(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 411 833**

(21) Número de solicitud: 201132151

(51) Int. Cl.:

C12Q 1/68 (2006.01)

(12) SOLICITUD DE PATENTE A2

(22) Fecha de presentación:

30.12.2011

(43) Fecha de publicación de la solicitud:

08.07.2013

(71) Solicitantes:

FUNDACIÓN PARA LA INVESTIGACIÓN BIOMÉDICA DEL HOSPITAL CLÍNICO SAN CARLOS (100.0%)
C/ Profesor Martín Lagos, s/n
28040 MADRID ES

(72) Inventor/es:

SANZ ORTEGA, Julián;
FERRER ALDEA, Milagros;
HERNÁNDEZ PRIETO, Susana;
ROMERA LÓPEZ, Alejandro;
PÉREZ-VILLAMIL SALGADO, Beatriz;
HERNANDO TRANCHO, Florentino;
GÓMEZ MARTÍNEZ, Ana María;
JARABO SARCEDA, Jose Ramón;
TORRES GARCÍA, Antonio José;
LÓPEZ GARCÍA-ASENJO, Jose Antonio;
GONZÁLEZ LARRIBA, Jose Luis;
PUENTE VÁZQUEZ, Javier y
DÍAZ-RUBIO GARCÍA, Eduardo

(74) Agente/Representante:

PONS ARIÑO, Ángel

(54) Título: **Método de pronóstico del carcinoma no microcítico de pulmón de estadio I o II.**

(57) Resumen:

Método de pronóstico del carcinoma no microcítico de pulmón de estadio I o II.

La presente invención se refiere a un método in vitro de obtención de datos Útiles para el pronóstico de cáncer de pulmón no microcítico de estadio I o II caracterizado por la detección y/o cuantificación de perfil de expresión génica de 50 biomarcadores. También se refiere a un método in vitro de pronóstico de cáncer de pulmón no microcítico de estadio I o II que comprende la detección y/o cuantificación de perfil de expresión génica de 50 biomarcadores. Además también se refiere al kit que comprende las sondas capaces de detectar dichos biomarcadores y su uso para la obtención de datos útiles para el pronóstico de cáncer de pulmón no microcítico de estadios I y II.

ES 2 411 833 A2

DESCRIPCIÓN

Método de pronóstico del carcinoma no microcítico de pulmón de estadio I o II.

La presente invención se refiere a un método *in vitro* de pronóstico en carcinoma no microcítico de pulmón de estadios I o II basado en la expresión diferencial de 50 genes. Mediante el método de la invención se diferencian pacientes con buen pronóstico y pacientes con mal pronóstico. La presente invención también se refiere a un kit que comprende un conjunto de sondas que reconocen los 50 genes de la invención. Por tanto, la invención se podría encuadrar en el campo de la medicina.

ESTADO DE LA TÉCNICA

El cáncer de pulmón es la primera causa de muerte por cáncer con una tasa anual de más de 1,1 millones de personas en todo el mundo, y con una tasa de supervivencia a cinco años de sólo el 15%. Aproximadamente el 80% de los casos diagnosticados se clasifican como carcinoma no microcítico de pulmón (CNMP) y el 20% restante corresponden a carcinoma microcítico de pulmón (CMP). En el CNMP, los tipos más frecuentes son el carcinoma epidermoide o escamoso y el adenocarcinoma.

El sistema de estadiaje TNM (7ª edición) basado en el tamaño del tumor (T), la afectación ganglionar (N) y la presencia de metástasis a distancia (M) es, en la actualidad, el factor pronóstico más utilizado en los pacientes con CNMP. En función de estos parámetros, los tumores se clasifican en: estadio I y estadio II (en ambos casos la enfermedad es localizada), estadio III (enfermedad localmente avanzada) y estadio IV (enfermedad metastásica) (Kligerman S. Amerian Journal of Roentgenology 2010. 194:562-573).

En estadios iniciales o tempranos (estadios I y II), la cirugía con intención curativa es el tratamiento de elección encontrándose en continua discusión el beneficio de la quimioterapia adyuvante para disminuir la elevada tasa de recurrencia posterior a la resección quirúrgica que oscila entre un 30-35% de los pacientes. En concreto, en estadios II, la quimioterapia adyuvante basada en platinos, como el cisplatino, ha demostrado mejorar la supervivencia de determinados subgrupos pero, por otro lado, existe un porcentaje de pacientes que a pesar de no recaer tras la cirugía reciben tratamiento adyuvante y que son por lo tanto pacientes tratados en exceso. Este sobretratamiento repercute en problemas en estos pacientes asociados a los efectos secundarios de dichos tratamientos. Respecto a los estadios I (que engloba a los subgrupos IA y IB), y según la guía de consenso elaborada por el “National Comprehensive Cancer Network” (NCCN) en 2011, en el subgrupo IA la quimioterapia adyuvante no está indicada, mientras que en los pacientes del subgrupo IB, sólo está recomendada en aquellos que cumplan factores de riesgo como pobre grado de diferenciación, invasión vascular, resección en cuña y márgenes mínimos. Por lo tanto, debido a la falta de precisión de los métodos actuales para definir el pronóstico de los estadios tempranos del CNMP, en la actualidad existen pacientes que reciben un tratamiento adyuvante que no les beneficia y también pacientes que no reciben un tratamiento adyuvante y que sin embargo tienen una alta probabilidad de recurrencia del tumor.

Actulamente, en el cáncer de pulmón no se conocen marcadores de probado valor pronóstico y predictivo que indiquen cúal será la progresión del paciente (Karapaniagiotou E, et al. Open Lung Cancer J 2009. 2: 24-30). En CNMP se han desarrollado estudios que utilizan plataformas de análisis masivo para la obtención de perfiles de expresión génica que puedan ser utilizadas como biomarcadores pronóstico. Los resultados obtenidos han sido dispares en cuanto a los genes a incluir en el biomarcador, quizás debido al uso de criterios diferentes en cuanto a la inclusión de pacientes en el estudio, la obtención de muestras, la elección de los estadios tumorales, la exclusión o no de subtipos histológicos de gran importancia en el CNMP, así como a la falta, en algunos casos, de validación independiente (Roepman P. et al. Clin Cancer Res 2009. 15:284-290; Chen HY et al. New Engl J Med 2007. 356(1):11-20); Raponi M et al. Cancer Res 2006 66:7466-7472; US20090062144; WO2010007093; Raz DJ et al. Clin Cancer Res 2008 14(17):5565–5570).

Por lo tanto existe la necesidad de desarrollar una herramienta alternativa que pueda ser usada clínicamente, que sea más efectiva que los factores de riesgo estándar en identificar aquellos pacientes completamente resecados y clasificados como estadio I que puedan beneficiarse de la quimioterapia adyuvante y que además permita identificar aquellos pacientes clasificados como estadios II que tengan bajo riesgo de recurrencia y en los que la quimioterapia no sería necesaria. Se requiere por lo tanto un método robusto que sea capaz de estratificar pacientes con CNMP en grupos de buen y mal pronóstico.

DESCRIPCIÓN DE LA INVENCIÓN

El problema técnico que resuelve la invención es el de proporcionar un método *in vitro* alternativo que determine el pronóstico del carcinoma no microcítico de pulmón (CNMP) en estadios iniciales para la obtención de un tratamiento personalizado del paciente.

En la presente invención se describe un método *in vitro* para el pronóstico de CNMP tanto de estadio I como de estadio II que se caracteriza por la detección y/o cuantificación de un producto de expresión del conjunto de 50 genes, que se muestran en la tabla 1 en la muestra biológica de un sujeto. La presente invención también se refiere al uso de los productos de expresión de dichos 50 genes como biomarcadores pronóstico de cáncer de CNMP de estadios I o II.

El método de la invención proporciona un predictor de 50 genes para estadios precoces de CNMP. La estrategia que se utilizó para la obtención de este predictor, comenzó por una detección y/o cuantificación de la expresión génica global de tumores de CNMP en estadios tempranos (I y II). En base a la expresión génica se realizó una clasificación molecular y una asociación con recidiva; la relación de los grupos moleculares con las variables histológicas y clínicas más importantes; la obtención de un predictor que identifica los grupos moleculares generados; la obtención de un predictor que diferencia un grupo de pacientes con buen pronóstico frente a un grupo de pacientes con mal pronóstico; y validación de los predictores con una serie externa. Finalmente, se observó que el método de la invención es útil para el pronóstico de CNMP. El predictor de la invención está constituido por 50 genes que se muestran en la tabla 1, de ahora en adelante, los denominados “50 genes de la invención”.

El término “predictor” se refiere en esta memoria a un perfil de expresión diferencial de genes o perfil de expresión génica.

Se entiende por “perfil de expresión génica” el perfil génico obtenido tras la cuantificación del producto de expresión de los genes de interés. Se entiende por “producto de expresión”, al ARN mensajero (ARNm), el ADN complementario (ADNc), el ARN complementario (ARNc) y/o la proteína producida por los genes de interés o biomarcadores, es decir, por los genes de la tabla 1, en una muestra biológica aislada.

El perfil de expresión de los genes se realiza, preferiblemente, determinando el nivel de ARNm derivado de su transcripción, previa extracción del ARN total presente en la muestra biológica aislada, lo cual puede realizarse mediante protocolos conocidos en el estado de la técnica. La determinación del nivel de ARNm derivado de la transcripción de los genes de la tabla 1 puede realizarse, por ejemplo, aunque sin limitarnos, mediante amplificación por reacción en cadena de la polimerasa (PCR), retrotranscripción en combinación con la reacción en cadena de la polimerasa (RT-PCR), RT-PCR cuantitativa, retrotranscripción en combinación con la reacción en cadena de la ligasa (RT-LCR), o cualquier otro método de amplificación de ácidos nucleicos; análisis en serie de la expresión génica (SAGE, SuperSAGE); *microarrays* de ADN o de ARN elaborados con oligonucleótidos o sondas sintetizados *in situ* mediante fotolitografía o por cualquier otro mecanismo; hibridación *in situ* utilizando sondas específicas marcadas con cualquier método de marcaje; mediante geles de electroforesis; mediante transferencia a membrana e hibridación con una sonda específica; mediante resonancia magnética nuclear o cualquier otra técnica de diagnóstico por imagen utilizando nanopartículas paramagnéticas o cualquier otro tipo de nanopartículas detectables funcionalizadas con anticuerpos o por cualquier otro medio. El perfil de expresión génica también podría obtenerse mediante la detección y/o cuantificación de las proteínas producto de la traducción del ARNm derivado de la transcripción de los genes de la tabla 1, mediante por ejemplo, pero sin limitarnos, inmunodetección por *inmuno blotting,* inmunohistoquímica, cromatografía o *microarrays.*

La presente invención podría referirse también a un método *in vitro* para el pronostico de CNMP tanto de estadio I como de estadio II que se caracteriza por la detección del número de copias en el ADN de los 50 genes que se muestran en la tabla 1, así como de las alteraciones epigenéticas como la hipermetilación del promotor de los genes o como de la alteración de la estabilidad del ARNm debido entre otros factores a modificaciones transcripcionales que afectan por ejemplo a la cola de Poli Adeninas. La presente invención también se refiere al uso de estas alteraciones de los 50 genes como biomarcadores pronóstico de cáncer de CNMP de estadios I o II.

Finalmente el perfil de expresión génica también podría obtenerse mediante la detección y/o cuantificación del número de copias de los genes presentes en la tabla 1, así como de los niveles de alteraciones epigenéticas como el nivel de metilación del promotor o de los niveles de estabilidad del mensajero de estos mismos genes. Esta detección podría llevarse a cabo, aunque sin limitarse mediante microarrays, CGH (Hibridación genómica comparada) o FISH (hibridación in situ fluorescente).

Esta invención también podría aplicarse para estadios avanzados (III y IV).

Por lo aquí descrito, un primer aspecto de la invención se refiere a un método *in vitro* de obtención de datos útiles para el pronóstico de CNMP en estadio I o II caracterizado por la detección y/o cuantificación del producto de expresión de los genes de la tabla 1 en la muestra biológica aislada de un sujeto. A partir de ahora nos referiremos a éste como al “método primero de la invención”.

El término “*in vitro”* se refiere a que el método de la invención se realiza fuera del cuerpo del sujeto.

El término “pronóstico” en la presente invención se refiere a la capacidad de detectar pacientes que presentan una alta o baja probabilidad de recidiva tras la cirugía. Una alta probabilidad de recidiva se asocia a un mal pronóstico mientras que una baja probabilidad de recidiva se asocia a un buen pronóstico. Se entiende por “recidiva” la reaparición de la enfermedad, en este caso de un cáncer de pulmón. Las expresiones “probabilidad de no recidiva” y “probabilidad de ILE (intervalo libre de enfermedad)” se usan indistintamente en la presente memoria.

El término “cáncer de pulmón no microcítico”, “carcinoma no microcítico de pulmón” (CNMP), “carcinoma de pulmón no microcítico” (CPNM), o cáncer pulmonar de células no pequeñas (en inglés “non-small cell lung cancer”, NSCLC) se

refiere a un tipo de cáncer o tumor de pulmón según clasificación histológica que comprende el subtipo carcinoma escamoso o epidermoide, adenocarcinoma, adenoescamoso, carcinoma sarcomatoide, y carcinoma de células grandes.

Se entiende por “estadio” la fase o la clasificación del cáncer de pulmón en base a la clasificación TNM. La clasificación TNM se refiere al tamaño del tumor (T), la afectación de ganglios linfáticos (N) y la afectación de otros órganos (M). El estadio I se refiere a los subestadios IA o IB. El subestadio IA se refiere a los tumores de pulmón de clasificación T1N0M0. El subestadio IB incluye los tumores de pulmón de clasificación T2aN0M0. El estadio II se refiere a cualquiera de los subestadios IIA o IIB. El subestadio IIA se refiere a los tumores de pulmón de clasificación T1N1M0, T2aN1M0 y T2bN0M0. El subestadio IIB incluye los tumores de pulmón de clasificación T2bN1M0 y T3N0M0. En la clasificación TNM, T1 se refiere a cuando el tumor ≤ 3 cm de dimensión máxima, está rodeado por tejido pulmonar o pleura visceral y sin invasión proximal al bronquio lobar en fibrobroncoscopia. El T1a es un tumor ≤ 2cm y el T1b es un tumor > 2cm y ≤ 3cm. T2 se refiere a un tumor > 3 cm de dimensión máxima y ≤ 7 cm o un tumor con al menos una de las siguientes características: infiltrar el bronquio principal a 2 cm o menos de la carina, invadir pleura visceral o asociarse con atelectasias o neumonitis obstructiva. T2a es un tumor > 3 cm y ≤ 5 cm y T2b es un tumor > 5 y ≤ 7 cm. T3 se refiere a un tumor > 7 cm o un tumor que afecta a la pared costal (incluidos los tumores de la cisura superior), diafragma, pleura mediastínica o pericardio; sin afectación del corazón, grandes vasos, tráquea, esófago, cuerpos vertebrales; o un tumor del bronquio principal a menos de 2 cm de la carina, sin infiltración de la misma; donde la atelectasia afecta a todo un pulmón y puede existir derrame pleural no maligno. N0 se refiere al tumor de pulmón sin afectación de los ganglios linfáticos. N1 se refiere al tumor que presenta afectación de los ganglios linfáticos peribronquiales o hiliares ipsilaterales o ambos. M0 se refiere al tumor de pulmón que no presenta metástasis a distancia.

En la presente invención los términos “estadios tempranos”, “estadios iniciales” o “estadios precoces” se refieren a estadio I o II de CNMP.

El término “genes de la tabla 1” se refiere a los genes descritos en la tabla 1 que se muestra a continuación.

Los términos “Identificador Entrez” o “ID Entrez” se refieren al número de referencia del gen en la base de datos de genes del NCBI (National Centre for Biotechnology Information, U.S. National Library of Medicine).

Tabla 1: listado de los 50 genes del predictor de la invención.

ID Entrez	Símbolo	Descripción
270	*AMPD1*	Adenosina monofosfato deaminasa
608	*TNFRSF17*	Miembro 17 de la superfamilia del receptor de necrosis tumoral
930	*CD19*	CD19
939	*CD27*	CD27
952	*CD38*	CD38
973	*CD79A*	Molécula alfa asociada a inmunoglobulina
974	*CD79B*	Molécula beta asociada a inmunoglobulina
3002	*GZMB*	Granzima B
3493	*IGHA1*	Inmunoglobulina pesada constante alfa 1
3494	*IGHA2*	Inmunoglobulina pesada constante alfa 2
3500	*IGHG1*	Inmunoglobulina pesada constante gamma 1
3512	*IGJ*	Polipéptido J de inmunoglobulina
3535	*IGL@*	Locus lambda de inmunoglobulina
3543	*IGLL1*	Polipéptido 1 "like" inmunoglobulina lambda
3662	*IRF4*	Factor 4 regulador del interferón
3782	*KCNN3*	Miembro 3 subfamilia N de canales de potasio activados por calcio
3887	*KRT81*	Keratina 81

4283	*CXCL9*	Ligando 9 del motivo CXC de quimioquinas
5368	*PNOC*	Preponocipeptina
5450	*POU2AF1*	Factor 1 asociado a POU clase 2
8419	*BFSP2*	Faquinina
9834	*KIAA0125*	KIAA0125
10563	*CXCL13*	Ligando 13 del motivo CXC de quimioquinas
11040	*PIM2*	Oncogén Pim2
26952	*SMR3A*	Proteína 3A regulada por glándula submaxilar
28904	*IGKV1D-8*	Cadena variable kappa 1D-8 de inmunoglobulina
51237	*MZB1*	Proteína especifica de las células B1 y B de la zona marginal
51303	*FKBP11*	Proteína 11 de unión a FK506
54900	*LAX1*	Adaptador 1 transmembrana de linfocitos
57699	*CPNE5*	Copina V
57823	*SLAMF7*	Miembro 7 de la familia SLAM
78986	*DUSP26*	Fosfatasa 26 especifica dual
79368	*FCRL2*	"like" receptor 2 Fc
80307	*FER1L4*	Pseudogen 4 "like" fer-1
83416	*FCRL5*	Receptor Fc "like” 5
84824	*FCRLA*	Receptor Fc "like” A
90925	*IGHV5-78*	Pseudogen región pesada variable de cadena pesada de inmunoglobulina 5-78
91319	*DERL3*	Miembro 3 de la familia de dominio parecido a Der-1
92154	*MTSS1L*	"like" supresor metástasis 1
126306	*JSRP1*	Proteína 1 de retículo sarcoplasmático
140947	*C5orf20*	Marco de lectura abierta 20 del cromosoma 5
150365	*MEI1*	Inhibidor de meiosis 1
221188	*GPR114*	Receptor 114 acoplado a proteína G
401847	*LOC401847*	Proteína hipotética LOC401847
642424	*LOC642424*	Región Walker "like" de la cadena kappa V-I de la inmunoglobulina
100132941	*LOC100132941*	Proteína hipotética LOC100132941
100133862	*LOC100133862*	Región V35 "like" de la cadena pesada V-I de la inmunoglobulina
100287723	*LOC100287723*	Región Walker "like" de la cadena kappa V-I de la inmunoglobulina
100290415	*IGHV1-24*	Inmunoglobulina pesada variable 1-24
100293440	*LOC100293440*	Similar a la cadena lambda de inmunoglobulina

A continuación se muestra una breve descripción de algunas de las funciones conocidas de los genes presentados en la tabla 1:

AMPD1: Cataliza la deaminación de la adenosina monofosfato (AMP) a inosina monofosfato (IMP) en el músculo esquelético y tiene un importante papel en el ciclo de las purinas.

TNFRSF17: Este receptor se expresa en linfocitos B maduros y es importante para el desarrollo de las células B y en la respuesta autoinmune. Tiene como ligando al miembro 13b de la superfamilia del factor de necrosis tumoral y activa el factor nuclear del potenciador del gen polipetídico de la cadena ligera Kappa en células B (NF-kappaB) y la proteína kinasa activada por mitógeno 8 (MAPK8/JNK). También se une a otros ligandos y envía señales de supervivencia celular y proliferación.

CD19*:* Molécula que se une al receptor de antígenos de los linfocitos B para disminuir el umbral de estimulación de los linfocitos a través de la estimulación por antígeno.

CD27: Miembro de la superfamilia del receptor del factor de necrosis tumoral. El receptor tiene la función de generar y mantener durante largo tiempo la inmunidad de las células T. El ligando CD70 se une a él y funciona en la activación de las células B y en la síntesis de inmunoglobulinas. Las proteínas adaptadoras denominadas Factor Asociado a Receptores de Factores de Necrosis Tumoral 2 y 5 (TRAF2 y TRAF5) median en este proceso. La proteína de unión a CD27 (SIVA) es una proteína proapoptótica que juega un importante papel en la apoptosis mediada por este receptor.

