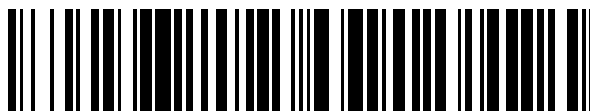


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 658 921**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **29.06.2012** **PCT/US2012/045139**

87 Fecha y número de publicación internacional: **10.01.2013** **WO13006503**

96 Fecha de presentación y número de la solicitud europea: **29.06.2012** **E 12807429 (1)**

97 Fecha y número de publicación de la concesión europea: **08.11.2017** **EP 2726635**

54 Título: **Ensayo de pronóstico multigénico para el cáncer de pulmón**

30 Prioridad:

01.07.2011 US 201161504063 P
02.07.2011 US 201161504193 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
12.03.2018

73 Titular/es:

THE REGENTS OF THE UNIVERSITY OF CALIFORNIA (50.0%)
1111 Franklin Street, 12th Floor
Oakland, CA 94607, US y
ENCORE CLINICAL, INC. (50.0%)

72 Inventor/es:

MANN, MICHAEL, J.;
JABLONS, DAVID, M.;
KRATZ, JOHANNES y
BERRYMAN, DAVID

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 658 921 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Ensayo de pronóstico multigénico para el cáncer de pulmón

5 Referencia a solicitudes relacionadas

La presente solicitud reivindica el beneficio de prioridad de la Solicitud de Patente Provisoria de Estados Unidos n.º de serie 61/504.063, presentada el 1 de julio de 2011, y la Solicitud de Patente Provisoria de Estados Unidos n.º de serie 61/504.193, presentada el 2 de julio de 2011.

10 Antecedentes de la invención

La probabilidad de una supervivencia a largo plazo para los pacientes con cáncer del pulmón está poco definida por la fase clínica y los hallazgos histopatológicos. Se ha publicado la identificación por micromatriz de genes identificados como pronósticos para el cáncer de pulmón, aunque existe una necesidad de un ensayo de reacción en cadena de la polimerasa (PCR) cuantitativo multigénico preciso que pueda predecir el riesgo de mortalidad entre los pacientes con cáncer de pulmón.

20 Breve resumen de la invención

En un aspecto, se proporciona un método para proporcionar un pronóstico para el cáncer de pulmón en un sujeto, comprendiendo el método las etapas de: (a) poner en contacto una muestra biológica del sujeto con reactivos que se unen de forma específica a un panel de biomarcadores que comprenden BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3, y (b) determinar si los marcadores se expresan de forma diferencial en la muestra o no; proporcionando de este modo un pronóstico para el cáncer de pulmón.

En una realización, el reactivo es un ácido nucleico. En otra realización, el reactivo es un oligonucleótido. En otra realización, el reactivo es un conjunto de cebadores de PCR. En otra realización, el reactivo es un anticuerpo.

En una realización, el cáncer de pulmón es cáncer de pulmón no escamoso. En otra realización, el cáncer de pulmón no escamoso está en estadio I. En otra realización, el cáncer de pulmón no escamoso está en estadio II. En otra realización, el cáncer de pulmón no escamoso está en estadio III. En otra realización, el cáncer de pulmón no escamoso está en estadio IV.

En una realización, la muestra es de tejido pulmonar o una biopsia de tumor pulmonar.

En una realización, el pronóstico proporciona una evaluación del riesgo. En algunas realizaciones, la evaluación del riesgo está basada en la mortalidad a 5 años. En algunas realizaciones, la evaluación del riesgo es una evaluación de riesgo alto, intermedio o bajo para la mortalidad a 5 años.

En un aspecto, se proporciona un kit, comprendiendo el kit reactivos que se unen de forma específica a un panel de biomarcadores que comprenden BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3. En una realización, el reactivo es un conjunto de transcriptasa inversa.

En aún otro aspecto, se proporciona un método de determinación del pronóstico de un sujeto que tiene un cáncer de pulmón mediante la medición en una muestra biológica de los niveles de metilación de un panel de biomarcadores que comprende BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3; en el que dicha muestra biológica se obtiene de dicho sujeto y dichos niveles de metilación son indicativos de dicho pronóstico.

En otro aspecto, se proporciona un informe, comprendiendo el informe un pronóstico de un sujeto que tiene un cáncer de pulmón, habiéndose determinado dicho pronóstico cuantificando en una muestra biológica del sujeto los niveles de expresión de un panel de biomarcadores que comprenden BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3; en el que dichos niveles de expresión son indicativos de dicho pronóstico.

En otro aspecto, se proporciona un método para determinar un plan de tratamiento, comprendiendo el método las etapas de: (a) poner en contacto una muestra biológica del sujeto con reactivos que se unen cada uno de forma específica a un miembro de un panel de biomarcadores que comprenden BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3, (b) determinar si los marcadores se expresan de forma diferencial en la muestra o no, (c) proporcionar un pronóstico para el cáncer de pulmón, (d) determinar una evaluación del riesgo para la mortalidad a 5 años basada en el pronóstico para el cáncer de pulmón y (e) idear un plan de tratamiento basado en la evaluación del riesgo.

En otro aspecto, se proporciona un método para proporcionar un pronóstico para el cáncer de pulmón en un sujeto, comprendiendo el método las etapas de: (a) poner en contacto una muestra biológica del sujeto con reactivos que se

- unen cada uno de forma específica a un miembro de un panel de biomarcadores que consiste en BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A, y (b) determinar si los biomarcadores se expresan de forma diferencial en la muestra o no; proporcionando de este modo un pronóstico para el cáncer de pulmón. En algunas realizaciones, la determinación de si los biomarcadores se expresan de forma diferencial en la muestra o no comprende adicionalmente normalizar los niveles de expresión de los biomarcadores con respecto a genes constitutivos seleccionados del grupo que consiste en ESD, TBP, YAP1 y cualquiera de las combinaciones de los mismos. En determinadas realizaciones, los niveles de expresión de los biomarcadores se normalizan frente al valor C_T (del inglés *cycle threshold*: ciclo umbral) promedio de los genes constitutivos. En una realización, BAG1, BRCA1, CDC6, CDK2AP1, FUT3, IL11 y RND3 indican en el sujeto una probabilidad aumentada de mortalidad y en los que ERBB3, LCK, SH3BGR y WNT3A indican en el sujeto una probabilidad disminuida de mortalidad.
- En una realización, los reactivos son ácidos nucleicos. En otra realización, los reactivos son oligonucleótidos. En otra realización, los reactivos son conjuntos de cebadores de PCR. En otra realización, los reactivos son anticuerpos.
- En una realización, el cáncer de pulmón es cáncer de pulmón no escamoso. En otra realización, el cáncer de pulmón no escamoso está en estadio I. En otra realización, el cáncer de pulmón no escamoso está en estadio II. En otra realización, el cáncer de pulmón no escamoso está en estadio III. En otra realización, el cáncer de pulmón no escamoso está en estadio IV.
- En una realización, la muestra es de un tumor extirpado de forma quirúrgica. En otra realización, la muestra procede de tejido pulmonar o de una biopsia de tumor pulmonar.
- En una realización, el pronóstico proporciona una evaluación del riesgo. En algunas realizaciones, la evaluación del riesgo está basada en la mortalidad a 5 años. En algunas realizaciones, la evaluación del riesgo es una evaluación de riesgo alto, intermedio o bajo para la mortalidad a 5 años.
- En otro aspecto, se proporciona un kit que comprende reactivos que se unen cada uno de forma específica a un miembro de un panel de biomarcadores que consiste en BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A. En determinadas realizaciones, los reactivos son conjuntos de transcriptasa inversa. En algunas realizaciones, el kit comprende adicionalmente genes constitutivos seleccionados del grupo que consiste en ESD, TBP, YAP1 y cualquier combinación de los mismos.
- En otro aspecto, se proporciona un método para proporcionar un pronóstico de un sujeto que tiene cáncer de pulmón, comprendiendo dicho método medir en una muestra biológica los niveles de metilación de un panel de biomarcadores que consisten en BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A; en el que dicha muestra biológica se obtiene de dicho sujeto y dichos niveles de metilación son indicativos de dicho pronóstico.
- En aún otro aspecto, se proporciona un informe que comprende un pronóstico de un sujeto que tiene un cáncer de pulmón, habiéndose determinado dicho pronóstico cuantificando en una muestra biológica del sujeto los niveles de expresión de un panel de biomarcadores que consisten en BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A; en el que dichos niveles de expresión son indicativos de dicho pronóstico. En algunas realizaciones, la cuantificación del sujeto de los niveles de expresión de un panel de biomarcadores en la muestra biológica comprende adicionalmente normalizar los niveles de expresión de los biomarcadores con respecto a genes constitutivos seleccionados del grupo que consiste en ESD, TBP, YAP1 y cualquiera de las combinaciones de los mismos. En determinadas realizaciones, los niveles de expresión de los biomarcadores se normalizan frente al valor C_T promedio de los genes constitutivos.
- En aún otro aspecto, se proporciona un método para determinar un plan de tratamiento, comprendiendo el método las etapas de: (a) poner en contacto una muestra biológica del sujeto con reactivos que se unen cada uno de forma específica a un miembro de un panel de biomarcadores que consiste en BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A, (b) determinar si los marcadores se expresan de forma diferencial en la muestra o no, (c) proporcionar un pronóstico para el cáncer de pulmón, (d) determinar una evaluación del riesgo para la mortalidad a 5 años basada en el pronóstico para el cáncer de pulmón, y (e) idear un plan de tratamiento basado en la evaluación del riesgo. En algunas realizaciones, la determinación de si los biomarcadores se expresan en la muestra de forma diferencial o no comprende adicionalmente normalizar los niveles de expresión de los biomarcadores con respecto a genes constitutivos seleccionados del grupo que consiste en ESD, TBP, YAP1 y cualquiera de las combinaciones de los mismos. En determinadas realizaciones, los niveles de expresión de los biomarcadores se normalizan frente al valor C_T promedio de los genes constitutivos.
- En aún otro aspecto, se proporciona un método para proporcionar un pronóstico para el cáncer de pulmón en un sujeto, comprendiendo el método las etapas de: (a) poner en contacto una muestra biológica del sujeto con reactivos que se unen cada uno de forma específica a un miembro de un panel de biomarcadores que consiste en BAG 1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A; (b) determinar una puntuación de riesgo del sujeto basada en los niveles de expresión de los biomarcadores en la muestra; y (c) proporcionar un pronóstico para el cáncer de pulmón basado en la puntuación de riesgo del sujeto. En algunas realizaciones, la

determinación de una puntuación de riesgo del sujeto basada en los niveles de expresión de los biomarcadores en la muestra comprende adicionalmente normalizar los niveles de expresión de los biomarcadores con respecto a genes constitutivos seleccionados del grupo que consiste en ESD, TBP, YAP1 y cualquiera de las combinaciones de los mismos. En determinadas realizaciones, los niveles de expresión de los biomarcadores se normalizan frente al valor C_T promedio de los genes constitutivos.

Breve descripción de las figuras

La Figura 1 muestra que el cáncer de pulmón es la causa más común de muerte por cáncer y muestra que para los cánceres en estadio 1 el pronóstico para la supervivencia a 5 años es aproximadamente del 60 %.

La Figura 2 muestra modelos genómicos de pronóstico.

La Figura 3 es un ejemplo de un análisis de supervivencia de Kaplan-Meier de pacientes de riesgo bajo, intermedio y alto, agrupados por la puntuación de riesgo utilizando el algoritmo del Ejemplo 1.

La Figura 4 es un cuadro que resume los genes de algoritmo utilizados en el algoritmo del Ejemplo 2.

La Figura 5A es un gráfico que ilustra la probabilidad de mortalidad a 5 años por la puntuación de riesgo, en donde las líneas discontinuas son intervalos de confianza (los IC) del 95 % y las líneas cortas perpendiculares al eje X son puntuaciones de riesgo individuales para cada paciente. La Figura 5B es un gráfico que ilustra el aumento de la razón de riesgos instantáneos (RRI) de la mortalidad global a 5 años por subgrupo para cada aumento gradual en la categoría de riesgo (por ejemplo, bajo a intermedio e intermedio a alto), en donde los tamaños de los cuadrados son proporcionales al tamaño del grupo y AJCC se refiere a American Joint Committee on Cancer.

La Figura 6A es un gráfico que ilustra la supervivencia global para toda la cohorte. La Figura 6B es un gráfico que ilustra la supervivencia específica para el cáncer de pulmón para toda la cohorte, en el que las muertes por cáncer que no es de pulmón están censuradas. La Figura 6C es un gráfico que ilustra la supervivencia global para 330 pacientes con enfermedad en estadio IA y IB de la American Joint Commission on Cancer, considerados como de riesgo bajo según los criterios patológicos convencionales (National Comprehensive Cancer Network).

La Figura 7 muestra gráficos que representan (A) la supervivencia global para toda la cohorte y la supervivencia en pacientes con enfermedad en (B) estadio I, (C) en estadio II y (D) en estadio III.

La Figura 8 es un diagrama de bloques que ilustra un sistema de ordenador ejemplar, en conformidad con diversas realizaciones.

La Figura 9 es un diagrama de flujo que ilustra aspectos de un método de acuerdo con una realización de la divulgación.

Descripción detallada de la invención

Introducción

La invención presenta la identificación de perfiles de expresión de determinados grupos de genes que permiten el pronóstico preciso de la mortalidad en el cáncer de pulmón en fase precoz. De forma ideal, una herramienta de pronóstico debería proporcionar una estratificación del riesgo precisa, debería ser factible desde el punto de vista clínico de emplear en la práctica cotidiana y debería ser económica. Tal ensayo sería de beneficio particular para los pacientes con CPNM no escamoso en estadio I o II extirpado de forma quirúrgica. El tratamiento habitual actual para la mayoría del CPNM no escamoso en estadio I es la lobulectomía y linfadenectomía de ganglios mediastínicos, sin quimioterapia adyuvante. Una mejor identificación de subconjuntos de pacientes con buen pronóstico podría permitir emplear procedimientos quirúrgicos menores con un potencial de supervivencia equivalente. A la inversa, los subconjuntos de estadio I con un mal pronóstico podrían seleccionarse para el tratamiento con una quimioterapia adyuvante para reducir el riesgo de metástasis a distancia utilizando los agentes de tratamiento habituales actuales. Además, los pacientes identificados con un mal pronóstico también podrían considerarse para la inclusión en ensayos clínicos que prueban estrategias novedosas y nuevos agentes terapéuticos. Teniendo en cuenta las actuales limitaciones de la quimioterapia en la enfermedad en estadio I, sería especialmente beneficioso un bioensayo que sea pronóstico y predictivo del beneficio de la quimioterapia. Por último, es probable que el CPNM no escamoso en estadio I sea de creciente importancia en el futuro. Aunque aproximadamente el 20-30 % de los pacientes actualmente diagnosticados con CPNM no escamoso son de estadio I, esta proporción probablemente crecerá debido al reciente advenimiento del crivado del cáncer del pulmón mediante tomografía computarizada.

Actualmente, a los pacientes con CPNM en estadio II se les recomienda someterse a quimioterapia adyuvante tras un intento de extirpación curativa. Sin embargo, el beneficio documentado de la quimioterapia para estos pacientes

en términos de mejora absoluta de la supervivencia a 5 años es pequeño. Como resultado, muchos pacientes renuncian a la quimioterapia, en particular a medida que se recuperan de su intento cirugía curativa. Un bioensayo que pueda asignar mejor el riesgo de recurrencia para los pacientes en estadio II puede, por lo tanto, mejorar el cumplimiento de las recomendaciones habituales actuales para la terapia adyuvante en pacientes que se encuentran en mayor riesgo de recurrencia. En un contexto experimental controlado, la terapia puede incluso suspenderse en pacientes que se encuentran en el riesgo más bajo de recurrencia incluso en el estadio II.

En una realización descrita en el presente documento, se desarrolló un ensayo basado en los patrones de expresión de 426 pacientes que se habían sometido a la extracción de un cáncer de pulmón escamoso (CPNM) de estadio I-IV en la Universidad de California, San Francisco. Se extrajo ARN de muestras de tejidos FFIP (fijados en formalina e incluidos en parafina) y se evaluaron los niveles de expresión para 11 genes diana relacionados con el pronóstico del paciente. Se buscó un ensayo que tendiese a asignar una puntuación de riesgo más alta en promedio a los pacientes que hubiesen fallecido que a los que habían sobrevivido a un período de seguimiento de 5 años. Los pacientes cuyas muestras recibieron las puntuaciones de riesgo más altas se considerarían en mayor riesgo de fallecer dentro de un período de 5 años después de la operación, mientras que los pacientes cuyas muestras recibieron una puntuación baja tendrían más probabilidades de haber sobrevivido durante este intervalo de tiempo después de sus operaciones.

El ensayo pronóstico se desarrolló correlacionando los patrones de expresión de BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR, WNT3A, que están relacionados con el pronóstico del paciente (en particular porque el pronóstico se relaciona con los resultados de supervivencia global a 5 años) utilizando el modelo de riesgos instantáneos proporcionales de Cox. El ensayo pronóstico proporciona un pronóstico para el cáncer de pulmón en un sujeto determinando en una muestra la expresión de BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A. La selección de cada uno de un panel de biomarcadores que incluye BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A, ofrece una gran mejora con respecto a la técnica anterior. El documento WO2008/151072 (The Regents of the University of California) describe métodos para el pronóstico de pacientes de cáncer de pulmón que están basados en el análisis de expresión de un conjunto de biomarcadores.

Después, se obtiene para cada paciente una puntuación del riesgo insertando los niveles de expresión de cada uno de los 11 genes pronósticos en un algoritmo de puntuación de riesgo. Los grupos de riesgo también se describen en el presente documento a base de estas puntuaciones de riesgo, colocando a los pacientes en distintas categorías de riesgo de acuerdo con su puntuación de riesgo. Por ejemplo, "Riesgo Bajo", "Riesgo Intermedio" o "Riesgo Alto".

Adicionalmente, se describe en el presente documento un kit de diagnóstico multigénico, compuesto de los marcadores descritos en el presente documento, que pueden utilizarse para proporcionar un pronóstico para pacientes de cáncer de pulmón, y de un informe que comprende un pronóstico de un sujeto que tiene cáncer de pulmón mediante la cuantificación de los niveles de expresión de los marcadores descritos en el presente documento.

Definiciones

"Cáncer de pulmón" se refiere en general a dos tipos principales de cáncer de pulmón categorizados por el tamaño y el aspecto de las células malignas: cáncer de pulmón no microcítico (aproximadamente el 80 % de los casos) y microcítico (aproximadamente el 20 % de los casos). "Cáncer de pulmón no microcítico" (CPNM) incluye al carcinoma escamoso. El adenocarcinoma de pulmón es el subtipo más común de CPNM y otros subtipos de cáncer de pulmón incluyen al carcinoma bronquioalveolar, el carcinoma de células grandes, el tumor carcinoide, el carcinoma adenoide quístico, el cilindroma y el carcinoma mucoepidermoide. En una realización, los cánceres de pulmón se estadifican de acuerdo con los estadios I-IV, siendo el I un estadio temprano y siendo el IV el más avanzado.

"Pronóstico" se refiere, por ejemplo, a la supervivencia global, la mortalidad a largo plazo y la supervivencia sin enfermedad. En una realización, mortalidad a largo plazo se refiere al fallecimiento dentro de los 5 años después del diagnóstico del cáncer de pulmón.

"Evaluación de riesgo" se refiere al riesgo relativo que afronta un individuo con respecto a la mortalidad. Por ejemplo, un pronóstico que proporciona una evaluación de alto riesgo para la mortalidad a 5 años, tiene una mayor probabilidad de mortalidad dentro de los 5 años que un individuo que tiene una evaluación de bajo riesgo para la mortalidad a 5 años. En una realización, el pronóstico para la mortalidad a largo plazo es "riesgo alto", por ejemplo, riesgo alto de mortalidad, "riesgo intermedio", por ejemplo, riesgo intermedio de mortalidad o "riesgo bajo", por ejemplo, riesgo bajo de mortalidad. El estadio del cáncer y el pronóstico pueden utilizarse para ajustar una terapia para el paciente que proporcione un mejor resultado, por ejemplo, terapia sistémica y cirugía, solo cirugía o terapia sistémica sola. La evaluación del riesgo puede dividirse como se desee, por ejemplo, en la mediana, en grupos terciarios, grupos cuaternarios, etc.

Otras formas de cáncer incluyen carcinomas, sarcomas, adenocarcinomas, linfomas, leucemias, etc., incluyendo los cánceres sólidos y linfocíticos, el cáncer de cabeza y cuello, por ejemplo, el cáncer de cavidad oral, faríngeo o de lengua, el cáncer de riñón, mama, riñón, vejiga, colon, ovario, próstata, páncreas, estómago, cerebro, cabeza y cuello, piel, uterino, de testículo, de esófago y de hígado, incluyendo hepatocarcinoma, linfoma, incluyendo linfomas no Hodgkin (por ejemplo, linfomas de Burkitt, microcítico y células grandes) y el linfoma de Hodgkin, leucemia y mieloma múltiple.

El término “marcador” se refiere a una molécula (normalmente una proteína, ácido nucleico, hidrato de carbono o lípido) que se expresa en la célula, expresada en la superficie de una célula cancerosa o secretada por una célula cancerosa, en comparación con una célula no cancerosa, y que es útil para el diagnóstico del cáncer, para proporcionar un pronóstico y para direccionar de forma preferencial un agente farmacológico hacia la célula cancerosa. A menudo, tales marcadores son moléculas que se sobreexpresan en un cáncer de pulmón u otras células cancerosas, en comparación con una célula no cancerosa, por ejemplo, una sobreexpresión de 1 vez, una sobreexpresión de 2 veces, una sobreexpresión de 3 veces o más, en comparación con una célula normal. Adicionalmente, un marcador puede ser una molécula que se sintetiza de forma inapropiada en la célula cancerosa, por ejemplo, una molécula que contenga delecciones, adiciones o mutaciones, en comparación con la molécula expresada en una célula normal. Como alternativa, tales biomarcadores son moléculas que se subexpresan en una célula cancerosa, en comparación con una célula no cancerosa, por ejemplo, una subexpresión de 1 vez, una subexpresión de 2 veces, una subexpresión de 3 veces o más. Adicionalmente, un marcador puede ser una molécula que se sintetiza de forma inapropiada en un cáncer, por ejemplo, una molécula que contiene delecciones, adiciones o mutaciones, en comparación con la molécula expresada en una célula normal.

El experto en la materia entenderá que los marcadores pueden utilizarse en combinación con otros marcadores o pruebas para cualquiera de los usos, por ejemplo, la predicción, el diagnóstico o el pronóstico del cáncer, divulgados en el presente documento.

“Muestra biológica” incluye cortes de tejidos tales como muestras de biopsia y autopsia, y cortes congelados tomados con fines histológicos. Tales muestras incluyen sangre y fracciones o productos hematológicos (por ejemplo, suero, plaquetas, glóbulos rojos y similares), esputo, lavado broncoalveolar, células cultivadas, por ejemplo, cultivos primarios, explantes y células transformadas, heces, orina, etc. Una muestra biológica se obtiene normalmente de un organismo eucariota, más preferentemente un mamífero tal como un primate, por ejemplo, un chimpancé o ser humano; vaca; perro; gato; un roedor, por ejemplo, cobaya, rata, ratón; conejo o un ave; reptil o pez.

Una “biopsia” se refiere al procedimiento de extraer una muestra de tejido para la evaluación diagnóstica o pronóstica, y a la propia muestra de tejido de ensayo. Para los métodos de diagnóstico y pronóstico de la presente invención puede aplicarse en la técnica cualquier técnica de biopsia conocida. La técnica de biopsia aplicada dependerá del tipo de tejido a evaluar (por ejemplo, pulmón, etc.), el tamaño y tipo del tumor, entre otros factores. Las técnicas de biopsia representativas incluyen, pero sin limitación, biopsia por escisión, biopsia por incisión, biopsia por punción, biopsia quirúrgica y biopsia de médula ósea. Una “biopsia por escisión” se refiere a la extracción de toda la masa tumoral con un pequeño margen de tejido normal que la rodee. Una “biopsia por incisión” se refiere a la extracción de tejido en cuña de dentro del tumor. Un diagnóstico o pronóstico hecho mediante la orientación por endoscopia o radiográfica requiere “biopsia biopsia por punción con aguja gruesa” o una “punción biopsia con aguja fina”, que en general obtiene una suspensión de células de dentro de un tejido diana. Las técnicas de biopsia se discuten, por ejemplo, en Harrison's Principles of Internal Medicine, Kasper, *et al.*, eds., 16^a ed., 2005, Capítulo 70 y en toda la Parte V.

