 9(12), páginas 535-541, todo el documento. | 1-12 |
| A | WO 2006138561 A2 (SOURCE MDX) 28.12.2006, página 3, línea 1 - página 4, línea 15. | 1-12 |
| A | US 20040018540 A1 (YAMAMURA et al.) 29.01.2004, página 1, párrafos 0005-0009. | 1-12 |

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:

Fecha de realización del informe
18.02.2010

Examinador
Mª D. García Grávalos

Página
1/4

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

C12Q1

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

Fecha de Realización de la Opinión Escrita: 18.02.2010

Declaración

| | | |
|--|-----------------------|-----------|
| Novedad (Art. 6.1 LP 11/1986) | Reivindicaciones 1-12 | SÍ |
| | Reivindicaciones | NO |
| Actividad inventiva (Art. 8.1 LP 11/1986) | Reivindicaciones 1-12 | SÍ |
| | Reivindicaciones | NO |

Se considera que la solicitud cumple con el requisito de **aplicación industrial**. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

Base de la Opinión:

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como ha sido publicada.

1. Documentos considerados:

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

| Documento | Número Publicación o Identificación | Fecha Publicación |
|-----------|--|-------------------|
| D01 | Iglesias AH. Journal of Neuroimmunology. 2004, Vol. 150, páginas 163-177. | 2004 |
| D02 | Christopher B. TRENDS in Molecular Medicine. 2003, Vol. 9(12), páginas 35-541. | 2003 |
| D03 | WO 2006138561 A2 | 28-12-2006 |
| D04 | US 20040018540 A1 | 29-01-2004 |

2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración

La presente solicitud divulga un método "in vitro" para determinar el pronóstico clínico de un paciente con esclerosis múltiple, comparando el valor de la expresión de los genes: KLHDC5, CASP2, EMID1, PRO1073, BTBD7, MGC2518, WDR20bis, NEK4, SYLT2, DOCK10, TTC10, PTPRC, CTLA4, con una tabla de probabilidades de buen o mal pronóstico de la enfermedad, junto con el valor de una variable clínica del grupo de EDSS y MSFC con una tabla de probabilidades condicionadas y asignar una probabilidad de pronóstico asociada a la expresión o variable (reivindicaciones 1-4). Se refiere también a un kit para identificación de los genes indicados y a su uso para diagnóstico de esclerosis múltiple en un paciente o para determinar el pronóstico de evolución de la enfermedad y/o eficacia en el tratamiento en un paciente afectado con esclerosis múltiple (reivindicaciones 5-12)

El documento D01 divulga un método, utilizando microarrays de ADN, para detección de la activación de la vía E2F en células mononucleares de sangre periférica de pacientes con esclerosis múltiple (ver todo el documento)

El documento D02 analiza, utilizando microarrays de ADN, la expresión de una serie de genes en lesiones cerebrales provocadas por una esclerosis múltiple, identificando una serie de genes como posibles dianas terapéuticas (ver todo el documento)

El documento D03 divulga un método de identificación de marcadores biológicos asociados a la esclerosis múltiple, mediante el análisis de los niveles de expresión de una serie de genes; así como el uso de los genes expresados para monitorización y tratamiento de la esclerosis múltiple, además de la evaluación de la inflamación inducida o relacionada con la enfermedad (ver página 3, línea 1 - página 4, línea 15)

El documento D04 divulga un método, utilizando microarrays de ADN, de evaluación de la eficacia del tratamiento con interferón B en un paciente con esclerosis múltiple, determinando los niveles de expresión de un cluster de genes específicos en leucocitos derivados de sangre periférica de un paciente con dicha enfermedad (ver página 1, párrafos 0005-0009)

1. NOVEDAD Y ACTIVIDAD INVENTIVA (Art. 6.1 y Art. 8.1 LP 11/1986)**1.1. REIVINDICACIONES 1-12**

El objeto técnico de la invención de la presente solicitud es un método, utilizando microarrays de ADN, para determinar el pronóstico clínico de un paciente con esclerosis múltiple comparando la expresión de los genes: KLHDC5, CASP2, EMID1, PRO1073, BTBD7, MGC2518, WDR20bis, NEK4, SYLT2, DOCK10, TTC10, PTPRC, CTLA4, junto con el valor de una variable clínica del grupo de EDSS y MSFC, asignando una probabilidad de pronóstico asociada a la expresión o a la variable; así como un kit para diagnóstico y/o pronóstico de la enfermedad.

Aunque en el Estado de la Técnica se encuentran documentos de patente y publicaciones relacionadas con el diagnóstico, pronóstico y evaluación de la esclerosis múltiple mediante técnicas con microarrays de ADN, evaluando la expresión de uno o varios genes seleccionados, en los documentos citados no se hace referencia a la expresión del conjunto de genes KLHDC5, CASP2, EMID1, PRO1073, BTBD7, MGC2518, WDR20bis, NEK4, SYLT2, DOCK10, TTC10, PTPRC, CTLA4, ni a su relación con la esclerosis múltiple.

En consecuencia, la invención, tal y como se recoge en las reivindicaciones 1-12, cumple con los requisitos de novedad y actividad inventiva (Art. 6.1 y Art. 8.1 LP 11/1986)