CD38*:* Es una ectoenzima multifuncional que se expresa en multitud de células y tejidos especialmente en leucocitos. CD38 también tiene funciones en la adhesión celular, transducción de señales y señalización por calcio.

CD79A *y* ***CD79B****:* codifican para las proteínas Ig-alpha e Ig-beta que son componentes del receptor antigénico de linfocitos B. Las moléculas Ig-alfa e Ig-beta son necesarias para la expresión y función de este receptor.

GZMB*:* Los linfocitos T citolíticos (CTL) y las células “natural killer” (NK) tienen la habilidad de reconocer, unir y lisar células diana específicas. La GZMB es crucial para la rápida inducción de la apoptosis de las células diana a través de la respuesta inmune generada por los linfocitos T citolíticos.

IGHA1 *e* ***IGHA2****:* Anticuerpo con una importante presencia en las secreciones mucosas y que representa la primera línea de defensa del organismo. Existen dos subclases Inmunoglobulina A1 (IgA1) e Inmunoglobulina (IgA2).

IGHG1*:* Este gen se encuentra traslocado en la leucemia linfocítica crónica de células B con el gen Ciclina D1 (CCND1) y en subclases de linfomas MALT (Tejido Linfoide Asociado a Mucosa) con los genes "LIM homeobox 4" (LHX4) y "Forkhead box P1" (FOXP1).

IGJ*:* Su función es unir dos monómeros o bien de Inmunoglobulina M (IgM) o bien de Inmunoglobulina A (IgA). También tiene la función de unir estas inmunoglobulinas al componente secretor.

IGL@: Cada molécula de inmunoglobulina tiene dos cadenas pesadas idénticas y dos cadenas ligeras idénticas. Hay dos clases de cadenas ligeras que son kappa y lambda. Este gen abarca el locus de la cadena ligera lambda que incluye el segmento V (variable), segmento J (unión) y segmento C (constante).

IGLL1*:* Es un gen de la superfamilia de las inmunoglobulinas que codifica para la cadena ligera sustitutiva del receptor de células preB. Mutaciones en este gen pueden producir deficiencia de células B o agammaglobulinemia.

IRF4*:* Pertenece a la familia de factores reguladores del interferón. Es específico de linfocitos y regula negativamente los receptores tipo Toll (o TLR), que es una molécula central en la activación de la respuesta inmune innata y adaptativa.

KCNN3*:* Regula la excitabilidad neuronal.

KRT81*:* Es un miembro de la familia de keratinas.

CXCL9*:* Su función no está bien definida pero parece que está implicado en el tráfico de células T.

PNOC*:* Es un neuropéptido que actúa como ligando endógeno del receptor "Opiate Receptor-Like 1" (ORL1).

POU2AF1*:* Es un coactivador específico de células B y su ausencia parece estar relacionada con defectos en el desarrollo de células B y con la falta de centros germinales.

BFSP2*:* también denominada faquinina, es una proteína estructural de filamentos del citoesqueleto. Junto a la filensina forma el BF ("beaded filament").

CXCL13*:* Promueve la migración de linfocitos B preferentemente frente a linfocitos T y macrófagos mediante estimulación con calcio.

PIM2*:* Es una serin/treonin/protein kinasa. Previene apoptosis y promueve supervivencia celular.

***SMR3A**:* Es un homólogo funcional del gen Vcsa1 ("Variable Coding Sequence A1"). Se ha asociado como un marcador de la disfunción eréctil asociada con etiología tanto diabética como no diabética.

***MZB1**:* Está asociada con las cadenas pesadas y ligeras de la inmunoglobulina tipo M (IgM), promoviendo el ensamblaje de la IgM y su secreción.

***FKBP11**:* Pertenece a la familia FKBP las cuales catalizan el plegamiento de los polipéptidos que contienen prolina. Su función es inhibida por FK506 y por rapamicina.

***LAX1**:* Un regulador negativo de la señalización de linfocitos.

***CPNE5**:* Proteína de unión a membrana dependiente de calcio que parece estar implicada en la regulación de fenómenos moleculares en la interfase de la membrana celular y en el citoplasma.

***SLAM7**:* Está implicada en la activación de células NK y en la regulación de la proliferación de linfocitos B durante la respuesta inmune.

***DUSP26**:* Está asociado con la inactivación de la Proteína Kinasa activada por mitógenos 1 y 3 (MAPK1 y MAPK3), así como con la inhibición de la proliferación de células epiteliales, lo que podría sugerir un papel como gen supresor de tumores.

***FCRL2**:* Forma parte de la superfamilia de receptores de inmunoglobulinas. Puede ser un marcador pronóstico de leucemia linfocítica crónica.

***FCRL5**:* También es parte de la superfamilia de receptores de inmunoglobulinas. Está implicado en el desarrollo de células B y en la linfomagénesis.

***FCRLA**:* Este receptor media la destrucción de los antígenos reconocidos por la Inmunoglobulina G (IgG). Es proteína selectiva de células B y puede estar implicada en su desarrollo.

***DERL3**:* Proteína que se ubica en el retículo endoplasmático con la función de degradar glicoproteínas mal plegadas.

***MTSS1L**:* Puede estar implicada en el empaquetamiento de la actina. Pertenece a la familia MTSS1 (Supresores de Metástasis Tipo 1).

***JSRP1**:* El retículo sarcoplasmático es un compartimento celular que controla la concentración de calcio intracelular y está implicado en las funciones de excitación-contracción de este compartimento celular. En ratones se ha visto que esta proteína interacciona con proteínas claves implicadas en estos procesos de excitación-contracción.

***C5orf20**:* Este gen se expresa en células dendríticas, que son potentes células presentadoras de antígenos implicadas en activar las células T nativas para iniciar la respuesta inmune específica de antígeno.

***MEI1**:* Defectos en su expresión están relacionados con parada en meiosis y se asocia a fenómenos de azoospermia.

***GPR114**:* Proteína G asociada a receptores con un extremo N terminal que contiene regiones ricas en serina/treonina. Se ha descrito su expresión en linfocitos citotóxicos.

***IGHV5-78, FER1L4, IGKV1D-8, KIAA0125, LOC401847, LOC642424, LOC100132941, LOC100133862, LOC100287723, IGHV1-24** y **LOC100293440**:* a día de hoy, todavía no se conoce la función de estos genes.

El término “muestra biológica” incluye, pero sin limitarnos, tejidos y/o fluidos biológicos de un individuo, obtenidos mediante cualquier método conocido por un experto en la materia que sirva para tal fin.

El término “sujeto” se refiere a un individuo, preferentemente humano, que ha sido diagnosticado de CNMP.

Una realización preferida del primer aspecto de la invención se refiere a un método que además comprende la comparación de los datos útiles obtenidos de la muestra biológica aislada de un nuevo sujeto, con los valores de expresión de referencia para los genes de la tabla 1 obtenidos de sujetos con CNMP estadio I o II en los que el pronóstico es conocido (muestra de referencia). La comparación permite la identificación del nuevo sujeto como un sujeto de buen pronóstico o de mal pronóstico. A partir de ahora, nos referiremos a este método como al “método segundo de la invención”.

El término “muestras de referencia” tal como se entiende en la presente invención se refiere, por ejemplo, pero sin limitarse, a las muestras obtenidas de individuos que presenten un perfil molecular conocido. Este perfil molecular puede ser de buen pronóstico o de mal pronóstico.

Un experto en la materia podría clasificar un nuevo paciente en el grupo de buen o en el grupo de mal pronóstico al comparar sus datos de expresión para los 50 genes de la invención con los datos de expresión para los 50 genes en las muestras de referencia. Estas muestras de referencia son un grupo de muestras de las que se conoce el perfil de

expresión de los 50 genes y la presencia o no de recidiva. Por ejemplo, pero sin limitarse, un nuevo sujeto cuyo perfil de expresión sea similar al grupo de referencia de buen pronóstico puede ser clasificado como perteneciente al grupo de buen pronóstico, el cual tiene una probabilidad media de ILE a los 3 años del 85% y/o a los 5 años del 79%. Por ejemplo, pero sin limitarse, un nuevo sujeto cuyo perfil de expresión sea similar al grupo de referencia de mal pronóstico puede ser clasificado como perteneciente al grupo de mal pronóstico, el cual tiene una probabilidad media de ILE a los 3 años del 62% y/o a los 5 años del 48%.

La determinación del pronóstico de nuevos pacientes diagnosticados con CNMP en estadios I o II implica la clasificación de esos pacientes en uno de los dos grupos de referencia previamente definidos: grupo de buen pronóstico o grupo de mal pronóstico. Estos grupos de referencia están constituidos por las muestras de referencia.

La comparación de los datos útiles obtenidos de la muestra biológica de un nuevo sujeto, con los valores de expresión de referencia para los genes de la tabla 1 obtenidos de sujetos con CNMP estadio I o II en los que el pronóstico es conocido (muestra de referencia), puede llevarse a cabo mediante cualquier método estadístico de predicción conocido en el estado de la técnica, como por ejemplo, pero sin limitarse, en cualquiera de los métodos descritos en Simon R. et al. J Clin Oncol 2005; 23:7332-41.

En una realización preferida del método segundo de la invención, la comparación se realiza mediante el método del centroide compacto más cercano. En adelante, el “método tercero de la invención”.

Se entiende como el “método del centroide compacto más cercano” (“nearest shrunken centroid method”) el método de clasificación descrito en Tibshirani R. et al. PNAS. 2002, 99:6567-6572 y aplicado a través de la herramienta Predicción de Análisis por Microarrays (“Prediction analysis of microarrays” o PAM). La herramienta “PAM” fue desarrollada por la Universidad de Standford y es de libre acceso.

La determinación del pronóstico de CNMP de estadios I o II puede establecerse, aunque sin limitarse, mediante la determinación de un “valor de referencia” para el grupo de buen pronóstico (valor 1) y de otro para el grupo de mal pronostico (valor 2). El pronóstico puede realizarse estimando la distancia entre los valores de expresión de la nueva muestra y los “valores de referencia” de cada uno de los dos grupos. Si la distancia entre la nueva muestra y el valor 1 es menor que la distancia entre la nueva muestra y el valor 2, se podrá determinar el pronóstico favorable. Por el contrario, si la distancia entre la nueva muestra y el valor 1 es mayor que la distancia entre la nueva muestra y valor 2, se podrá determinar el pronóstico desfavorable.

Los valores de referencia de cada grupo, se pueden calcular en base a los valores de expresión de los 50 genes en las muestras de la matriz de referencia o “matriz de desarrollo” y vendrán expresados por tanto mediante un vector de 50 componentes. El cálculo del valor de referencia de cada grupo (en nuestro caso el grupo de buen pronóstico y el grupo de mal pronóstico), se obtiene de sumar al valor promedio global de todas las muestras, un segundo factor definido como la distancia (estadístico “t”) entre el valor promedio de expresión de los 50 genes de dicho grupo con respecto al valor promedio de expresión de los 50 genes de todas las muestras incluidas en la matriz de entrenamiento. Los datos del segundo factor serán estandarizados teniendo en cuenta, la variabilidad de expresión de cada uno de los 50 genes dentro del grupo analizado y teniendo en cuenta un valor de convergencia Δ que permite evaluar el poder predictivo de cada uno de los genes. Se entiende como distancia entre dos muestras, grupos o subtipos, la cuantificación de sus diferencias de expresión.

Aunque el valor final de referencia o “shrunken centroid” obtenido para cada grupo se basa en los valores de expresión, su valor real es adimensional y no es directamente proporcional a los datos de fluorescencia inicialmente obtenidos en cada muestra. Dicho valor de referencia, en cada grupo, contiene 50 componentes, una por cada uno de los genes analizados.

Una vez calculados los valores de referencia para cada grupo, el método del “nearest shrunken centroid”, es capaz de asignar nuevas muestras (que en nuestro caso conforman la matriz de validación) a cada uno de los grupos definidos. La distancia entre la nueva muestra y cada uno de los grupos es relativa a la diferencia entre los valores de expresión de los 50 genes en la muestra nueva con respecto a las componentes del centroide compacto (“shrunken centroid”) que representan cada grupo. La cuantificación de las distancias podrían medirse, aunque sin limitarse, mediante la distancia euclidea (Tibshirani R. Diagnosis of multiple cancer types by shrunken centroids of gene expression. PNAS 2002; 99(10):6567-72). Como se mencionó con anterioridad, la nueva muestra será asignada al grupo del que se encuentre a una menor distancia.

Por todo lo aquí descrito, un segundo aspecto de la invención se refiere a un método *in vitro* para el pronóstico del CNMP de estadio I o II caracterizado por:

a. la detección y cuantificación del producto de expresión de los genes de la tabla 1 en una muestra de referencia;

b. el cálculo de un valor de referencia (valor 1) para cada producto de expresión de los genes de la tabla 1 en las muestras de referencia de pronóstico favorable (grupo de buen pronóstico) y el cálculo de un

valor de referencia (valor 2) en las muestras de referencia de pronóstico desfavorable (grupo de mal pronóstico) mediante el uso del método del centroide más cercano;

c. la detección y cuantificación del producto de expresión de los genes de la tabla 1 en la muestra biológica de un nuevo sujeto en el que el pronóstico es desconocido (muestra de estudio);

d. la comparación mediante el uso del método de clasificación del centroide compacto más cercano de los valores obtenidos en la detección y cuantificación del producto de expresión de los genes de la tabla 1 en la muestra de estudio con los valores de referencia obtenidos en los grupos de buen y mal pronóstico.

e. la asociación de la muestra de estudio al grupo de buen pronóstico o al grupo de mal pronóstico según lo establecido en el método del centroide compacto más cercano.

En adelante este método se denominará “método cuarto de la invención”.

Una realización preferida del método cuarto de la invención se refiere al método donde el método del centroide compacto más cercano se lleva a cabo a través de la aplicación de Predicción de Análisis de Microarrays (PAM).

Una realización preferida del primer y del segundo aspecto de la invención, se refiere al método donde la muestra de referencia y las muestras de estudio han sido previamente normalizadas antes de la comparación.

Se entiende por “normalización” la utilización de una muestra control que sirva para eliminar variaciones experimentales entre las distintas muestras.

Otra realización preferida del primer y del segundo aspecto de la invención, se refiere al método que además comprende la detección y/o cuantificación de al menos un producto de expresión de los genes descritos en la tabla 2.

Otra realización preferida del primer y del segundo aspecto de la invención, se refiere al método donde el producto de expresión es ARN mensajero. Una realización aún más preferida se refiere al método donde la detección y/o cuantificación del ARN mensajero se realiza mediante microarrays. Una realización también más preferida se refiere al método donde la detección y/o cuantificación del ARN mensajero se realiza mediante RT-PCR.

Otra realización preferida del primer y del segundo aspecto de la invención, se refiere al método donde el producto de expresión es una proteína. Una realización aún más preferida se refiere al método donde la detección y/o cuantificación de la proteína se realiza mediante inmuno blotting, inmunohistoquímica, cromatografía o microarrays.

La detección y cuantificación del producto de expresión (ARNm, ARN complementario obtenido a partir de ADNc, ADN complementario o proteína) se puede realizar utilizando los métodos conocidos por el experto en la materia. Por ejemplo, determinando el nivel de ARNm derivado de su transcripción, previa extracción del ARN total presente en la muestra biológica aislada, lo cual puede realizarse mediante protocolos conocidos en el estado de la técnica. Para ello la muestra biológica aislada puede tratarse física o mecánicamente para romper el tejido o las estructuras celulares y liberar los componentes intracelulares a una solución acuosa u orgánica para preparar los ácidos nucleicos para un posterior análisis. Los ácidos nucleicos se extraen de la muestra por procedimientos conocidos por el experto en la materia y comercialmente disponibles. La determinación del nivel de ARNm derivado de la transcripción de los genes de la tabla 1 puede realizarse, por ejemplo, aunque sin limitarnos, mediante amplificación por reacción en cadena de la polimerasa (PCR), retrotranscripción en combinación con la reacción en cadena de la polimerasa (RT-PCR), RT-PCR cuantitativa, retrotranscripción en combinación con la reacción en cadena de la ligasa (RT-LCR), o cualquier otro método de amplificación de ácidos nucleicos; análisis en serie de la expresión génica (SAGE, SuperSAGE); microarrays, micromatrices o chips de ADN elaborados con oligonucleótidos depositados por cualquier mecanismo *o* elaborados con oligonucleótidos sintetizados *in situ* mediante fotolitografía o por cualquier otro mecanismo; hibridación *in situ* utilizando sondas específicas marcadas con cualquier método de marcaje; mediante geles de electroforesis; mediante transferencia a membrana e hibridación con una sonda específica; mediante resonancia magnética nuclear o cualquier otra técnica de diagnóstico por imagen utilizando nanopartículas paramagnéticas o cualquier otro tipo de nanopartículas detectables funcionalizadas con anticuerpos o por cualquier otro medio.

En la presente invención se demuestra que la detección y cuantificación del ARNm total de una muestra biológica de un sujeto con CNMP de estadios I o II es útil para el pronóstico de dicha enfermedad. Por todo ello en una realización preferida de este aspecto de la invención el producto de expresión detectado y cuantificado es ARNm.

Por ello, otra realización preferida del primer aspecto de la invención se refiere a un método donde el producto de expresión es ARNm.

Se entiende por “microarray” (microarray de expresión, chip o micromatriz) al conjunto de sondas (oligonucleótidos o ADNc) dispuestas de manera ordenada sobre una superficie sólida, que permite analizar simultáneamente la expresión del genoma completo de un organismo. Cada una de las sondas representa específicamente un gen determinado al poseer una secuencia complementaria al ARNm transcrito por dicho gen, posibilitando así, la medición de los niveles de expresión de todos los genes que conforman el genoma al mismo tiempo y en un único experimento. Para la utilización

de microarrays y obtención de datos a partir de ellos, la fase experimental de los microarrays puede constar de los pasos que se describen a continuación. En primer lugar, el ARN total se retrotranscribe usando como cebador un cebador específico para mensajero (PolidT) y una enzima retrotranscriptasa. Utilizando como molde el ADNc de doble cadena obtenido anteriormente, se sintetizó el ARNc, a la vez que se llevaba a cabo el proceso de amplificación y marcaje de la muestra. El ARNc marcado obtenido se purificó mediante columnas. El ARNc es fragmentado en secuencias mas pequeñas e hibridado al microarray. Dicho proceso de hibridación se lleva a cabo en un horno de hibridación durante un periodo largo de tiempo. En este proceso el ARNc marcado se une de manera específica a los oligonucleótidos sintetizados en el microarrays. Posteriormente el microarray es lavado para eliminar todo el ARNc excedente no unido a los oligonucleótidos.

De acuerdo con la presente invención el producto de expresión, preferiblemente ARNm o ADNc o ARN complementario (ARNc) obtenido a partir de ADNc, puede ser marcado o etiquetado mediante técnicas bien conocidas en el estado de la técnica. Etiquetas detectables incluyen, por ejemplo, isótopos radiactivos, etiquetas fluorescentes, etiquetas quimioluminiscentes, etiquetas bioluminiscentes o etiquetas enzimáticas. Las etiquetas fluorescentes pueden ser distintas en el caso del marcaje del producto de expresión de la muestra biológica y del producto expresión de la muestra control.

Por otra parte, la detección y cuantificación también se pueden realizar mediante RT-PCR, por lo que otra realización preferida del primer aspecto de la invención se refiere a un método según la reivindicación 7 donde la detección y/o cuantificación del ARNm se realiza mediante RT-PCR o preferiblemente mediante RT-PCR a tiempo real. El proceso de RT-PCR se puede llevar a cabo en dos fases:

- Retrotranscripción: se produce la unión entre un cebador y el ARNm mediante un proceso de incubación conjunta de ambos productos. Seguidamente se produce la retrotranscripción propiamente dicha utilizando enzimas de transcripción inversa.

- PCR posterior: se produce la amplificación del ADNc obtenido en la fase anterior mediante la técnica de reacción en cadena de la polimerasa (PCR). Para cada muestra y para cada transcrito de los genes analizados se llevará a cabo la reacción de PCR de manera individualizada. Este proceso implica la repetición cíclica de 3 fases: fase de desnaturalización del ADNc, fase de unión específica del oligonucleótido del gen en estudio a la hebra del ADNc desnaturalizado y fase de elongación a partir del oligonucleótido unido mediante la que se sintetizará una hebra nueva de ADNc. Al tratarse de un proceso que se mide en tiempo real, es necesario usar una molécula fluorescente para monitorizar lo que sucede a lo largo del proceso.