Los términos “sobreexpresar”, “sobreexpresión” o “sobreexpresado” se refieren de forma indistinta a una proteína o ácido nucleico (ARN) que se traduce o transcribe a un nivel detectablemente mayor, habitualmente en una célula cancerosa, en comparación con una célula normal. Los términos incluyen la sobreexpresión debida a la transcripción, el procesamiento postranscripcional, la traducción, el procesamiento postraduccional, la localización celular (por ejemplo, orgánulo, citoplasma, núcleo, superficie celular) y la estabilidad del ARN y las proteínas, en comparación con una célula normal. La sobreexpresión puede detectarse utilizando técnicas convencionales para la detección de ARNm (es decir, RT-PCR, PCR, hibridación) o proteínas (es decir, ELISA, técnicas inmunohistoquímicas). La sobreexpresión puede ser del 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % o más en comparación con una célula normal. En determinados casos, la sobreexpresión es de 1 vez, 2 veces, 3 veces, 4 veces o más niveles de transcripción o traducción mayor, en comparación con una célula normal.

Los términos “subexpresar”, “subexpresión” o “subexpresado”, o “regulado de forma negativa” se refieren indistintamente a una proteína o ácido nucleico que se traduce o transcribe en una célula cancerosa a un nivel detectablemente inferior, en comparación con una célula normal. El término incluye la subexpresión debida a la transcripción, el procesamiento postranscripcional, la traducción, el procesamiento postraduccional, la localización celular (por ejemplo, orgánulo, citoplasma, núcleo, superficie celular) y la estabilidad del ARN y las proteínas, en comparación con un control. La subexpresión puede detectarse utilizando técnicas convencionales para la detección de ARNm (es decir, RT-PCR, PCR, hibridación) o proteínas (es decir, ELISA, técnicas inmunohistoquímicas). La subexpresión puede ser del 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % o menos, en comparación con

un control. En determinados casos, la subexpresión es 1 vez, 2 veces, 3 veces, 4 veces o más niveles de transcripción o traducción más baja, en comparación con un control.

La expresión "expresado de forma diferencial" o "regulado de forma diferencial" se refiere en general a una proteína o ácido nucleico que se sobreexpresa (se regula de forma positiva) o se subexpresa (se regula de forma negativa) en una muestra, en comparación con al menos una otra muestra, en general un paciente de cáncer, comparado con una muestra de tejido no canceroso en el contexto de la presente invención.

"Tratamiento terapéutico" y "terapias para el cáncer" se refieren a la quimioterapia, terapia hormonal, radioterapia, inmunoterapia y terapia biológica (dirigida).

Por "cantidad o dosis terapéuticamente eficaz" o "cantidad o dosis suficiente" se entiende una dosis que produce los efectos para los cuales se administra. La dosis exacta dependerá del fin del tratamiento y la determinará un experto en la materia utilizando técnicas conocidas (véase, por ejemplo, Lieberman, *Pharmaceutical Dosage Forms* (vols. 1-3, 1992); Lloyd, *The Art, Science and Technology of Pharmaceutical Compounding* (1999); Pickar, *Dosage Calculations* (1999) y Remington: *The Science and Practice of Pharmacy*, 20ª Edición, 2003, Gennaro, Ed., Lippincott, Williams & Wilkins).

"Ácido nucleico" se refiere a desoxirribonucleótidos o ribonucleótidos, y polímeros de los mismos, en forma mono o bicatenaria y a sus complementarios. El término abarca ácidos nucleicos que contienen análogos de nucleótido conocidos, o restos o uniones del esqueleto modificados, que son sintéticos, de origen natural o de origen no natural, que tienen propiedades de unión similares al ácido nucleico de referencia y que se metabolizan de forma similar a los nucleótidos de referencia. Los ejemplos de tales análogos incluyen, sin limitación, fosforotioatos, fosforamidatos, metilfosfonatos, metilfosfonatos quirales, 2-O-metil ribonucleótidos, ácidos nucleicos peptídicos (los ANP).

A menos que se indique otra cosa, una secuencia de ácido nucleico particular también abarca de forma implícita variantes modificadas de forma conservativa de la misma (por ejemplo, sustituciones de codones degenerados) y las secuencias complementarias, así como la secuencia indicada de forma explícita. En concreto, las sustituciones de codones degenerados puede conseguirse generando secuencias en las que la tercera posición de uno o más codones seleccionados (o todos) se sustituye con restos de bases mixtas y/o de desoxiinosina (Batzer *et al.*, *Nucleic Acid Res.* 19: 5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260: 2605 - 2608 (1985), Rossolini *et al.*, *Mol. Cell. Probes* 8: 91-98 (1994)). La expresión ácido nucleico se utiliza indistintamente con un gen, ADNc, ARNm, oligonucleótido y polinucleótido.

Una secuencia de ácido nucleico particular también abarca de forma implícita "variantes de corte y empalme" y secuencias de ácido nucleico que codifican formas truncadas de una proteína. De forma similar, una proteína particular codificada por un ácido nucleico abarca de forma implícita cualquier proteína codificada por una variante de corte y empalme, o una forma truncada de un ácido nucleico. "Variantes de corte y empalme", como sugiere el nombre, son productos del corte y empalme alternativo de un gen. Después de la transcripción, un transcrito de ácido nucleico inicial puede cortarse y empalmarse de forma que los productos de corte y empalme distintos (o alternos) del ácido nucleico codifican distintos polipéptidos. Los mecanismos para la producción de variantes de corte y empalme varían, pero incluyen el corte y empalme alterno de los exones. Los polipéptidos alternos obtenidos del mismo ácido nucleico por transcripción más allá del codón finalizador también están abarcados en esta definición. Cualquiera de los productos de una reacción de corte y empalme, incluyendo las formas recombinantes de los productos de corte y empalme, están incluidos en esta definición. Los ácidos nucleicos pueden estar truncados en el extremo 5' o en el extremo 3'. Los polipéptidos pueden estar truncados en el extremo N terminal o el extremo C terminal. Las versiones truncadas de secuencias de ácido nucleico o polipéptido pueden ser de origen natural o crearse de forma recombinante.

Los términos "polipéptido", "péptido" y "proteína" se utilizan indistintamente en el presente documento para referirse a un polímero de restos de aminoácido. Los términos se aplican a polímeros de aminoácidos en los que uno o más restos de aminoácido son un mimético químico artificial de un correspondiente aminoácido de origen natural, así como a polímeros de aminoácidos de origen natural y polímeros de aminoácidos que no son de origen natural.

El término "aminoácido" se refiere a aminoácidos de origen natural o sintéticos, así como a análogos de aminoácido o miméticos de aminoácido que actúan de una manera similar a los aminoácidos de origen natural. Los aminoácidos de origen natural son los que codifica el código genético, así como los aminoácidos que se modifican posteriormente, por ejemplo, hidroxiprolina, γ -carboxiglutamato y O-fosfoserina. Análogos de aminoácido se refiere a compuestos que tienen la misma estructura química básica que el aminoácido de origen natural, es decir, un carbono que está unido a un hidrógeno, un grupo carboxilo, un grupo amino y un grupo R, por ejemplo, homoserina, norleucina, sulfóxido de metionina, metil sulfonio de metionina. Tales análogos tienen grupos R modificados (por ejemplo, norleucina) o los esqueletos peptídicos modificados, pero conservan la misma estructura química básica que un aminoácido de origen natural. Miméticos de aminoácido se refiere a compuestos químicos que tienen una estructura que es distinta de la estructura química general de un aminoácido, pero que actúa de una manera similar a un aminoácido de origen natural.

Los aminoácidos pueden denominarse en el presente documento por sus símbolos de tres letras comúnmente conocidos o por los símbolos de una letra recomendados por la Comisión de Nomenclatura Bioquímica IUPAC-IUB. Asimismo, los nucleótidos pueden denominarse por sus códigos de letra única comúnmente aceptados.

5 “Variantes modificadas de forma conservativa” se aplica a las secuencias de aminoácidos y de ácidos nucleicos. Con respecto a las secuencias de ácidos nucleico particulares, variantes modificadas de forma conservativa se refiere a los ácidos nucleicos que codifican secuencias de aminoácidos idénticas o esencialmente idénticas o, cuando el ácido nucleico no codifica una secuencia de aminoácidos, para secuencias esencialmente idénticas. Debido a la degeneración del código genético, un gran número de ácidos nucleicos funcionalmente idénticos
10 codifican una dada proteína. Por ejemplo, todos los codones GCA, GCC, GCG y GCU codifican el aminoácido alanina. Por lo tanto, en cada posición en donde una alanina está especificada por un codón, el codón puede modificarse a cualquiera de los codones correspondientes descritos sin alterar el polipéptido codificado. Tales variaciones de ácido nucleico son “variaciones silenciosas”, que son una especie de variaciones modificadas de forma conservativa. En el presente documento, cada secuencia de ácido nucleico que codifica un polipéptido
15 también describe cada posible variación silenciosa del ácido nucleico. Un experto en la materia reconocerá que cada codón en un ácido nucleico (excepto AUG, que habitualmente es el único codón para la metionina, y TGG, que habitualmente es el único codón para el triptófano) puede modificarse para producir una molécula funcionalmente idéntica. Por consiguiente, cada variación silenciosa de un ácido nucleico que codifica un polipéptido está implícita en cada secuencia descrita, con respecto al producto de expresión pero no con respecto a las secuencias de sondas
20 reales.

Como para las secuencias de aminoácidos, un experto en la materia reconocerá que las sustituciones, deleciones o adiciones individuales en una secuencia de ácido nucleico, péptido, polipéptido o proteína, que modifique, añada o delecione un aminoácido individual o un pequeño porcentaje de aminoácidos en la secuencia codificada, es una
25 “variante modificada de forma conservativa” en donde la modificación da como resultado la sustitución de un aminoácido con un aminoácido similar desde el punto de vista química. Son bien conocidas en la técnica las tablas de sustituciones conservativas que proporcionan aminoácidos similares desde el punto de vista funcional. Tales variantes modificadas de forma conservativa se suman a, y no excluyen, variantes polimórficas, homólogos interespecie y alelos de la invención.

30 Los siguientes ocho grupos contienen cada uno aminoácidos que son sustituciones conservativas el uno para el otro: 1) Alanina (A), Glicina (G); 2) Ácido aspártico (D), Ácido glutámico (E); 3) Asparragina (N), Glutamina (Q); 4) Arginina (R), Lisina (K); 5) Isoleucina (I), Leucina (L), Metionina (M), Valina (V); 6) Fenilalanina (F), Tirosina (Y), Triptófano (W); 7) Serina (S), Treonina (T) y 8) Cisteína (C), Metionina (M). Véase, por ejemplo, Creighton, *Proteins*
35 (1984).

Un “marcador” o una “fracción detectable” es una composición detectable por medios espectroscópicos, fotoquímicos, bioquímicos, inmunoquímicos, químicos u otros medios físicos. Por ejemplo, los marcadores útiles incluyen ³²P, colorantes fluorescentes, reactivos electrodenso, enzimas (por ejemplo, como se utiliza habitualmente
40 en un ELISA), biotina, digoxigenina o haptenos, y proteínas que pueden fabricarse para que sean detectables, por ejemplo, incorporando un radiomarcador en el péptido, o utilizarse para detectar anticuerpos reactivos de forma específica con el péptido.

El término “recombinante” cuando se utiliza en referencia, por ejemplo, a una célula o a un ácido nucleico, proteína o vector, indica que la célula, ácido nucleico, proteína o vector, se ha modificado mediante la introducción de un ácido nucleico heterólogo o proteína, o por la alteración de un ácido nucleico o proteína nativo, o que la célula se obtiene de una célula así modificada. Por lo tanto, por ejemplo, las células recombinantes expresan genes que no se encuentran dentro de la forma nativa (no recombinante) de la célula o expresan genes nativos que, de otro modo, se expresan de forma anómala, se subexpresan o no se expresan en absoluto.
50

La frase “condiciones rigurosas de hibridación” se refiere a condiciones en las que una sonda hibridará con su subsecuencia diana, normalmente en una mezcla compleja de ácidos nucleicos, pero no con otras secuencias. Las condiciones rigurosas son dependientes de secuencia y serán distintas en circunstancias distintas. Las secuencias más largas hibridan de forma específica a temperaturas más altas. Una guía extensa para la hibridación de ácidos nucleicos se encuentra en Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, “Overview of principles of hybridization and the strategy of nucleic acid assays” (1993). En general, las condiciones rigurosas se seleccionan para que sean aproximadamente 5-10 °C menores que el punto de fusión térmico (T_m) para la secuencia específica a una fuerza iónica definida pH. La T_m es la temperatura (en una fuerza iónica definida, pH y concentración de ácidos nucleicos) a la que el 50 % de las sondas complementarias con la diana hibridan con la secuencia diana en el equilibrio (dado que las secuencias diana están presentes en exceso, a la T_m, en el equilibrio el 50 % de las sondas están ocupadas). Las condiciones rigurosas también pueden lograrse con la adición de agentes desestabilizantes tales como formamida. Para la hibridación selectiva o específica, una señal positiva es al menos dos veces el fondo, preferentemente 10 veces la hibridación de fondo. Las condiciones de hibridación rigurosa ejemplares pueden ser las siguientes: formamida al 50 %, SSC 5x y SDS al 1 %, incubando a
60 42 °C, o SSC 5x, SDS al 1 %, incubando a 65 °C, con un lavado en SSC 0,2x y SDS al 0,1 % a 65 °C.
65

Los ácidos nucleicos que no hibridan entre sí en condiciones rigurosas son todavía sustancialmente idénticos si los polipéptidos que codifican son sustancialmente idénticos. Esto se produce, por ejemplo, cuando se crea una copia de un ácido nucleico utilizando la degeneración de codón máxima permitida por el código genético. En tales casos, los ácidos nucleicos normalmente hibridan en condiciones de hibridación moderadamente rigurosas. Las

5 “condiciones de hibridación moderadamente rigurosas” ejemplares incluyen una hibridación en un tampón de formamida al 40 %, NaCl 1 M, SDS al 1 % a 37 °C y un lavado en SSC 1X a 45 °C. Una hibridación positiva es al menos dos veces el fondo. Los expertos en la materia reconocerán fácilmente que pueden utilizarse condiciones de hibridación y lavado alternativas para proporcionar condiciones de rigurosidad similar. Las directrices adicionales para determinar los parámetros de hibridación se proporcionan en numerosas referencias, por ejemplo, en Current

10 Protocols in Molecular Biology, ed. Ausubel, *et al.*, citado anteriormente.

Para la PCR, una temperatura de aproximadamente 36 °C es típica para la amplificación con baja rigurosidad, aunque las temperaturas de apareamiento pueden variar entre aproximadamente 32 °C y 48 °C, dependiendo de la longitud del cebador. Para una amplificación por PCR de rigurosidad alta, es típica una temperatura de

15 aproximadamente 62 °C, aunque las temperaturas de apareamiento de alta rigurosidad pueden variar de aproximadamente 50 °C a aproximadamente 65 °C, dependiendo de la longitud y especificidad del cebador. Las condiciones típicas de ciclo, tanto para las amplificaciones de alta como baja rigurosidad, incluyen una fase de desnaturalización de 90 °C-95 °C durante 30 s - 2 min, una fase de apareamiento que dura 30 s - 2 min y una fase de extensión de aproximadamente 72 °C durante 1 - 2 min. Los protocolos y directrices para las reacciones de

20 amplificación de rigurosidad baja y alta se proporcionan, por ejemplo, en Innis *et al.* (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, Inc. N. Y.

“Anticuerpo” se refiere a un polipéptido que comprende una región marco conservada procedente de un gen de inmunoglobulina o fragmentos del mismo, que se une de forma específica y reconoce a un antígeno. Los genes de

25 inmunoglobulina identificados incluyen los genes de la región constante kappa, lambda, alfa, gamma, delta, épsilon y mu, así como la gran cantidad de genes de la región variable de inmunoglobulina. Las cadenas ligeras se clasifican ya sea como kappa o lambda. Las cadenas pesadas se clasifican como gamma, mu, alfa, delta o épsilon, que a su vez definen las clases de inmunoglobulina, IgG, IgM, IgA, IgD e IgE, respectivamente. Normalmente, la región de unión al antígeno de un anticuerpo será la más crítica en la especificidad y afinidad de unión. Los anticuerpos

30 pueden ser policlonales o monoclonales, obtenerse de suero, un hibridoma o pueden clonarse de forma recombinante, y también pueden ser quiméricos, primatizados o humanizados.

Una unidad estructural de inmunoglobulina (anticuerpo) ejemplar comprende un tetrámero. Cada tetrámero está compuesto de dos parejas idénticas de cadenas polipeptídicas, teniendo cada pareja una cadena “ligera”

35 (aproximadamente 25 kDa) y una “pesada” (aproximadamente 50-70 kDa). El extremo N de cada cadena define una región variable de aproximadamente 100 a 110, o más, aminoácidos responsables principalmente del reconocimiento del antígeno. Las expresiones cadena ligera variable (V_L) y cadena pesada variable (V_H) se refieren a estas cadenas ligera y pesada respectivamente.

Los anticuerpos existen, por ejemplo, como inmunoglobulinas intactas o como una serie de fragmentos bien caracterizados producidos mediante digestión con diversas peptidasas. Por lo tanto, por ejemplo, la pepsina digiere un anticuerpo por debajo de las uniones disulfuro en la región bisagra para producir $F(ab)_2$, un dímero de Fab que es él mismo una cadena ligera unida a V_H-C_H1 por un enlace disulfuro. El $F(ab)_2$ puede reducirse en condiciones

45 suaves hasta romper el enlace disulfuro en la región bisagra, convirtiendo de este modo el dímero $F(ab)_2$ en un monómero Fab'. El monómero Fab' es esencialmente el Fab con parte de la región bisagra (véase Fundamental Immunology (Paul Ed., 3ª ed., 1993). Aunque se definen varios fragmentos de anticuerpo en términos de la digestión de un anticuerpo intacto, un experto en la materia apreciará que tales fragmentos pueden sintetizarse *de novo* ya sea de forma química o utilizando la metodología de ADN recombinante. Por lo tanto, el término anticuerpo, como se usa en el presente documento, incluye también a los fragmentos de anticuerpo producidos por la modificación de

50 anticuerpos completos o los sintetizados *de novo* utilizando las metodologías de ADN recombinante (por ejemplo, Fv monocatenario) o los identificados utilizando bibliotecas de presentación en fagos (véase, por ejemplo, McCafferty *et al.*, Nature 348: 552-554 (1990)).

En una realización, el anticuerpo se conjuga a una fracción “efectora”. La fracción efectora puede ser un número

55 variado de moléculas, incluyendo fracciones de marcaje tales como marcadores radiactivos o marcadores fluorescentes, o puede ser una fracción terapéutica. En un aspecto, el anticuerpo modula la actividad de la proteína.

Los ácidos nucleicos de los genes expresados de forma diferencial de la presente invención o los polipéptidos que codifican, se refiere a todas las formas de ácidos nucleicos (por ejemplo, pre-ARNm, ARNm) o proteínas, sus

60 variantes polimórficas, alelos, mutantes y homólogos interespecie que (según corresponda a ácido nucleico o proteína): (1) tengan una secuencia de aminoácidos que tenga más de aproximadamente el 60 % de identidad de secuencia de aminoácidos, el 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, preferentemente el 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % o más identidad de secuencia de aminoácidos, preferentemente a lo largo de una región de al menos aproximadamente 25, 50, 100, 200, 500, 1000 o más aminoácidos, con un polipéptido codificado

65 por un ácido nucleico al que se hace referencia o con una secuencia de aminoácidos descrita en el presente documento; (2) se unan de forma específica a anticuerpos, por ejemplo anticuerpos policlonales, generados frente a

un inmunógeno que comprende una secuencia de aminoácidos a la que se hace referencia, a fragmentos inmunogénicos de la misma y variantes modificadas de forma conservativa de la misma; (3) hibridan de forma específica en condiciones rigurosas de hibridación con un ácido nucleico que codifica una secuencia de aminoácidos a la que se hace referencia y con las variantes modificadas de forma conservativa del mismo; (4) tienen una

5 secuencia de ácido nucleico que tiene más de aproximadamente el 95 %, preferentemente más de aproximadamente el 96 %, 97 %, 98 %, 99 % o más de identidad de secuencia de nucleótidos, preferentemente a lo largo de una región de al menos aproximadamente 25, 50, 100, 200, 500, 1000 o más nucleótidos, con una

10 secuencia de ácido nucleico a la que se hace referencia. Una secuencia de polinucleótido o polipéptido procede normalmente de un mamífero, incluyendo, pero sin limitación, un primate, por ejemplo, un ser humano; roedor, por ejemplo, rata, ratón, hámster; vaca, cerdo, caballo, oveja o cualquier mamífero. Los ácidos nucleicos y proteínas de la invención incluyen moléculas tanto de origen natural como recombinantes. Las formas truncadas y cortadas y empalmadas de forma alternativa de estos antígenos están incluidas en la definición.

La frase “que se une de forma específica (o de forma selectiva)”, cuando se hace referencia a una proteína, ácido

15 nucleico, anticuerpo o compuesto de molécula pequeña, se refiere a una reacción de unión que es determinativa de la presencia de la proteína o ácido nucleico, tal como los genes expresados de forma diferencial de la presente invención, a menudo en una población heterogénea de proteínas o ácidos nucleicos, u otros productos biológicos. En el caso de los anticuerpos, en las condiciones de inmunoensayo designadas, un anticuerpo especificado puede unirse a una proteína particular al menos dos veces el fondo y, de forma más típica, más de 10 a 100 veces el fondo.

20 La unión específica a un anticuerpo en tales condiciones requiere un anticuerpo que se selecciona por su especificidad por una proteína particular. Por ejemplo, los anticuerpos policlonales pueden seleccionarse para obtener solo los anticuerpos policlonales que sean inmunorreactivos de forma específica con el antígeno seleccionado y no con otras proteínas. Esta selección puede lograrse sustrayendo los anticuerpos que reaccionan de forma cruzada con otras moléculas. Puede utilizarse una diversidad de formatos de inmunoensayo para

25 seleccionar anticuerpos inmunorreactivos de forma específica con una proteína particular. Por ejemplo, los inmunoensayos de ELISA en fase sólida se utilizan de forma rutinaria para seleccionar anticuerpos inmunorreactivos de forma específica con una proteína (véase, por ejemplo, Harlow & Lane, *Antibodies, A Laboratory Manual* (1988) para una descripción de los formatos de inmunoensayo y condiciones que pueden utilizarse para determinar la inmunorreactividad específica).

La frase “efectos funcionales”, en el contexto de los ensayos para analizar compuestos que modulan una proteína

30 marcadora, incluye la determinación de un parámetro que está de forma indirecta o directa bajo la influencia de un biomarcador de la invención, por ejemplo, uno químico o fenotípico. Por lo tanto, un efecto funcional incluye la actividad de unión a ligando, la activación o represión transcripcional, la capacidad de proliferar de las células, la

35 capacidad de migrar, entre otros. “Efectos funcionales” incluye actividades *in vitro*, *in vivo* y *ex vivo*.