La cuantificación por otro lado también se puede realizar determinando el nivel de proteína derivado de la traducción de los ARNm transcritos a partir de los 50 genes de la invención. Esta cuantificación proteica se puede realizar mediante cualquier método conocido por un experto en la materia que sirva para tal fin, como por ejemplo, pero sin limitarnos, métodos de inmunodetección (como *western blot*, ELISA, inmunohistoquímica, inmunocitoquímica, inmunofluorescencia), métodos basados en marcajes isobáricos (como iTRAQ -*isobaric Tag for Relative and Absolute Quantitation*-, o ICAT -*Isotope-Coded Affinity Tag*-) o en marcajes isotópicos (como SILAC -*Stable Isotopes Labeling by Amino Acids in Cell Culture*-) o basados en marcajes fluorescentes (como 2D-DIGE -*Difference in Gel Electrophoresis*-), así como métodos basados en espectrometría de masas (MRM, -*Multiple Reaction Monitoring*-) . Por todo ello en otra realización preferida de este aspecto de la invención es el método donde el producto de expresión es una proteína.

Otra realización preferida del primer y segundo aspecto de la presente invención se refiere al método donde la detección y/o cuantificación de la proteína se realiza mediante inmuno blotting, inmunohistoquímica, cromatografía o arrays de expresión de proteínas.

Los términos “secuencia de aminoácidos” o “proteína” se usan aquí de manera intercambiable, y se refieren a una forma polimérica de aminoácidos de cualquier longitud, que pueden estar, o no, química o bioquímicamente modificados. El término “residuo” corresponde a un aminoácido.

Otra realización preferida del primer y segundo aspectos de la presente invención se refiere al método donde la muestra biológica se selecciona de la lista que comprende: tejido, sangre, plasma, suero, linfa, lavado broncoalveolar o fluido ascítico.

Otra realización también preferida del primer y segundo aspectos de la presente invención se refiere al método donde la muestra biológica es fresca, congelada, fijada o fijada y embebida en parafina.

Otra realización preferida del primer y segundo aspectos de la invención se refiere a un método donde el sujeto es un humano.

Un tercer aspecto de la invención se refiere al uso *in vitro* de los productos de expresión de los genes de la tabla 1 como marcador pronóstico de cáncer de pulmón no microcítico de estadio I o II.

Un cuarto aspecto de la invención se refiere a un kit que comprende las sondas que reconocen el ARN mensajero, producto de la expresión de los genes de la tabla 1, o el ARNc o ADNc a dicho ARNm, o anticuerpos que reconocen una

proteína producto de expresión de los genes de la tabla 1. La cuantía de sondas utilizadas para cada gen puede variar en número. Preferiblemente el kit comprende sondas, que consisten en las sondas que reconocen el ARN mensajero producto de la expresión de los genes de la tabla 1. Más preferiblemente las sondas son las secuencias descritas como SEQ ID NO: 1 a SEQ ID NO: 66 y que reconocen específicamente los 50 genes de la tabla 1. En adelante nos referiremos a este kit como al "kit primero de la invención".

Una realización preferida del cuarto aspecto de la invención se refiere al kit que además comprende al menos una sonda o un anticuerpo que reconoce un producto de expresión de los genes de la tabla 2. En adelante nos referiremos a este kit como al "kit segundo de la invención".

Otra realización preferida del cuarto aspecto de la invención se refiere a que el kit puede comprender al menos una retrotransciptasa, o una ARN polimerasa o un fluoróforo. Por lo que una realización preferida del tercer aspecto de la invención se refiere a un kit que además comprende al menos unos de los reactivos seleccionados de la lista que comprende: retrotranscriptasa, una ARN polimerasa o un fluoróforo. Además el kit puede comprender una mezcla de deoxinucleótidos tri-fosfato (dNTPs), una mezcla de nucleótidos tri-fosfato (NTPs), deoxiribonucleasa (DNasa), inhibidores de la ribonucleasa (RNasa), Dithiothreitol (DTT), pirofosfatasa inorgánica (PPi) y los tampones necesarios para las enzimas proporcionadas en el kit.

Además, la presente invención también se refiere al kit donde las sondas o los anticuerpos están preferiblemente situados en un soporte sólido, por ejemplo, pero sin limitarse, cristal, plástico, tubos, placas multipocillo, membranas, o cualquier otro soporte conocido. Por lo que una realización preferida del cuarto aspecto de la invención se refiere a un kit donde las sondas o los anticuerpos están preferiblemente situadas en un soporte sólido.

Un quinto aspecto de la invención se refiere al uso del kit del cuarto aspecto de la invención para la obtención de datos útiles para el pronóstico de CNMP estadios I o II. Además, la obtención de datos puede ser útil para la administración de tratamiento adyuvante, por ejemplo quimioterapia. Por lo que también se refiere al uso del kit primero de la invención para la evaluación de la necesidad de suministrar dicho tratamiento.

A lo largo de la descripción y las reivindicaciones la palabra "comprende" y sus variantes no pretenden excluir otras características técnicas, aditivos, componentes o pasos. Para los expertos en la materia, otros objetos, ventajas y características de la invención se desprenderán en parte de la descripción y en parte de la práctica de la invención. Los siguientes ejemplos y figuras se proporcionan a modo de ilustración, y no se pretende que sean limitativos de la presente invención.

DESCRIPCION DE LAS FIGURAS

FIG. 1. Muestra la probabilidad de ILE en los dos subtipos histológicos principales de CNMP. Curva de Kaplan-Meier que muestra la probabilidad de ILE de los dos subtipos histológicos principales del CNMP, adenocarcinoma y carcinoma escamoso en la matriz de desarrollo. ILE, intervalo libre de enfermedad; p, es la probabilidad asociada a que las diferencias encontradas en el ILE entre los subgrupos analizados sean debidas al azar.

FIG. 2. Muestra la probabilidad de intervalo libre de enfermedad en estadios I y II. . Curva de Kaplan-Meier que muestra la probabilidad de ILE para estadios I y II de CNMP en la matriz de desarrollo. ILE, intervalo libre de enfermedad; p, es la probabilidad asociada a que las diferencias encontradas en el ILE entre los subgrupos analizados sean debidas al azar.

FIG. 3. Muestra la agrupación jerárquica de las muestras de la matriz de desarrollo analizadas en función de su perfil molecular con 3.232 genes. Se muestra la agrupación ("clustering") jerárquica de 84 muestras con 3.232 genes (ver filtrado 3 del ejemplo 1) según el método descrito en Quackenbush J. Nat Rev Genet. 2001;2(6):418-27. Las muestras están diferenciadas en función del subtipo histológico: línea continua, subtipo adenocarcinoma; línea rayada, subtipo escamoso; línea continua terminada en *, otros subtipos de CNMP. Se define "perfil molecular": como el conjunto de datos genómicos (en nuestro casos niveles de expresión del ARNm) capaz de caracterizar e identificar un sujeto o muestra. Los subtipos moleculares encontrados muestran una clara asociación con los subtipos histológicos de los tumores.

FIG. 4. Muestra la probabilidad de ILE en función de los grupos moleculares obtenidos en la matriz de desarrollo a partir de 3.232 genes. Curva de Kaplan-Meier que muestra la probabilidad de ILE de los dos subtipos moleculares principales del CNMP encontrados en la matriz de desarrollo. ILE, intervalo libre de enfermedad; p, es la probabilidad asociada a que las diferencias encontradas en el ILE entre los subgrupos analizados sean debidas al azar.

FIG. 5. Muestra la agrupación jerárquica de las muestras de la matriz de desarrollo analizadas en función de su perfil molecular con 2.160 genes. Análisis del patrón de expresión génica global de los tumores de la matriz de desarrollo para la obtención de grupos moleculares utilizando el listado de 2.160 genes (ver filtrado 4 del ejemplo 1). **A**, se muestra la agrupación molecular ("clustering") jerárquica de 84 muestras con 2.160 genes. **B**, agrupación perfeccionada por el método de "k-means" descrito en Quackenbush J. Nat Rev Genet. 2001;2(6):418-27. En ambos casos resulta en tres grupos moleculares (Grupo 1, 2 y 3) o "clusters".

FIG. 6. Muestra la probabilidad de ILE en los tres grupos moleculares obtenidos en función de su perfil molecular con 2.160 genes en la matriz de desarrollo. Curva de Kaplan-Meier que muestra la probabilidad de ILE de los tres grupos moleculares obtenidos utilizando el listado de 2.160 genes y la técnica de "k-means". ILE, intervalo libre de enfermedad; p, es la probabilidad asociada a que las diferencias encontradas en el ILE entre los subgrupos analizados sean debidas al azar. (x), indica el número de muestras que hay en cada uno de los grupos analizados.

FIG. 7. Muestra la probabilidad de ILE en las muestras de la matriz de validación de acuerdo a la clasificación de 3 grupos moleculares. Curva de Kaplan-Meier que muestra la probabilidad de ILE para las muestras de la matriz de validación (serie externa, Roepman et al.) agrupadas en función de los perfiles moleculares (Grupo 1, Grupo 2 y Grupo 3) previamente observados en la matriz de desarrollo y definidos a través de un predictor de 1.000 genes generado con la aplicación "PAM". ILE, intervalo libre de enfermedad; p, es la probabilidad asociada a que las diferencias encontradas en el ILE entre los subgrupos analizados sean debidas al azar. (x), indica el número de muestras que hay que hay en cada uno de los grupos analizados.

FIG. 8. Muestra la probabilidad de ILE en las muestras de la matriz de validación de acuerdo a la clasificación establecida mediante el predictor de 50 genes. Curva de Kaplan-Meier que muestra la probabilidad de ILE de los dos grupos moleculares obtenidos en la matriz de validación utilizando el predictor de 50 genes. ILE, intervalo libre de enfermedad; p, es la probabilidad asociada a que las diferencias encontradas en el ILE entre los subgrupos analizados son debidas al azar. (x) indica el número de muestras que hay en cada una de las ramas de la curva de Kaplan-Meier.

FIG. 9. Probabilidad de ILE en las muestras de la matriz de validación de acuerdo a la clasificación establecida mediante el predictor de 50 genes de manera independiente para estadios I y II. Curva de Kaplan-Meier que muestra la probabilidad de ILE de los dos grupos moleculares obtenidos en la matriz de validación con el predictor de 50 genes generado con la aplicación "PAM" para: **A**, el estadio I, y **B**, estadio II. ILE, intervalo libre de enfermedad; p, es la probabilidad asociada a que las diferencias encontradas en el ILE entre los subgrupos analizados son debidas al azar. (x) indica el número de muestras que hay en cada una de las ramas de la curva de Kaplan-Meier.

EJEMPLOS

Los siguientes ejemplos específicos que se proporcionan en este documento de patente sirven para ilustrar la naturaleza de la presente invención. Estos ejemplos se incluyen solamente con fines ilustrativos y no han de ser interpretados como limitaciones a la invención que aquí se reivindica. Por tanto, los ejemplos descritos más adelante ilustran la invención sin limitar el campo de aplicación de la misma.

Ejemplo 1: Obtención del predictor de 50 genes

1.1.- Materiales y métodos

1.1.1 Selección de pacientes

En este estudio se han incluido 84 pacientes (12 mujeres y 72 varones con media de edad de 66,5 -rango de 36-82 años-) diagnosticados en estadios iniciales (60 pacientes estadio I y 24 pacientes estadio II) de CNMP durante los años 2001 a 2008 en el Hospital Clinico San Carlos (HCSC) de Madrid. Todos los pacientes cumplieron los siguientes criterios de inclusión: pacientes con tumores completamente resecados, sin afectación de ganglios mediastínicos, sin tratamiento quimioterápico y de los cuales existiera material tumoral congelado en el biobanco del HCSC perteneciente al subprograma RETICS del Instituto de Salud Carlos III (número de expediente RD090076/0102). Los datos recogidos para el estudio se dividen en datos clínicos del paciente (edad de diagnóstico, sexo y hábito tabáquico) y datos histológicos del tumor (subtipo histológico, tamaño tumor, estadio tumoral –7ª Clasificación TNM (Kligerman S. American Journal of Roentgenology 2010. 194:562-573)–, grado de diferenciación, queratinización, presencia de linfocitos polimorfonucleares –PMN-, afectación ganglionar, mutaciones de *k-ras*, necrosis, estroma tumoral, inflamación crónica, presencia de linfocitos intratumorales –TIL–, localización por lóbulos pulmonares y tipo de recidiva –loco regional o a distancia–).

1.1.2. Muestras tumorales. Extracción y purificación del ARN.

Siguiendo el protocolo de congelación de las muestras incluidas en el biobanco del HCSC, los tumores de CNMP se recogieron inmediatamente después de la cirugía y se congelaron y almacenaron a -80ºC. Se llevó a cabo la revisión histopatológica de los tumores congelados con el fin de que todos los pacientes incluidos en el estudio tuvieran una representación tumoral como mínimo del 70% en la muestra utilizada. Paralelamente, se recogieron de estos mismos pacientes, muestras de parénquima pulmonar no tumoral que también fueron congeladas siguiendo el mismo protocolo. El ARN proveniente de estas últimas muestras se utilizó para crear la muestra control (un pool de ARN de tejidos normales). En todos los casos, el ácido ribonucleico (ARN o RNA) total fue extraído directamente de las muestras congeladas utilizando Trizol® y un homogeneizador de tejidos. Posteriormente fue tratado con DNAsa y cuantificado en el espectofotómetro NanoDrop ND-1000®. La calidad del ARN extraído se midió en Bioanalyzer 2100® mediante el RIN (o Número de Integridad del ARN) y únicamente las muestras con una buena calidad de ARN (RIN > 7,5), fueron incluidas para el estudio.

1.1.3. Perfil de expresión por microarrays.

El perfil de expresión de los 84 tumores se determinó utilizando microarrays de oligonucleótidos de genoma completo de Agilent® (G4112F) siguiendo el protocolo suministrado por el fabricante. Brevemente, se utilizó doble marcaje, con cianina-5 (Cy5) para cada uno de los 84 tumores incluidos en el estudio y con cianina-3 (Cy3) para la muestra control, compuesta por un "pool" de 42 muestras de parénquima no tumoral de pulmón. Esta muestra control se introdujo en cada uno de los experimentos (la misma en todos ellos) para poder identificar y corregir las variaciones técnicas introducidas durante la fase experimental del análisis. Tras esta corrección (denominada normalización) el dato generado es el ratio entre la fluorescencia del tumor y la muestra control.

Durante las etapas de marcaje e hibridación se incluyeron los "SpikeIns", que son 10 transcritos control sintetizados *in vitro* que derivan del transcriptoma del Adenovirus E1A, que no interaccionan con el ARNm humano y cuya concentración inicial es conocida. El conocimiento "a priori" de la concentración inicial de cada uno de los "SpikeIns", nos permite predecir a qué nivel de fluorescencia deberían emitir estos transcritos una vez hibridados en el microarray y por tanto poder utilizarlos como control de calidad de la fase experimental.

Los microarrays fueron escaneados y cuantificados usando el escáner de Agilent® y el programa Feature Extraction® (10.7.3) respectivamente. Para la normalización de los datos extraídos se utilizó la técnica Lowess o "Locally Weighted Scatterplot Smoothing". (*Cleveland WS: Journal of the American statistical Association 1979, 74:829-836; Cleveland WS, et al. Journal of the American Statistical Association 1988, 83:596-610.)*

1.1.4. Análisis de datos

Para la obtención del método de la invención, se partió de un listado inicial de 41.000 sondas presentes en el microarrays de oligonucleótidos de genoma completo de Agilent®. A partir de un proceso de filtrado se llegó hasta una clasificación molecular que finalmente derivó en la creación del predictor de la invención compuesto por sólo 50 genes. El método se desarrolló siguiendo los siguientes pasos de filtrado:

1.- Filtrado por "flags": exclusión de sondas con baja fluorescencia o con problemas durante el proceso de hibridación en más de un 10% de las muestras. El nuevo listado incluía 24.617 sondas.

2.- Promedio de las sondas con el mismo identificador con el objetivo de trabajar con valores de expresión únicos para cada gen. El nuevo listado incluía 17.881 genes.

3.- Filtrado por expresión: selección de genes con una variación de expresión al menos de 3 veces respecto a la mediana de ese gen en al menos el 10% de las muestras. El nuevo listado incluyó un total de 3.232 genes (Fig. 3). Una vez generados los grupos moleculares a partir de este listado de 3.232 genes, se evaluó la clasificación molecular obtenida para conocer si existía o no asociación con el intervalo libre de enfermedad (ILE) (Fig. 4).

4.- Filtrado histológico: se eliminaron los genes que caracterizan las diferencias histológicas entre los principales subtipos histológicos del CNMP (adenocarcinoma y carcinoma escamoso). Para ello, se seleccionaron los genes diferencialmente expresados (p-valor < 0,01 y diferencia de expresión > 1,5) y el listado generado (1.072 genes) se excluyó del listado inicial (3.232 genes). Se genera por tanto un listado de 2.160 genes que se utilizan para la clasificación molecular final de los 84 tumores.

La estrategia utilizada para el descubrimiento de los grupos moleculares consistió en aplicar en primer lugar un método de análisis no supervisado, agrupamiento o "clustering" jerárquico (Fig. 5A), y a continuación un perfeccionamiento de los grupos moleculares obtenidos mediante un segundo método, método de k-Means (Fig. 5B), el cual permite disminuir la heterogeneidad intra-grupo y aumentar la variabilidad inter-grupo. El listado de 2.160 genes se usa para construir inicialmente la clasificación molecular (que tiene 3 grupos). Una vez generados estos grupos moleculares, se evaluó la clasificación molecular obtenida para conocer si existía o no asociación con el intervalo libre de enfermedad (ILE) (Fig. 6). El ILE se define como el tiempo que transcurre desde la fecha de la cirugía hasta que se confirma la recidiva del paciente.

En el análisis estadístico se han utilizado curvas de Kaplan-Meier y el test log-rank para evaluar la probabilidad de cada subtipo molecular respecto a la recidiva (Clark TG. British Journal of Cancer 2003. 89: 232-238). Además, con el método de regresión proporcional de Cox se calcula el "hazard ratio" para los grupos moleculares.

Asimismo, se realizó un análisis de las vías moleculares que se encuentran alteradas de manera significativa entre los grupos moleculares obtenidos. Se llevó a cabo utilizando la herramienta GSEA ("Gene Set Enrichment Análisis" o análisis de enriquecimiento de conjuntos de genes) (Subramanian A *et al.* PNAS 2005 102 (43) 15545-15550 y Mootha VK *et al.* Nat Gen 2003). Sólo se evaluaron las vías moleculares con una representación mínima de más de 15 genes y se utilizaron 100.000 permutaciones para asegurar los resultados. Para obtener los resultados de GSEA se partió del listado original de 17.881 genes ya que cuando se analizan vías moleculares conviene incluir todos los genes disponibles que cumplan los controles de calidad (17.881 genes), ya que diferencias no significativas de expresión en un

grupo de genes pueden sin embargo ser claves, para definir qué caminos de señalización ("pathways") están alterados entre los grupos.

1.1.5. Validación de los 3 grupos moleculares en una serie externa.

Para la validación de la clasificación molecular obtenida, se utilizó la matriz de datos publicada por el grupo de Roepman y colaboradores (Roepman P et al. Clin Cancer Res 2009. 15:284-290). La matriz de validación incluye los datos de expresión de 162 pacientes diagnosticados de los mismos subtipos histológicos que los de la invención.

El termino "matriz de entrenamiento" o "matriz de desarrollo" se refiere a las muestras del biobanco del HCSC (n= 84). El término "matriz de validación" se refiere al conjunto de muestras publicado por Roepman *et al* utilizado para la validación de la clasificación molecular. Por "matriz" se entiende el conjunto de datos de expresión obtenidos en una serie de pacientes mediante microarrays.

Para la validación se ha generado una matriz de datos común que incluye 246 muestras (84 de la matriz de desarrollo + 162 de la matriz de validación) cada una de ellas con 17.881 genes. Con la matriz de desarrollo se obtuvo un predictor, mediante la aplicación PAM (Análisis de Predicción de Microarray) (Tibshirani R. *et al.* PNAS 2002; 99(10):6567-72) que fue evaluado en la matriz de validación estudiando su asociación, mediante la curva de Kaplan-Meier, con el ILE. El modelo de regresión proporcional Cox se utilizó para confirmar el poder pronóstico de nuestro predictor.

1.1.6. Obtención y validación del Predictor de 50 genes.

Los 3 grupos moleculares generados mediante el filtrado histológico, se agruparon en 2 grupos, grupo de buen pronóstico o grupo 3 y grupo de mal pronostico o grupo 1+2, debido a la similitud pronóstica de ambos grupos moleculares. Así, con dos grupos y partiendo de 2.160 genes se seleccionan, mediante la aplicación de PAM, 50 genes (los genes de la invención que se muestran en la tabla 1) capaces de clasificar nuevas muestras en base a estos dos grupos pronóstico en CNMP de estadios I o II. En base a este predictor de 50 genes, las muestras de la matriz de validación fueron clasificadas en el grupo de buen pronóstico o en el de mal pronóstico. Las curvas de Kaplan-Meier y el modelo de regresión proporcional de Cox se utilizaron para validar el poder pronóstico de nuestro predictor (Figs.8, 9A y 9B).

1.1.7. Explicación del análisis con PAM (Tibshirani R. et al. PNAS. 2002, 99:6567-6572).

Para ejemplificar esta descripción, utilizaremos como ejemplo la creación del predictor de 50 genes para dos grupos moleculares (buen y mal pronóstico) mencionados en el apartado anterior. Así pues, usando como herramienta de clasificación la aplicación PAM, el proceso de clasificación pronóstica, requiere como punto de partida el cálculo de un "valor de referencia" para cada uno de los dos grupos. Estos "valores de referencia" se obtienen a partir de las muestras de los pacientes que conforman la denominada "matriz de entrenamiento" o "matriz de desarrollo" y de los que "a priori" se conoce su clasificación (pues fueron con ellas con las que se definió lo que era el grupo de buen y mal pronóstico). A partir de los pacientes del grupo de buen pronóstico obtendremos el "valor 1 de referencia" y a partir de los pacientes del grupo de mal pronóstico obtendremos el "valor 2 de referencia". Cada uno de los valores de referencia vendrá expresado como un vector de 50 componentes (una por cada uno de los genes de la invención) y se calculará como la suma de dos subvectores cada uno de ellos expresados también con 50 componentes. El primer subvector es común para los dos valores de referencia mientras que el segundo es específico para cada uno de los dos valores de referencia que se quieren calcular. El primer subvector consta de 50 componentes, cada una de las cuales corresponde al valor medio de expresión de uno de los 50 genes a lo largo de todas las muestras que conforman la matriz de entrenamiento o desarrollo independientemente del grupo en el que se encuentren clasificadas (es decir los 84 tumores de nuestra matriz). El segundo subvector también vendrá definido por 50 componentes (cada una de las cuales representa un gen) que vendrán definidas por un estadistico "t" que compara para dicho gen las diferencias entre el primer subvector y el valor medio de expresión de ese gen en las muestras incluidas en el grupo para el que se quiere calcular el valor de referencia (o bien el grupo de buen pronostico (29 muestras) o bien el grupo de mal pronóstico (55 muestras)). Los datos del segundo subvector serán estandarizados teniendo en cuenta, la variabilidad de expresión de cada uno de los 50 genes dentro del grupo analizado y teniendo en cuenta un valor de convergencia Δ que permite evaluar el poder predictivo de cada uno de los genes. Las transformaciones mencionadas harán que aunque el "valor de referencia" o "shrunken centroid" obtenido para cada grupo se basa en valores de expresión, su valor real sea adimensional y no sea un reflejo de los datos de fluorescencia iniciales de cada muestra. Una vez calculado el "valor de referencia" o "shrunken centroid" para cada grupo, el PAM es capaz de asignar las nuevas muestras, que en este ejemplo conformaron la matriz de validación (162 muestras), a cada uno de los grupos previamente definidos. La aplicación de esta invención para conocer el pronóstico de los nuevos pacientes se realiza calculando la distancia entre los valores de expresión de los 50 genes de la nueva muestra con respecto a las 50 componentes del "valor de referencia" o "shrunken centroid" de cada grupo. Si la distancia entre la nueva muestra y el "valor 1 de referencia" es menor que la distancia entre la nueva muestra y el "valor 2 de referencia", se podrá determinar el pronóstico favorable para el nuevo paciente. Por el contrario, si la distancia entre la nueva muestra y el "valor 1 de referencia" es mayor que la distancia entre la nueva muestra y "valor 2 de referencia", se podrá determinar el pronóstico desfavorable para el nuevo paciente. Durante estos últimos cálculos también se introducen factores que corrigen el resultado teniendo en cuenta la variabilidad de expresión dentro de los grupos y la probabilidad de pertenecer a un determinado grupo teniendo en cuenta su tamaño muestral con respecto al de la población analizada. La cuantificación de las distancias se mide utilizando la distancia euclídea.

1.2.- RESULTADOS

1.2.1. Análisis de asociación del ILE con las variables clínicas e histopatológicas.

Se llevó a cabo un primer análisis estadístico para comprobar si existía una asociación entre las variables histopatológicas más importantes en el manejo rutinario del CNMP (la clasificación histológica del tumor, el estadio, etc.), con el ILE. Las curvas de Kaplan-Meier obtenidas no mostraron una asociación estadísticamente significativa del ILE con el tipo histopatológico (Fig. 1), el estadio (Fig. 2) o con cualquier otra variable analizada (datos no mostrados). Solamente la presencia de mutaciones en el gen *K-Ras* mostró una tendencia hacia la asociación con un peor pronóstico (p=0,07).

1.2.2. Grupos moleculares a partir de 3.232 genes.

Mediante el método de clustering jerárquico (centrado de Pearson y Average linkage (Quackenbush J. Nat Rev Genet. 2001;2(6):418-27) se identifican dos subtipos moleculares principales que muestran una clara asociación con los subtipos histológicos más representados en nuestra serie, separando molecularmente, los tumores del subtipo adenocarcinoma de los tumores del subtipo escamoso (Fig. 3). Estos 2 subtipos moleculares no muestran diferencias estadísticamente significativas con el ILE (p = 0,350) (Fig. 4).

A la vista de estos resultados, concluimos que los grupos moleculares obtenidos utilizando el listado de 3.232 genes (que son los genes que varían su expresión al menos de 3 veces respecto a la mediana de ese gen en al menos el 10% de las muestras; paso 3 del filtrado anteriormente explicado) se encuentran condicionados por la histología de los tumores. Es importante destacar que no existen diferencias estadísticamente significativas en el tiempo de recidiva cuando se comparan ambos grupos moleculares y recordar que tampoco existían cuando se comparaban ambos grupos clasificados según criterios histológicos.

Teniendo en cuenta que el criterio oncológico para el manejo de los pacientes de CNMP nos indica que la histología de los tumores sólo es importante en la enfermedad metastásica (estadio IV) y sólo en relación con el tratamiento indicado, excluimos del listado inicial de 3.232 genes aquellos que caracterizan las diferencias histológicas de los 84 tumores mediante un filtrado que incluía: T-Test a p<0,01 con corrección para comparaciones múltiples de Benjamini and Hochberg (B&H) (Benjamini Y and Hochberg Y. Journal of the Royal Statistical Society. 1995) y una diferencia de expresión de más de 1,5 veces. Los genes que cumplieron estos criterios de filtrado fueron excluidos, resultado un listado de 2.160 genes que se utilizaron para la obtención de la clasificación molecular y que son los genes incluidos en las tablas 1 y 2.

1.2.3. Grupos moleculares con 2.160 genes. Asociación con ILE.

Tras la agrupación de los 84 pacientes según el perfil de expresión génica utilizando el listado de 2.160 genes y el método de clustering jerárquico (Fig. 5A) posteriormente perfeccionado por el método de k-means, se obtuvieron 3 grupos moleculares que se denominaron como Grupo 1, Grupo 2 y Grupo 3 (Fig. 5B).

Estos tres grupos se asociaron de manera estadísticamente significativa con el ILE (log-rank p=0,004), mostrando en la curva de Kaplan-Meier, 2 grupos moleculares de mal pronóstico respecto a la recidiva (Grupo 1 y Grupo 2) y un grupo molecular de buen pronóstico (Grupo 3) (Fig. 6). El “Hazard ratio” (HR, es decir, el riesgo o probabilidad de recaída que tiene un grupo con respecto a otro) de los grupos de mal pronóstico frente al grupo de buen pronóstico es de 6,4 para el Grupo 1 (IC 95%: 1,8-22,3; p = 0,004) y de 4,9 para el Grupo 2 (IC 95%: 1,4-17,8; p = 0,014). No existe diferencia estadísticamente significativa para el riesgo entre el Grupo 1 y el Grupo 2 (p=0,526).

a)- Análisis multivariante

En este análisis se incluyeron las mutaciones para *k-ras* por presentar tendencia (p=0,07) para la asociación con el ILE y la clasificación por Estadio ya que es el principal factor pronóstico para el CNMP.

Después de ajustar por Estadio y por el estatus de *K-ras*, el modelo multivariante de riesgos proporcionales de Cox confirmó la clasificación molecular como factor pronóstico independiente para evaluar el riesgo de recidiva (HR Grupo 1 vs. 3 = 11.170; 95% CI: 2,9 a 43,4; p = 4,9E-04; HR Grupo 2 vs. 3 = 7,521; 95% CI: 2,0 a 28,8; p = 0,003); HR Grupo 1 vs. 2= no significativo).

b)-. Estudio de vías moleculares.

Se observó que la clasificación molecular en 3 grupos estaba relacionada con la implicación de vías moleculares relacionadas con sistema inmune como la vía de Células T, Células B, Inflamación y respuesta Th1 que diferencian el Grupo 3 del Grupo 2 y especialmente el Grupo 3 del Grupo 1. Por otro lado, la alteración de genes implicados en vías

de ciclo celular y mecanismos de reparación del ADN confiere las principales diferencias biológicas entre el Grupo 2 y el Grupo 1.

c)- Análisis estadístico de las variables clínicas e histológicas incluidas en el estudio.

Respecto a las variables clínicas de los pacientes incluidas en el estudio, el hábito tabáquico se asoció de manera estadísticamente significativa con la clasificación molecular obtenida (p=0,002). En el caso de las variables histológicas del tumor, la afectación ganglionar (p=0,041), a pesar de tener solamente 3 pacientes diagnosticados con N1, y la inflamación crónica (p=0,001) también se asocian de manera estadísticamente significativa con los subtipos moleculares.

d)- Validación en serie externa y obtención de predictor para 3 grupos moleculares.

Utilizando la matriz de desarrollo (84 tumores) se obtuvo, mediante el uso de PAM, un primer predictor de 1.000 genes que identificaba los pacientes en los 3 grupos moleculares, dos de mal pronóstico (grupo 1 y grupo 2) y uno de buen pronóstico (grupo 3). Para la evaluación del poder pronóstico de dicho predictor, se utilizaron los datos de los 162 tumores de la matriz de validación. Estas muestras fueron clasificadas en los 3 grupos moleculares utilizando dicho predictor (1.000 genes). La curva de Kaplan-Meier para las muestras de la matriz de validación reveló una asociación estadísticamente significativa de estos tres grupos moleculares con el ILE (log-rank p=0,022) (Fig. 7). El “Hazard Ratio” (HR) de los grupos de mal pronóstico frente al de buen pronóstico es de 2,4 veces para el Grupo 1 (p=0,012) y de 2,5 veces para el Grupo 2 (p=0,019).

1.2.4. Obtención del predictor de 50 genes.

Como se observó con anterioridad en los resultados obtenidos en la matriz de desarrollo, de los tres grupos obtenidos mediante análisis de expresión génica, el comportamiento del Grupo 1 y el Grupo 2 es similar respecto a la recidiva, no existiendo diferencia estadística significativa para el riesgo entre estos dos grupos (p=0,526). Por ello, ambos grupos se englobaron en uno sólo y se generó un segundo predictor de 50 genes, mediante PAM, para diferenciar pacientes de mal pronóstico (Grupo 1 y 2) y pacientes de buen pronóstico (Grupo 3). En la tabla 3 se incluye el valor del centroide compacto (“shrunken centroid”) para los grupos de buen y mal pronóstico obtenidos con las muestras de la matriz de desarrollo. Este segundo predictor engloba los denominados "50 genes de la invención" (ver tabla 1) y la evaluación del poder pronóstico del mismo se llevó a cabo de nuevo en la matriz de validación, obteniendo las curvas de Kaplan-Meier que muestran una asociación estadísticamente significativa de los dos grupos obtenidos con el ILE (log-rank p=0,001) (Fig. 8). El HR para el Grupo de mal pronóstico es de 3,4 frente al de buen pronóstico (IC 95%: 1,6-7,3; p=0,001).

Tabla 3. Muestra los valores de referencia de los 50 genes obtenidos con las muestras de la matriz de desarrollo para cada uno de los grupos. Estos valores serán utilizados para clasificar las nuevas muestras en grupo de buen o mal pronóstico.

ID Entrez	**Símbolo**	**Buen Pronóstico (“Shrunken centroid”)**	**Mal Pronóstico (“Shrunken centroid”)**
3493	*IGHA1*	0,549	-0,29
100287723	*LOC100287723*	0,539	-0,284
642424	*LOC642424*	0,536	-0,282
100132941	*LOC100132941*	0,52	-0,274
401847	*LOC401847*	0,517	-0,273
3500	*IGHG1*	0,504	-0,266
90925	*LOC90925*	0,487	-0,257
3543	*IGLL1*	0,481	-0,254
973	*CD79A*	0,443	-0,234
100290415	*LOC100290415*	0,44	-0,232
100133862	*LOC100133862*	0,436	-0,23
608	*TNFRSF17*	0,433	-0,228
26952	*SMR3A*	0,428	-0,226

3535	*IGL@*	0,428	-0,226
51237	*MGC29506*	0,428	-0,226
100293440	*LOC100293440*	0,427	-0,225
3887	*KRT81*	0,417	-0,22
28904	*IGKV1D-8*	0,407	-0,214
57823	*SLAMF7*	0,403	-0,212
91319	*DERL3*	0,386	-0,204
57699	*CPNE5*	0,381	-0,201
54900	*LAX1*	0,379	-0,2
150365	*MEI1*	0,378	-0,199
9834	*KIAA0125*	0,372	-0,196
5450	*POU2AF1*	0,364	-0,192
84824	*FCRLA*	0,346	-0,183
83416	*FCRL5*	0,344	-0,181
92154	*MTSS1L*	0,344	-0,181
3662	*IRF4*	0,333	-0,176
3782	*KCNN3*	0,333	-0,176
11040	*PIM2*	0,326	-0,172
939	*CD27*	0,322	-0,17
3494	*IGHA2*	0,32	-0,169
5368	*PNOC*	0,316	-0,166
79368	*FCRL2*	0,312	-0,164
51303	*FKBP11*	0,294	-0,155
8419	*BFSP2*	0,291	-0,153
270	*AMPD1*	0,289	-0,152
10563	*CXCL13*	0,286	-0,151
974	*CD79B*	0,281	-0,148
930	*CD19*	0,28	-0,148
3512	*IGJ*	0,279	-0,147
952	*CD38*	0,265	-0,14
78986	*DUSP26*	0,242	-0,128
4283	*CXCL9*	0,23	-0,121
80307	*FER1L4*	0,228	-0,12
221188	*GPR114*	0,228	-0,12
126306	*JSRP1*	0,22	-0,116

140947	*C5orf20*	0,215	-0,113
3002	GZMB	0,208	-0,11

1.2.5. Utilidad del predictor de 50 genes en CNMP separados por estadio.

Una de las principales críticas aparecidas en el estado del arte (Subramanian J. et al. J Natl Cancer Inst 2010;102:1–11) respecto a la utilidad de los predictores generados para el CNMP es que es necesario demostrar su utilidad para predecir el pronóstico de los pacientes de manera independiente del estadio en el que se clasificaron. Para ello, separamos los 162 pacientes de la matriz de validación en pacientes clasificados en estadio I (110 pacientes) y clasificados como estadio II (52 pacientes). Se utilizó el predictor de 50 genes para obtener los grupos moleculares de alto y bajo riesgo (es decir, de mal y buen pronóstico, respectivamente) y se estudió su asociación con el ILE mediante las curvas de Kaplan-Meier. Tanto en estadios I por separado (Fig. 9 A) como en estadios II (Fig. 9 B) se observó una asociación estadísticamente significativa de los grupos con el ILE (p=0,013 y p=0,029 respectivamente) y los HR del grupo de mal pronóstico respecto del de buen pronóstico fueron en el estadio I de 3,2 (IC 95%:1,2-8,3; p=0,018) y en el estadio II de 3,5 (IC 95%:1,1-12; p=0,041).

1.2.6. Sensibilidad y especificidad del predictor de 50 genes.

Los valores de sensibilidad y especificidad del predictor para la clasificación de las muestras en los grupos moleculares identificados se muestran en la tabla 4.

Tabla 4. Sensibilidad, Especificidad, Valor Predictivo Positivo (VPP) y Valor Predictivo Negativo (VPN) para la clasificación molecular con el predictor de 50 genes.

Grupos moleculares	**Sensibilidad**	**Especificidad**	**VPP**	**VPN**
Grupo buen pronóstico	0,966	0,891	0,824	0,98
Grupo mal pronóstico	0,891	0,966	0,98	0,824

Por lo tanto, y en base a los resultados mostrados, la presente invención demuestra la utilidad del método de la invención, así como del uso de los 50 genes descritos en la tabla 1 como marcadores pronóstico del CNMP de estadios I o II.

A continuación se muestra la tabla 2 a la que se ha hecho referencia previamente. Cuando el “ID Entrez” no está indicado o es “-“ o “- -“ se trata de genes de los que no hay información en la base de datos NCBI. En el símbolo del gen se ha indicado el nombre de la sonda del array utilizado.

ID Entrez	**Símbolo**	**ID Entrez**	**Símbolo**	**ID Entrez**	**Símbolo**
2	*A2M*	586	*BCAT1*	1004	*CDH6*
12	*SERPINA3*	590	*BCHE*	1028	*CDKN1C*
13	*AADAC*	595	*CCND1*	1029	*CDKN2A*
49	*ACR*	623	*BDKRB1*	1030	*CDKN2B*
55	*ACPP*	624	*BDKRB2*	1033	*CDKN3*
58	*ACTA1*	640	*BLK*	1036	*CDO1*
59	*ACTA2*	641	*BLM*	1047	*CLGN*
72	*ACTG2*	643	*CXCR5*	1066	*CES1*
104	*ADARB1*	650	*BMP2*	1073	*CFL2*
124	*ADH1A*	654	*BMP6*	1080	*CFTR*

126	*ADH1C*	658	*BMPR1B*	1111	*CHEK1*
136	*ADORA2B*	660	*BMX*	1117	*CHI3L2*
152	*ADRA2C*	687	*KLF9*	1118	*CHIT1*
154	*ADRB2*	699	*BUB1*	1138	*CHRNA5*
176	*ACAN*	718	*C3*	1160	*CKMT2*
216	*ALDH1A1*	721	*C4B*	1191	*CLU*
220	*ALDH1A3*	730	*C7*	1236	*CCR7*
231	*AKR1B1*	760	*CA2*	1264	*CNN1*
241	*ALOX5AP*	762	*CA4*	1268	*CNR1*
249	*ALPL*	767	*CA8*	1277	*COL1A1*
275	*AMT*	820	*CAMP*	1278	*COL1A2*
284	*ANGPT1*	828	*CAPS*	1281	*COL3A1*
290	*ANPEP*	857	*CAV1*	1289	*COL5A1*
301	*ANXA1*	858	*CAV2*	1290	*COL5A2*
306	*ANXA3*	862	*RUNX1T1*	1291	*COL6A1*
348	*APOE*	863	*CBFA2T3*	1292	*COL6A2*
356	*FASLG*	873	*CBR1*	1295	*COL8A1*
360	*AQP3*	874	*CBR3*	1296	*COL8A2*
362	*AQP5*	890	*CCNA2*	1299	*COL9A3*
367	*AR*	891	*CCNB1*	1300	*COL10A1*
374	*AREG*	898	*CCNE1*	1302	*COL11A2*
384	*ARG2*	909	*CD1A*	1305	*COL13A1*
395	*ARHGAP6*	911	*CD1C*	1306	*COL15A1*
430	*ASCL2*	925	*CD8A*	1307	*COL16A1*
443	*ASPA*	948	*CD36*	1311	*COMP*
444	*ASPH*	956	*ENTPD3*	1356	*CP*
445	*ASS1*	969	*CD69*	1359	*CPA3*
463	*ZFHX3*	970	*CD70*	1363	*CPE*
467	*ATF3*	971	*CD72*	1380	*CR2*
482	*ATP1B2*	978	*CDA*	1381	*CRABP1*
501	*ALDH7A1*	991	*CDC20*	1382	*CRABP2*
563	*AZGP1*	1000	*CDH2*	1400	*CRMP1*