Por “determinar los efectos funcionales” se entiende ensayar un compuesto que aumenta o disminuye un parámetro que está de forma indirecta o directa bajo la influencia de un biomarcador de la invención, por ejemplo, medir los

40 efectos físicos y químicos o fenotípicos. Tales efectos funcionales pueden medirse por cualquier medio conocido para los expertos en la materia, por ejemplo, cambios en características espectroscópicas (por ejemplo, fluorescencia, absorbancia, índice de refracción); propiedades hidrodinámicas (por ejemplo, forma), cromatográficas o de solubilidad de la proteína; ensayos de unión a ligando, por ejemplo, unión a anticuerpos; medición de

45 marcadores inducibles o de la activación transcripcional del marcador; medición de los cambios de la actividad enzimática; la capacidad de aumentar o disminuir la proliferación celular, la apoptosis, la detención del ciclo celular, medición de cambios en los marcadores de superficie celular. Los efectos funcionales pueden evaluarse mediante muchos medios conocidos para los expertos en la materia, por ejemplo, microscopía para medidas cuantitativas y

50 cualitativas de modificaciones en las características morfológicas, medición de cambios en los niveles de ARN o proteína para otros genes expresados en tejido placentario, medición de la estabilidad del ARN, identificación de la expresión de un gen aguas abajo o indicador (CAT, luciferasa, β -gal, GFP y similares), por ejemplo, a través de quimioluminiscencia, fluorescencia, reacciones colorimétricas, unión a anticuerpos, marcadores inducibles, etc.

“Inhibidores”, “activadores” y “moduladores” de los marcadores se utilizan para referirse a moléculas activadoras, inhibidoras o moduladoras identificadas utilizando ensayos *in vitro* e *in vivo* de biomarcadores para el cáncer. Los

55 inhibidores son compuestos que, por ejemplo, se unen, bloquean la actividad de forma parcial o total, disminuyen, impiden, retrasan la activación, inactivan, desensibilizan o regulan de forma negativa la actividad o expresión de biomarcadores para el cáncer. “Activadores” son compuestos que aumentan, abren, activan, facilitan, potencian la activación, sensibilizan, agonizan o regulan de forma positiva la actividad de biomarcadores para el cáncer, por ejemplo, los agonistas. Los inhibidores, activadores o moduladores también incluyen versiones genéticamente

60 modificadas de biomarcadores para el cáncer, por ejemplo, versiones con actividad modificada, así como ligandos de origen natural y sintético, antagonistas, agonistas, anticuerpos, péptidos, péptidos cíclicos, ácidos nucleicos, moléculas antisentido, ribozimas, moléculas de iARN y ARNip, moléculas orgánicas pequeñas y similares. Tales ensayos para inhibidores y activadores incluyen, por ejemplo, la expresión de biomarcadores para el cáncer *in vitro*, en células o extractos celulares, la aplicación de supuestos compuestos moduladores y, después, la determinación de los efectos funcionales sobre la actividad, como se describe anteriormente.

Las muestras o ensayos que comprenden biomarcadores para el cáncer que se tratan con un posible activador, inhibidor o modulador se comparan con muestras de control sin el inhibidor, activador o modulador para examinar el grado de inhibición. A las muestras de control (no tratadas con inhibidores) se les asigna un valor relativo de actividad de proteína del 100 %. La inhibición de los biomarcadores para el cáncer se logra cuando el valor de actividad con respecto al control es aproximadamente del 80 %, preferentemente del 50 %, más preferentemente del 25-0 %. La activación de los biomarcadores para el cáncer se logra cuando el valor de actividad con respecto control (no tratado con activadores) es del 110 %, más preferentemente del 150 %, más preferentemente del 200-500 % (es decir, dos a cinco veces mayor con respecto al control), más preferentemente del 1000-3000 % mayor.

La expresión "compuesto de prueba" o "candidato farmacológico", o "modulador", o equivalentes gramaticales como se utiliza en el presente documento, describe cualquier molécula, ya sea de origen natural o sintética, por ejemplo, una proteína, oligopéptido (por ejemplo, de aproximadamente 5 a aproximadamente 25 aminoácidos de longitud, preferentemente de aproximadamente 10 a 20 o 12 a 18 aminoácidos de longitud, preferentemente de 12, 15 o 18 aminoácidos de longitud), molécula orgánica pequeña, polisacárido, péptido, péptido circular, lípido, ácido graso, ARNip, polinucleótido, oligonucleótido, etc., a probar por la capacidad de modular de forma directa o indirecta los biomarcadores para el cáncer. El compuesto de prueba puede estar en forma de una biblioteca de compuestos de prueba, tal como una biblioteca combinatoria o aleatorizada que proporcione una variedad suficiente de diversidad. Los compuestos de prueba se unen, de forma opcional, a un compañero de fusión, por ejemplo, compuestos de direccionamiento, compuestos de rescate, compuestos de dimerización, compuestos estabilizadores, compuestos direccionables y otras fracciones funcionales. De forma convencional, se generan nuevas entidades químicas con propiedades útiles identificando un compuesto de prueba (llamado "compuesto de partida") con alguna propiedad o actividad conveniente, por ejemplo, actividad inhibidora, creando variantes del compuesto de partida, y evaluando la propiedad y actividad de los compuestos variantes. A menudo se emplean para tal análisis métodos de selección de alto rendimiento (SAR).

Una "molécula orgánica pequeña" se refiere a una molécula orgánica, ya sea de origen natural o sintética, que tiene un peso molecular de más de aproximadamente 50 daltons y menos de aproximadamente 2.500 daltons, preferentemente menos de aproximadamente 2.000 daltons, preferentemente entre aproximadamente 100 y aproximadamente 1000 daltons, más preferentemente entre aproximadamente 200 y aproximadamente 500 daltons.

Métodos de pronóstico

La presente invención proporciona métodos para predecir o proporcionar un pronóstico para el cáncer de pulmón detectando la expresión de un panel de marcadores expresados de forma diferencial en el cáncer. El panel incluye los genes que codifican BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A. La predicción y el pronóstico implican determinar el nivel de un panel de polinucleótidos biomarcadores para el cáncer de pulmón o los correspondientes polipéptidos en un paciente o muestra de paciente, y comparar después el nivel con un valor de referencia o intervalo. Normalmente, el valor de referencia es representativo de los niveles del polinucleótido o ácido nucleico en una persona sana que no padece, o que está destinada a desarrollar, cáncer de pulmón, medido utilizando una muestra biológica tal como una biopsia de pulmón o una muestra de un fluido corporal. La variación de los niveles de un polinucleótido o los polipéptidos correspondientes de la invención con respecto al intervalo de referencia (ya sea hacia arriba o hacia abajo), indica que el paciente tiene un riesgo aumentado de mortalidad a largo plazo.

El algoritmo utilizado para calcular una puntuación de evaluación del riesgo en un método divulgado en el presente documento puede agrupar los valores del nivel de expresión de genes, y la puntuación del riesgo puede obtenerse de cualquier algoritmo conocido en la técnica. Los ejemplos proporcionados en el presente documento emplean algoritmos ejemplares que pueden utilizarse para desarrollar una evaluación del riesgo. Los algoritmos son conjuntos de reglas para describir la evaluación del riesgo de cáncer de pulmón utilizando la expresión del panel de genes descrito en el presente documento. El conjunto de reglas puede estar definido exclusivamente de forma algebraica pero también puede incluir puntos de decisión alternativos o múltiples que requieran un conocimiento específico del campo, la interpretación de expertos u otros indicadores clínicos. Se pueden desarrollar muchos algoritmos que pueden proporcionar distintas evaluaciones del riesgo utilizando perfiles de expresión del panel de genes descritos en el presente documento. Por ejemplo, las puntuaciones de riesgo de un individuo pueden generarse utilizando el modelo de riesgos instantáneos proporcionales de Cox. También puede determinarse una categorización de pronóstico del individuo utilizando un modelo estadístico o un algoritmo de aprendizaje automático, que calcula la probabilidad de recurrencia basada en los perfiles de expresión génica del individuo.

A base de la determinación de un riesgo, los individuos pueden distribuirse en grupos de riesgo (por ejemplo, terciles o cuartiles) a base de en un valor seleccionado de la puntuación de riesgo, en donde todos los individuos con valores en un dado intervalo pueden clasificarse como que pertenecen a un grupo de riesgo particular. Por lo tanto, los valores escogidos definirán grupos de riesgo de pacientes con, respectivamente, mayor o menor riesgo. Los grupos de riesgo pueden clasificarse adicionalmente en distintos intervalos de mortalidad, por ejemplo, en una mortalidad a 6 meses, a 1 año, a 2 años, a 3 años, a 4 años, a 5 años, a 10 años, a 25 años. Los grupos de riesgo pueden clasificarse adicionalmente en distintos intervalos de sucesos asociados con cáncer de pulmón, los que pueden incluir, pero sin limitación, probabilidad de metástasis, recurrencia, etc.

En el presente documento se exponen diversas estrategias tecnológicas para la determinación de los niveles de expresión del panel de genes, incluyendo, pero sin limitación, RT-PCR, micromatrices, secuenciación de alto rendimiento, análisis en serie de la expresión génica (SAGE, acrónimo de *serial analysis of gene expression*) y Expresión Génica Digital (DGE, forma siglada de *digital gene expression*). El nivel de expresión de cada gen puede determinarse en relación a diversas características de los productos de expresión del gen, incluyendo exones, intrones, epítomos de proteína y actividad de proteína.

En una realización preferente, se utiliza PCR en tiempo real o de transcripción inversa cuantitativa (RTPCR) para examinar la expresión de los once marcadores del panel, utilizando ARN procedente de una muestra biológica, tal como un tejido tumoral. No se necesita microdissección. La extracción de ARN puede realizarse mediante cualquier método conocido para los expertos en la materia, por ejemplo, métodos que implican la digestión de tejido con proteinasa K y la precipitación de ácidos nucleicos basada en alcoholes, el tratamiento con ADNasa para digerir el ADN contaminante, la purificación de ARN utilizando la tecnología de membrana de gel de sílice, métodos que utilizan kits disponibles en el mercado tal como Trizol y RNeasy, o cualquier combinación de los mismos. La RT-PCR en tiempo real puede realizarse por cualquier método conocido para los expertos en la materia, por ejemplo, PCR en tiempo real Taqman utilizando los ensayos de Applied Biosystem. La expresión génica se calcula con respecto al ARN de pulmón normal agrupado y la expresión se normaliza con respecto a genes constitutivos. Los expertos en la materia seleccionan cebadores de oligonucleótido adecuados. En una realización, el ensayo se utiliza para cánceres en estadio I, estadio II, estadio III o estadio IV. En una realización, la muestra de tejido procede de un tumor extirpado de forma quirúrgica.

En una realización, los biomarcadores de ARN se examinan utilizando moléculas de unión de ácido nucleico como sondas, oligonucleótidos, matrices de oligonucleótidos y cebadores, que detectan la expresión diferencial de ARN en muestras de pacientes. En una realización, se utiliza RT-PCR de acuerdo con métodos convencionales conocidos en la técnica. En otra realización, para detectar ácidos nucleicos y variantes de los mismos pueden utilizarse ensayos de RT-PCR cuantitativos tales como los que utilizan los ensayos Taqman® disponibles de, por ejemplo, Applied Biosystems. En otras realizaciones, para detectar ácidos nucleicos pueden utilizarse micromatrices de ácidos nucleicos. También está incluido en el ámbito de la presente invención el análisis de los ácidos nucleicos puede conseguirse utilizando técnicas de rutina tales como análisis de Northern o cualquier otro método basado en hibridación con una secuencia de ácido nucleico que sea complementaria con una porción de la secuencia codificante del marcador (por ejemplo, hibridación por transferencia por ranuras). Los reactivos que se unen a biomarcadores de ácido nucleico seleccionados pueden prepararse de acuerdo con métodos conocidos para el experto en la materia o pueden adquirirse en el mercado.

Las técnicas de amplificación por PCR aplicables se describen en, por ejemplo, Ausubel *et al.* e Innis *et al.*, citado anteriormente. Los métodos generales de hibridación de ácidos nucleicos se describen en Anderson, "Nucleic Acid Hybridization", BIOS Scientific Publishers, 1999. La amplificación o hibridación de una pluralidad de secuencias de ácidos nucleicos (por ejemplo, ADN genómico, ARNm o ADNc) puede también realizarse a partir de la secuencia de ARNm o ADNc dispuesta en una micromatriz. Los métodos de micromatriz se describen en general en Hardiman, "Microarrays Methods and Applications: Nuts & Bolts", DNA Press, 2003; y Baldi *et al.*, "DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modeling," Cambridge University Press, 2002.

El análisis de los marcadores de ácido nucleico puede realizarse utilizando técnicas conocidas en la técnica incluyendo, pero sin limitación, análisis de secuencia y análisis electroforético. Los ejemplos no limitativos de análisis de secuencia incluyen la secuenciación de Maxam-Gilbert, la secuenciación de Sanger, la secuenciación de ADN capilar, la secuenciación por ciclo térmico (Sears *et al.*, *Biotechniques*, 13: 626-633 (1992)), la secuenciación en fase sólida (Zimmerman *et al.*, *Methods Mol. Cell Biol.*, 3: 39-42 (1992)), la secuenciación con espectrometría de masas tal como la desorción e ionización por láser asistida por matriz-tiempo de vuelo-espectrometría de masas (MALDI-TOF/MS; Fu *et al.*, *Nat. Biotechnol.*, 16: 381-384 (1998)) y la secuenciación por hibridación. Chee *et al.*, *Science*, 274: 610-614 (1996); Drmanac *et al.*, *Science*, 260: 1649-1652 (1993); Drmanac *et al.*, *Nat. Biotechnol.*, 16: 54-58 (1998). Los ejemplos no limitativos de análisis electroforético incluyen electroforesis en gel plano, tal como electroforesis en gel de agarosa o poliacrilamida, electroforesis capilar y electroforesis en gel en gradiente desnaturizante.

En otro aspecto, los reactivos de anticuerpos pueden utilizarse en ensayos para detectar los niveles de expresión de los biomarcadores proteicos de la invención en muestras de paciente, que utilizan cualquiera de una serie de inmunoensayos conocidos para los expertos en la materia. Las técnicas y protocolos de inmunoensayos se describen en general en Price y Newman, "Principles and Practice of Immunoassay", 2ª edición, Grove's Dictionaries, 1997; y Gosling, "Immunoassays: A Practical Approach", Oxford University Press, 2000. Puede utilizarse una diversidad de técnicas de inmunoensayo, incluyendo los inmunoensayos competitivos y no competitivos. Véase, por ejemplo, Self *et al.*, *Curr. Opin. Biotechnol.*, 7: 60-65 (1996). El término inmunoensayo abarca técnicas que incluyen, pero sin limitación, inmunoensayos enzimáticos (EIA) tales como la técnica de inmunoensayo multiplicado por enzimas (EMIT), el ensayo inmunoabsorbente ligado a enzimas (ELISA), el ELISA de captura con anticuerpo IgM (MAC ELISA) y el inmunoensayo enzimático de micropartículas (MEIA); los inmunoensayos con electroforesis capilar (CEIA); radioinmunoensayos (RIA); ensayos inmunoradiométricos (IRMA); inmunoensayos de polarización fluorescente (FPIA) y ensayos de quimioluminiscencia (CL). Si se desea, tales

inmunoensayos pueden estar automatizados. Los inmunoensayos también pueden utilizarse de forma conjunta con inmunofluorescencia inducida por láser. Véase, por ejemplo, Schmalzing *et al.*, *Electrophoresis*, 18: 2184-93 (1997); Bao, J. *Chromatogr. B. Biomed. Sci.*, 699: 463-80 (1997). Los inmunoensayos con liposomas, tales como los inmunoensayos de inyección de flujo por liposomas y los inmunosensores de liposomas, también son adecuados para su uso en la presente invención. Véase, por ejemplo, Rongen *et al.*, *J. Immunol. Methods*, 204: 105-133 (1997). Además, son adecuados para su uso en los métodos de la presente invención los ensayos de nefelometría, en los que la formación de los complejos de proteína/anticuerpo dan como resultado una dispersión de la luz aumentada que se convierte en una señal de frecuencia máxima como una función de la concentración del marcador. Los ensayos de nefelometría están disponibles en el mercado de Beckman Coulter (Brea, CA; Kit n.º 449430) y pueden realizarse utilizando un analizador nefelómetro de Behring (Fink *et al.*, *J. Clin. Chem. Clin. Biochem.*, 27: 261-276 (1989).

Los niveles de expresión de los genes pronósticos y/o predictivos pueden medirse en un tejido tumoral. Por ejemplo, el tejido tumoral se obtiene tras la extirpación o extirpación quirúrgica del tumor, o por biopsia del tumor. El nivel de expresión de los genes pronósticos y/o predictivos también pueden medirse en células tumorales recuperadas a partir de un sitio distante del tumor, por ejemplo células tumorales en circulación o un fluido corporal.

En los ensayos descritos en el presente documento puede utilizarse una fracción detectable (detección directa o indirecta). Puede utilizarse una amplia diversidad de fracciones detectables, dependiendo la elección del marcador de la sensibilidad necesaria, la facilidad de conjugación con el anticuerpo, los requisitos de estabilidad y las disposiciones disponibles sobre instrumentación y evacuación. Las fracciones detectables adecuadas incluyen, pero sin limitación, radionúclidos, colorantes fluorescentes (por ejemplo, fluoresceína, isotiocianato de fluoresceína (FITC), Oregon Green™, rodamina, rojo Texas, isotiocianato de tetrarrodamina (TRITC), Cy3, Cy5, etc.), marcadores fluorescentes (por ejemplo, proteína verde fluorescente (GFP), ficoeritrina, etc.), compuestos fluorescentes autoinactivados que se activan mediante proteasas asociadas a tumor, enzimas (por ejemplo, luciferasa, peroxidasa de rábano picante, fosfatasa alcalina, etc.), nanopartículas, biotina, digoxigenina, metales y similares.

Es adecuado para la detección sensible, no radiactiva de los niveles de proteínas un ensayo quimioluminiscente que utiliza un anticuerpo quimioluminiscente específico para el ácido nucleico. También es adecuado un anticuerpo marcado con fluorocromo. Los ejemplos de fluorocromos incluyen, pero sin limitación, DAPI, fluoresceína, Hoechst 33258, R-ficocianina, B-ficoeritrina, R-ficoeritrina, rodamina, Rojo Texas y lisamina. Los marcadores indirectos incluyen diversas enzimas bien conocidas en la técnica, tales como peroxidasa de rábano picante (HRP), fosfatasa alcalina (AP), β -galactosidasa, ureasa y similares. Puede utilizarse un sistema de detección por peroxidasa de rábano picante, por ejemplo, con el sustrato cromogénico tetrametilbenzidina (TMB), que produce un producto soluble en presencia de peróxido de hidrógeno que es detectable a 450 nm. Puede utilizarse un sistema de detección de fosfatasa alcalina con el sustrato cromogénico p-nitrofenil fosfato, por ejemplo, que produce un producto soluble fácilmente detectable a 405 nm. De forma similar, puede utilizarse un sistema de detección de β -galactosidasa con el sustrato cromogénico o-nitrofenil- β -D-galactopiranosido (ONPG), que produce un producto soluble detectable a 410 nm. Puede utilizarse un sistema de detección de ureasa con un sustrato tal como urea-púrpura de bromocresol (Sigma Immunochemicals, St. Louis, MO).

Puede analizarse una señal procedente de un marcador directo o indirecto, por ejemplo, utilizando un espectrofotómetro para detectar color de un sustrato cromogénico; un contador de radiación para detectar radiación tal como un contador gamma para la detección de ^{125}I o un fluorómetro para detectar fluorescencia en presencia de luz de una determinada longitud de onda. Para la detección de anticuerpos ligados a enzimas puede hacerse un análisis cuantitativo utilizando un espectrofotómetro tal como un lector de microplacas EMAX (Molecular Devices; Menlo Park, CA) en conformidad con las instrucciones del fabricante. Si se desea, los ensayos de la presente invención pueden automatizarse o realizarse de forma robótica, y puede detectarse de forma simultánea la señal procedente de múltiples muestras.

Los anticuerpos pueden inmovilizarse en una diversidad de soportes sólidos, tal como partículas de una matriz magnética o cromatográfica, la superficie de una placa de ensayo (por ejemplo, pocillos de microtitulación), trozos de un material o membrana de sustrato sólido (por ejemplo, plástico, nylon, papel) y similar. Puede prepararse una tira de ensayo recubriendo el anticuerpo o una pluralidad de anticuerpos en una matriz sobre un soporte sólido. Esta tira puede sumergirse después en la muestra de ensayo y procesarse rápidamente a través de etapas de lavado y detección, para generar una señal que se puede medir, tal como una mancha coloreada.

Los formatos físicos útiles comprenden superficies que tienen una pluralidad de emplazamientos discretos y evaluables para la detección de una pluralidad de distintos marcadores. Tales formatos incluyen micromatrices y determinados dispositivos capilares. Véase, por ejemplo, Ng *et al.*, *J. Cell Mol. Med.*, 6: 329-340 (2002); la Patente de Estados Unidos n.º 6.019.944. En estas realizaciones, cada emplazamiento discreto de la superficie puede comprender anticuerpos para inmovilizar uno o más marcadores para la detección en cada emplazamiento. Las superficies pueden, de forma alternativa, comprender una o más partículas discretas (por ejemplo, micropartículas o nanopartículas) inmovilizadas en emplazamientos discretos de una superficie, en donde las micropartículas comprenden anticuerpos para inmovilizar uno o más marcadores para la detección.

El análisis puede llevarse a cabo en una diversidad de formatos físicos. Por ejemplo, podría utilizarse el uso de placas de microtitulación o de automatización para facilitar el procesamiento de un gran número de muestras de ensayo. Como alternativa, podrían desarrollarse formatos de muestra única para facilitar el diagnóstico o el pronóstico en el momento oportuno.

Como alternativa, los anticuerpos o las sondas de ácido nucleico de la invención pueden aplicarse a cortes de biopsias de pacientes inmovilizados sobre portaobjetos de microscopio. La tinción por anticuerpo o el patrón de hibridación *in situ* resultante puede visualizarse utilizando uno cualquiera de una diversidad de métodos microscópicos fluorescentes conocidos en la técnica.

En otro formato, los diversos marcadores de la invención también proporcionan reactivos para la formación de imágenes *in vivo* tales como, por ejemplo, la formación de imágenes de reactivos marcados que detectan los ácidos nucleicos o proteínas codificadas de los biomarcadores de la invención. Para los fines de formación de imágenes *in vivo*, los reactivos que detectan la presencia de proteínas codificadas por biomarcadores para el cáncer, tales como anticuerpos, pueden marcarse utilizando un marcador apropiado, tal como un marcador fluorescente.

El panel de 11 genes descrito en el presente documento (BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A) proporciona un ensayo pronóstico más preciso para el cáncer de pulmón que otras combinaciones de genes conocidas actualmente en la técnica. El éxito de las distintas combinaciones de genes en la asignación de individuos a las categorías de riesgo alto o bajo puede compararse utilizando el área bajo la curva del análisis de la característica operativa del receptor (AUROC, forma siglada de *area under the receiver operating characteristic*). La AUROC es un estadístico sumatorio que captura la capacidad de distintos modelos de discriminar de forma precisa entre grupos de riesgo (en este caso, el riesgo de muerte dentro de los 5 años de la extirpación quirúrgica de un tumor de cáncer de pulmón).

El análisis AUROC puede realizarse utilizando diversos programas informáticos conocidos en la técnica, incluyendo, pero sin limitación, el paquete de programas informáticos de computación estadística R y el paquete de programas informáticos de computación STATA. Un experto en la materia reconocerá que distintos paquetes de programas informáticos de computación pueden producir distintos estadísticos c. Un estadístico c de AUROC mayor indica una firma génica pronóstica más precisa.

Los valores de expresión génica contribuyen de forma positiva a un riesgo de mortalidad del paciente (representado por una razón de riesgos instantáneos mayor de 1,0) o de forma negativa a un riesgo de mortalidad del paciente (representado por una razón de riesgos instantáneos menor de 1,0). Por lo tanto, cada gen puede ya sea aumentar el riesgo de mortalidad de un paciente o disminuir el riesgo de mortalidad de un paciente. Los genes de riesgo son genes para los que el aumento de la expresión está asociado con un riesgo mayor de muerte, mientras que los genes protectores son genes para los que un aumento de la expresión está asociado con un riesgo menor de muerte. En algunas realizaciones, BAG1, BRCA1, CDC6, CDK2AP1, FUT3, IL11 y RND3 del panel de once genes son genes de riesgo que indican una probabilidad aumentada de la mortalidad del sujeto, mientras que ERBB3, LCK, SH3BGR y WNT3A son genes protectores que indican una probabilidad disminuida de mortalidad del sujeto.