ID Entrez	**Símbolo**	**ID Entrez**	**Símbolo**	**ID Entrez**	**Símbolo**
1404	*HAPLN1*	1946	*EFNA5*	2346	*FOLH1*

1415	*CRYBB2*	1948	*EFNB2*	2352	*FOLR3*
1468	*SLC25A10*	1958	*EGR1*	2353	*FOS*
1490	*CTGF*	1960	*EGR3*	2354	*FOSB*
1501	*CTNND2*	1969	*EPHA2*	2358	*FPR2*
1511	*CTSG*	1999	*ELF3*	2444	*FRK*
1513	*CTSK*	2006	*ELN*	2525	*FUT3*
1524	*CX3CR1*	2009	*EML1*	2537	*IFI6*
1571	*CYP2E1*	2026	*ENO2*	2539	*G6PD*
1573	*CYP2J2*	2034	*EPAS1*	2568	*GABRP*
1594	*CYP27B1*	2041	*EPHA1*	2571	*GAD1*
1602	*DACH1*	2042	*EPHA3*	2579	*GAGE7*
1608	*DGKG*	2043	*EPHA4*	2591	*GALNT3*
1634	*DCN*	2059	*EPS8*	2619	*GAS1*
1668	*DEFA3*	2114	*ETS2*	2621	*GAS6*
1673	*DEFB4A*	2115	*ETV1*	2627	*GATA6*
1674	*DES*	2122	*MECOM*	2628	*GATM*
1690	*COCH*	2152	*F3*	2633	*GBP1*
1728	*NQO1*	2153	*F5*	2635	*GBP3*
1734	*DIO2*	2160	*F11*	2668	*GDNF*
1735	*DIO3*	2162	*F13A1*	2669	*GEM*
1740	*DLG2*	2172	*FABP6*	2681	*GGTA1*
1748	*DLX4*	2191	*FAP*	2719	*GPC3*
1755	*DMBT1*	2192	*FBLN1*	2731	*GLDC*
1756	*DMD*	2194	*FASN*	2735	*GLI1*
1763	*DNA2*	2202	*EFEMP1*	2736	*GLI2*
1776	*DNASE1L3*	2205	*FCER1A*	2781	*GNAZ*
1789	*DNMT3B*	2206	*MS4A2*	2786	*GNG4*
1805	*DPT*	2237	*FEN1*	2791	*GNG11*
1809	*DPYSL3*	2239	*GPC4*	2841	*GPR18*
1837	*DTNA*	2246	*FGF1*	2842	*GPR19*
1842	*ECM2*	2247	*FGF2*	2843	*GPR20*
1843	*DUSP1*	2254	*FGF9*	2850	*GPR27*
1848	*DUSP6*	2258	*FGF13*	2852	*GPER*
1852	*DUSP9*	2273	*FHL1*	2857	*GPR34*
1869	*E2F1*	2274	*FHL2*	2869	*GRK5*

1870	*E2F2*	2277	*FIGF*	2905	*GRIN2C*
1879	*EBF1*	2294	*FOXF1*	2920	*CXCL2*
1896	*EDA*	2295	*FOXF2*	2921	*CXCL3*
1906	*EDN1*	2327	*FMO2*	2922	*GRP*
1907	*EDN2*	2328	*FMO3*	2936	*GSR*
1945	*EFNA4*	2335	*FN1*	2939	*GSTA2*

ID Entrez	Símbolo	ID Entrez	Símbolo	ID Entrez	Símbolo
2944	*GSTM1*	3400	*ID4*	3824	*KLRD1*
2952	*GSTT1*	3429	*IFI27*	3833	*KIFC1*
2953	*GSTT2*	3434	*IFIT1*	3834	*KIF25*
2999	*GZMH*	3437	*IFIT3*	3851	*KRT4*
3001	*GZMA*	3455	*IFNAR2*	3880	*KRT19*
3003	*GZMK*	3458	*IFNG*	3892	*KRT86*
3005	*H1F0*	3479	*IGF1*	3898	*LAD1*
3006	*HIST1H1C*	3481	*IGF2*	3902	*LAG3*
3007	*HIST1H1D*	3486	*IGFBP3*	3908	*LAMA2*
3008	*HIST1H1E*	3488	*IGFBP5*	3914	*LAMB3*
3013	*HIST1H2AD*	3489	*IGFBP6*	3925	*STMN1*
3024	*HIST1H1A*	3559	*IL2RA*	3934	*LCN2*
3042	*HBM*	3569	*IL6*	3945	*LDHB*
3043	*HBB*	3577	*CXCR1*	3953	*LEPR*
3045	*HBD*	3579	*CXCR2*	3957	*LGALS2*
3049	*HBQ1*	3598	*IL13RA2*	3998	*LMAN1*
3053	*SERPIND1*	3604	*TNFRSF9*	4015	*LOX*
3067	*HDC*	3620	*IDO1*	4017	*LOXL2*
3117	*HLA-DQA1*	3624	*INHBA*	4023	*LPL*
3131	*HLF*	3625	*INHBB*	4033	*LRMP*
3149	*HMGB3*	3627	*CXCL10*	4045	*LSAMP*
3161	*HMMR*	3640	*INSL3*	4057	*LTF*
3164	*NR4A1*	3653	*IPW*	4062	*LY6H*
3169	*FOXA1*	3663	*IRF5*	4068	*SH2D1A*
3199	*HOXA2*	3667	*IRS1*	4069	*LYZ*
3201	*HOXA4*	3679	*ITGA7*	4071	*TM4SF1*
3204	*HOXA7*	3690	*ITGB3*	4093	*SMAD9*

3212	*HOXB2*	3694	*ITGB6*	4105	*MAGEA6*
3216	*HOXB6*	3699	*ITIH3*	4117	*MAK*
3225	*HOXC9*	3706	*ITPKA*	4118	*MAL*
3235	*HOXD9*	3730	*KAL1*	4128	*MAOA*
3240	*HP*	3751	*KCND2*	4129	*MAOB*
3250	*HPR*	3768	*KCNJ12*	4131	*MAP1B*
3270	*HRC*	3772	*KCNJ15*	4133	*MAP2*
3292	*HSD17B1*	3775	*KCNK1*	4137	*MAPT*
3294	*HSD17B2*	3777	*KCNK3*	4151	*MB*
3303	*HSPA1A*	3778	*KCNMA1*	4166	*CHST6*
3316	*HSPB2*	3783	*KCNN4*	4199	*ME1*
3357	*HTR2B*	3800	*KIF5C*	4223	*MEOX2*
3371	*TNC*	3816	*KLK1*	4233	*MET*
3373	*HYAL1*	3821	*KLRC1*	4239	*MFAP4*
3397	*ID1*	3822	*KLRC2*	4241	*MFI2*

ID Entrez	Símbolo	ID Entrez	Símbolo	ID Entrez	Símbolo
4248	*MGAT3*	4929	*NR4A2*	5563	*PRKAA2*
4256	*MGP*	4935	*GPR143*	5569	*PKIA*
4311	*MME*	4953	*ODC1*	5570	*PKIB*
4316	*MMP7*	5021	*OXTR*	5577	*PRKAR2B*
4318	*MMP9*	5023	*P2RX1*	5578	*PRKCA*
4325	*MMP16*	5029	*P2RY2*	5602	*MAPK10*
4332	*MNDA*	5046	*PCSK6*	5625	*PRODH*
4356	*MPP3*	5050	*PAFAH1B3*	5627	*PROS1*
4477	*MSMB*	5054	*SERPINE1*	5630	*PRPH*
4481	*MSR1*	5064	*PALM*	5655	*KLK10*
4490	*MT1B*	5087	*PBX1*	5727	*PTCH1*
4495	*MT1G*	5099	*PCDH7*	5730	*PTGDS*
4499	*MT1M*	5122	*PCSK1*	5740	*PTGIS*
4502	*MT2A*	5125	*PCSK5*	5743	*PTGS2*
4585	*MUC4*	5144	*PDE4D*	5753	*PTK6*
4599	*MX1*	5152	*PDE9A*	5783	*PTPN13*
4602	*MYB*	5154	*PDGFA*	5787	*PTPRB*

4605	*MYBL2*	5157	*PDGFRL*	5789	*PTPRD*
4606	*MYBPC2*	5167	*ENPP1*	5794	*PTPRH*
4629	*MYH11*	5172	*SLC26A4*	5801	*PTPRR*
4653	*MYOC*	5187	*PER1*	5806	*PTX3*
4674	*NAP1L2*	5196	*PF4*	5816	*PVALB*
4703	*NEB*	5225	*PGC*	5874	*RAB27B*
4739	*NEDD9*	5226	*PGD*	5909	*RAP1GAP*
4744	*NEFH*	5272	*SERPINB9*	5918	*RARRES1*
4751	*NEK2*	5324	*PLAG1*	5961	*PRPH2*
4774	*NFIA*	5327	*PLAT*	5990	*RFX2*
4792	*NFKBIA*	5328	*PLAU*	5996	*RGS1*
4810	*NHS*	5339	*PLEC*	5999	*RGS4*
4811	*NID1*	5347	*PLK1*	6019	*RLN2*
4824	*NKX3-1*	5348	*FXYD1*	6036	*RNASE2*
4837	*NNMT*	5350	*PLN*	6098	*ROS1*
4851	*NOTCH1*	5376	*PMP22*	6192	*RPS4Y1*
4856	*NOV*	5396	*PRRX1*	6236	*RRAD*
4857	*NOVA1*	5427	*POLE2*	6275	*S100A4*
4862	*NPAS2*	5460	*POU5F1*	6285	*S100B*
4881	*NPR1*	5468	*PPARG*	6286	*S100P*
4883	*NPR3*	5471	*PPAT*	6288	*SAA1*
4884	*NPTX1*	5502	*PPP1R1A*	6289	*SAA2*
4885	*NPTX2*	5507	*PPP1R3C*	6300	*MAPK12*
4902	*NRTN*	5521	*PPP2R2B*	6327	*SCN2B*
4907	*NT5E*	5549	*PRELP*	6335	*SCN9A*

ID Entrez	**Símbolo**	**ID Entrez**	**Símbolo**	**ID Entrez**	**Símbolo**
6339	*SCNN1D*	6820	*SULT2B1*	7472	*WNT2*
6343	*SCT*	6941	*TCF19*	7475	*WNT6*
6354	*CCL7*	6999	*TDO2*	7481	*WNT11*
6355	*CCL8*	7010	*TEK*	7504	*XK*
6356	*CCL11*	7015	*TERT*	7538	*ZFP36*
6357	*CCL13*	7018	*TF*	7691	*ZNF132*
6358	*CCL14*	7022	*TFAP2C*	7694	*ZNF135*

6361	*CCL17*	7025	*NR2F1*	7704	*ZBTB16*
6362	*CCL18*	7031	*TFF1*	7762	*ZNF215*
6363	*CCL19*	7033	*TFF3*	7781	*SLC30A3*
6364	*CCL20*	7035	*TFPI*	7837	*PXDN*
6366	*CCL21*	7039	*TGFA*	7849	*PAX8*
6373	*CXCL11*	7042	*TGFB2*	7850	*IL1R2*
6374	*CXCL5*	7045	*TGFBI*	7941	*PLA2G7*
6376	*CX3CL1*	7048	*TGFBR2*	7976	*FZD3*
6387	*CXCL12*	7056	*THBD*	7991	*TUSC3*
6401	*SELE*	7058	*THBS2*	8000	*PSCA*
6403	*SELP*	7076	*TIMP1*	8013	*NR4A3*
6414	*SEPP1*	7078	*TIMP3*	8038	*ADAM12*
6423	*SFRP2*	7083	*TK1*	8076	*MFAP5*
6424	*SFRP4*	7093	*TLL2*	8111	*GPR68*
6440	*SFTPC*	7102	*TSPAN7*	8115	*TCL1A*
6442	*SGCA*	7122	*CLDN5*	8140	*SLC7A5*
6447	*SCG5*	7137	*TNNI3*	8190	*MIA*
6495	*SIX1*	7138	*TNNT1*	8208	*CHAF1B*
6505	*SLC1A1*	7139	*TNNT2*	8277	*TKTL1*
6518	*SLC2A5*	7153	*TOP2A*	8284	*KDM5D*
6540	*SLC6A13*	7177	*TPSAB1*	8302	*KLRC4*
6583	*SLC22A4*	7216	*TRO*	8313	*AXIN2*
6588	*SLN*	7262	*PHLDA2*	8322	*FZD4*
6590	*SLPI*	7296	*TXNRD1*	8325	*FZD8*
6615	*SNAI1*	7345	*UCHL1*	8326	*FZD9*
6623	*SNCG*	7356	*SCGB1A1*	8335	*HIST1H2AB*
6652	*SORD*	7365	*UGT2B10*	8339	*HIST1H2BG*
6662	*SOX9*	7367	*UGT2B17*	8357	*HIST1H3H*
6678	*SPARC*	7368	*UGT8*	8360	*HIST1H4D*
6689	*SPIB*	7373	*COL14A1*	8364	*HIST1H4C*
6691	*SPINK2*	7379	*UPK2*	8365	*HIST1H4H*
6695	*SPOCK1*	7391	*USF1*	8366	*HIST1H4B*
6696	*SPP1*	7422	*VEGFA*	8368	*HIST1H4L*
6769	*STAC*	7431	*VIM*	8395	*PIP5K1B*
6790	*AURKA*	7450	*VWF*	8404	*SPARCL1*

ID Entrez	Símbolo	ID Entrez	Símbolo	ID Entrez	Símbolo
8406	*SRPX*	9077	*DIRAS3*	9833	*MELK*
8434	*RECK*	9084	*VCY*	9890	*LPPR4*
8436	*SDPR*	9086	*EIF1AY*	9899	*SV2B*
8437	*RASAL1*	9088	*PKMYT1*	9914	*ATP2C2*
8464	*SUPT3H*	9118	*INA*	9915	*ARNT2*
8483	*CILP*	9148	*NEURL*	9934	*P2RY14*
8490	*RGS5*	9156	*EXO1*	9940	*DLEC1*
8492	*PRSS12*	9172	*MYOM2*	9956	*HS3ST2*
8503	*PIK3R3*	9200	*PTPLA*	9957	*HS3ST1*
8516	*ITGA8*	9201	*DCLK1*	10024	*TROAP*
8528	*DDO*	9284	*NPIP*	10076	*PTPRU*
8537	*BCAS1*	9289	*GPR56*	10082	*GPC6*
8547	*FCN3*	9369	*NRXN3*	10085	*EDIL3*
8612	*PPAP2C*	9388	*LIPG*	10100	*TSPAN2*
8622	*PDE8B*	9401	*RECQL4*	10103	*TSPAN1*
8638	*OASL*	9420	*CYP7B1*	10112	*KIF20A*
8639	*AOC3*	9429	*ABCG2*	10135	*NAMPT*
8644	*AKR1C3*	9447	*AIM2*	10158	*PDZK1IP1*
8653	*DDX3Y*	9452	*ITM2A*	10164	*CHST4*
8660	*IRS2*	9455	*HOMER2*	10170	*DHRS9*
8685	*MARCO*	9481	*SLC25A27*	10186	*LHFP*
8693	*GALNT4*	9486	*CHST10*	10202	*DHRS2*
8727	*CTNNAL1*	9508	*ADAMTS3*	10203	*CALCRL*
8786	*RGS11*	9547	*CXCL14*	10231	*RCAN2*
8792	*TNFRSF11A*	9586	*CREB5*	10234	*LRRC17*
8835	*SOCS2*	9590	*AKAP12*	10253	*SPRY2*
8840	*WISP1*	9610	*RIN1*	10265	*IRX5*
8847	*DLEU2*	9612	*NCOR2*	10290	*SPEG*
8857	*FCGBP*	9625	*AATK*	10309	*CCNO*
8876	*VNN1*	9633	*MTL5*	10319	*LAMC3*
8908	*GYG2*	9636	*ISG15*	10332	*CLEC4M*
8914	*TIMELESS*	9639	*ARHGEF10*	10335	*MRVI1*

8938	*BAIAP3*	9700	*ESPL1*	10351	*ABCA8*
8942	*KYNU*	9718	*ECE2*	10361	*NPM2*
8968	*HIST1H3F*	9720	*CCDC144A*	10365	*KLF2*
8970	*HIST1H2BJ*	9721	*GPRIN2*	10371	*SEMA3A*
9033	*PKD2L1*	9732	*DOCK4*	10417	*SPON2*
9037	*SEMA5A*	9735	*KNTC1*	10418	*SPON1*
9038	*TAAR5*	9737	*GPRASP1*	10439	*OLFM1*
9060	*PAPSS2*	9770	*RASSF2*	10446	*LRRN2*
9068	*ANGPTL1*	9787	*DLGAP5*	10457	*GPNMB*
9071	*CLDN10*	9832	*JAKMIP2*	10462	*CLEC10A*

ID Entrez	Símbolo	ID Entrez	Símbolo	ID Entrez	Símbolo
10512	*SEMA3C*	22983	*MAST1*	26108	*PYGO1*
10529	*NEBL*	23037	*PDZD2*	26150	*RIBC2*
10537	*UBD*	23066	*CAND2*	26153	*KIF26A*
10570	*DPYSL4*	23089	*PEG10*	26166	*RGS22*
10578	*GNLY*	23194	*FBXL7*	26172	*LOC26172*
10579	*TACC2*	23209	*MLC1*	26206	*SPAG8*
10580	*SORBS1*	23213	*SULF1*	26232	*FBXO2*
10606	*PAICS*	23236	*PLCB1*	26253	*CLEC4E*
10615	*SPAG5*	23242	*COBL*	26256	*CABYR*
10631	*POSTN*	23314	*SATB2*	26271	*FBXO5*
10643	*IGF2BP3*	23327	*NEDD4L*	26298	*EHF*
10656	*KHDRBS3*	23373	*CRTC1*	26353	*HSPB8*
10669	*CGREF1*	23414	*ZFPM2*	26470	*SEZ6L2*
10733	*PLK4*	23460	*ABCA6*	26577	*PCOLCE2*
10742	*RAI2*	23475	*QPRT*	26579	*MYEOV*
10744	*PTTG2*	23541	*SEC14L2*	26585	*GREM1*
10874	*NMU*	23562	*CLDN14*	26872	*STEAP1*
10875	*FGL2*	23587	*C17orf81*	26960	*NBEA*
10878	*CFHR3*	23657	*SLC7A11*	26974	*ZNF285A*
10891	*PPARGC1A*	23704	*KCNE4*	27074	*LAMP3*
10894	*LYVE1*	23743	*BHMT2*	27111	*SDCBP2*
10942	*PRSS21*	23767	*FLRT3*	27112	*FAM155B*

10964	*IFI44L*	23768	*FLRT2*	27122	*DKK3*
10974	*C10orf116*	24137	*KIF4A*	27123	*DKK2*
10984	*KCNQ1OT1*	24141	*C20orf103*	27129	*HSPB7*
11005	*SPINK5*	25791	*NGEF*	27145	*FILIP1*
11012	*KLK11*	25805	*BAMBI*	27147	*DENND2A*
11013	*TMSB15A*	25817	*FAM19A5*	27151	*CPAMD8*
11015	*KDELR3*	25833	*POU2F3*	27156	*RTDR1*
11065	*UBE2C*	25840	*METTL7A*	27165	*GLS2*
11082	*ESM1*	25878	*MXRA5*	27237	*ARHGEF16*
11096	*ADAMTS5*	25884	*CHRDL2*	27285	*TEKT2*
11130	*ZWINT*	25890	*ABI3BP*	27286	*SRPX2*
11169	*WDHD1*	25891	*PAMR1*	27299	*ADAMDEC1*
11197	*WIF1*	25894	*PLEKHG4*	27303	*RBMS3*
11226	*GALNT6*	25945	*PVRL3*	27306	*HPGDS*
11227	*GALNT5*	26002	*MOXD1*	27344	*PCSK1N*
11254	*SLC6A14*	26027	*ACOT11*	27345	*KCNMB4*
22801	*ITGA11*	26040	*SETBP1*	27443	*CECR2*
22885	*ABLIM3*	26049	*FAM169A*	28231	*SLCO4A1*
22915	*MMRN1*	26053	*AUTS2*	28999	*KLF15*
22949	*PTGR1*	26074	*C20orf26*	29070	*CCDC113*