En la determinación del estadístico c de AUROC, a los genes de riesgo se les asigna un valor positivo en el algoritmo de ensayo, de forma que sus valores de expresión conducen a un aumento de la puntuación del riesgo, lo que corresponde a riesgo mayor de muerte dentro de 5 años. A los genes protectores se les asigna un valor negativo en el algoritmo de ensayo, de forma que sus valores de expresión conducen a una disminución de la puntuación del riesgo, lo que corresponde a un riesgo menor de muerte dentro de 5 años. Adicionalmente, cada uno de los valores de expresión génica puede ponderarse para representar la contribución relativa de los genes al riesgo de mortalidad del paciente. Debe entenderse que coeficientes grandes representan genes que son muy importantes en la determinación del resultado del paciente, mientras que coeficientes más pequeños representan genes que contribuyen menos a la determinación del resultado del paciente. Los valores ponderados de la combinación de genes de riesgo y de protección en el panel de once genes descrito en el presente documento pueden utilizarse para calcular el estadístico c de AUROC.

Informes

En otro aspecto, la invención presenta un informe que indica un pronóstico de un sujeto con cáncer. El informe puede, por ejemplo, estar en un formato electrónico o en papel. El informe puede incluir información básica del paciente, incluyendo un identificador de paciente (por ejemplo, el nombre del sujeto, el número de seguridad social, el número de seguro médico o un número generado de forma aleatoria), características físicas del sujeto (por ejemplo, edad, peso o sexo), el nombre del médico solicitante, la fecha en que se generó el pronóstico y la fecha de recolección de la muestra. El pronóstico informado puede estar relacionado con la probabilidad de supervivencia para un determinado período de tiempo, la probabilidad de respuesta a determinados tratamientos dentro de un determinado período de tiempo (por ejemplo, tratamientos quimioterapéuticos o quirúrgicos) y/o la probabilidad de recurrencia del cáncer. El pronóstico informado puede estar en forma de una probabilidad porcentual de supervivencia para un determinado período de tiempo, una probabilidad porcentual de respuesta favorable al tratamiento (la respuesta favorable puede definirse, por ejemplo, como contracción del tumor o enlentecimiento del

crecimiento tumoral) o la recurrencia durante un período de tiempo definido (por ejemplo, probabilidad del 20 % de supervivencia durante un período de cinco años). El pronóstico informado puede, como alternativa, estar en forma de una puntuación calculada. Una puntuación mayor o menor, por ejemplo, puede ser indicativa de un pronóstico favorable. En otra realización, el pronóstico informado puede ser una descripción general de la probabilidad de supervivencia, la respuesta al tratamiento o la recurrencia durante un período de tiempo (por ejemplo, muy probable, probable o improbable de sobrevivir durante cinco años). En otra realización, el pronóstico informado puede estar en forma de un gráfico. Además de los niveles de expresión génica, el pronóstico informado también puede tener en cuenta características adicionales del sujeto (por ejemplo, edad, estadio del cáncer, género, tratamiento anterior, estado físico, salud cardiovascular y salud mental).

Además de un pronóstico, el informe puede incluir de forma optativa datos sin procesar referentes al nivel de expresión de BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A.

Composiciones, kits y sistemas integrados

En el presente documento se describen composiciones, kits y sistemas integrados para la práctica de los ensayos descritos en el presente documento, que utilizan anticuerpos específicos para los polipéptidos o ácidos nucleicos específicos para los polinucleótidos de la invención.

Los kits para llevar a cabo los ensayos de diagnóstico de la invención normalmente incluyen una sonda que comprende un anticuerpo o secuencia de ácido nucleico que se une de forma específica a los polipéptidos o polinucleótidos de la invención, y un marcador para detectar la presencia de la sonda. Este kit puede incluir varios anticuerpos o secuencias de polinucleótido que codifican polipéptidos de la invención, por ejemplo, un cóctel de anticuerpos que reconocen las proteínas codificadas por los biomarcadores de la invención.

Planes de tratamiento

Después de un pronóstico que proporciona una evaluación de riesgo bajo, intermedio o alto de mortalidad a 5 años, puede idearse un método para determinar un plan de tratamiento. Por ejemplo, una vez que se ha determinado la clase de evaluación de riesgo, se puede desarrollar un plan de tratamiento específico para el grupo de riesgo. Por ejemplo, para un individuo que tiene un perfil de expresión de los 11 genes descritos en el presente documento indicativo de una evaluación de riesgo alto de mortalidad a 5 años, un profesional sanitario puede utilizar un tratamiento más agresivo. Para un individuo que tiene un perfil de expresión de los 11 genes descritos en el presente documento indicativo de una evaluación de riesgo bajo de mortalidad, un profesional sanitario puede utilizar un tratamiento menos agresivo. Para un individuo que tiene un perfil de expresión de los 11 genes descritos en el presente documento indicativo de una evaluación de riesgo intermedio de mortalidad, un profesional sanitario puede utilizar un tratamiento que no sea tan agresivo como el caso de una evaluación de riesgo alto, pero más agresivo que en el caso de una evaluación de riesgo bajo de mortalidad a 5 años.

Sistema implementado por ordenador

La Figura 8 es un diagrama de bloques que ilustra un sistema de ordenador 100, sobre el que pueden implementarse las realizaciones de las presentes enseñanzas. En diversas realizaciones, el sistema de ordenador 100 puede incluir un bus 102 u otro mecanismo de comunicación para comunicar información, y un procesador 104 acoplado con el bus 102 para procesar información. En diversas realizaciones, el sistema de ordenador 100 puede incluir también una memoria 106, que puede ser una memoria de acceso directo (RAM) u otro dispositivo de almacenamiento dinámico, acoplado al bus 102 para determinar los llamados de bases e instrucciones a ejecutar por el procesador 104. La memoria 106 también puede utilizarse para almacenar variables temporales u otra información intermedia durante la ejecución de las instrucciones a ejecutar por el procesador 104. En diversas realizaciones, el sistema de ordenador 100 puede incluir adicionalmente una memoria solo de lectura (ROM) 108 u otro dispositivo de almacenamiento estático acoplado al bus 102, para almacenar información estática e instrucciones para el procesador 104. Puede proporcionarse y acoplarse al bus 102 un dispositivo de almacenamiento 110, tal como un disco magnético o un disco óptico, para almacenar información e instrucciones.

En diversas realizaciones, el sistema de ordenador 100 puede acoplarse a través del bus 102 a una pantalla 112, tal como un tubo de rayos catódicos (CRT) o pantalla de cristal líquido (LCD), para presentar información al usuario del ordenador. Puede acoplarse al bus 102 un dispositivo de entrada 114, que incluye claves alfanuméricas y otras, para comunicar información y selecciones de comando al procesador 104. Otro tipo de dispositivo de entrada para el usuario es un control de cursor 116, tal como un ratón, un ratón de bola o teclas para la dirección del cursor, para comunicar información de dirección y selección de comandos al procesador 104 y para controlar el movimiento del cursor en la pantalla 112. Este dispositivo de entrada normalmente tiene dos grados de libertad en dos ejes, un primer eje (es decir, x) y un segundo eje (es decir, y), que permite al dispositivo especificar posiciones en un plano.

Un sistema de ordenador 100 puede realizar las presentes enseñanzas. De forma concordante con determinadas implementaciones de las presentes enseñanzas, el sistema de ordenador 100 puede proporcionar resultados en respuesta al procesador 104, que ejecuta una o más secuencias de una o más instrucciones contenidas en la

memoria 106. Tales instrucciones pueden leerse en la memoria 106 de otro medio leíble por ordenador, tal como un dispositivo de almacenamiento 110. La ejecución de las secuencias de instrucciones contenidas en la memoria 106 puede provocar que el procesador 104 realice los procedimientos descritos en el presente documento. De forma alternativa, puede utilizarse una circuitería cableada en lugar de, o en combinación con, las instrucciones del software para implementar las presentes enseñanzas. Por lo tanto, las implementaciones de las presentes enseñanzas no se limitan a ninguna combinación específica de circuitería de hardware y software.

Ejemplos

Los siguientes ejemplos se ofrecen para ilustrar, pero no para limitar, la invención reivindicada.

Ejemplo 1: Ensayo pronóstico para determinar la puntuación de riesgo en pacientes con cáncer de pulmón no microcítico (CPNM)

Se desarrolló un ensayo pronóstico basado en los patrones de expresión de 426 pacientes que se habían sometido a la extirpación de un cáncer de pulmón no microcítico (CPNM) en estadio I-IV en la Universidad de California, San Francisco, para generar una puntuación de riesgo útil desde el punto de vista clínico. Se buscó un ensayo que tendiera a asignar una puntuación de riesgo mayor en promedio para los pacientes que habían sucumbido a su cáncer que para los que habían sobrevivido el período de seguimiento. Los pacientes cuya muestra recibió puntuaciones de riesgo más altas se considerarían en mayor riesgo de fallecimiento dentro de un período de 5 años después de la operación, mientras que los pacientes cuyas muestras recibieron una puntuación baja tendrían más probabilidades de haber sobrevivido durante este intervalo de tiempo después de sus operaciones.

Se extrajo ARN de las muestras de tejidos FFIP y se evaluaron los niveles de expresión para los genes diana relacionados con el pronóstico del paciente. Después, se desarrolló un ensayo pronóstico correlacionando los patrones de expresión de los 11 genes diana relacionados con el pronóstico del paciente (*BAG1*, *BRCA1*, *CDC6*, *CDK2AP1*, *ERBB3*, *FUT3*, *IL11*, *LCK*, *RND3*, *SH3BGR*, *WNT3A*) con los resultados de supervivencia global a 5 años utilizando el modelo de riesgos instantáneos proporcionales de Cox. 337 pacientes tenían CPNM de células no escamosas. Una combinación de validación cruzada múltiple de 10 veces que maximiza el índice de concordancia de supervivencia, así como el modelo de riesgos instantáneos proporcionales de Cox penalizado con L2 produjeron coeficientes para cada uno de los 11 genes diana. Estos coeficientes penalizados con L2 se muestran en la Tabla 1.

Tabla 1: Coeficientes para cada uno de los predictores de supervivencia marcadores

Predictor	Coeficiente
b_cotXbag	-0,0023688
b_cotXbrc	-0,1460735
b_cotXcdc	-0,0833502
b_cotXcdk	-0,1865318
b_cotXerb	0,0663762
b_cotXfut	-0,0345802
b_cotXil1	-0,0138303
b_cotXlck	0,2098769
b_cotXrnd	-0,0884906
b_cotXsh3	0,1982098
b_cotXwnt	0,1185592
b_cscXcdk	-0,2019083
b_cscXerb	0,2257878
b_cscXfut	0,0264782
b_cscXlck	0,2306624

Utilizando este modelo, se obtuvo una puntuación de riesgo para cada paciente insertando los niveles de expresión de cada uno de los 11 genes pronósticos en el siguiente algoritmo de puntuación de riesgo. Las puntuaciones de riesgo se muestran en la Tabla 2. Esta puntuación de riesgo es una puntuación de riesgo continuo, con un intervalo entre 1 y 100 en la cohorte de desarrollo del algoritmo de la UCSF. Una puntuación de riesgo creciente se correlaciona con cambios crecientes de la mortalidad dentro de los 5 años desde la fecha de la extirpación del tumor. La razón de riesgos instantáneos de la puntuación de riesgo continuo se muestra en la Tabla 3. La RRI de la puntuación de riesgo como una variable continua es 1,184, que corresponde a un aumento del 18 % en el riesgo de muerte dentro de 5 años desde la extirpación quirúrgica para cada punto de aumento en la puntuación de riesgo (IC del 95 %: 1,123 a 1,248, $p < 0,0005$).

Tabla 2: Puntuación de riesgo

Puntuación de riesgo = $e^{\wedge}$ (
-0,0023688*cot*ddct BAG1-1 +
-0,1460735*cot*ddct BRCA1-1 +
-0,0833502*cot*ddct CDC6-3 +
-0,1865318*cot*ddct CDK2AP1-2
+
0,0663762*cot*ddct ERBB3-1 +
-0,0345802*cot*ddct FUT3-1 +
-0,0138303*cot*ddct IL11-1 +
0,2098769*cot*ddct LCK-3 +
-0,0884906*cot*ddct RND3-3 +
0,1982098*cot*ddct SH3BGR-6 +
0,1185592*cot*ddct WNT3A-1 +
-0,2019083*csc*ddct CDK2AP1-2
+
0,2257878*csc*ddct ERBB3-1 +
0,0264782*csc*ddct FUT3-1 +
0,2306624*csc*ddct LCK3)

Tabla 3: Razón de riesgos instantáneos de la puntuación de riesgo continua

Regresión de Cox – método Breslow para empates					
N.º de sujetos	=	426	Número de obs.	=	426
N.º de fallos	=	187			
Tiempo en riesgo	=	17366,98426			
			LR chi2(1)	=	28,10
Log de probabilidad	=	-1058,448	Prob > chi2	=	0,0000
t	Razón de riesgos inst.	Err. Típ.	z	P> [z]	[Intervalo de conf. del 95 %]
puntuación	1,183853	,0319197	6,26	0,000	1,122916 1,248097

- 5 Los grupos de riesgo también se desarrollaron a bases de estas puntuaciones de riesgo, colocando a los pacientes en distintas categorías de riesgo de acuerdo con su puntuación de riesgo. Por ejemplo, se obtuvieron los puntos de corte por tercil colocando a pacientes en los grupos de “riesgo bajo”, “riesgo intermedio” o “riesgo alto” a base de sus puntuaciones de riesgo. Para la cohorte del desarrollo del algoritmo de la UCSF de 426 pacientes, estos puntos de corte de tercil en los 337 pacientes con CPNM no escamoso fueron de 2,7440550 para el punto de corte para riesgo
- 10 bajo y de 3,8221440 para el punto de corte para el riesgo alto.

Para evaluar la precisión de estas asignaciones de grupo de riesgo, se generaron curvas de supervivencia de Kaplan-Meier para los grupos de riesgo. Los resultados de la supervivencia de Kaplan-Meier de estos grupos de riesgo se muestran en la Figura 3 y las Tablas 4-9.

Tabla 4: Tendencia de la prueba de rango logarítmico para los pacientes de riesgo bajo, intermedio y alto agrupados por puntuación de riesgo en toda la cohorte de desarrollo del algoritmo de la UCSF de 426.

	Inic.		Función	Error		
Tiempo	Total	Fallo	de supervivencia	Típ.	[Int. de	Conf. del 95 %]
tercil de riesgo=0						
0	0	0	1,0000	.	.	
12	138	5	0,9648	0,0155	0,9175	0,9852
24	125	13	0,8732	0,0279	0,8064	0,9182
36	111	4	0,8448	0,0304	0,7740	0,8950
48	87	10	0,7629	0,0369	0,6810	0,8264
60	72	3	0,7338	0,0391	0,6480	0,8019
tercil de riesgo=1						
0	0	0	1,0000	.	.	
12	140	5	0,9653	0,0153	0,9186	0,9854
24	116	23	0,8055	0,0330	0,7309	0,8614
36	101	9	0,7417	0,0366	0,6617	0,8056
48	70	13	0,6356	0,0417	0,5478	0,7108
60	50	7	0,5673	0,0445	0,4754	0,6490
tercil de riesgo=2						
0	0	0	1,0000	.	.	
12	104	37	0,7357	0,0373	0,6543	0,8008
24	73	31	0,5143	0,0422	0,4287	0,5933
36	52	16	0,3966	0,0416	0,3150	0,4769
48	44	3	0,3722	0,0414	0,2917	0,4526
60	28	8	0,2956	0,0409	0,2182	0,3769
Nota: La función de supervivencia se calcula sobre los datos completos y se evalúa en los momentos indicados; no se calcula a partir de los agregados que se muestran a la izquierda.						

Prueba de rango logarítmico para la igualdad de funciones de supervivencia		
	Sucesos	Sucesos
tercil de riesgo	observados	esperados
0	35	72,61
1	57	67,92
2	95	46,47
Total	187	187,00
chi2(2) = 72,61		
Pr>chi2 = 0,0000		

Prueba para la tendencia de las funciones de supervivencia		
chi2(1) =		64,88
Pr>chi2 =		0,0000

Tabla 5: Tendencia de la prueba de rango logarítmico para pacientes de riesgo bajo, intermedio y alto agrupados por puntuación de riesgo en los 278 pacientes con adenocarcinoma en la cohorte de desarrollo de algoritmo de la UCSF.

	Inic.		Función	Error		
Tiempo	Total	Fallo	de supervivencia	Típ.	[Int. de	Conf. del 95 %)
tercil de riesgo=0						
0	0	0	1,0000	.	.	.
12	99	1	0,9899	0,0100	0,9305	0,9986
24	92	7	0,9192	0,0274	0,8449	0,9587
36	81	3	0,8884	0,0317	0,8075	0,9366
48	61	7	0,8042	0,0418	0,7063	0,8724
60	52	1	0,7890	0,0437	0,6877	0,8608
tercil de riesgo=1						
0	0	0	1,0000	.	.	.
12	95	1	0,9895	0,0105	0,9276	0,9985
24	79	15	0,8314	0,0384	0,7397	0,8931
36	70	7	0,7562	0,0442	0,6563	0,8308
48	46	7	0,6649	0,0508	0,5549	0,7537
60	33	5	0,5841	0,0561	0,4662	0,6846
tercil de riesgo=2						
0	0	0	1,0000	.	.	.
12	66	19	0,7738	0,0456	0,6686	0,8493
24	45	21	0,5238	0,0545	0,4122	0,6239
36	30	11	0,3867	0,0538	0,2820	0,4900
48	26	1	0,3728	0,0536	0,2691	0,4764
tercil de riesgo=2						
60	15	4	0,3052	0,0537	0,2045	0,4117
Nota: La función de supervivencia se calcula sobre los datos completos y se evalúa en los momentos indicados; no se calcula a partir de los agregados que se muestran a la izquierda.						

Prueba de rango logarítmico para la igualdad de funciones de supervivencia		
	Sucesos	Sucesos
tercil de riesgo	observados	esperados
0	19	45,88
1	35	39,78

Prueba de rango logarítmico para la igualdad de funciones de supervivencia		
2	56	24,34
Total		
	110	110,00
chi2(2) = 58,13		
Pr>chi2 = 0,0000		
Prueba para la tendencia de las funciones de supervivencia		
chi2(1) = 52,45		
Pr>chi2 = 0,0000		

Tabla 6: Tendencia de la prueba de rango logarítmico para pacientes de riesgo bajo, intermedio y alto agrupados por puntuación de riesgo en los 89 pacientes con carcinoma escamoso en la cohorte de desarrollo de algoritmo de la UCSF

Inic.	Función		Error			
Tiempo	Total	Fallo	de supervivencia	Típ.	[Int. de	Conf. del 95 %]
tercil de riesgo=0						
0	0	0	1,0000			
12	28	3	0,9000	0,0548	0,7212	0,9666
24	23	5	0,7333	0,0807	0,5369	0,8567
36	22	0	0,7333	0,0807	0,5369	0,8567
48	19	2	0,6600	0,0877	0,4593	0,8010
60	13	2	0,5844	0,0926	0,3838	0,7398
tercil de riesgo=1						
0	0	0	1,0000			
12	28	3	0,9000	0,0548	0,7212	0,9666
24	22	6	0,7000	0,0837	0,5026	0,8312
36	18	2	0,6296	0,0889	0,4311	0,7754
48	14	4	0,4815	0,0939	0,2914	0,6484
60	10	1	0,4444	0,0937	0,2593	0,6144
tercil de riesgo=2						
0	0	0	1,0000			
12	20	10	0,6552	0,0883	0,4541	0,7973
24	14	6	0,4483	0,0923	0,2652	0,6157
36	11	2	0,3793	0,0901	0,2087	0,5490
48	10	0	0,3793	0,0901	0,2087	0,5490
60	7	2	0,2950	0,0876	0,1400	0,4687
Nota: La función de supervivencia se calcula sobre los datos completos y se evalúa en los momentos indicados; no se calcula a partir de los agregados que se muestran a la izquierda.						

Prueba de rango logarítmico para la igualdad de funciones de supervivencia		
	Sucesos	Sucesos
tercil de riesgo	observados	esperados
0	12	18.64
1	16	17.14
2	20	12.22
Total	48	48,00
chi2(2) = 7,46		
Pr>chi2 = 0,0240		
Prueba para la tendencia de las funciones de supervivencia		
chi2(1) = 6,99		
Pr>chi2 = 0,0082		

Tabla 7: Tendencia de la prueba de rango logarítmico para pacientes de riesgo bajo, intermedio y alto agrupados por puntuación de riesgo en los 267 pacientes con CPNM en estadio I en la cohorte de desarrollo de algoritmo de la UCSF.

Tiempo	Inic. Total	Fallo	Función de supervivencia	Error Típ.	[Int. de	Conf. del 95 %]
tercil de riesgo=0						
0	0	0	1,0000	.		
12	98	5	0,9510	0,0214	0,8862	0,9793
24	90	8	0,8725	0,0330	0,7907	0,9239
36	81	1	0,8627	0,0341	0,7793	0,9163
48	66	6	0,7959	0,0409	0,7011	0,8635
60	54	1	0,7821	0,0425	0,6846	0,8527
tercil de riesgo=1						
0	0	0	1,0000	.		
12	94	4	0,9588	0,0202	0,8939	0,9843
24	78	15	0,8040	0,0403	0,7101	0,8702
36	70	5	0,7512	0,0441	0,6521	0,8258
48	50	8	0,6565	0,0497	0,5495	0,7440
60	32	7	0,5559	0,0548	0,4421	0,6554
tercil de riesgo=2						
0	0	0	1,0000	.		
12	55	14	0,7941	0,0490	0,6773	0,8725
24	43	12	0,6176	0,0589	0,4915	0,7212
36	31	8	0,4924	0,0615	0,3676	0,6057
48	28	1	0,4760	0,0616	0,3518	0,5901
60	15	7	0,3367	0,0625	0,2185	0,4588
Nota: La función de supervivencia se calcula sobre los datos completos y se evalúa en los momentos indicados; no se calcula a partir de los agregados que se muestran a la izquierda.						

Prueba de rango logarítmico para la igualdad de funciones de supervivencia		
	Sucesos	Sucesos
tercil de riesgo	observados	esperados
0	21	43,59
1	39	37,53
2	42	20,88
Total	102	102,00
chi2(2) = 33,33		
Pr>chi2 = 0,0000		
Prueba para la tendencia de las funciones de supervivencia		
chi2(1) = 32,36		
Pr>chi2 = 0,0000		

Tabla 8: Tendencia de la prueba de rango logarítmico para pacientes de riesgo bajo, intermedio y alto agrupados por puntuación de riesgo en los 71 pacientes con CPNM en estadio II en la cohorte de desarrollo de algoritmo de la UCSF.

	Inic.		Función	Error		
Tiempo	Total	Fallo	de supervivencia	Típ.	[Int. de	Conf. del 95 %]
tercil de riesgo=0						
0	0	0	1,0000			
12	0	0	1,0000			
24	13	2	0,8571	0,0935	0,5394	0,9622
36	13	0	0,8571	0,0935	0,5394	0,9622
48	9	1	0,7619	0,1224	0,4209	0,9181
60	8	0	0,7619	0,1224	0,4209	0,9181
tercil de riesgo=1						
0	0	0	1,0000			
12	0	0	1,0000			
24	19	4	0,8182	0,0822	0,5853	0,9276
36	17	2	0,7273	0,0950	0,4910	0,8671
48	12	4	0,5455	0,1062	0,3207	0,7239
60	10	0	0,5455	0,1062	0,3207	0,7239
tercil de riesgo=2						
0	0	0	1,0000	.		
12	28	8	0,7714	0,0710	0,5946	0,8785
24	17	11	0,4571	0,0842	0,2890	0,6105
36	13	4	0,3429	0,0802	0,1934	0,4979
48	11	1	0,3117	0,0788	0,1682	0,4666
60	7	1	0,2771	0,0772	0,1402	0,4324
Nota: La función de supervivencia se calcula sobre los datos completos y se evalúa en los momentos indicados; no se calcula a partir de los agregados que se muestran a la izquierda.						

Prueba de rango logarítmico para la igualdad de funciones de supervivencia		
	Sucesos	Sucesos
Tercil de riesgo	observados	esperados
0	3	9,24
1	10	14,10
2	25	14,66
Total	38	38,00
chi2(2) = 12,89		
Pr>chi2 = 0,0016		
Prueba para la tendencia de las funciones de supervivencia		
chi2(1) = 12,04		
Pr>chi2 = 0,0005		

Tabla 9: Tendencia de la prueba de rango logarítmico para pacientes de riesgo bajo, intermedio y alto agrupados por puntuación de riesgo en los 69 pacientes con CPNM en estadio III en la cohorte de desarrollo de algoritmo de la UCSF.

Tiempo	Inic. Total	Fallo	Función de supervivencia	Error Típ.	[Int. de	Conf. del 95 %]
tercil de riesgo=0						
0	0	0	1,0000	.	.	.
12	0	0	1,0000	.	.	.
24	19	2	0,9000	0,0671	0,6560	0,9740
36	15	2	0,8000	0,0894	0,5511	0,9198
48	12	1	0,7333	0,1039	0,4680	0,8810
60	8	2	0,5926	0,1230	0,3204	0,7862
tercil de riesgo=1						
0	0	0	1,0000	.	.	.
12	22	1	0,9545	0,0444	0,7187	0,9935
24	18	4	0,7727	0,0893	0,5374	0,8985
36	13	2	0,6818	0,0993	0,4462	0,8338
48	8	1	0,6198	0,1079	0,3768	0,7911
60	7	0	0,6198	0,1079	0,3768	0,7911
tercil de riesgo=2						
0	0	0	1,0000	.	.	.
12	17	11	0,5926	0,0946	0,3863	0,7499
24	13	4	0,4444	0,0956	0,2556	0,6175
36	8	4	0,2963	0,0879	0,1406	0,4703
48	5	1	0,2469	0,0860	0,1024	0,4238
60	4	0	0,2469	0,0860	0,1024	0,4238
Nota: La función de supervivencia se calcula sobre los datos completos y se evalúa en los momentos indicados; no se calcula a partir de los agregados que se muestran a la izquierda.						

Prueba de rango logarítmico para la igualdad de funciones de supervivencia		
	Sucesos	Sucesos
tercil de riesgo	observados	esperados
0	7	12.78
1	8	12.66
2	20	9.56
Total	35	35,00
chi2(2) = 16,03		
Pr>chi2 = 0,0003		
Prueba para la tendencia de las funciones de supervivencia		
chi2(1) = 12,16		
Pr>chi2 = 0,0005		

Ejemplo 2: Ensayo de 11 genes para predecir la supervivencia en el cáncer de pulmón no microcítico no escamoso extirpado

- 5 Se desarrolló un ensayo de 14 genes que utiliza un análisis por PCR cuantitativa de tejidos fijados en formalina incluidos en parafina (FFIP) con una cohorte de 361 pacientes con CPNM no escamoso extraído en la Universidad de California, San Francisco (UCSF). Este ensayo incluyó once biomarcadores (BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A) y tres genes constitutivos (ESD, TBP, YAP1).
- 10 Este ensayo, desarrollado y llevado a cabo en un laboratorio independiente certificado por Clinical Laboratory Improvement Amendments (Enmiendas de Mejora de Laboratorios Clínicos) (CLIA) fue validado después por la Kaiser Permanente Division of Research (División de Investigación de Kaiser Permanente) (KPDOR) con un estudio con ocultación diseñado en una cohorte de 433 pacientes con CPNM no escamoso en estadio I extirpado en los hospitales del sistema de Kaiser Permanente Northern California (CA, EE.UU.). La KPDOR comparó de forma
- 15 independiente los resultados del ensayo con los resultados reales de los pacientes. También se realizó una validación a gran escala independiente e internacional de este ensayo de pronóstico molecular en una cohorte de 1006 pacientes chinos que se habían sometido a la extirpación de un CPNM de fase temprana en una de las varias instituciones que participan en el China Clinical Trials Consortium (CCTC) (Consortio de Ensayos Clínicos de China).
- 20 Los pacientes eran elegibles para entrar en el estudio como parte de la cohorte de entrenamiento si se habían sometido a una extirpación quirúrgica de un CPNM no escamoso en la UCSF con intención curativa, entre el 1 de enero de 1997 y el 31 de diciembre de 2007.
- 25 Los pacientes eran aptos para ser incluidos en la cohorte de validación de Kaiser Permanente si se habían sometido a la extirpación completa de un CPNM no escamoso en estadio I de la American Joint Commission on Cancer (Comisión Conjunta Americana sobre Cáncer) por una estadificación clínica y patológica en una instalación de Kaiser Permanente del Norte de California entre el 1 de enero de 1998 y el 31 de diciembre de 2005.
- 30 Los pacientes eran aptos para ser incluidos en la cohorte de validación del CCTC si se habían sometido a un intento de extirpación curativa de un CPNM no escamoso en estadio I-III de la American Joint Commission on Cancer, en el First Affiliated Hospital of Guangzhou Medical College (Guangdong, China), Sun Yat-sen University Cancer Center en Guangzhou (Guangdong, China) o el Shanghai Pulmonary Hospital (Shanghai, China) entre el 1 de enero de 2000 y el 31 de diciembre de 2008.

Los criterios de exclusión para los pacientes en ya sea la cohorte de entrenamiento o la de validación eran los siguientes: bloques de tejido ausentes o inadecuados (es decir, un tumor que ocupa <25 % del área de superficie tisular), fallecimiento dentro de los 30 días desde la extirpación, tratamiento preoperatorio con quimioterapia (solo cohortes de validación), márgenes positivos en el análisis de anatomía patológica (solo cohortes de validación) y un segundo cáncer (excluyendo los carcinomas cutáneos de células basales y escamosas) diagnosticado dentro de los 3 años desde el diagnóstico de cáncer de pulmón (solo cohorte de validación del CCTC). La información sobre las variables clínicas, el seguimiento y la causa de fallecimiento se obtuvieron a partir de una revisión de las historias clínicas. El estado vital y la fecha de fallecimiento se establecieron mediante la revisión de las historias clínicas y se verificaron por fuentes que incluían el Kaiser Permanente Northern California Cancer Registry, California Death Records, Social Security Death Master File y el contacto directo con el paciente o su familia.

Preparación y análisis de las muestras

Se utilizaron seis cortes de FFIP de 10 micrómetros por muestra. Las muestras se limpiaron de parafina mediante el uso de xileno y después se incubaron con proteinasa K (Kit de Purificación de ARN MasterPure, Epicentre, Madison, WI) a 65 °C durante 2 horas. La precipitación de proteínas y la precipitación de ácidos nucleicos basada en alcoholes se realizaron utilizando el kit de purificación de ARN MasterPure (Epicentre, Madison, WI). Los extractos de ARN se trataron con ADNasa y se purificaron utilizando columnas de centrifugación de membrana de gel de sílice (Kit RNEasy Micro, Qiagen, Valencia, CA).

Para controlar la degradación del ARN que pudiera producirse en las muestras de FFIP, se midió la cantidad y calidad del ARN utilizando un espectrofotómetro Nanodrop (Thermo Scientific, Wilmington, DE). El ARN extraído se sometió a transcripción inversa (Kit iScript Select cDNA Synthesis, BioRad Laboratories, Hercules, CA) utilizando cebadores específicos de gen. Los cebadores específicos de gen eran versiones truncadas en 9-13 meros de los cebadores de PCRq inversos optimizados para una temperatura de apareamiento de 42 °C. El ADNc se sometió a 10 ciclos de preamplificación (Mezcla Maestra TaqMan PreAmp, Applied Biosystems, Carlsbad, CA) antes de la PCRq.

Se utilizaron ensayos de PCR cuantitativa TaqMan (BioSearch Technologies, Novato, CA) diseñados a petición del cliente para su uso con el ARN extraído de los tejidos FFIP para cuantificar la expresión de ARN utilizando química FAST en un Sistema de PCR en tiempo real 7900HT Fast (Applied Biosystems, Carlsbad, CA). Se diseñaron ensayos de PCR cuantitativa TaqMan específicos para FFIP para tener como objetivo amplicones de 65-85 pares de bases que traspasaban los límites exón-exón, que evitaban las estructuras de molde y las homologías cruzadas (Beacon Designer 5.0, Premier Biosoft, Palo Alto, CA). Todas las secuencias de los cebadores se sometieron a búsqueda por BLAST frente al genoma humano (NCBI ref_assembly 37.1) para asegurar la especificidad de diana. Los cebadores sintetizados se analizaron para la concentración óptima de los cebadores y la disociación de producto individual utilizando curvas de fusión por verde SYBR.

Todas las mediciones de la expresión de ARN se normalizaron con respecto a ARN disponible en el mercado extraído de muestras de pulmón normal congeladas agrupadas (Laboratorios Clontech, Mountain View, CA) y la expresión relativa para cada gen diana se calculó utilizando el método de C_T comparativo. El valor promedio de C_T de los tres genes constitutivos ESD, TBP e YAP1 se utilizó para normalizar la expresión génica y calcular los valores de delta C_T .

Desarrollo de un algoritmo de pronóstico

Se evaluaron en la cohorte de entrenamiento de la UCSF once genes diana relacionados con cáncer (BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A) y tres genes de referencia (ESD, TBP y YAP1). Véase la Figura 4.

El modelo de riesgos instantáneos proporcionales de Cox penalizado con L2 (paquete R glmnet v1.5.3) fue la herramienta analítica principal utilizada para desarrollar los coeficientes en el algoritmo pronóstico, utilizando los valores de expresión relativos de los once genes diana en la cohorte de FFIP de la UCSF. La cantidad de penalización L2 aplicada se determinó utilizando una validación cruzada de 10 veces. Se generó una puntuación de riesgo continuo para cada sujeto, a base de los coeficientes del modelo. Se utilizó un modelo de riesgo lineal convencional en el que cada nivel de expresión génica relativo (delta-delta C_T) se multiplicó por su coeficiente; estos productos se sumaron después, dando como resultado una puntuación de riesgo única sin procesar.

Puntuación de Riesgo Sin Procesar = delta-delta C_T BAG1 * coeficiente de modelo BAG1 + delta-delta C_T BRCA1 * coeficiente de modelo BRCA1 + delta-delta C_T CDC6 * coeficiente de modelo CDC6 + delta-delta C_T CDK2AP1 * coeficiente de modelo CDK2AP1 + delta-delta C_T ERBB3 * coeficiente de modelo ERBB3 + delta-delta C_T FUT3 * coeficiente de modelo FUT3 + delta-delta C_T IL11 * coeficiente de modelo IL11 + delta-delta C_T LCK * coeficiente de modelo LCK + delta-delta C_T RND3 * coeficiente de modelo RND3 + delta-delta C_T SH3BGR * coeficiente de modelo SH3BGR + delta-delta C_T WNT3A * coeficiente de modelo WNT3A.

Después, la puntuación de riesgo sin procesar se ajustó a escala utilizando una función lineal basada en las puntuaciones de riesgo sin procesar mínima y máxima, produciendo una única puntuación de riesgo integrada que estaba incluida en una escala de 1-100.

$$5 \quad \text{Puntuación de Riesgo} = 39,39747 * \text{Puntuación de Riesgo Sin Procesar} - 16,94965 + 1$$

10 Las puntuaciones de riesgo pronosticadas resultantes se dividieron en los percentiles 33º y 67º para generar grupos de riesgo bajo, intermedio y alto. En la Tabla 10 a continuación se proporciona una tabla completa de los coeficientes del algoritmo, los coeficientes de ajuste de escala y los valores de los puntos de corte de las categorías de riesgo.

Tabla 10: Resumen de los coeficientes del algoritmo, coeficientes de ajuste de escala y los valores de los puntos de corte de las categorías de riesgo obtenidos de la cohorte de entrenamiento de la UCSF

	Valor
Coeficientes de los genes	
BAG1	-0,0023688
BRCA1	-0,1460735
CDC6	-0,0833502
CDK2AP1	-0,1865318
ERBB3	0,0663762
FUT3	-0,0345802
IL11	-0,0138303
LCK	0,2098769
RND3	-0,0884906
SH3RGR	0,1982098
WTN3A	0,1185592
Coeficientes de ajuste de escala	
Pendiente	39,39747
Intersección	-16,94965
Valores de puntos de corte de las categorías de riesgo	
Punto de corte de bajo-intermedio	23,81960
Punto de corte de intermedio-alto	36,87494

15 Se generaron las puntuaciones de riesgo y las categorías de riesgo se asignaron utilizando los mismos coeficientes
 20 génicos, los coeficientes de ajuste de escala y los valores de los puntos de corte en las cohortes de validación. Dado que los límites de la escala de puntuación de riesgo eran 1 y 100, a base del intervalo de la cohorte de entrenamiento de la UCSF, cualquier puntuación de riesgo de menos de 1 se asignó a una puntuación de riesgo de 1 mientras que cualquier puntuación de riesgo mayor de 100 se asignó a la puntuación de riesgo de 100.

Robustez y validación del ensayo

25 El ensayo molecular se desarrolló, se especificó de forma completa y se validó de forma analítica en un laboratorio certificado por el CLIA antes del inicio del estudio de validación clínica con ocultación. La concentración por nanodrop y los valores de los puntos de corte de pureza para el ARN se determinaron de forma empírica evaluando más de 400 muestras cuyos valores de expresión de genes constitutivos sin procesar promedio estaban incluidos dentro de un intervalo de C_T sin procesar preespecificado. La concentración, la proporción 260/280 y los puntos de corte de la proporción 260/230 se determinaron eliminando el 2,5 % más bajo para cada parámetro individual
 30 clasificado de menor a mayor y luego tomando la siguiente medida más alta como el número de punto de corte aceptable. Además, se calculó para cada muestra el C_T del gen constitutivo sin procesar promedio. Solo se incluyeron en el estudio las muestras cuyos valores de expresión sin procesar promedio de los genes constitutivos estaban incluidos dentro de un intervalo preespecificado. Para obtener este intervalo, se estudió la expresión de 442 muestras de cáncer de pulmón FFIP obtenidas de la UCSF. El valor de C_T del gen constitutivo promedio para estas
 35 muestras varió entre 17,79-37,76, con un valor medio de 23,94 y una desviación típica de 2,66. El intervalo preespecificado se determinó teniendo en cuenta el C_T medio del gen constitutivo más o menos tres desviaciones típicas.

Cada placa de PCR se procesó con un control positivo (ARN extraído de muestras de pulmón normal congeladas y agrupadas, disponible en el mercado (Laboratorios Clontech, Mountain View, CA) así como con un control negativo (sin molde). Cada muestra se procesó con un ensayo TaqMan diseñado para detectar contaminación genómica. El análisis repetido en distintos días y en distintos cortes del ensayo de pronóstico molecular de las muestras FFIP del mismo tumor demostró una alta reproducibilidad de la puntuación de riesgo, con una desviación típica promedio de 2,18 unidades (intervalo entre 0,83-4,62) en una escala de 100 unidades.

Los criterios de inclusión de más de 25% de tumor se obtuvieron empíricamente por categorización de muestras en un 25-50 % de tumor, 50-75 % de tumor y > 75 % de tumor. La razón de riesgos instantáneos del grupo de riesgo alto en comparación con el grupo de riesgo bajo para cada una de estas categorías porcentuales se muestra a continuación en la Tabla 1.

Tabla 11: Razón de riesgos instantáneos de la categoría de riesgo alto por tumor porcentual

Tumor Porcentual	RRI de categoría de riesgo ^s	IC del 95 %	valor p
25-50 %	2,39	1,65-3,47	<0,0001
50-75 %	2,31	1,57-3,28	<0,0001
>75 %	2,40	1,52-3,79	<0,0001

^sModelado como una variable continua

Análisis estadístico

La supervivencia global desde el momento de la extirpación se eligió como el criterio de valoración principal. Un criterio de valoración secundario en la cohorte de validación de Kaiser fue la mortalidad específica por cáncer de pulmón. El predictor principal evaluado fue la categoría de riesgo asignada por el ensayo molecular. Se compararon con el resultado otras covariables importantes, incluyendo la edad, sexo, antecedentes como fumador, histopatología, tamaño tumoral y fase de la enfermedad, mediante el uso de un modelo de riesgos instantáneos proporcionales de Cox de una variable o de múltiples variables. Se realizaron pruebas de razón de verosimilitudes de Wald y anidadas para el modelado de una variable y de múltiples variables, respectivamente, para evaluar la significación estadística. Las pruebas de proporción de probabilidad anidadas son más apropiadas para los modelos de múltiples variables debido a que examinan si la adición de una nueva variable, tal como la prueba molecular, ofrece una mejora en el ajuste por encima de las variables clínicas convencionales tales como la edad, sexo y tamaño tumoral. Se utilizaron un análisis estratificado de Kaplan-Meier con un conjunto de datos con censura por la derecha y la prueba de rango logarítmico para la tendencia para evaluar la asociación entre la categoría de riesgo y los criterios de valoración principal y secundario. Para todos los análisis estadísticos, se consideró como estadísticamente significativo un α de dos colas preespecificado de 0,05. El área bajo la curva del análisis de la característica operativa del receptor dependiente del tiempo (AUROC) se calculó con el paquete survcomp (versión 1.1.6) en R; las diferencias en las AUROC se analizaron mediante el modelo de riesgos instantáneos proporcionales de Cox de múltiples variables y se compararon mediante el uso de las AUROC integradas con la prueba de la suma de rangos de Wilcoxon.

Se realizó un cálculo de la potencia de las cohortes de entrenamiento de la UCSF y de validación de Kaiser Permanente utilizando las siguientes suposiciones: β de 0,9, desviación típica de la población de categoría de riesgo de 0,8, una razón de riesgos instantáneos de 1,5 para el grupo de riesgo alto, probabilidad de retirada de 0,2 y probabilidades de sucesos de 0,4 (cohorte de entrenamiento de la UCSF en estadio I-III) y 0,3 (cohorte de validación de Kaiser en estadio I). El cálculo de la potencia dio como resultado un tamaño de muestra estimado de 313 y 417 pacientes para las cohortes de UCSF y Kaiser Permanente respectivamente. Como la proporción de la extracción de ARN satisfactoria no se conocía antes del inicio del estudio, se asumió una tasa de fallo del 10 %, lo que produjo un tamaño de muestra final de 344 pacientes para la cohorte de entrenamiento de la UCSF.

Los análisis se realizaron con los lenguajes de programación R29 (versión 2.12.2 para Macintosh) y Stata/MP (versión 11).

Resultados

Se identificó un total de 399 pacientes que se habían sometido a extirpación de un CPNM no escamoso en la UCSF durante el período de estudio; de estos, 361 cumplían los criterios de inclusión en la cohorte de entrenamiento. 460 pacientes de Kaiser del Norte de California se habían sometido a la extirpación de un CPNM no escamoso en fase, de quienes 433 cumplían los criterios de inclusión en la cohorte de validación independiente. Se identificaron 1006 pacientes en las instituciones del CCTC que cumplían los criterios de inclusión en ese estudio de validación. Las características clínicas y patológicas relevantes de estos pacientes se muestran a continuación en la Tabla 12. La tasa de extracción de ARN satisfactoria fue alta en las tres cohortes.

Tabla 12: Características clínicas y patológicas de los pacientes

	Cohorte de entrenamiento de la UCSF	Cohorte de validación de Kaiser	Cohorte de validación del CCTC
Bloques FFIP disponibles	361	433	1006
Recuperación de ARN satisfactoria	337 (93 %)	420 (97 %)	967 (96 %)
Edad en el momento de la extirpación (años, media [DT])	67,4 (11)	66,6 (9)	58,3 (11)
Sexo (femenino)	200 (59 %)	229 (55 %)	366 (38 %)
Antecedentes como fumador			
Sí	224 (66 %)	355 (85 %)	492 (49 %)
No	57 (17 %)	36 (9 %)	403 (40 %)
Desconocido	56 (17 %)	29 (7 %)	72 (7 %)
Seguimiento de los supervivientes (meses; mediana [amplitud intercuartílica])	64,0 (45,6-88,9)	106,0 (88,0-125,0)	53-4 (37,4-68,0)
Muertes a los 5 años desde la extirpación	139 (41 %)	179 (43 %)	406 (42 %)
Histopatología			
Adenocarcinoma	278 (83 %)	325 (77 %)	881 (88 %)
Células grandes	17 (5 %)	15 (4 %)	17 (2 %)
Mixto	10 (3 %)	15 (4 %)	46 (5 %)
CPNM (si no se especifica otra cosa)	32 (10 %)	65 (16 %)	23 (2 %)
Estadio			
I	223 (66 %)	420 (100 %)	471 (47 %)
Ia	152 (45 %)	285 (68 %)	239 (24 %)
Ib	71 (21 %)	135 (32 %)	232 (23 %)
II	41 (12 %)	0	222 (22 %)
IIa	14 (4 %)	0	69 (7 %)
IIb	27 (8 %)	0	153 (15 %)
III	58 (17 %)	0	266 (26 %)
IIIa*	32 (10 %)	0	247 (25 %)
IIIb	26 (8 %)	0	19 (2 %)
IV	9 (3 %)	0	0
Indeterminado	6 (2 %)	0	8 (1 %)

Los datos son n (%), a menos que se establezca otra cosa. CCTC= China Clinical Trials Consortium. FFIP= fijado en formalina, incluido en parafina. CPNM= cáncer de pulmón microcítico. UCSF= University of California, San Francisco.

5 Durante la rigurosa validación técnica y el establecimiento del ensayo, el análisis de expresión génica candidato y la comparación con los resultados de los pacientes mostraron ser similares en los grupos de bloques de tejido en los que el tumor ocupaba ya sea 25-50 %, 50-75 % o más del 75 % del área de superficie tisular.

10 Se calcularon las puntuaciones de riesgo individuales para cada paciente en la cohorte de entrenamiento de la UCSF. Las puntuaciones de riesgo más altas estaban asociadas de forma positiva con una probabilidad aumentada de mortalidad a los 5 años. Con referencia a las Figuras 5A y 5B, debería entenderse que una razón de riesgos instantáneos mayor a 1 implica que más pacientes en el grupo de riesgo alto están muriendo en cualquier momento en comparación con el grupo de riesgo bajo, mientras que una proporción de riesgos instantáneos menor a 1 significa que están muriendo en cualquier momento menos pacientes en el grupo de riesgo alto en comparación con el grupo de riesgo bajo.

15 Para identificar mejor a los pacientes en el riesgo más alto o el más bajo, se obtuvieron valores de corte que definen los grupos de riesgo bajo, riesgo intermedio y riesgo alto dividiendo en terciles las puntuaciones de riesgo de la cohorte de entrenamiento. Con referencia otra vez a las Figuras 5A y 5B, se observó un riesgo más alto de mortalidad con cada aumento en la categoría de riesgo en casi cada subgrupo de la cohorte de entrenamiento de la

UCSF y la correlación de la puntuación de riesgo con el resultado clínico fue muy similar en las muestras con 25-50 %, 50-75 % y >75 % de área de superficie tumoral.

Después de que el ensayo se especificara de forma completa, el laboratorio certificado por el CLIA realizó la validación técnica. El KPDOR y el CCTC realizaron después validaciones con ocultación independientes del ensayo validado de forma técnica. El KPDOR envió las muestras al laboratorio certificado por el CLIA para el análisis con ocultación y la asignación a la categoría de riesgo, y la realización del ensayo. Por el contrario, con la razón empírica (1:1:1) de pacientes de riesgo bajo, riesgo intermedio y riesgo alto en la cohorte de entrenamiento, se identificó una mayor proporción de pacientes de riesgo alto en la cohorte de validación de Kaiser Permanente.