ID Entrez	**Símbolo**	**ID Entrez**	**Símbolo**	**ID Entrez**	**Símbolo**
29089	*UBE2T*	51678	*MPP6*	55086	*CXorf57*
29091	*STXBP6*	51705	*EMCN*	55107	*ANO1*
29126	*CD274*	51744	*CD244*	55118	*CRTAC1*
29128	*UHRF1*	51751	*HIGD1B*	55137	*FIGN*
29775	*CARD10*	51761	*ATP8A2*	55138	*FAM90A1*
29798	*C2orf27A*	51804	*SIX4*	55143	*CDCA8*
29802	*VPREB3*	53342	*IL17D*	55165	*CEP55*
29931	*LOH3CR2A*	53358	*SHC3*	55203	*LGI2*
29948	*OSGIN1*	53616	*ADAM22*	55228	*PNMAL1*
29993	*PACSIN1*	53832	*IL20RA*	55231	*CCDC87*
30001	*ERO1L*	54206	*ERRFI1*	55258	*THNSL2*
30846	*EHD2*	54360	*CYTL1*	55273	*TMEM100*

50486	*G0S2*	54361	*WNT4*	55286	*C4orf19*
50487	*PLA2G3*	54437	*SEMA5B*	55304	*SPTLC3*
50507	*NOX4*	54463	*FAM134B*	55344	*PLCXD1*
50509	*COL5A3*	54478	*FAM64A*	55351	*STK32B*
50617	*ATP6V0A4*	54551	*MAGEL2*	55353	*LAPTM4B*
50636	*ANO7*	54596	*L1TD1*	55359	*STYK1*
50863	*NTM*	54621	*VSIG10*	55366	*LGR4*
51083	*GAL*	54674	*LRRN3*	55388	*MCM10*
51087	*YBX2*	54682	*MANSC1*	55450	*CAMK2N1*
51090	*PLLP*	54739	*XAF1*	55510	*DDX43*
51129	*ANGPTL4*	54756	*IL17RD*	55521	*TRIM36*
51162	*EGFL7*	54829	*ASPN*	55545	*MSX2P1*
51196	*PLCE1*	54830	*NUP62CL*	55576	*STAB2*
51208	*CLDN18*	54847	*SIDT1*	55698	*RADIL*
51284	*TLR7*	54852	*PAQR5*	55714	*ODZ3*
51297	*PLUNC*	54855	*FAM46C*	55765	*C1orf106*
51299	*NRN1*	54869	*EPS8L1*	55786	*ZNF415*
51302	*CYP39A1*	54894	*RNF43*	55789	*DEPDC1B*
51308	*REEP2*	54906	*C10orf18*	55799	*CACNA2D3*
51311	*TLR8*	54922	*RASIP1*	55872	*PBK*
51316	*PLAC8*	54933	*RHBDL2*	55966	*AJAP1*
51348	*KLRF1*	54959	*ODAM*	56000	*NXF3*
51364	*ZMYND10*	54979	*HRASLS2*	56121	*PCDHB15*
51450	*PRRX2*	55034	*MOCOS*	56122	*PCDHB14*
51454	*GULP1*	55040	*EPN3*	56126	*PCDHB10*
51554	*CCRL1*	55061	*SUSD4*	56127	*PCDHB9*
51560	*RAB6B*	55064	*C9orf68*	56128	*PCDHB8*
51561	*IL23A*	55065	*GPR172B*	56129	*PCDHB7*
51659	*GINS2*	55083	*KIF26B*	56131	*PCDHB4*
51673	*TPPP3*	55084	*SOBP*	56143	*PCDHA5*

ID Entrez	**Símbolo**	**ID Entrez**	**Símbolo**	**ID Entrez**	**Símbolo**
56147	*PCDHA1*	57758	*SCUBE2*	79071	*ELOVL6*
56164	*STK31*	57817	*HAMP*	79092	*CARD14*

56253	*CRTAM*	58189	*WFDC1*	79148	*MMP28*
56256	*SERTAD4*	58494	*JAM2*	79153	*GDPD3*
56477	*CCL28*	59269	*HIVEP3*	79173	*C19orf57*
56603	*CYP26B1*	59271	*C21orf63*	79191	*IRX3*
56649	*TMPRSS4*	59272	*ACE2*	79258	*MMEL1*
56673	*C11orf16*	59277	*NTN4*	79365	*BHLHE41*
56675	*NRIP3*	59285	*CACNG6*	79370	*BCL2L14*
56937	*PMEPA1*	59353	*TMEM35*	79413	*ZBED2*
56944	*OLFML3*	60437	*CDH26*	79589	*RNF128*
56967	*C14orf132*	63027	*SLC22A23*	79611	*ACSS3*
56977	*STOX2*	63876	*PKNOX2*	79614	*C5orf23*
56992	*KIF15*	63951	*DMRTA1*	79618	*HMBOX1*
57101	*ANO2*	64072	*CDH23*	79623	*GALNT14*
57110	*HRASLS*	64073	*C19orf33*	79633	*FAT4*
57188	*ADAMTSL3*	64093	*SMOC1*	79642	*ARSJ*
57194	*ATP10A*	64094	*SMOC2*	79645	*EFCAB1*
57210	*SLC45A4*	64108	*RTP4*	79682	*MLF1IP*
57214	*KIAA1199*	64131	*XYLT1*	79692	*ZNF322A*
57235	*KIAA0485*	64167	*ERAP2*	79723	*SUV39H2*
57335	*ZNF286A*	64220	*STRA6*	79739	*TTLL7*
57348	*TTYH1*	64283	*RGNEF*	79772	*MCTP1*
57393	*TMEM27*	64288	*ZNF323*	79774	*GRTP1*
57402	*S100A14*	64332	*NFKBIZ*	79801	*SHCBP1*
57452	*GALNTL1*	64344	*HIF3A*	79819	*WDR78*
57460	*PPM1H*	64377	*CHST8*	79825	*CCDC48*
57482	*KIAA1211*	64388	*GREM2*	79841	*AGBL2*
57484	*RNF150*	64399	*HHIP*	79844	*ZDHHC11*
57493	*HEG1*	64506	*CPEB1*	79852	*EPHX3*
57501	*KIAA1257*	64699	*TMPRSS3*	79875	*THSD4*
57514	*ARHGAP31*	64757	*MOSC1*	79901	*CYBRD1*
57535	*KIAA1324*	64849	*SLC13A3*	79919	*C2orf54*
57537	*SORCS2*	64866	*CDCP1*	79931	*TNIP3*
57608	*KIAA1462*	64919	*BCL11B*	79953	*TMEM90B*
57616	*TSHZ3*	65243	*ZNF643*	79957	*PAQR6*
57631	*LRCH2*	65997	*RASL11B*	79974	*C7orf58*

57633	*LRRN1*	66000	*TMEM108*	79977	*GRHL2*
57670	*KIAA1549*	78989	*COLEC11*	79987	*SVEP1*
57683	*ZDBF2*	78995	*C17orf53*	79992	*C6orf59*
57705	*WDFY4*	79006	*METRN*	79993	*ELOVL7*
57717	*PCDHB16*	79019	*CENPM*	80034	*CSRNP3*

ID Entrez	Símbolo	ID Entrez	Símbolo	ID Entrez	Símbolo
80054	*LOC80054*	83988	*NCALD*	85004	*RERG*
80144	*FRAS1*	83992	*CTTNBP2*	85016	*C11orf70*
80150	*ASRGL1*	84000	*TMPRSS13*	85453	*TSPYL5*
80178	*C16orf59*	84057	*MND1*	85455	*DISP2*
80206	*FHOD3*	84059	*GPR98*	85462	*FHDC1*
80235	*PIGZ*	84069	*PLEKHN1*	89765	*RSPH1*
80237	*ELL3*	84073	*MYCBPAP*	89795	*NAV3*
80258	*EFHC2*	84152	*PPP1R1B*	89796	*NAV1*
80303	*EFHD1*	84171	*LOXL4*	89822	*KCNK17*
80310	*PDGFD*	84189	*SLITRK6*	89872	*AQP10*
80326	*WNT10A*	84206	*MEX3B*	89958	*C9orf140*
80705	*TSGA10*	84215	*ZNF541*	90050	*FAM181A*
80723	*TMEM22*	84258	*SYT3*	90362	*FAM110B*
80726	*KIAA1683*	84417	*C2orf40*	90865	*IL33*
80759	*KHDC1*	84419	*C15orf48*	91057	*CCDC34*
80760	*ITIH5*	84448	*ABLIM2*	91120	*ZNF682*
81029	*WNT5B*	84553	*C6orf168*	91133	*L3MBTL4*
81035	*COLEC12*	84623	*KIRREL3*	91543	*RSAD2*
81501	*TM7SF4*	84624	*FNDC1*	91614	*DEPDC7*
81557	*MAGED4B*	84627	*ZNF469*	91683	*SYT12*
81575	*APOLD1*	84632	*AFAP1L2*	91687	*CENPL*
81578	*COL21A1*	84649	*DGAT2*	91851	*CHRDL1*
81831	*NETO2*	84667	*HES7*	91977	*MYOZ3*
81931	*ZNF93*	84676	*TRIM63*	92291	*CAPN13*
83439	*TCF7L1*	84688	*C9orf24*	92304	*SCGB3A1*
83461	*CDCA3*	84696	*ABHD1*	92312	*MEX3A*
83481	*EPPK1*	84706	*GPT2*	92340	*C17orf72*

83540	*NUF2*	84707	*BEX2*	92747	*C20orf114*
83543	*AIF1L*	84708	*LNX1*	92949	*ADAMTSL1*
83604	*TMEM47*	84709	*C4orf49*	93082	*NEURL3*
83648	*FAM167A*	84803	*AGPAT9*	93099	*DMKN*
83657	*DYNLRB2*	84830	*C6orf105*	93273	*LEMD1*
83661	*MS4A8B*	84842	*HPDL*	93517	*SDR42E1*
83690	*CRISPLD1*	84848	*MGC16121*	93986	*FOXP2*
83849	*SYT15*	84870	*RSPO3*	94274	*PPP1R14A*
83853	*ROPN1L*	84900	*RNFT2*	96626	*LIMS3*
83869	*TTTY14*	84935	*C13orf33*	112476	*PRRT2*
83872	*HMCN1*	84952	*CGNL1*	113115	*FAM54A*
83879	*CDCA7*	84953	*MICALCL*	113130	*CDCA5*
83888	*FGFBP2*	84962	*JUB*	113146	*AHNAK2*
83903	*GSG2*	84966	*IGSF21*	113278	*C20orf54*
83959	*SLC4A11*	84969	*TOX2*	113730	*KLHDC7B*

ID Entrez	**Símbolo**	**ID Entrez**	**Símbolo**	**ID Entrez**	**Símbolo**
114569	*MAL2*	126129	*CPT1C*	142683	*ITLN2*
114787	*GPRIN1*	126567	*C2CD4C*	143662	*MUC15*
114800	*CCDC85A*	127495	*LRRC39*	144193	*AMDHD1*
114818	*KLHL29*	127602	*DNAH14*	144347	*FAM101A*
114827	*FHAD1*	128153	*SPATA17*	144406	*WDR66*
114880	*OSBPL6*	128229	*C1orf182*	144501	*KRT80*
114899	*C1QTNF3*	128602	*C20orf85*	145258	*GSC*
114905	*C1QTNF7*	128611	*ZNF831*	145773	*FAM81A*
114907	*FBXO32*	128872	*HMGB3L1*	146212	*KCTD19*
115361	*GBP4*	129790	*C7orf13*	146330	*FBXL16*
115362	*GBP5*	130271	*PLEKHH2*	146434	*ZNF597*
115572	*FAM46B*	130367	*SGPP2*	146760	*RTN4RL1*
115908	*CTHRC1*	130733	*TMEM178*	146845	*WDR16*
116028	*C16orf75*	131177	*FAM3D*	147166	*TRIM16L*
116039	*OSR2*	131450	*CD200R1*	147372	*CCBE1*
116159	*CYYR1*	131578	*LRRC15*	147463	*ANKRD29*
116211	*TM4SF19*	131873	*COL6A6*	147525	*C18orf18*

116328	*C8orf34*	132671	*SPATA18*	147645	*VSIG10L*
116372	*LYPD1*	132864	*CPEB2*	147744	*TMEM190*
116442	*RAB39B*	133022	*TRAM1L1*	147906	*DACT3*
116832	*RPL39L*	133690	*CAPSL*	147920	*IGFL2*
116969	*ART5*	134147	*CMBL*	147968	*CAPN12*
117156	*SCGB3A2*	134265	*AFAP1L1*	148229	*ATP8B3*
117157	*SH2D1B*	134466	*LOC134466*	148252	*DIRAS1*
117166	*WFIKKN1*	135112	*NCOA7*	148418	*SAMD13*
117247	*SLC16A10*	135398	*C6orf141*	148641	*SLC35F3*
117248	*GALNTL2*	135932	*TMEM139*	149111	*CNIH3*
118491	*TTC18*	136306	*SVOPL*	149351	*LOC149351*
118663	*BTBD16*	137872	*ADHFE1*	149483	*CCDC17*
118932	*ANKRD22*	137902	*PXDNL*	150248	*C22orf15*
119391	*GSTO2*	137994	*LETM2*	150378	*FLJ30901*
119587	*CPXM2*	139065	*SLITRK4*	150696	*PROM2*
120071	*GYLTL1B*	139324	*HDX*	150763	*GPAT2*
120376	*C11orf93*	139886	*SPIN4*	151174	*LOC151174*
120892	*LRRK2*	140032	*RPS4Y2*	151176	*FAM132B*
122402	*TDRD9*	140578	*CHODL*	151242	*PPP1R1C*
122622	*ADSSL1*	140597	*TCEAL2*	151246	*SGOL2*
124093	*CCDC78*	140706	*C20orf160*	151473	*SLC16A14*
124220	*ZG16B*	140733	*MACROD2*	151507	*MSL3L2*
124602	*KIF19*	140766	*ADAMTS14*	151827	*LRRC34*
124739	*USP43*	140809	*SRXN1*	151887	*CCDC80*
124976	*SPNS2*	140862	*ISM1*	151888	*BTLA*

ID Entrez	Símbolo	ID Entrez	Símbolo	ID Entrez	Símbolo
152078	*C3orf55*	200634	*KRTCAP3*	256076	*COL29A1*
153478	*PLEKHG4B*	200810	*ALG1L*	256435	*ST6GALNAC3*
153572	*IRX2*	200958	*MUC20*	256691	*MAMDC2*
153579	*BTNL9*	201161	*CENPV*	259173	*ALS2CL*
154313	*C6orf165*	201456	*FBXO15*	259266	*ASPM*
154796	*AMOT*	201799	*TMEM154*	259307	*IL4I1*
155368	*WBSCR27*	202333	*CMYA5*	260293	*CYP4X1*

157310	*PEBP4*	202915	*TMEM184A*	260436	*C4orf7*
157313	*CDCA2*	203111	*C8orf47*	266977	*GPR110*
157506	*RDH10*	219348	*PLAC9*	282679	*AQP11*
157570	*ESCO2*	219595	*FOLH1B*	283120	*H19*
157773	*C8orf48*	219736	*STOX1*	283208	*P4HA3*
157869	*C8orf84*	219790	*RTKN2*	283358	*B4GALNT3*
157927	*C9orf62*	220136	*CCDC11*	283481	*LOC283481*
158295	*MGC24103*	220164	*DOK6*	283518	*KCNRG*
158326	*FREM1*	220359	*TIGD3*	283663	*LOC283663*
158511	*CSAG1*	220594	*LOC220594*	284013	*VMO1*
159371	*TMEM20*	220963	*SLC16A9*	284047	*CCDC144B*
160364	*CLEC12A*	221002	*RASGEF1A*	284119	*PTRF*
160365	*CLECL1*	221150	*SKA3*	284307	*ZIK1*
160622	*GRASP*	221303	*FAM162B*	284339	*TMEM145*
161725	*OTUD7A*	221357	*GSTA5*	284367	*SIGLECP3*
162632	*LOC162632*	221424	*C6orf154*	284654	*RSPO1*
162963	*ZNF610*	221476	*PI16*	284716	*RIMKLA*
163255	*ZNF540*	221687	*RNF182*	284889	*LOC284889*
163782	*KANK4*	221806	*VWDE*	284904	*SEC14L4*
164832	*LONRF2*	221914	*GPC2*	284992	*CCDC150*
165055	*CCDC138*	221935	*SDK1*	285016	*FAM150B*
167359	*MGC42105*	221981	*THSD7A*	285141	*LOC285141*
167465	*ZNF366*	222171	*PRR15*	285489	*DOK7*
168002	*DACT2*	222235	*FBXL13*	285755	*PPIL6*
169792	*GLIS3*	245972	*ATP6V0D2*	286046	*XKR6*
170679	*PSORS1C1*	246100	*CTAG1A*	286527	*TMSB15B*
171024	*SYNPO2*	252995	*FNDC5*	286749	*STON1-GTF2A1L*
196740	*C10orf72*	253152	*EPHX4*	286827	*TRIM59*
198437	*C20orf201*	253264	*LOC253264*	338339	*CLEC4D*
199675	*C19orf59*	253650	*ANKRD18A*	338382	*RAB7B*
199713	*NLRP7*	253982	*ASPHD1*	339145	*FAM92B*
199786	*FAM129C*	254910	*LCE5A*	339184	*CCDC144NL*
200162	*SPAG17*	255480	*LOC255480*	339390	*CLEC4G*
200172	*SLFNL1*	255631	*COL24A1*	339524	*LOC339524*
200315	*APOBEC3A*	255743	*NPNT*	339768	*ESPNL*

ID Entrez	Símbolo	ID Entrez	Símbolo	ID Entrez	Símbolo
339778	*C2orf70*	401233	*LOC401233*	728882	*FAM182B*
339804	*C2orf74*	401237	*FLJ22536*	729085	*FAM198A*
340286	*FAM183B*	401474	*SAMD12*	729680	*LOC729680*
340542	*BEX5*	402483	*tcag7.907*	729983	*LOC729983*
340719	*NANOS1*	439949	*LOC439949*	100113384	*SNORD123*
342035	*GLDN*	440068	*CARD17*	100128511	*LOC100128511*
342979	*PALM3*	440461	*LOC440461*	100130967	*LOC100130967*
344148	*NCKAP5*	440556	*FLJ42875*	100131170	*LOC100131170*
346689	*KLRG2*	440585	*FAM183A*	100131541	*UNQ6228*
347744	*C6orf52*	440712	*C1orf186*	100131564	*LOC100131564*
347902	*AMIGO2*	440792	*LOC440792*	100131733	*LOC100131733*
348751	*LOC348751*	440823	*MIAT*	100131997	*FAM27E3*
352961	*HCG26*	441027	*TMEM150C*	100132247	*LOC100132247*
353189	*SLCO4C1*	441168	*FAM26F*	100132885	*LOC100132885*
374569	*ASPG*	441208	*LOC441208*	100133161	*LOC100133161*
387066	*SNHG5*	441295	*OR2A9P*	100133941	*CD24*
387328	*ZNF322B*	441317	*FAM90A7*	100134018	*hCG_1774568*
387758	*FIBIN*	441376	*C8orf85*	100271835	*LIMS3-LOC440895*
387763	*AG2*	441631	*TSPAN11*	100288985	*LOC100288985*
387885	*CCDC42B*	441795	*LOC441795*	100289058	*LOC100289058*
387914	*SHISA2*	445328	*ARHGEF5L*	100293193	*LOC100293193*
388115	*C15orf52*	474354	*LRRC18*	—	A_32_P149011
388125	*C2CD4B*	619279	*ZNF704*	—	A_32_P208200
388135	*C15orf59*	642236	*LOC642236*	—	A_24_P315862
388242	*LOC388242*	642477	*FLJ39632*	—	A_23_P121234
388585	*HES5*	642521	*LOC642521*	—	A_32_P148122
388630	*LOC388630*	644186	*C22orf41*	—	A_32_P115277
388815	*C21orf34*	644246	*LOC644246*	—	A_32_P33213
389136	*VGLL3*	645027	*EVPLL*	—	A_32_P124773
389432	*SAMD5*	645277	*LOC645277*	—	A_32_P105465
389458	*LOC389458*	645722	*LOC645722*	—	A_24_P5994
389493	*LOC389493*	650655	*ABCA17P*	—	A_23_P28307

389831	*LOC389831*	653140	*C2orf84*	—	A_24_P890995
389834	*LOC389834*	653509	*SFTPA1*	—	A_32_P210106
391267	*C21orf81*	654433	*LOC654433*	—	A_24_P298228
399668	*NCRNA00086*	654790	*PCP4L1*	—	A_24_P702749
399744	*LOC399744*	723790	*HIST2H2AA4*	—	A_24_P307375
399948	*C11orf92*	727832	*GOLGA6L6*	—	A_24_P930088
399959	*LOC399959*	727916	*LOC727916*	—	A_24_P659980
400043	*LOC400043*	728215	*FAM155A*	—	A_32_P115749
400456	*LOC400456*	728242	*XAGE2B*	—	A_32_P181826
400566	*C17orf97*	728613	*LOC728613*	—	A_32_P172578