Este hallazgo podría atribuirse a la supervivencia global a 5 años más baja de la cohorte de validación de Kaiser en estadio I (56,4 %) en comparación con la cohorte de entrenamiento de la UCSF en estadio I (61,9 %). En la cohorte de validación de Kaiser Permanente, el análisis de supervivencia de Kaplan-Meier mostró una supervivencia a 5 años del 71,4 % (CI del 95 % 60,5-80,0) en el grupo de riesgo bajo, del 58,3 % (48,9-66,6) en el grupo de riesgo intermedio y del 49,2 % (42,2-55,8) en el grupo de riesgo alto. Un análisis de sensibilidad que excluía a los 18 pacientes de la cohorte de validación de Kaiser Permanente que recibieron quimioterapia adyuvante proporcionó resultados de supervivencia a 5 años que eran muy similares: el 70,0 % (58,7-78,8) en el grupo de riesgo bajo, el 58,2 % (48,5-66,7) en el grupo de riesgo intermedio y el 48,9 % (41,7-55,6) en el grupo de riesgo alto ($p_{\text{tend}} = 0,0006$). La supervivencia específica para cáncer de pulmón a 5 años fue del 84,6 % (74,4-91,0) en el grupo de riesgo bajo, del 70,3 % (60,6-78,0) en el grupo de riesgo intermedio y del 63,3 % (55,8-69,8) en el grupo de riesgo alto. También se realizó un análisis de supervivencia de Kaplan-Meier para los pacientes con enfermedad en estadio I en la cohorte de validación de Kaiser que no tenían un criterio de la National Comprehensive Cancer Network (NCCN) de riesgo alto. Este grupo incluía a todos los pacientes con enfermedad en estadio IA y un subgrupo de pacientes con enfermedad de estadio IB. La supervivencia global a 5 años para los pacientes en este subgrupo (con la estadificación del riesgo según los resultados del ensayo molecular) fue del 72,7 % (61,3-81,3) en el grupo de riesgo bajo, del 59,0 % (48,9-67,8) en el grupo de riesgo intermedio y del 50,4 % (42,0-58,3) en el grupo de riesgo alto.

Con referencia a la Figura 6A, la supervivencia mediana se observó como de 113 meses en el grupo de riesgo bajo, de 91 meses en el grupo de riesgo intermedio y de 59 meses en el grupo de riesgo alto. Con referencia a la Figura 6B, no se observó que en algún grupo de riesgo se alcanzara la supervivencia mediana específica para cáncer de pulmón. La tasa de incidencia de mortalidad fue de 2,7 cada 100 personas-años en el grupo de riesgo bajo, 5,0 cada 100 personas-años en el grupo de riesgo intermedio y 6,6 cada 100 personas-años en el grupo de riesgo alto. Con referencia a la Figura 6C, la supervivencia mediana fue de 113 meses en el grupo de riesgo bajo, de 88 meses en el grupo de riesgo intermedio y de 70 meses en el grupo de riesgo alto.

En la cohorte de validación de Kaiser Permanente, la categoría de riesgo (alto), la edad y el sexo eran predictores estadísticamente significativos de mortalidad en un análisis de una variable, como se observa a continuación en la Tabla 13.

Tabla 13: Modelo de riesgos instantáneos proporcionales de Cox para la mortalidad global a 5 años en la cohorte de validación de Kaiser Permanente

Categoría del riesgo*	Análisis de una variable		Análisis de múltiples variables	
	Razón de riesgos instantáneos (IC del 95 %)	Valor p de la prueba de Wald	Razón de riesgos instantáneos (IC del 95 %)	Valor p de la prueba de razón de verosimilitudes
Riesgo alto	2,16 (1,39-3,36)	0,0007	2,04 (1,28-3,26)	0,0016
Riesgo intermedio	1,60 (0,98-2,60)	0,0610	1,66 (1,00-2,74)	0,00436
Edad >65 años	1,55 (1,14-2,10)	0,0054	1,66 (1,21-2,29)	0,0016
Sexo (femenino)	0,55 (0,41-0,74)	0,0001	0,67 (0,49-0,92)	0,0123
Nunca fumador	0,59 (0,32-1,09)	0,0917	0,83 (0,44-1,55)	0,5438
Histopatología [†]				
Carcinoma de células grandes	0,96 (0,42-2,17)	0,9139	0,64 (0,26-1,59)	0,3038
Mixto	0,98 (0,43-2,22)	0,9615	0,85 (0,37-1,95)	0,6932
CPNM (si no se especifica otra cosa)	1,13 (0,89-1,93)	0,1659	1,18 (0,79-1,75)	0,4320
Tamaño tumoral > 4 cm	1,42 (0,97-2,07)	0,0697	1,10 (0,73-1,66)	0,6435

CPNM = cáncer de pulmón no microcítico. *Comparado con el grupo de bajo riesgo. [†]Comparado con adenocarcinoma.

El análisis de múltiples variables (ajustando para edad, sexo, antecedentes como fumador, histopatología y tamaño tumoral > 4 cm) mostró que ambos agrupamientos de riesgo alto y riesgo intermedio así como la edad y el sexo fueron predictores estadísticamente significativos de mortalidad, como se observa a continuación en la Tabla 14.

5 **Tabla 14:** Modelo de riesgos instantáneos proporcionales de Cox de múltiples variables para la mortalidad global a 5 años en la cohorte de validación de Kaiser Permanente excluyendo la prueba molecular

	RRI	IC del 95 %	Valor p de la prueba RV
Edad>65	1,59	1,26-2,20	0,0036
Sexo femenino	0,65	0,47-0,89	0,0064
Nunca fumador	0,74	0,40-1,38	0,3294
Histopatología [§]			
Carcinoma de células grandes	0,76	0,31-1,87	0,5346
Mixto	1,00	0,44-2,27	0,9933
CPNM si no se especifica otra cosa	1,26	0,85-1,88	0,2555
Tamaño tumoral > 4 cm	1,21	0,81-1,80	0,3608

[§]Comparado con adenocarcinoma

10 En la cohorte del CCTC, la mortalidad a 5 años después de la extirpación completa de un CPNM no escamoso por grupo de riesgo (definido de acuerdo con los resultados del ensayo molecular) fue la siguiente: el 74,1 % (66,0-80,6) en el grupo de riesgo bajo, el 57,4 % (48,3-65,5) en el grupo de riesgo intermedio y el 44,6 % (40,2-48,9) en el grupo de riesgo alto, como se observa en la Figura 7.

15 La supervivencia mediante fue de 101,1 meses en el grupo de riesgo bajo, de 77,2 en los grupos de riesgo intermedio y de 43,1 meses en el grupo de riesgo alto. Se sugiere una mejora con el uso del ensayo pronóstico de 14 genes en comparación con el uso de la estadificación tradicional por la separación estadísticamente significativa de las curvas de supervivencia de Kaplan-Meier para la supervivencia global a 5 años entre los pacientes de riesgo bajo, riesgo intermedio y riesgo alto en el análisis de subgrupo de pacientes con enfermedad de distinto estadio (véase la Figura 7): enfermedad en estadio I (riesgo bajo = 83,0 % [73,8-89,1], riesgo intermedio = 67,7 % [54,8-77,7]; riesgo alto = 64,6 % [57,9-70,5]), enfermedad en estadio II (riesgo bajo = 54,2 % [30,1-73,2]; riesgo intermedio = 45,8 % [26,2-63,4]; riesgo alto = 38,1 % [29,4-46,8]) y enfermedad en estadio III (riesgo bajo = 53,3 % [32,6-70,3]; riesgo intermedio = 43,3 % [27,2-58,5]; riesgo alto = 24,0 % [17,5-30,9]). El modelado de riesgos instantáneos

20 proporcionales de Cox de una variable indicó que el sexo (masculino), los antecedentes como fumador, la histopatología de células grandes y mixtas y el estadio de la enfermedad tenían todos un efecto negativo en la supervivencia en la cohorte del CCTC, como se observa a continuación en la Tabla 15.

25

Tabla 15: Modelo de riesgos instantáneos proporcionales de Cox para la mortalidad global a 5 años en la cohorte de validación del China Clinical Trials Consortium

	Análisis de una variable		Análisis de múltiples variables	
	Razón de riesgos instantáneos (IC del 95 %)	Valor p de la prueba de Wald	Razón de riesgos instantáneos (IC del 95 %)	Valor p de la prueba de Wald
Categoría del riesgo*				
Riesgo alto	3,07 (2,21-4,25)	0,0001	2,37 (1,63-3,43)	<0,0001
Riesgo intermedio	1,87 (1,26-2,77)	0,0019	1,60(1,03-2,49)	0,0354
Edad >65 años	1,11(0,90-1,37)	0,3337	1,19 (0,94-1,49)	0,1493
Sexo (femenino)	0,78 (0,63-0,95)	0,0150	0,93 (0,70-1,23)	0,6057
Nunca fumador	0,70 (0,56-0,86)	0,0009	0,84 (0,64-1,10)	0,1986
Histopatología [†]				
Carcinoma de células grandes	2,12 (1,09-4,11)	0,0259	1,68 (0,83-3,41)	0,1831
Mixto	1,67 (1,12-2,48)	0,0118	1,07 (0,69-1,64)	0,7714
CPNM (si no se especifica otra cosa)	0,89 (0,44-1,80)	0,7528	0,77 (0,36-1,63)	0,4759
Estadio [‡]	1,44 (1,35-1,53)	<0,0001	1,43 (1,33-1,53)	<0,0001

CPNM = cáncer de pulmón no microcítico. *Comparado con el grupo de bajo riesgo. [†]Comparado con adenocarcinoma. [‡]Modelado como una variable continua

Sin embargo no se observó que alguno de estos factores tuviese un efecto tan grande sobre la supervivencia como la designación en la categoría de riesgo alto de acuerdo con el ensayo molecular. El análisis de múltiples variables mostró que la designación de riesgo alto y riesgo intermedio siguió siendo un predictor estadísticamente significativo de la supervivencia incluso después del ajuste para la edad, sexo, antecedentes como fumador, histopatología y estadio de la enfermedad. Véanse a continuación las Tablas 16-18.

5

Tabla 16: Tabulación de los factores de riesgo por categoría de riesgo en la cohorte de validación de Kaiser Permanente

	Riesgo Bajo	Riesgo Intermedio	Riesgo Alto	Valor P
Número de pacientes	85	119	216	
Edad en el momento de la extirpación	66,6 (9,8) [§]	67,1 (8,9) [§]	66,2 (9,3) [§]	0,7114*
Sexo				
Masculino	28 (32,9)	51 (42,9)	112 (51,9)	0,0098
Femenino	57 (67,1)	68 (57,1)	104 (48,1)	
Antecedentes como fumador				
Sí	68 (180,0)	100 (84,0)	187 (86,6)	0,1096
No	12 (14,1)	10 (8,4)	14 (6,5)	
Histopatología				
Adenocarcinoma	77 (90,6)	95 (79,8)	153 (70,8)	0,0070
Células grandes	0	2 (1,7)	13 (6,0)	
Mixto	2 (2,4)	3 (2,5)	10 (4,6)	
CPNM si no se especifica otra cosa	6 (7,1)	19 (16,0)	40 (18,5)	
Tamaño Tumoral (cm)	2,3 (1,4) [§]	2,6 (1,6) [§]	3,4 (1,8) [§]	<0,0001*
Estadio				
1a	70 (82,4)	95 (79,8)	120 (55,6)	<0,0001
1b	15 (17,6)	24 (20,2)	96 (44,4)	

Los números entre paréntesis representan el porcentaje de pacientes en cada categoría de riesgo excepto en donde se indique.

[§]Media de la cohorte (desviación típica)

*Prueba de ANOVA

Tabla 17: Modelo de riesgos instantáneos proporcionales de Cox de múltiples variables para la mortalidad global a 5 años en la cohorte de validación del China Clinical Trials Consortium excluyendo la prueba molecular

	RRI	CI del 95 %	Valor p de la prueba de RV
Edad >65	1,18	0,94-1,49	0,1532
Sexo femenino	0,93	0,70-1,22	0,5917
Nunca fumador	0,74	0,56-0,97	0,0260
Histopatología [§]			
Carcinoma de células grandes	1,79	0,89-3,63	0,1374
Mixto	1,15	0,75-1,77	0,5221
CPNM si no se especifica otra cosa	0,77	0,36-1,63	0,4707
Fase ^{§§}	1,48	1,38-1,58	<0,0001

[§]Comparado con Adenocarcinoma

^{§§}Modelado como una variable continua

Tabla 18: Tabulación de los factores de riesgo por categoría de riesgo en la cohorte de validación del China Clinical Trials Consortium

	Riesgo bajo	Riesgo intermedio	Riesgo alto	Valor <i>P</i>
Número de Pacientes	193	172	602	
Edad en el momento de la extirpación	57,6 (11,4) [§]	58,5 (10,7) [§]	58,5 (10,6) [§]	0,5529*
Sexo				
Masculino	98 (50,8)	86 (50,0)	417 (69,3)	<0,0001
Femenino	95 (49,2)	86 (50,0)	185 (30,7)	
Historial como Fumador				
Sí	70 (38,3)	68 (43,9)	354 (63,6)	<0,0001
No	113 (61,7)	87 (56,1)	203 (36,4)	
Histopatología				
Adenocarcinoma	186 (96,4)	159 (92,4)	536 (89,0)	0,0032
Células Grandes	0	4 (2,3)	13 (2,2)	
Mixto	4 (2,1)	3 (1,7)	43 (7,1)	
CPNM si no se especifica otra cosa	3 (1,6)	6 (3,5)	10 (1,7)	
Tamaño tumoral (cm)	2,8 (1,4) [§]	3,4 (1,7) [§]	4,3 (4,1) [§]	<0,0001*
Estadio				
I	130 (68,4)	95 (55,6)	246 (41,1)	<0,0001
II	34 (17,9)	32 (18,7)	156 (26,1)	
III	26 (13,7)	44 (25,7)	196 (32,8)	

Los números entre paréntesis representan el porcentaje de pacientes en cada categoría de riesgo excepto en donde se indique.

[§]Media de la cohorte (desviación típica)

*Prueba de ANOVA

- Además del análisis de múltiples variables, se realizó un análisis AUROC dependiente del tiempo para probar si el ensayo molecular proporciona información pronóstica más útil que solo la estadificación convencional. La AUROC es una medida de la discriminación de un test pronóstico y coincide con el estadístico c. Se han utilizado los criterios de la NCCN, que identifican tales pacientes como en estadio IB más al menos uno de los siguientes factores de riesgo: tumores poco diferenciados, invasión vascular, extirpación en cuña, márgenes mínimos, tumores mayores de 4 cm de diámetro, implicación de la pleura visceral y estado desconocido de ganglios linfáticos. En la cohorte de validación de Kaiser Permanente la suma del ensayo molecular proporcionó una mejor discriminación del riesgo que solo los criterios de riesgo del NCCN, demostrado por una AUROC más grande (estadístico c de 0,60 frente a 0,54; $p < 0,0001$). No estaban disponibles los datos completos para todos los criterios para el estadio I de riesgo alto de la NCCN en los pacientes de la cohorte del CCTC. Por lo tanto, el análisis AUROC en esta cohorte se centró en 471 pacientes con enfermedad en estadio I; la adición del ensayo molecular a solo la estadificación convencional aumentó, de forma similar, la AUROC para este grupo, de forma coherente con una mejor discriminación en la predicción del riesgo por la suma del ensayo molecular (estadístico c de 0,61 frente a 0,56; $p < 0,0001$).

Discusión

- Se observó que el ensayo basado en PCR cuantitativa utilizado en el presente Ejemplo identifica de forma fiable a pacientes con CPNM no escamoso de estadio temprano en riesgo alto de mortalidad tras la extirpación quirúrgica, discriminando tales pacientes con mayor precisión que el uso del criterio del NCCN solo. Este ejemplo demuestra la implementación de una plataforma con extracción de ARN interpretable a partir de tejido incluido en parafina fijado con formalina, la realización del ensayo en uno de los estudios en un laboratorio que era independiente del laboratorio en el que se desarrolló el ensayo, los tamaños muy grandes de las cohortes de validación independientes y la disparidad potencialmente grande entre los acervos genéticos de una de las cohortes de validación y los de la cohorte de entrenamiento original utilizada para el desarrollo del ensayo.
- El ensayo molecular utilizado en el presente Ejemplo proporciona una prueba más precisa para la definición de subconjuntos de pacientes con CPNM no escamoso y resultados estadísticamente heterogéneos. Este ensayo se validó de forma independiente en una cohorte americana extrahospitalaria grande para mejorar la estratificación del riesgo en pacientes con enfermedad en estadio I. A la vista de la gran magnitud de la crisis de la sanidad pública debida al cáncer de pulmón en la China, la validación adicional de este ensayo molecular en una gran población de chinos aumenta adicionalmente su posible efecto. Este ensayo proporciona una diferenciación pronóstica de

pacientes con enfermedad en fase temprana y podría ser útil en la identificación de la aplicación más apropiada de las directrices de tratamiento para mejorar los resultados clínicos.

Ejemplo 3: Comparación de la AUROC para el ensayo de 11 genes frente a otros ensayos con distintas combinaciones génicas

El presente ejemplo predice el éxito del ensayo de 11 genes en comparación con otras combinaciones de genes en la asignación de pacientes a las categorías de riesgo alto o bajo utilizando el área bajo la curva del análisis de la característica operativa del receptor (AUROC).

El análisis AUROC se realizó en muestras de ARN extraídas de 337 pacientes con cáncer de pulmón en estadio I-IV. Como se observa a continuación en la Tabla 19, se observó que el ensayo de 11 genes es superior a las otras combinaciones de genes probadas en el presente ejemplo, como se refleja por un estadístico de c de AUROC mayor.

Tabla 19: Valores de AUROC

Conjunto de genes	AUROC
BAG1 BRCA1 CDC6 CDK2AP1 ERBB3 FUT3 IL11 LCK RND3 SH3BGR WNT3A	0,7215
BAG1 BRCA1 CDC6 CDK2AP1 ERBB3 IL11 LCK	0,6321
BAG1 BRCA1 CDC6 RND3	0,5889
BAG1 BRCA1 CDC6 FUT3 IL11 RND3 SH3BGR	0,6438
BAG1 BRCA1 CDC6 CDK2AP1 FUT3 IL11 RND3 SH3BGR	0,6466
BAG1 BRCA1 CDC6 CDK2AP1 EMX2 FUT3 IL11 LCK RND3 SH3BGR SIX3	0,6758
BAG1 BRCA1 CDC6 CDK2AP1 ERBB3 IL11 RND3 SH3BGR	0,6489
BAG1 BRCA1 CDC6 CDK2AP1 EMX2 FUT3 IL11 RND3 SH3BGR SIX3 WNT3A	0,6829
BAG1 BRCA1 CDC6 CDK2AP1 EMX2 ERBB3 FUT3 IL11 LCKRND3 SIX3	0,6658
BAG1 BRCA1 CDC6 CDK2AP1 ERBB3 FUT3 IL11 RND3	0,6382
BAG1 BRCA1 CDC6 CDK2AP1 FUT3 IL11 LCKRND3 SH3BGR	0,6767
BAG1 BRCA1 CDC6	0,6280
BAG1 BRCA1 CDC6 CDK2AP1 FUT3 IL11 RND3	0,6321
BAG1 BRCA1 CDC6 CDK2AP1 ERBB3 FUT3 IL11 RND3 SH3BGR	0,6531
BAG1 BRCA1 CDC6 CDK2AP1 ERBB3 FUT3 IL11 LCK RND3 SH3BGR	0,6815
BAG1 BRCA1 CDC6 CDK2AP1 FUT3 IL11 WNT3A	0,6735
BAG1 BRCA1 CDC6 CDK2AP1 EMX2 ERBB3 FUT3 IL11 LCKRND3 SH3BGR SIX3 WNT3A	0,6413
CDK2AP1 FUT3 IL11 RND3	0,6276
BAG1 BRCA1 CDC6 CDK2AP1 EMX2 ERBB3 FUT3 IL11 LCKRND3 SH3BGR SIX3	0,6823
FUT3 IL11 RND3	0,5883
BAG1 CDC6 FUT3 IL11 WNT3A	0,6770
BAG1 BRCA1 CDC6 FUT3 RND3	0,6298
BAG1 BRCA1 CDC6 FUT3 IL11 RND3 SH3BGR	0,6438
BAG1 BRCA1 CDC6 CDK2AP1 ERBB3 IL11 RND3	0,6321
BAG1 BRCA1 CDC6 CDK2AP1 EMX2 ERBB3 FUT3 IL11 RND3 SH3BGR SIX3 WNT3A	0,6841
BAG1 BRCA1 CDC6 CDK2AP1 ERBB3 FUT3 IL11 RND3 WNT3A	0,6704
BAG1 BRCA1 CDC6 CDK2AP1 ERBB3 FUT3 IL11 RND3 SH3BGR WNT3A	0,6842
ERBB3 LCK RND3 WNT3A	0,6861
ERBB3 LCK RND3	0,5821
BAG1 CDC6 FUT3 IL11 RND3	0,6379

Selección de la propiedad de riesgo génico

Se analizó el ensayo de 11 genes para determinar qué genes conferían riesgo y qué genes conferían protección utilizando el modelo de riesgos instantáneos proporcionales de Cox. Estas determinaciones se enumeran para cada gen en la Tabla 20.

Tabla 20: Propiedad de riesgo de los genes del ensayo de 11 genes

Gen	Razón de riesgos instantáneos	Propiedad
BAG1	1,002	Riesgo
BRCA1	1,157	Riesgo
CDC6	1,087	Riesgo
CDK2AP1	1,205	Riesgo
ERBB3	0,936	Protección
FUT3	1,035	Riesgo
IL11	1,014	Riesgo
LCK	0,811	Protección
RND3	1,093	Riesgo
SH3BGR	0,820	Protección
WNT3A	0,888	Protección

10 *Selección del coeficiente de los genes*

Los valores de expresión de los 11 Genes (BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A) pueden combinarse en número infinito de formas para proporcionar un número único que sea representativo del riesgo de mortalidad del paciente. Cada uno de los valores de la expresión de los genes se ponderaron (como se representa mediante un coeficiente) para representar la contribución relativa del gen a un riesgo de mortalidad del paciente. Los coeficientes para cada uno de los genes del ensayo de 11 genes se enumeran a continuación en la Tabla 21.

Tabla 21: Coeficientes para los genes en el ensayo de 11 genes

Gen	Coeficiente	Propiedad	AUROC de la puntuación de riesgo
BAG1	-0,0024	Riesgo	0,7215
BRCA1	-0,1461	Riesgo	
CDC6	-0,0834	Riesgo	
CDK2AP1	-0,1865	Riesgo	
ERBB3	0,0664	Protección	
FUT3	-0,0346	Riesgo	
IL11	-0,0138	Riesgo	
LCK	0,2099	Protección	
RND3	-0,0885	Riesgo	
SH3BGR	0,1982	Protección	
WNT3A	0,1186	Protección	

Análisis AUROC

Se midió la AUROC de la puntuación de riesgo para el ensayo de 11 genes en una cohorte de muestras de ensayo procedentes de 337 pacientes que se habían sometido a extirpación de cáncer de pulmón, midiendo sus puntuaciones de riesgo y comparando la asignación de riesgo con los resultados reales de la supervivencia a 5 años de los pacientes. El estadístico c de AUROC, que utilizó los coeficientes de la Tabla 21 anterior, fue de 0,7215. A base del análisis AUROC, se observó que el ensayo de 11 genes supera a otros modelos (enumerados con sus estadísticos c de AUROC a continuación en las Tablas 22-31), en los cuales se utilizaron distintos coeficientes para ponderar las contribuciones de los 11 genes.