ID Entrez	Símbolo	ID Entrez	Símbolo	ID Entrez	Símbolo
—	A_32_P23810	—	A_23_P211468	—	A_24_P827096
—	A_32_P164917	—	A_32_P50973	—	A_32_P132194
—	A_24_P233560	—	A_24_P936252	—	A_32_P186865
—	A_23_P159435	—	A_32_P406142	—	A_24_P92174
—	A_32_P96752	—	A_32_P15288	—	A_23_P21907
—	A_24_P67494	—	A_24_P204727	—	A_32_P73991
—	A_32_P72351	—	A_32_P146844	—	A_24_P221903
—	A_24_P683861	—	A_32_P157504	—	A_32_P108033
—	A_24_P643776	—	A_32_P227921	—	A_32_P88587
—	A_32_P162862	—	A_24_P263786	—	A_24_P64442
—	A_24_P170136	—	A_32_P17343	—	A_32_P61298
—	A_23_P170719	—	A_24_P110487	—	A_32_P226941
—	A_24_P590547	—	A_24_P490109	—	A_32_P226356
—	A_24_P213134	—	A_32_P125402	—	A_24_P822869
—	A_32_P194563	—	A_32_P35486	—	A_32_P99533
—	A_24_P247074	—	A_32_P198295	—	A_24_P101742
—	A_24_P395415	—	A_23_P43649	—	A_32_P100947
—	A_32_P32463	—	A_32_P227496	—	A_23_P250516
—	A_32_P59277	—	A_24_P494425	—	A_32_P212258
—	A_24_P110618	—	A_24_P604784	—	A_32_P132748
—	A_32_P5205	—	A_32_P126259	—	A_24_P558135
—	A_32_P90812	—	A_32_P32043	—	A_23_P135123

—	A_32_P122715	—	A_32_P229818	—	A_24_P212024
—	A_32_P155512	—	A_32_P118657	—	A_24_P517918
—	A_24_P538459	—	A_24_P15388	—	A_23_P116195
—	A_24_P144346	—	A_24_P204374	—	A_32_P112100
—	A_24_P178444	—	A_24_P92683	—	A_24_P478940
—	A_32_P214395	—	A_24_P589028	—	A_24_P177844
—	A_23_P44053	—	A_32_P117322	—	A_24_P241996
—	A_32_P152255	—	A_24_P929818	—	A_32_P117186
—	A_23_P21260	—	A_32_P151747	—	A_32_P233799
—	A_32_P101623	—	A_32_P109078	—	A_24_P707102
—	A_32_P188921	—	A_24_P484699	—	A_23_P32793
—	A_32_P164971	—	A_32_P43664	—	A_24_P358474
—	A_24_P795230	—	A_24_P608268	—	A_32_P42224
—	A_24_P267814	—	A_24_P24371	—	A_24_P100684
—	A_32_P103678	—	A_23_P111525	—	A_32_P164477
—	A_24_P333306	—	A_32_P214565	—	A_32_P110485
—	A_32_P215556	—	A_32_P1166	—	A_24_P460763
—	A_23_P72252	—	A_32_P153361	—	A_32_P39944
—	A_24_P497843	—	A_24_P315941	—	A_24_P845631
—	A_24_P916853	—	A_23_P158868	—	A_32_P166733

ID Entrez	Símbolo	ID Entrez	Símbolo	ID Entrez	Símbolo
—	A_32_P157465	—	A_24_P927205	—	A_32_P39440
—	A_32_P115122	—	A_32_P196918	—	A_23_P166508
—	A_24_P203418	—	A_24_P315854	—	A_32_P35303
—	A_23_P21249	—	A_23_P435390	—	A_32_P33802
—	A_24_P203886	—	A_32_P182458	—	A_32_P173922
—	A_32_P30874	—	A_32_P155984	—	A_23_P87421
—	A_32_P108516	—	A_24_P478556	—	A_32_P40673
—	A_32_P106615	—	A_32_P21459	—	A_32_P164916
—	A_24_P626951	—	A_32_P161755	—	A_24_P745352
—	A_24_P306814	—	A_24_P8257	—	A_32_P139163
—	A_24_P940469	—	A_32_P189034	—	A_23_P390206
—	A_24_P592060	—	A_32_P122951	—	A_32_P68055

—	A_24_P42453	—	A_24_P565110	—	A_24_P825969
—	A_24_P914479	—	A_32_P226768	—	A_24_P816844
—	A_24_P401124	—	A_32_P157927	—	A_32_P142602
—	A_24_P372189	—	A_32_P114918	—	A_23_P170713
—	A_32_P46456	—	A_23_P140884	—	A_24_P418744
—	A_24_P592591	—	A_23_P9997	—	A_32_P43959
—	A_24_P384119	—	A_32_P330691	—	A_24_P16214
—	A_24_P923271	—	A_23_P98671	—	A_32_P77102
—	A_32_P49552	—	A_24_P813550	—	A_24_P246825
—	A_32_P108748	—	A_32_P47166	—	A_23_P84791
—	A_24_P16337	—	A_24_P213325	—	A_32_P55414
—	A_24_P733308	—	A_24_P871940	—	A_32_P15081
—	A_24_P76868	—	A_24_P677559	—	A_24_P930327
—	A_24_P878419	—	A_23_P217187	—	A_32_P188993
—	A_24_P934162	—	A_23_P56868	—	A_23_P361654
—	A_24_P592871	—	A_24_P932887	—	A_32_P229493
—	A_23_P96191	—	A_32_P177595	—	A_24_P333357
—	A_32_P169550	—	A_23_P170830	—	A_23_P35546
—	A_24_P484904	—	A_32_P68076	—	A_32_P12372
—	A_24_P639701	—	A_32_P212406	—	A_24_P24053
—	A_32_P165990	—	A_32_P877	—	A_23_P112957
—	A_23_P20328	—	A_32_P121537	—	A_24_P341546
—	A_32_P214054	—	A_23_P21393	—	A_32_P43914
—	A_24_P932388	—	A_32_P226073	—	A_24_P621701
—	A_32_P106864	—	A_32_P224586	—	A_24_P75994
—	A_32_P45375	—	A_24_P152345	—	A_24_P665185
—	A_24_P376139	—	A_24_P127462	—	A_32_P44099
—	A_24_P361816	—	A_32_P38313	—	A_32_P30004
—	A_24_P919283	—	A_24_P519504	—	A_24_P916965
—	A_24_P844100	—	A_24_P229447	—	A_23_P8812

ID Entrez	Símbolo	ID Entrez	Símbolo	ID Entrez	Símbolo
—	A_32_P202066	—	A_24_P904845	—	A_24_P514678
—	A_32_P221641	—	A_32_P228886	—	A_24_P750327

—	A_32_P9924	—	A_23_P210158	—	A_24_P358321
—	A_23_P408363	—	A_23_P136753	—	A_32_P72541
—	A_24_P576445	—	A_32_P213678	—	A_32_P213418
—	A_32_P97513	—	A_32_P30075	—	A_32_P214969
—	A_32_P74814	—	A_24_P7750	—	A_23_P159163
—	A_24_P195974	—	A_24_P135841	—	A_24_P504939
—	A_32_P722809	—	A_24_P169713	—	A_24_P862083
—	A_24_P318990	—	A_24_P341126	—	A_32_P199824
—	A_32_P52227	—	A_32_P194704	—	A_32_P115518
—	A_32_P42895	—	A_24_P161764	—	A_24_P16004
—	A_32_P224345	—	A_32_P102581	—	A_32_P82179
—	A_32_P226205	—	A_23_P124313	—	A_23_P21800
—	A_32_P36552	—	A_32_P193240	—	A_32_P79041
—	A_32_P133038	—	A_32_P131998	—	A_24_P298805
—	A_32_P157671	—	A_32_P174214	—	A_24_P925422
—	A_32_P76602	—	A_32_P1291	—	A_32_P227657
—	A_32_P46171	—	A_23_P11980	—	A_32_P53603
—	A_24_P33341	—	A_32_P209582	—	A_32_P117185
—	A_32_P104448	—	A_32_P127019	—	A_32_P29149
—	A_32_P144852	—	A_32_P197720	—	A_24_P357847
—	A_32_P48149	—	A_24_P375691	—	A_32_P201434
—	A_24_P195454	—	A_24_P927639	—	A_32_P134634
—	A_24_P384604	—	A_32_P207885	—	A_32_P230253
—	A_32_P224888	—	A_32_P121234	—	A_32_P171253
—	A_32_P130536	—	A_24_P264143	—	A_32_P185766
—	A_32_P15829	—	A_24_P451992	—	A_32_P71858
—	A_32_P2883	—	A_32_P509964	—	A_32_P132766
—	A_23_P3083	—	A_23_P58137	—	A_32_P8732
—	A_24_P703642	—	A_24_P867201	—	A_24_P671115
—	A_24_P417352	—	A_32_P215143	—	A_32_P75867
—	A_24_P109921	—	A_32_P35575	—	A_24_P139761
—	A_32_P3932	—	A_32_P216888	—	A_32_P72611
—	A_23_P32583	—	A_24_P101226	—	A_32_P226525
—	A_32_P49959	—	A_32_P105940	—	A_32_P61522
—	A_23_P149441	—	A_32_P93352	—	A_32_P179686

—	A_24_P465799	—	A_32_P13823	—	A_23_P350782
—	A_24_P472081	—	A_32_P89480	—	A_32_P429876
—	A_32_P184268	—	A_24_P716405	—	A_32_P167111
—	A_24_P173566	—	A_24_P928306	—	A_32_P133090
—	A_32_P12703	—	A_32_P51988	—	A_23_P357504

ID Entrez	Símbolo	ID Entrez	Símbolo	ID Entrez	Símbolo
—	A_32_P47285	—	A_23_P136026	—	A_32_P71476
—	A_32_P76137	—	A_24_P15550	—	A_24_P767725
—	A_24_P379629	—	A_24_P185516	—	A_24_P272146
—	A_24_P358606	—	A_24_P281605	—	A_24_P128361
—	A_32_P159192	—	A_32_P65589	—	A_32_P200144
—	A_32_P42329	—	A_24_P659202	—	A_32_P111072
—	A_24_P23327	—	A_24_P649327	—	A_24_P323298
—	A_32_P194423	—	A_24_P110242	—	A_32_P177843
—	A_32_P12065	—	A_32_P85593	—	A_23_P417363
—	A_24_P741023	—	A_32_P94685	—	A_24_P67063
—	A_23_P147224	—	A_32_P232413		
—	A_32_P232682	—	A_24_P178834		
—	A_32_P214860	—	A_32_P65022		
—	A_24_P367100	—	A_32_P88163		
—	A_23_P147578	—	A_24_P475753		
—	A_24_P488083	—	A_32_P44210		
—	A_24_P923854	—	A_32_P215866		
—	A_24_P33055	—	A_32_P176036		
—	A_23_P259763	—	A_24_P918044		
—	A_24_P900555	—	A_23_P339954		
—	A_24_P326084	—	A_32_P13337		
—	A_24_P68008	—	A_32_P91186		
—	A_23_P251002	—	A_24_P255415		
—	A_23_P113056	—	A_24_P306905		
—	A_24_P524452	—	A_32_P74477		
—	A_24_P288915	—	A_32_P159289		
—	A_32_P184039	—	A_32_P168727		

—	A_32_P35947	—	A_32_P209104
—	A_24_P845223	—	A_32_P134679
—	A_32_P78488	—	A_23_P388146
—	A_32_P121978	—	A_24_P204574
—	A_32_P212373	—	A_32_P154445
—	A_32_P116488	—	A_32_P206308
—	A_32_P144281	—	A_24_P179107
—	A_24_P895836	—	A_24_P398370
—	A_24_P418216	—	A_32_P219581
—	A_32_P35668	—	A_32_P175557
—	A_23_P73328	—	A_32_P30434
—	A_32_P77742	—	A_32_P327750
—	A_32_P191895	—	A_32_P109835
—	A_32_P101653	—	A_32_P185530
—	A_24_P315594	—	A_32_P23731

REIVINDICACIONES

1. Método *in vitro* de obtención de datos útiles para el pronóstico de cáncer de pulmón no microcítico de estadio I o II caracterizado por la detección y/o cuantificación del producto de expresión de los genes de la tabla 1 en la muestra biológica aislada de un sujeto.

2. Método según la reivindicación 1 que además comprende la comparación de los datos útiles con valores de expresión de referencia para el producto de expresión de los genes de la tabla 1 en cáncer de pulmón no microcítico de estadio I o II obtenidos de sujetos en los que el pronóstico es conocido (muestra de referencia) para identificación del sujeto como un sujeto de buen pronóstico o de mal pronóstico.

3. Método según las reivindicaciones 1 o 2 donde la comparación se realiza mediante el método del centroide compacto más cercano.

4. Método *in vitro* para el pronóstico de cáncer de pulmón no microcítico de estadio I o II caracterizado por:

   a. la detección y cuantificación del producto de expresión de los genes de la tabla 1 en una muestra de referencia;

   b. el cálculo de un valor de referencia (valor 1) para cada producto de expresión de los genes de la tabla 1 en las muestras de referencia de pronóstico favorable (grupo de buen pronóstico) y el cálculo de un valor de referencia (valor 2) en las muestras de referencia de pronóstico desfavorable (grupo de mal pronóstico) mediante el uso del método del centroide más cercano;

   c. la detección y cuantificación del producto de expresión de los genes de la tabla 1 en la muestra biológica de un nuevo sujeto en el que el pronóstico es desconocido (muestra de estudio);

   d. la comparación mediante el uso del método de clasificación del centroide compacto más cercano de los valores obtenidos en la detección y cuantificación del producto de expresión de los genes de la tabla 1 en la muestra de estudio con los valores de referencia obtenidos en los grupos de buen y mal pronóstico.

   e. la asociación de la muestra de estudio al grupo de buen pronóstico o al grupo de mal pronóstico según lo establecido en el método del centroide compacto más cercano.

5. Método según la reivindicación 4 donde el método del centroide más cercano se lleva a cabo a través de la aplicación de Predicción de Análisis de Microarrays (PAM).

6. Método según cualquiera de las reivindicaciones 1 a 5 donde la muestra de referencia y las muestras de estudio han sido previamente normalizadas antes de la comparación.

7. Método según cualquiera de las reivindicaciones 1 a 6 que además comprende la detección y/o cuantificación de al menos un producto de expresión de los genes descritos en la tabla 2.

8. Método según cualquiera de las reivindicaciones 1 a 7 donde el producto de expresión es ARN mensajero.

9. Método según la reivindicación 8 donde la detección y/o cuantificación del ARN mensajero se realiza mediante microarrays.

10. Método según la reivindicación 8 donde la detección y/o cuantificación del ARN mensajero se realiza mediante RT-PCR.

11. Método según cualquiera de las reivindicaciones 1 a 7 donde el producto de expresión es una proteína.

12. Método según la reivindicación 11 donde la detección y/o cuantificación de la proteína se realiza mediante inmuno blotting, inmunohistoquímica, cromatografía o microarrays.

13. Método según cualquiera de las reivindicaciones 1 a 12 donde la muestra biológica se selecciona de la lista que comprende: tejido, sangre, plasma, suero, linfa, lavado broncoalveolar o fluido ascítico.

14. Método según cualquiera de las reivindicaciones 1 a 13 donde la muestra biológica es fresca, congelada, fijada o fijada y embebida en parafina.

15. Método según cualquiera de las reivindicaciones 1 a 14 donde el sujeto es un humano.

16. Uso *in vitro* de los productos de expresión de los genes de la tabla 1 como marcador pronóstico de cáncer de pulmón no microcítico de estadio I o II.

17. Kit que comprende sondas que consisten en las sondas que reconocen el ARN mensajero, producto de la expresión de los genes de la tabla 1, o el ADN complementario o ARN complementario a dicho ARN mensajero, o anticuerpos que reconocen una proteína producto de expresión de los genes de la tabla 1.

18. Kit según la reivindicación 17 que comprende sondas, que consisten en las sondas que reconocen el ARN mensajero producto de la expresión de los genes de la tabla 1.

19. Kit según la reivindicación 15 donde las sondas son las secuencias SEQ ID NO: 1 a SEQ IDNO: 66.

20. Kit según cualquiera de las reivindicaciones 17 a 19 que además comprende al menos una sonda o un anticuerpo que reconoce un producto de expresión de los genes de la tabla 2.

21. Kit según la reivindicación 20 que comprende al menos una sonda que reconoce un producto de expresión de los genes de la tabla 2.

22. Kit según cualquiera de las reivindicaciones 17 a 21 que además comprende al menos unos de los reactivos seleccionados de la lista que comprende: una retrotranscriptasa, una ARN polimerasa o un fluoróforo.

23. Kit según cualquiera de las reivindicaciones 17 a 22 donde las sondas están situadas en un soporte sólido.

24. Uso del kit según las reivindicaciones 17 a 23 para la obtención de datos útiles para el pronóstico del carcinoma de pulmón no microcítico de estadios I o II.

FIG.1

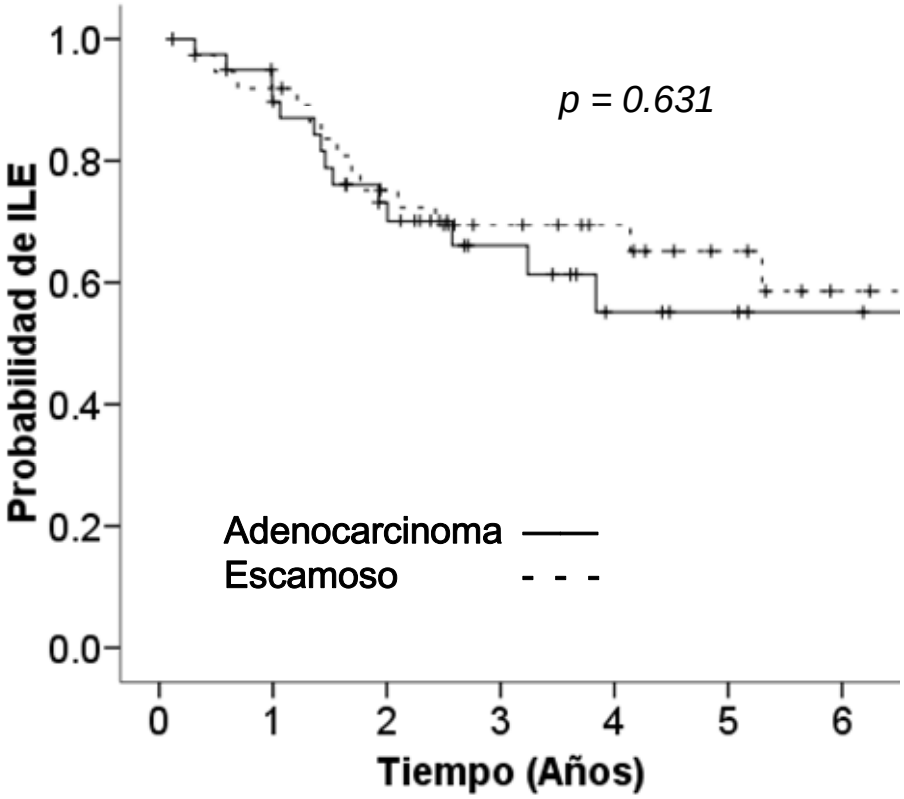

1.0
0.8
0.6
0.4
0.2
0.0
Probabilidad de ILE
p = 0.631
Adenocarcinoma
Escamoso
0
1
2
3
4
5
6
Tiempo (Años)