Tabla 22: Modelo alternativo n.º 1

Gen	Coeficiente	Propiedad	AUROC de la puntuación de riesgo
BAG1	-0,4806	Riesgo	0,6755
BRCA1	-0,2392	Riesgo	
CDC6	-0,4906	Riesgo	
CDK2AP1	-0,0386	Riesgo	
ERBB3	0,4989	Protección	
FUT3	-0,3274	Riesgo	
IL11	-0,2656	Riesgo	
LCK	0,4071	Protección	
RND3	-0,2710	Riesgo	
SH3BGR	0,1232	Protección	
WNT3A	0,3385	Protección	

Tabla 23: Modelo n.º 2

Gen	Coeficiente	Propiedad	AUROC de la puntuación de riesgo
BAG1	-0,0718	Riesgo	0,6715
BRCA1	-0,1588	Riesgo	
CDC6	-0,4781	Riesgo	
CDK2AP1	-0,1202	Riesgo	
ERBB3	0,2770	Protección	
FUT3	-0,2002	Riesgo	
IL11	-0,2237	Riesgo	
LCK	0,2451	Protección	
RND3	-0,4197	Riesgo	
SH3BGR	0,1902	Protección	
WNT3A	0,2869	Protección	

Tabla 24: Modelo n.º 3

Gen	Coeficiente	Propiedad	AUROC de la puntuación de riesgo
BAG1	-0,1630	Riesgo	0,6751
BRCA1	-0,0783	Riesgo	
CDC6	-0,4656	Riesgo	
CDK2AP1	-0,2018	Riesgo	
ERBB3	0,0551	Protección	
FUT3	-0,0729	Riesgo	
IL11	-0,1818	Riesgo	
LCK	0,0832	Protección	
RND3	-0,0685	Riesgo	
SH3BGR	0,2572	Protección	
WNT3A	0,2354	Protección	

Tabla 25: Modelo n.º 4

Gen	Coeficiente	Propiedad	AUROC de la puntuación de riesgo
BAG1	-0,4308	Riesgo	0,6385
BRCA1	-0,1066	Riesgo	
CDC6	-0,2033	Riesgo	
CDK2AP1	-0,1515	Riesgo	
ERBB3	0,2706	Protección	
FUT3	-0,4905	Riesgo	
IL11	-0,4776	Riesgo	
LCK	0,1602	Protección	
RND3	-0,2419	Riesgo	
SH3BGR	0,4835	Protección	
WNT3A	0,2145	Protección	

Tabla 26: Modelo n.º 5

Gen	Coeficiente	Propiedad	AUROC de la puntuación de riesgo
BAG1	-0,0220	Riesgo	0,6419
BRCA1	-0,0262	Riesgo	
CDC6	-0,1908	Riesgo	
CDK2AP1	-0,2331	Riesgo	
ERBB3	0,0487	Protección	
FUT3	-0,3632	Riesgo	
IL11	-0,4358	Riesgo	
LCK	0,4983	Protección	
RND3	-0,3906	Riesgo	
SH3BGR	0,0506	Protección	
WNT3A	0,1630	Protección	

Tabla 27: Modelo n.º 6

Gen	Coeficiente	Propiedad	AUROC de la puntuación de riesgo
BAG1	-0,0690	Riesgo	0,6652
BRCA1	-0,0436	Riesgo	
CDC6	-0,1157	Riesgo	
CDK2AP1	-0,2227	Riesgo	
ERBB3	0,2175	Protección	
FUT3	-0,0998	Riesgo	
IL11	-0,1845	Riesgo	
LCK	0,0266	Protección	
RND3	-0,2833	Riesgo	
SH3BGR	0,4528	Protección	
WNT3A	0,3538	Protección	

Tabla 28: Modelo n.º 7

Gen	Coeficiente	Propiedad	AUROC de la puntuación de riesgo
BAG1	-0,1602	Riesgo	0,6787
BRCA1	-0,4631	Riesgo	
CDC6	-0,1032	Riesgo	
CDK2AP1	-0,3042	Riesgo	
ERBB3	0,4957	Protección	
FUT3	-0,4726	Riesgo	
IL11	-0,1426	Riesgo	
LCK	0,3646	Protección	
RND3	-0,4320	Riesgo	
SH3BGR	0,0198	Protección	
WNT3A	0,3023	Protección	

Tabla 29: Modelo n.º 8

Gen	Coeficiente	Propiedad	AUROC de la puntuación de riesgo
BAG1	-0,2513	Riesgo	0,6882
BRCA1	-0,3827	Riesgo	
CDC6	-0,0907	Riesgo	
CDK2AP1	-0,3858	Riesgo	
ERBB3	0,2738	Protección	
FUT3	-0,3453	Riesgo	
IL11	-0,1007	Riesgo	
LCK	0,2027	Protección	
RND3	-0,0808	Riesgo	
SH3BGR	0,0869	Protección	
WNT3A	0,2507	Protección	

Tabla 30: Modelo n.º 9

Gen	Coeficiente	Propiedad	AUROC de la puntuación de riesgo
BAG1	-0,3425	Riesgo	0,6717
BRCA1	-0,3023	Riesgo	
CDC6	-0,0782	Riesgo	
CDK2AP1	-0,4674	Riesgo	
ERBB3	0,0519	Protección	
FUT3	-0,2181	Riesgo	
IL11	-0,0588	Riesgo	
LCK	0,0407	Protección	
RND3	-0,2296	Riesgo	
SH3BGR	0,1539	Protección	
WNT3A	0,1992	Protección	

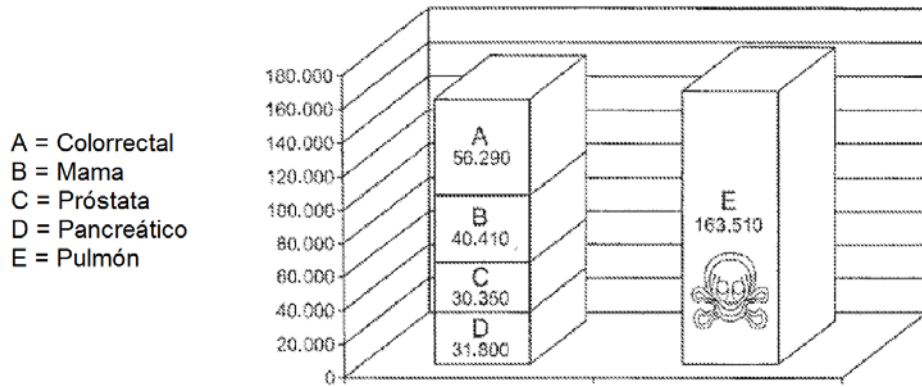
Tabla 31: Modelo n.º 10

Gen	Coeficiente	Propiedad	AUROC de la puntuación de riesgo
BAG1	-0,4336	0,3686239	0,6797
BRCA1	-0,2218	Riesgo	
CDC6	-0,0657	Riesgo	
CDK2AP1	-0,0490	Riesgo	
ERBB3	0,3301	Protección	
FUT3	-0,0909	Riesgo	
IL11	-0,0169	Riesgo	
LCK	0,3788	Protección	
RND3	-0,3783	Riesgo	
SH3BGR	0,2209	Protección	
WNT3A	0,1477	Protección	

REIVINDICACIONES

1. Un método para proporcionar un pronóstico para el cáncer de pulmón no microcítico (CPNM) en un sujeto, comprendiendo el método las etapas de:
 - (a) poner en contacto una muestra biológica del sujeto con reactivos que se unen de forma específica a cada miembro de un panel de biomarcadores que consiste en BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A; y
 - (b) determinar una puntuación de riesgo del sujeto basada en los niveles de ácido nucleico de expresión de los biomarcadores en la muestra o determinar si los biomarcadores se expresan de forma diferencial o no a nivel de ácido nucleico en la muestra; y
 - (c) proporcionar un pronóstico para el CPNM basado en la puntuación de riesgo del sujeto o en la expresión diferencial.
2. Un método de determinación de un plan de tratamiento para un sujeto, comprendiendo el método las etapas de:
 - (a) poner en contacto una muestra biológica del sujeto con reactivos que se unen de forma específica a cada miembro de un panel de biomarcadores que consiste en BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A;
 - (b) determinar si los marcadores se expresan de forma diferencial o no a nivel de ácido nucleico en la muestra;
 - (c) proporcionar un pronóstico para el CPNM;
 - (d) determinar una evaluación de riesgo para la mortalidad a 5 años basada en el pronóstico para el CPNM; y
 - (e) idear un plan de tratamiento basado en la evaluación de riesgo.
3. El método de la reivindicación 1, en el que la expresión aumentada de BAG1, BRCA1, CDC6, CDK2AP1, FUT3, IL11 o RND3 contribuye a una puntuación de riesgo aumentada, y en el que la expresión aumentada de ERBB3, LCK, SH3BGR o WNT3A contribuye a una puntuación de riesgo disminuida, en el que una puntuación de riesgo aumentada indica un riesgo aumentado de mortalidad a largo plazo del sujeto.
4. El método de una cualquiera de las reivindicaciones 1 a 3, en el que el método comprende adicionalmente normalizar los niveles de expresión de los biomarcadores con respecto a genes constitutivos seleccionados del grupo que consiste en ESD, TBP, YAP1 y cualquier combinación de los mismos, y, de forma opcional, los niveles de expresión de los biomarcadores se normalizan frente al valor de C_T promedio de los genes constitutivos.
5. El método de una cualquiera de las reivindicaciones 1 a 4, en el que los reactivos son ácidos nucleicos, oligonucleótidos o conjuntos de cebadores de PCR.
6. El método de una cualquiera de las reivindicaciones 1 a 5, en el que el CPNM es CPNM no escamoso.
7. El método de la reivindicación 6, en el que:
 - (a) el CPNM no escamoso es de estadio I;
 - (b) el CPNM no escamoso es de estadio II;
 - (c) el CPNM no escamoso es de estadio III; o
 - (d) el CPNM no escamoso es de estadio IV.
8. El método de una cualquiera de las reivindicaciones 1 a 7, en el que la muestra es de un tumor extirpado de forma quirúrgica, de tejido pulmonar o una biopsia de tumor pulmonar.
9. El método de una cualquiera de las reivindicaciones 1 a 8, en el que el pronóstico proporciona una evaluación de riesgo y opcionalmente la evaluación de riesgo es para una mortalidad a 5 años.
10. El método de la reivindicación 9, en el que el pronóstico proporciona:
 - a) una evaluación de riesgo alto para la mortalidad a 5 años;
 - b) una evaluación de riesgo intermedio para la mortalidad a 5 años; o
 - c) una evaluación de riesgo bajo para la mortalidad a 5 años.
11. Un kit que comprende reactivos que se unen de forma específica a cada miembro de un panel de biomarcadores que consiste en BAG1, BRCA1, CDC6, CDK2AP1, ERBB3, FUT3, IL11, LCK, RND3, SH3BGR y WNT3A, en el que los reactivos son conjuntos de cebadores de PCR.
12. Uso del kit de la reivindicación 11 en el método de una cualquiera de las reivindicaciones 1-10.

Cáncer de pulmón: causa más común de muerte por cáncer
 Los fallecimientos por cáncer pulmón superan
 ahora la mortalidad combinada de los siguientes cuatro cánceres letales



Fuente: Fallecimientos estimados para 2005 de la American Cancer Society

Estadio I

Supervivencia a 5 años del 60 %

Tratamiento habitual: solo extracción quirúrgica

Información clínica insuficiente para predecir el resultado

Figura 1

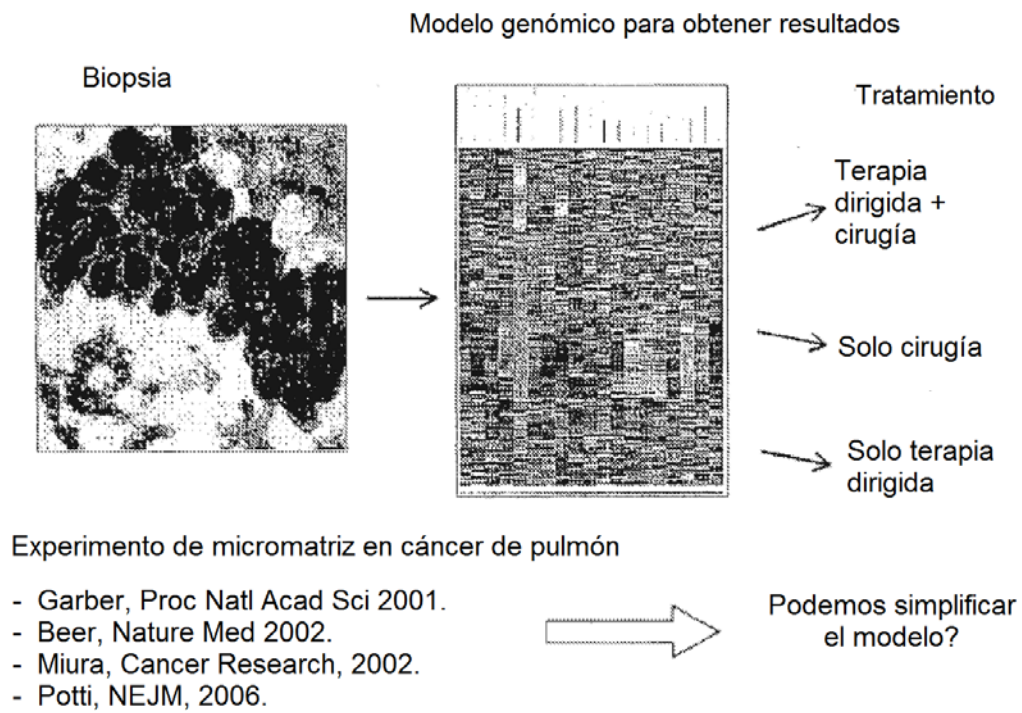


Figura 2

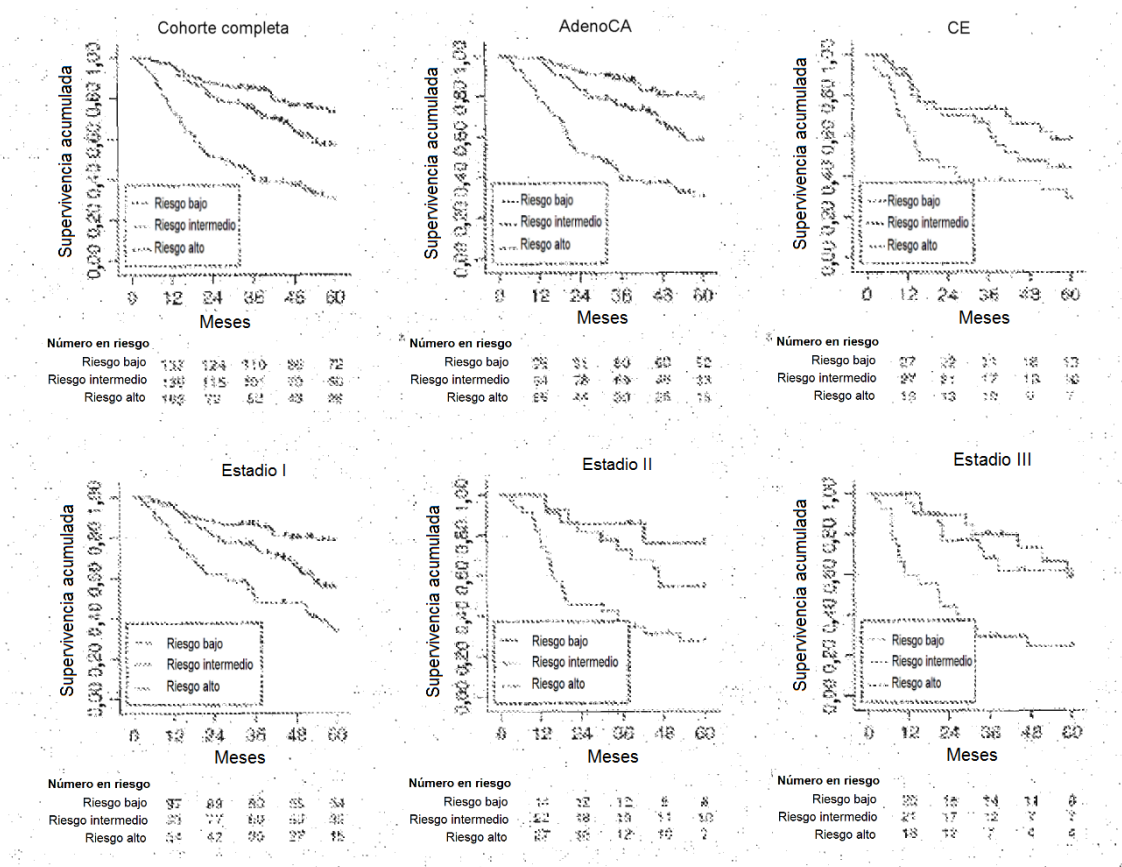


Figura 3

	Nombre del gen	Cromosoma	Secuencia de referencia	Emplazamiento de la proteína	Funciones y rutas biológicas relevantes	Papel del algoritmo
<i>BAG1</i>	Atanogen asociado con BCL2	9p12	NM_004323	Citoplasma	Apoptosis, señalización ligada al receptor de superficie celular	Pronóstico
<i>BRCA1</i>	Cáncer de mama 1, inicio temprano	17q21-q24	NM_007294	Núcleo	Inducción de apoptosis; ubiquitinación de proteínas; regulación de los puntos de control de daño de ADN de la transición G2/N; regulación de la reparación de ADN	Pronóstico
<i>CDC6</i>	Homólogo de ciclo de división celular 6	17q21.3	NM_001254	Núcleo	Regulación de la transcripción implicada en la fase G1/S del ciclo celular mitótico; regulación de la proliferación celular	Pronóstico
<i>CDK2AP1</i>	Proteína 1 asociada a a la quinasa dependiente de ciclina 2	12q24	NM_004642	Núcleo	Regulación de la fase S del ciclo celular mitótico; replicación de ADN dependiente de ADN	Pronóstico
<i>ERBB3</i>	Homólogo 3 del oncogén vírico de leucemia eritroblástica V-erb-b2	12q13	NM_001982	Membrana citoplasmática	Regulación de la adhesión celular; señalización del receptor de la proteína tirosina quinasa transmembrana, regulación de la cascada de fosfoinositida-3-quinasa; regulación de la proliferación celular	Pronóstico
<i>FUT3</i>	Fucosiltransferasa 3	19q13.3	NM_000149	Citoplasma	Metabolismo de hidratos de carbono; glucosilación de aminoácidos proteicos	Pronóstico
<i>IL11</i>	Interleucina 11	19q13.3-q13.4	NM_000641	Extracelular	Regulación de la proliferación celular; diferenciación de linfocitos B; diferenciación de megacariocitos; regulación de la fosforilación de peptidil serina; regulación de la cascada MAPKKK	Pronóstico
<i>LCK</i>	Proteína tirosina quinasa específica de linfocitos	1q34.3	NM_001042771	Citoplasma	Inducción de apoptosis; activación de la actividad caspasa; regulación de la señalización del receptor de linfocitos T; migración de leucocitos	Pronóstico
<i>RND3</i>	GTPasa 3 de la familia Rho	2q23.3	NM_005168	Citoplasma	Adhesión celular; transducción de señal mediada por GTPasa pequeña; organización del citoesqueleto de actina	Pronóstico
<i>SH3BGR</i>	Proteína rica en ácido glutámico de unión al dominio SH3	21q22.3	NM_007341	Citoplasma	Ensamblaje de complejos proteicos	Pronóstico
<i>WNT3A</i>	Familia de sitio de integración de MMTV de tipo Wingless, miembro 3A	1q42	NM_033131	Extracelular	Señalización del receptor canónico Wnt; desarrollo de organismos multicelulares; regulación de la proliferación y diferenciación celulares; regulación de la translocación nuclear de la proteína catenina	Pronóstico
<i>ESD</i>	Esterasa D	13q14.1-q14.2	NM_001984	Citoplasma	Reciclado de ácidos siálicos; hidrólisis de serina	Referencia
<i>TBP</i>	Proteína de unión a la caja TATA	6q27	NM_003194	Núcleo	Inicio de la transcripción a partir del promotor de la ARN polimerasa II; elongación de ARN a partir de promotor de la ARN polimerasa II; transcripción a partir de promotor del ARN polimerasa III	Referencia
<i>YAP1</i>	Proteína asociada a Yes 1	11q13	NM_006106	Núcleo	Cascada de señalización de hippo; regulación de la transcripción; inhibición por contacto	Referencia