FIG.2

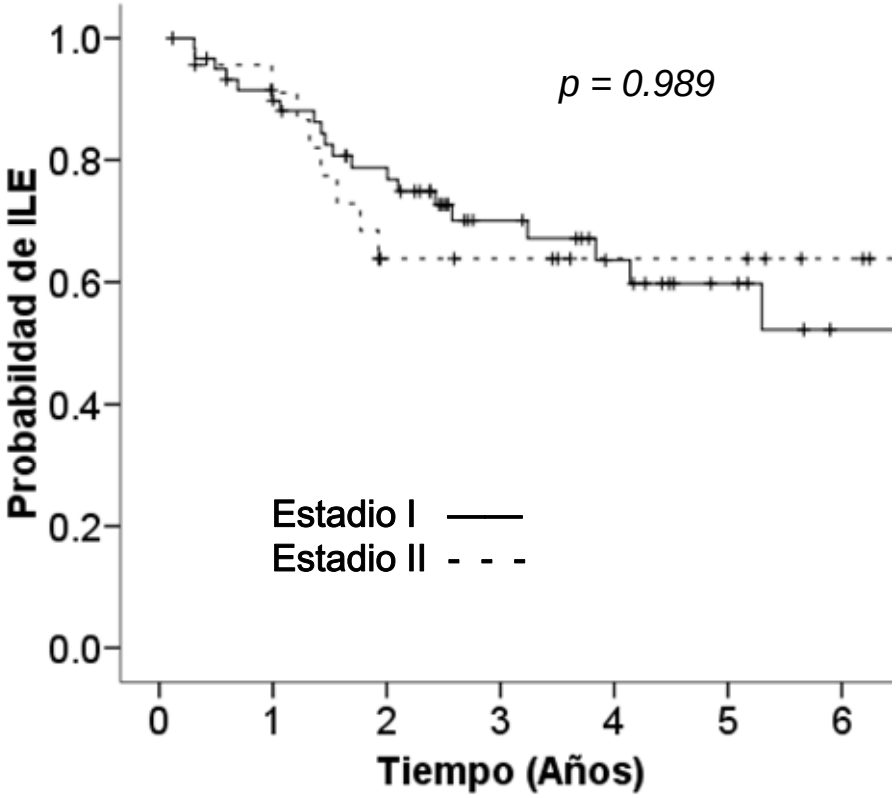

1.0
0.8
0.6
0.4
0.2
0.0
Probabilidad de ILE
p = 0.989
Estadio I
Estadio II
0
1
2
3
4
5
6
Tiempo (Años)

FIG.3

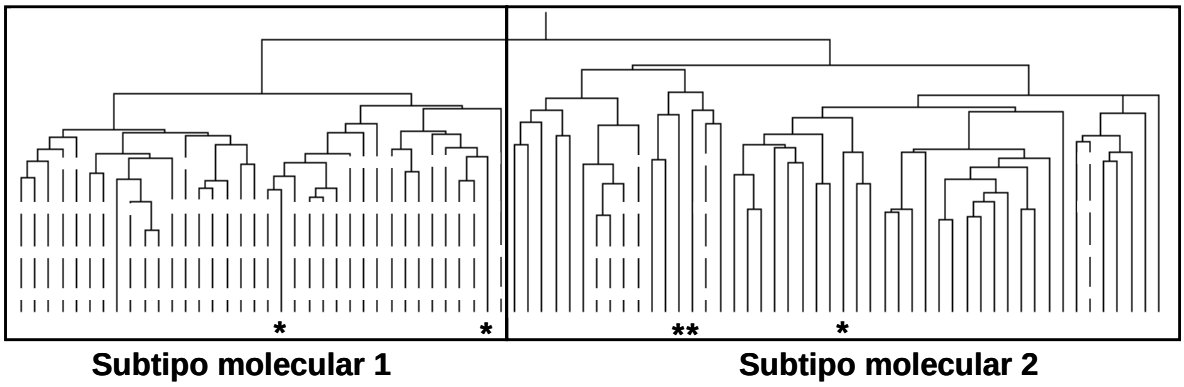

*
*
**
*
Subtipo molecular 1
Subtipo molecular 2

FIG.4

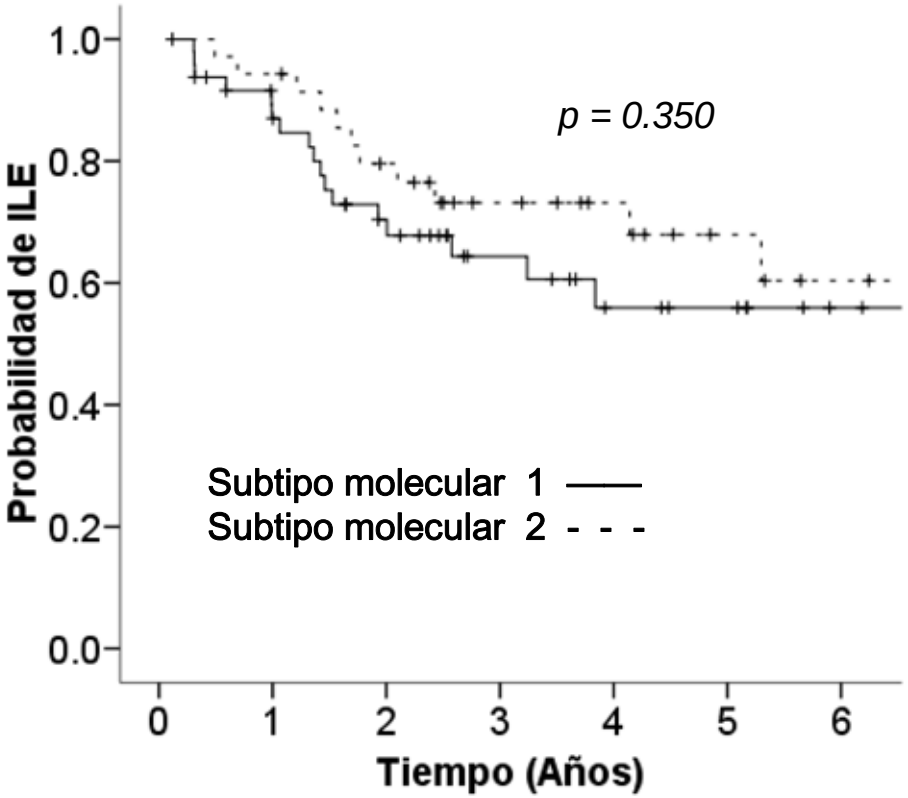

1.0
0.8
0.6
0.4
0.2
0.0
Probabilidad de ILE
p = 0.350
Subtipo molecular 1
Subtipo molecular 2
0
1
2
3
4
5
6
Tiempo (Años)

FIG.5 A

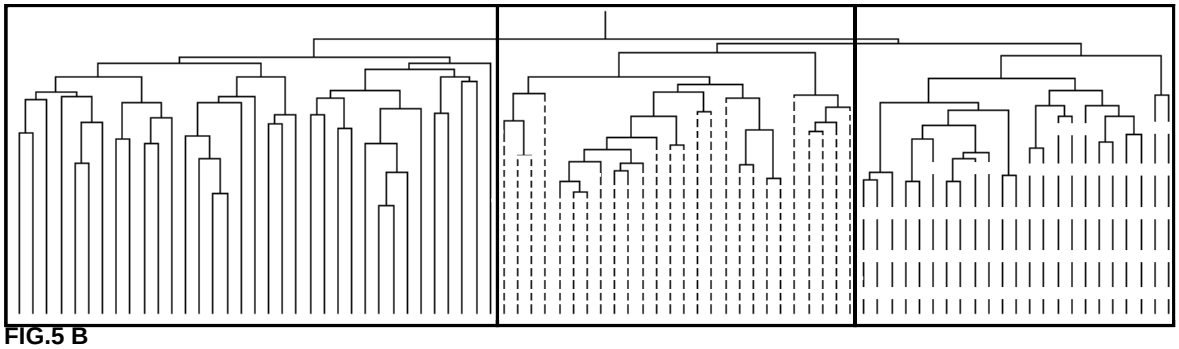


FIG.5 B

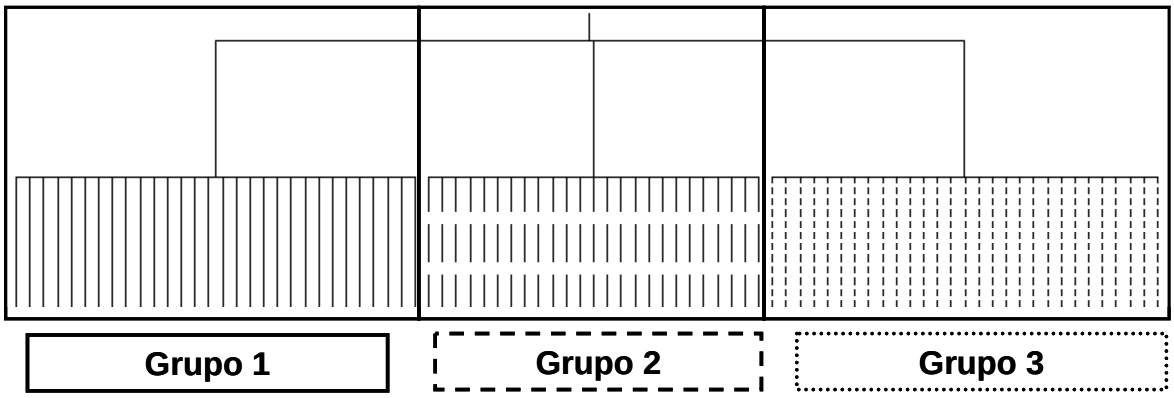


FIG 6

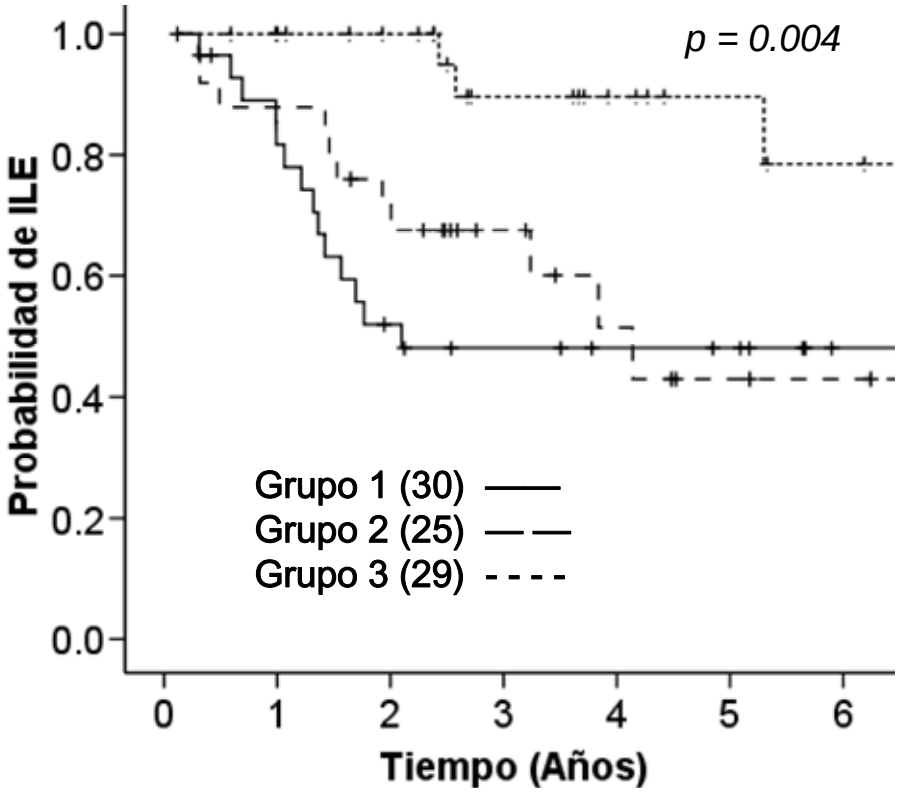


FIG.7

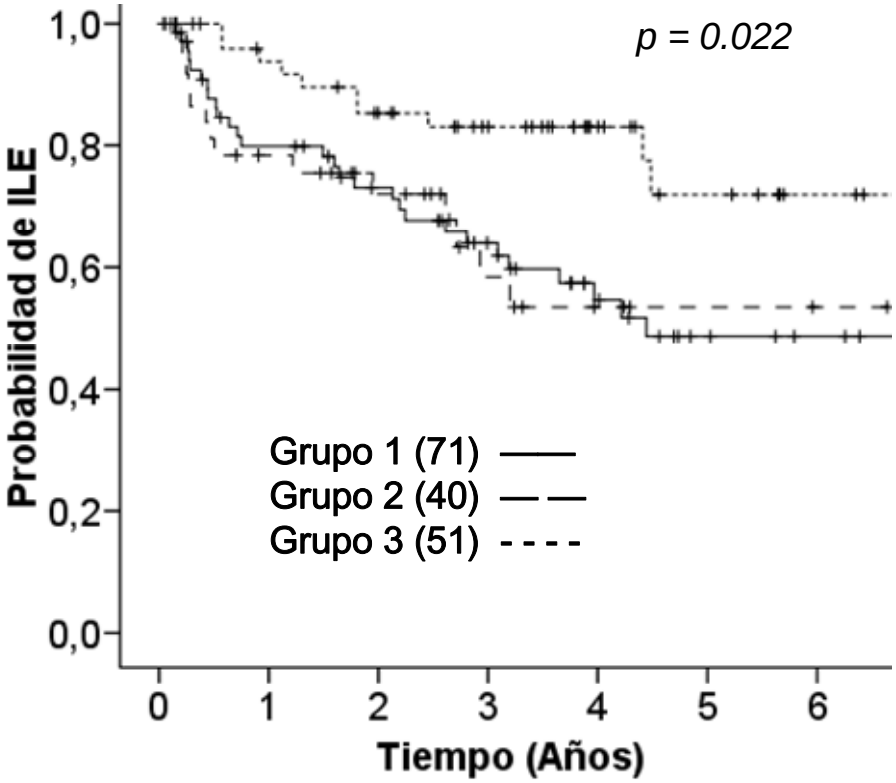

p = 0.022
Probabilidad de ILE
1,0
0,8
0,6
0,4
0,2
0,0
Grupo 1 (71)
Grupo 2 (40)
Grupo 3 (51)
0
1
2
3
4
5
6
Tiempo (Años)

FIG.8

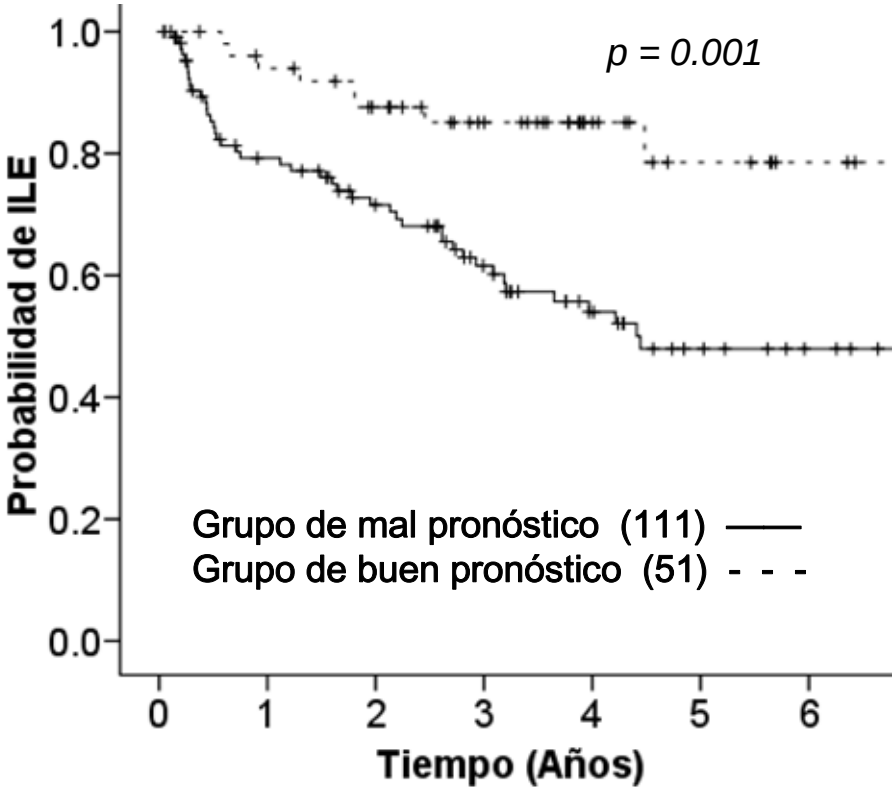

p = 0.001
Probabilidad de ILE
1.0
0.8
0.6
0.4
0.2
0.0
Grupo de mal pronóstico (111)
Grupo de buen pronóstico (51)
0
1
2
3
4
5
6
Tiempo (Años)

FIG.9 A

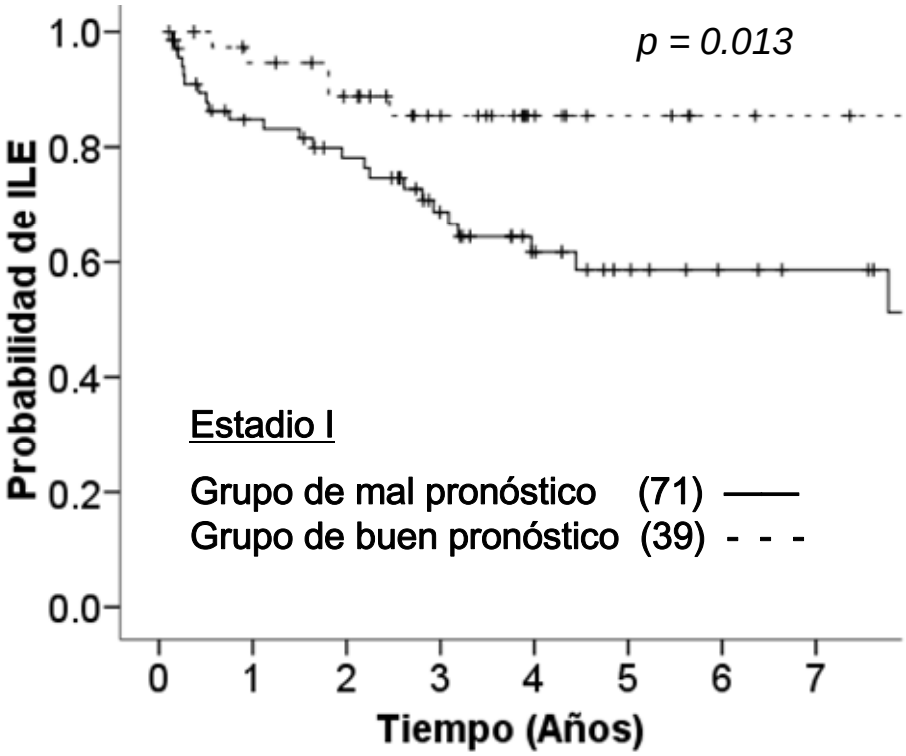
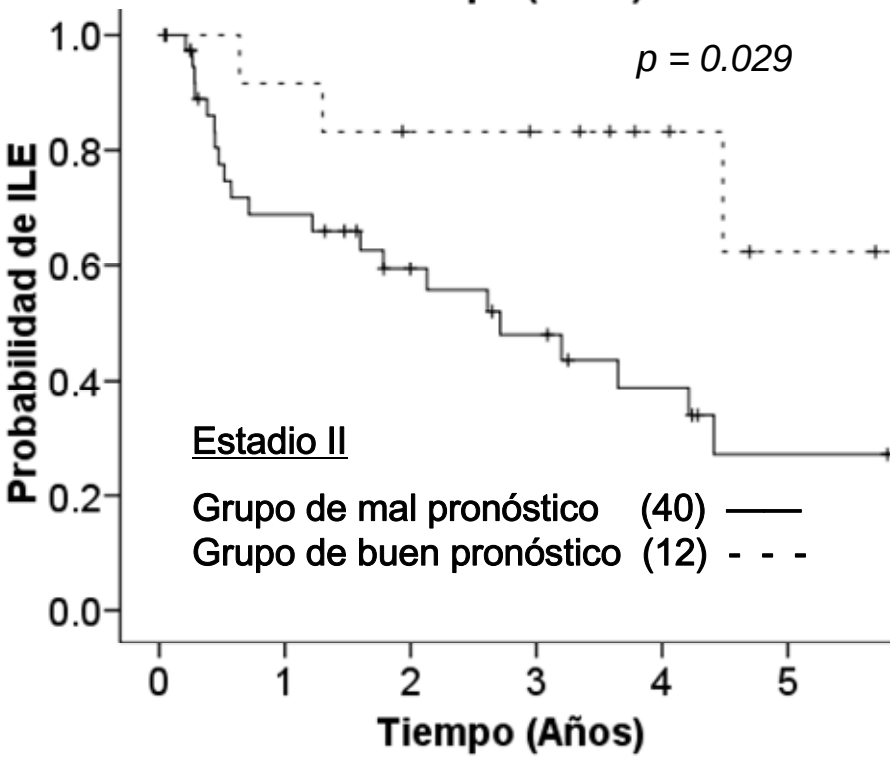

p = 0.013
Probabilidad de ILE
1.0
0.8
0.6
0.4
0.2
0.0
Estadio I
Grupo de mal pronóstico (71)
Grupo de buen pronóstico (39)
0
1
2
3
4
5
6
7
Tiempo (Años)

FIG.9 B


p = 0.029
Probabilidad de ILE
1.0
0.8
0.6
0.4
0.2
0.0
Estadio II
Grupo de mal pronóstico (40)
Grupo de buen pronóstico (12)
0
1
2
3
4
5
Tiempo (Años)