Figura 4

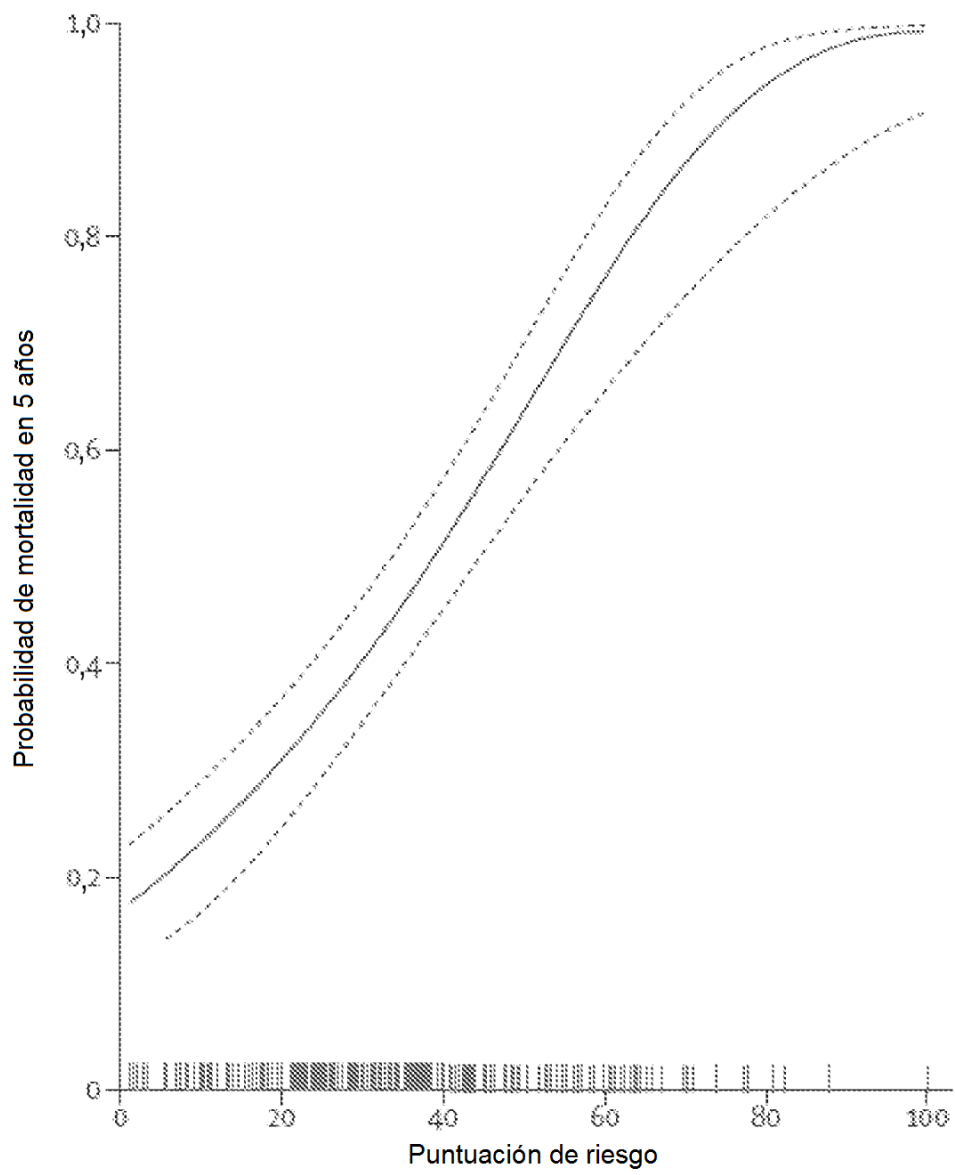


Figura 5A

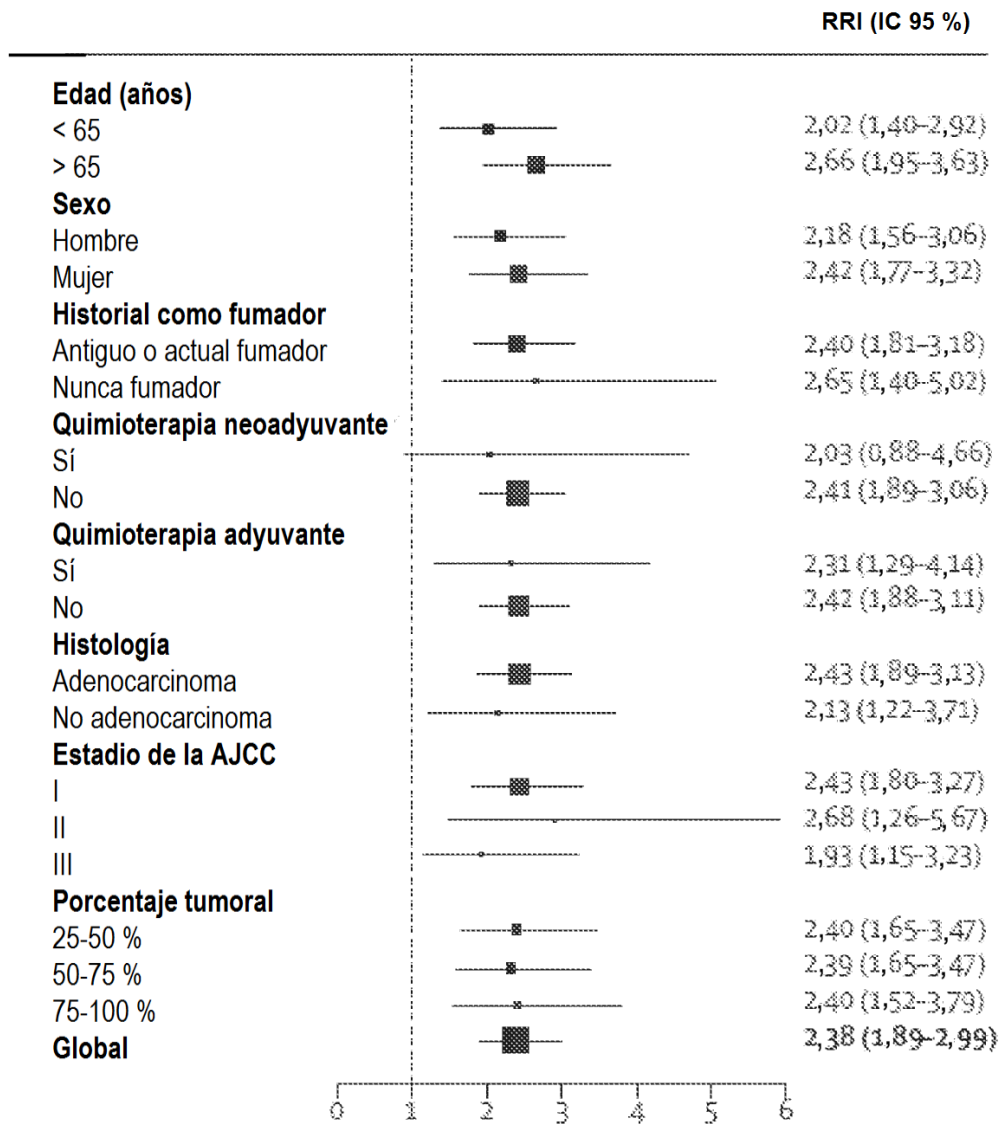


Figura 5B

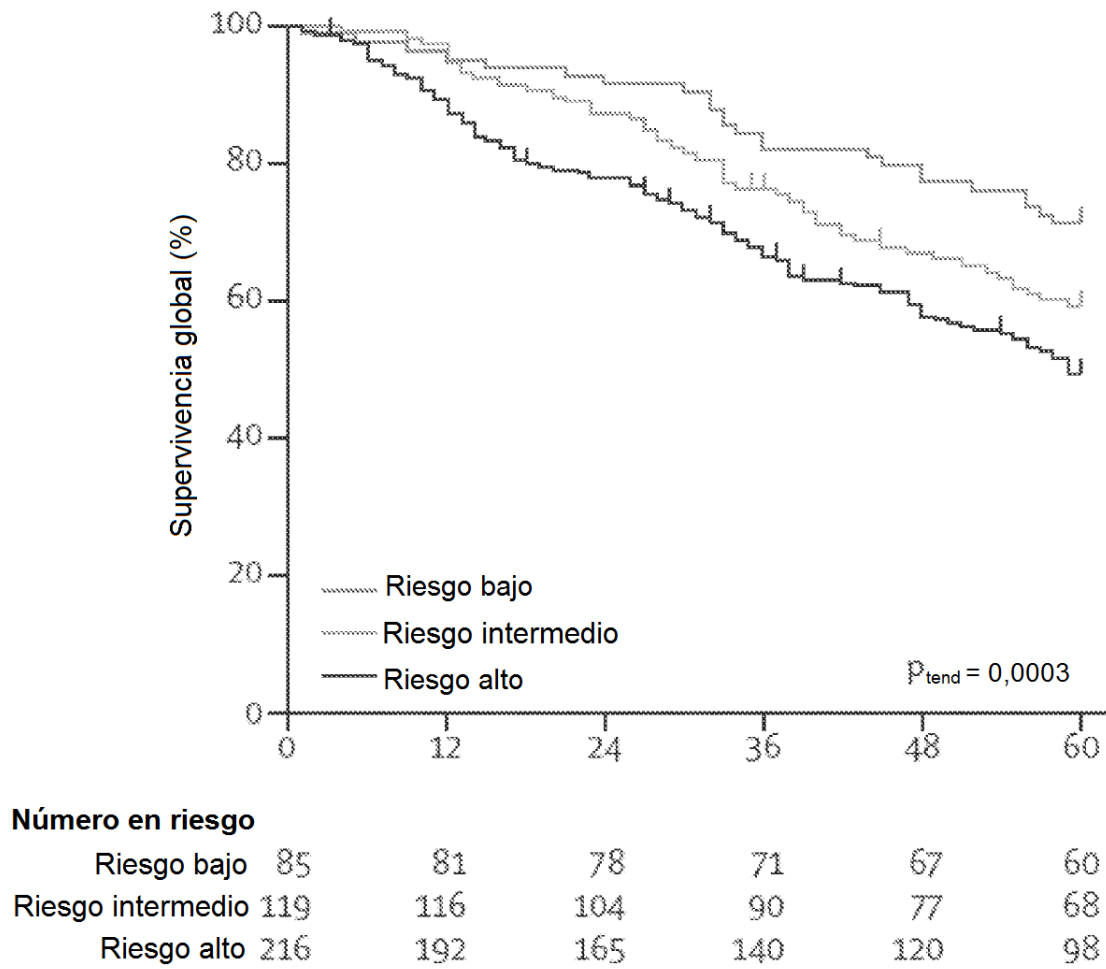


Figura 6A

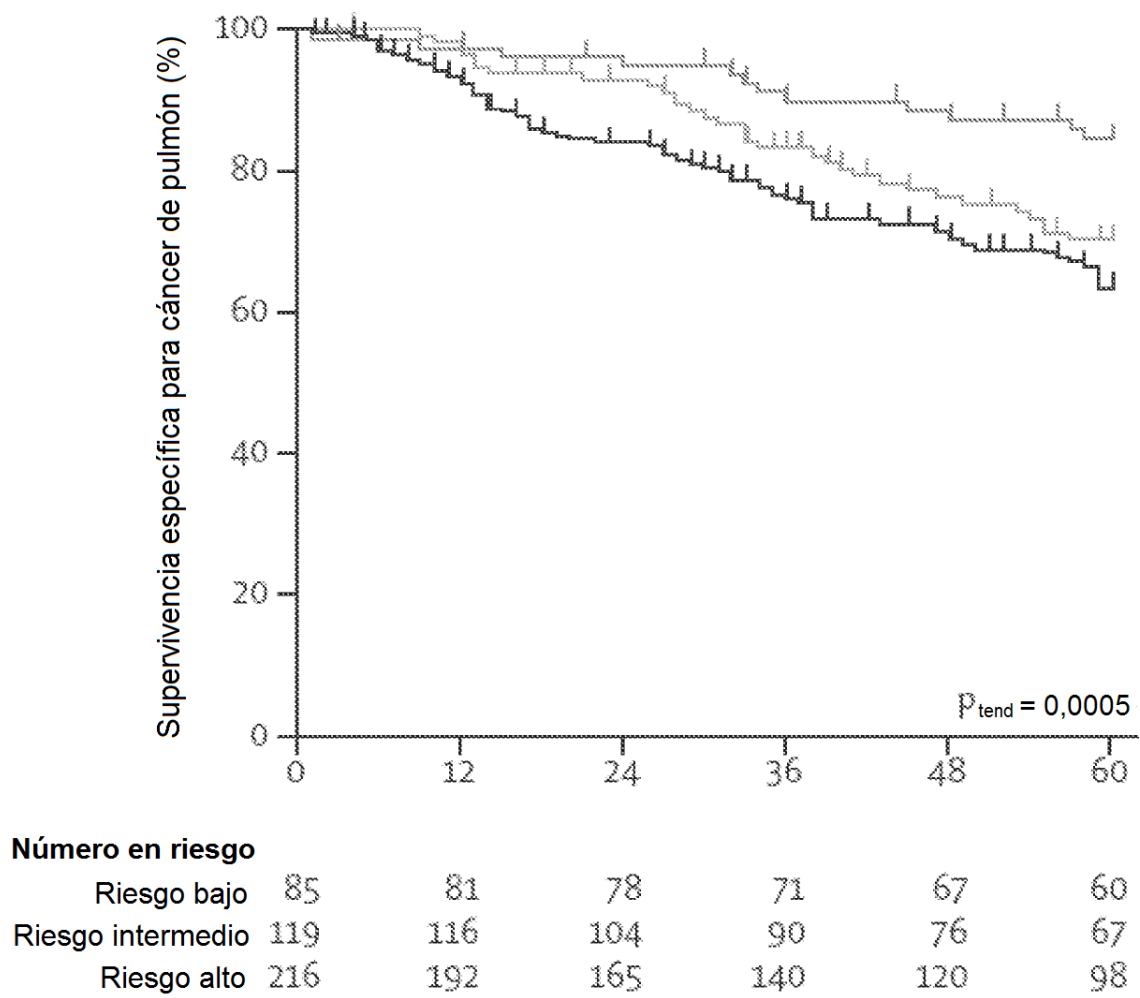


Figura 6B

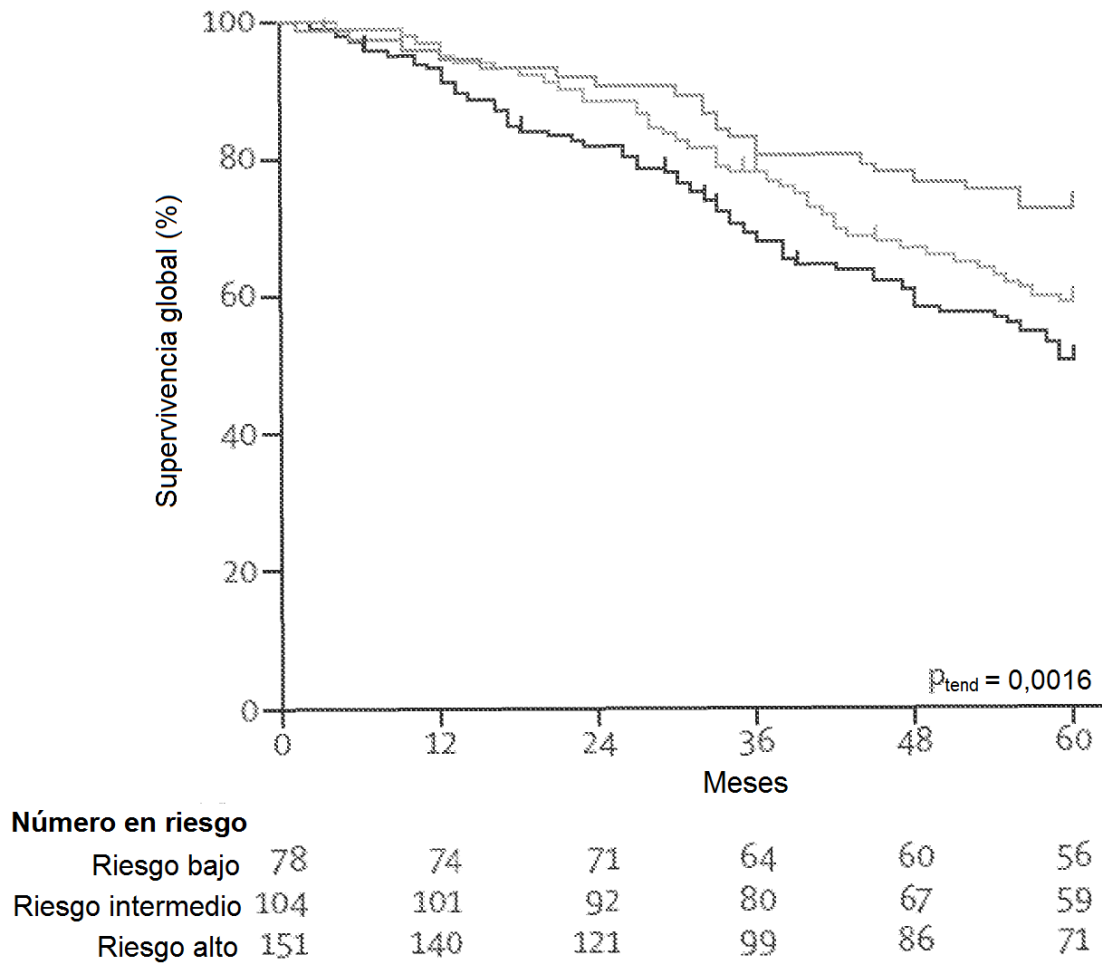


Figura 6C

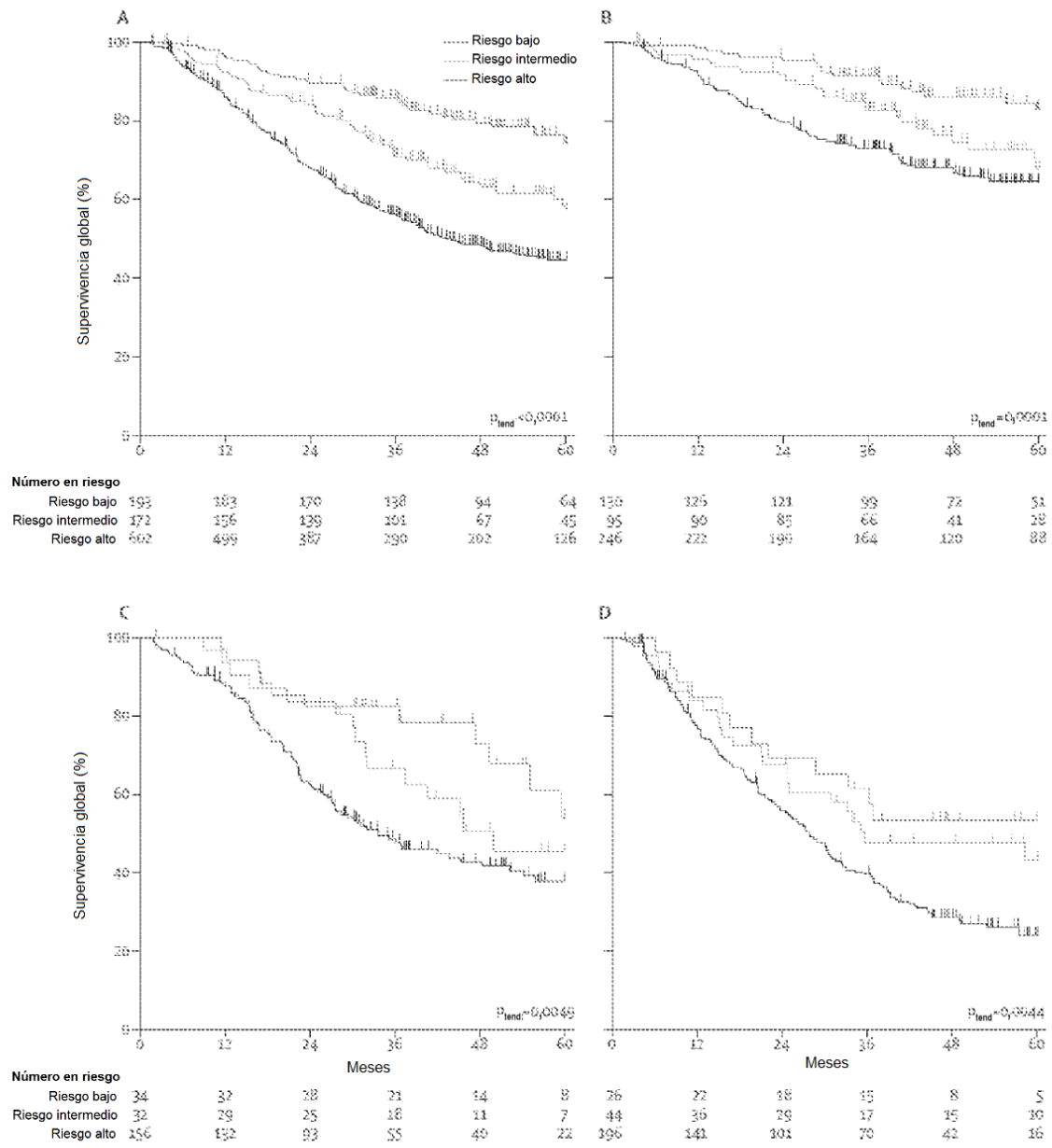


Figura 7

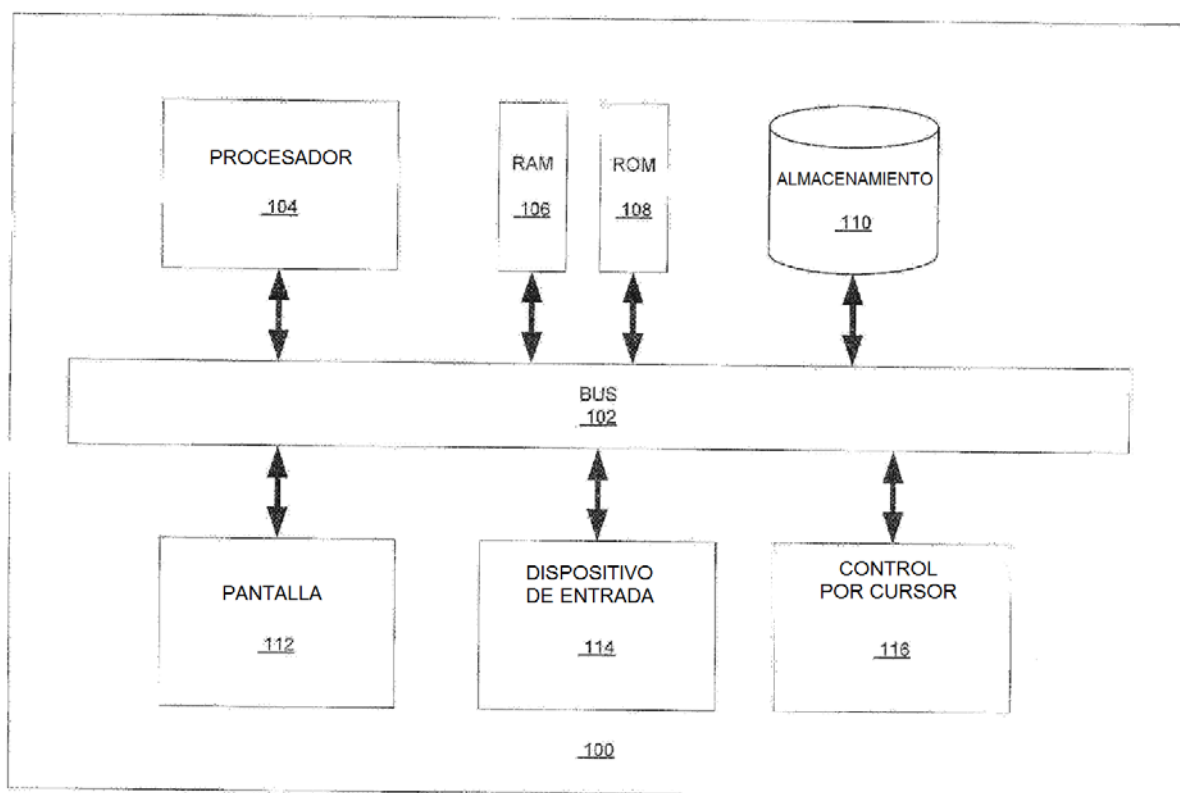


Figura 8

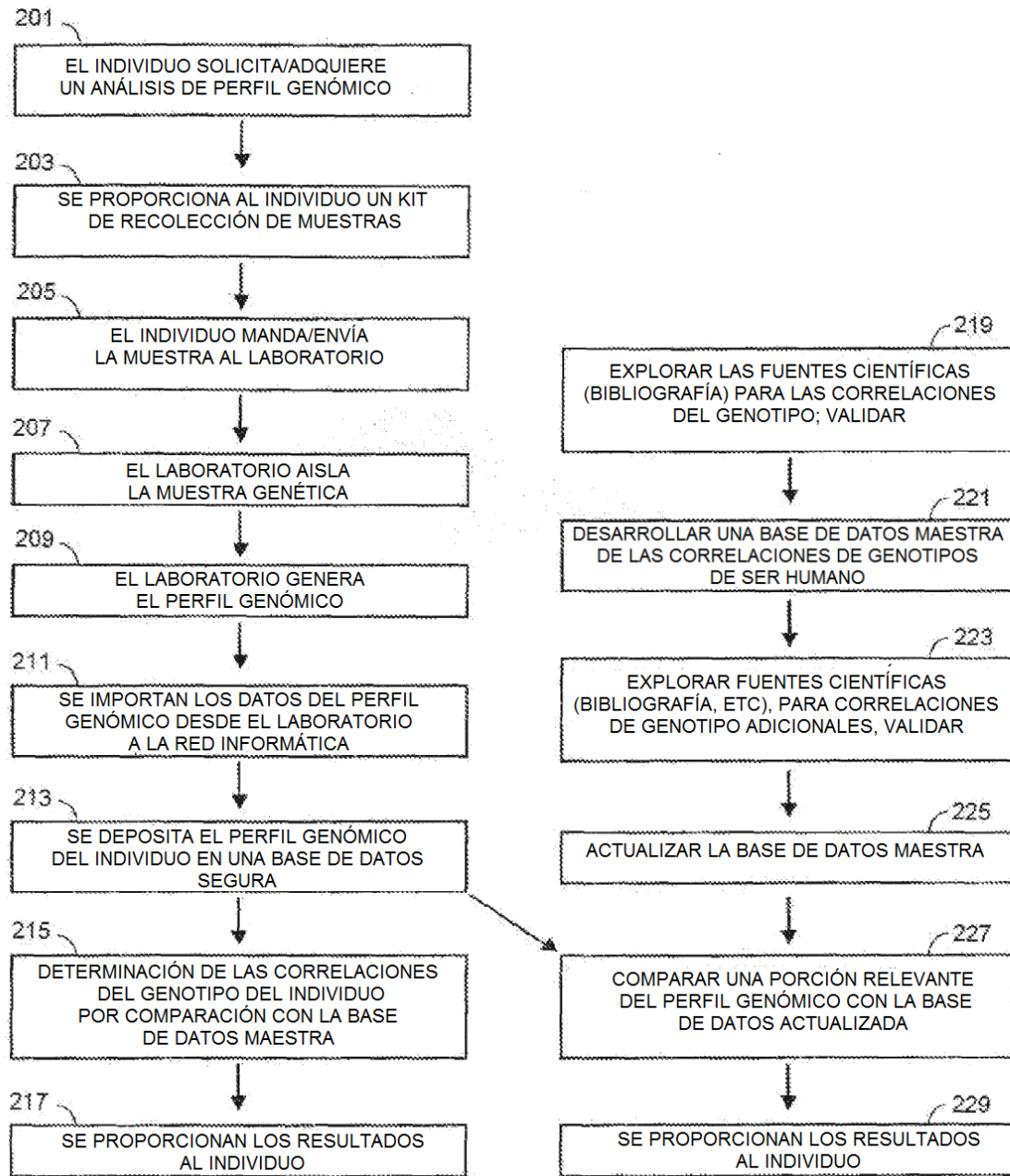


Figura 9