

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 633 596**

51 Int. Cl.:

G01N 33/574 (2006.01)

C07K 14/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **27.06.2008** **E 13166996 (2)**

97 Fecha y número de publicación de la concesión europea: **31.05.2017** **EP 2650683**

54 Título: **Métodos y combinaciones de marcadores para detectar la predisposición al cáncer de pulmón**

30 Prioridad:

29.06.2007 US 771727

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

22.09.2017

73 Titular/es:

**ABBOTT MOLECULAR INC. (100.0%)
1300 East Touhy Avenue
Des Plaines, IL 60018, US**

72 Inventor/es:

**COLPITTS, TRACEY;
RUSSEL, ERIC L.;
FROST, STEPHEN;
RAMIREZ, JAVIER;
SINGH, BHAWANI y
RUSSEL, JOHN C.**

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 633 596 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

Métodos y combinaciones de marcadores para detectar la predisposición al cáncer de pulmón

5 Antecedentes de la invención

El cáncer de pulmón es el segundo cáncer más común tanto en hombres como mujeres en los Estados Unidos, con una estimación de 172.500 nuevos casos que se prevé que se diagnosticarán durante 2005 (estadísticas de la Sociedad Americana del Cáncer). Es la causa más común de muerte por cáncer en ambos sexos, esperándose más de 163.000 muertes relacionadas con el cáncer de pulmón en 2005. El cáncer de pulmón es también un problema sanitario importante en otras áreas del mundo. En la Unión Europea, se producen aproximadamente 135.000 casos cada año (Informe Génesis, Febrero 1995). También está aumentando rápidamente la incidencia en Europa Central y Oriental donde los hombres tienen las tasas de consumo de tabaco más alta del mundo (véase, T. Reynolds, J. Natl. Cancer Inst. 87: 1348-1349 (1995)). El tabaco solo es responsable de más del 90 % de todos los casos de cáncer de pulmón, tráquea, y bronquios (véase, CPMCnet, Guide to Clinical Preventive Services). La Agencia Internacional de Investigación del Cáncer de la Organización Mundial de la Salud estimó que, en 2002, en todo el mundo, había 1.352.000 casos de cáncer de pulmón con 1.179.000 muertes debido a la enfermedad.

El estadio temprano del cáncer de pulmón se puede detectar por radiografía torácica y examen citológico del esputo, sin embargo, estos procedimientos no tienen la suficiente sensibilidad para utilizarse rutinariamente como ensayos de detección en individuos asintomáticos. Los potenciales problemas técnicos que pueden limitar la sensibilidad de la radiografía torácica incluyen la técnica subóptima, exposición insuficiente, y posicionamiento y cooperación del paciente (véase, T.G. Tape, et al., Ann. Intern. Med. 104: 663-670 (1986)). Los radiólogos a menudo no se ponen de acuerdo en la interpretación de las radiografías torácicas y más del 40 % de estas son significativas o potencialmente significativas (véase, P.G. Herman, et al., Chest 68: 278-282 (1975)). Las interpretaciones de falsos negativos son la causa de la mayoría de los errores y los resultados no concluyentes necesitan un ensayo de seguimiento para su clarificación (véase T.G. Tape et al., supra).

La citología del esputo es menos sensible que la radiografía torácica en la detección precoz del cáncer de pulmón (véase, The National Cancer Institute Cooperative Early Lung Cancer Detection Program, Am. Rev. Resp. Dis. 130: 565-567 (1984)). Los factores que afectan a la capacidad de la citología de esputo para diagnosticar el cáncer de pulmón incluyen la capacidad del paciente para producir suficiente esputo, el tamaño del tumor, la proximidad del tumor a las vías aéreas importantes, el tipo histológico del tumor, y la experiencia y entrenamiento del citopatólogo (véase, R.J. Ginsberg et al. En: Cancer: Principles and Practice of Oncology, cuarta Edición, pp. 673-723, Philadelphia, PA: J/B. Lippincott Co. (1993)).

La mayoría de los nuevos cánceres de pulmón se detectarán cuando la enfermedad ya se ha diseminado más allá del pulmón. En los Estados Unidos solamente el 16 % de los nuevos cánceres de pulmón de células no pequeñas se detectan en un estado localizado cuando la supervivencia a 5 años es mayor del 49,7 %. Por el contrario, el 68 % de los nuevos casos se detectan cuando la enfermedad ya se ha diseminado localmente o ha metastatizado a sitios lejanos que tienen tasas de supervivencia a 5 años del 18,5 % y el 1,8 %, respectivamente. De manera similar, el 80 % de los cánceres de pulmón de células pequeñas que se detectan por primera vez se descubren con invasión local o metástasis lejana que tienen tasas de supervivencia a los 5 años del 9,5 % y 1,7 %, respectivamente (véase, Stat Bite, J. Natl. Cancer Inst. 87:1662 (1995)). Estas estadísticas muestran que los procedimientos actuales están fallando en la detección del cáncer de pulmón en un estadio temprano, tratable, de la enfermedad y que se necesitan métodos mejorados de detección y tratamiento para reducir la mortalidad.

Los métodos que se utilizan más frecuentemente para controlar los pacientes de cáncer de pulmón tras la terapia primaria son las visitas clínicas, los rayos X de tórax, el recuento sanguíneo completo, el ensayo de la función hepática y la tomografía computarizada de tórax (CT). La detección de recurrencias por el seguimiento regular, sin embargo, no afecta mucho al modo de tratamiento y el tiempo de supervivencia total lo que lleva a la conclusión de que los métodos actuales de seguimientos no son eficaces respecto al coste (véase, K.S. Naunhcin et al., Ann. Thorac. Surg. 60:1612-1616 (1995); G. L. Walsh et al., Ann. Thorac. Surg. 60: 1563-1572 (1995)).

Más recientemente, se ha hecho un re-evaluación del uso de la tomografía computarizada (CT) para detectar personas asintomáticas que tienen alto riesgo de cáncer de pulmón. C. L. Henschke et al., (Clin. Imaging 28:317-321 (2004)) comunican dos estudios que indicaban que la exploración por CT puede detectar cáncer de pulmón asintomático sin generar demasiado falsos positivos. J. Gohagan et al., (Chest 126:114-121 (2004)) evaluaron un protocolo de prueba para un estudio aleatorio que comparaba los rayos X de tórax con CT espiral de dosis bajas y concluyeron que era fiable en un estudio clínico aleatorizado para explorar el cáncer de pulmón. Sin embargo, aunque está implementado en la práctica clínica, el coste de la detección por CT sería alto y el número de falsos positivos que da lugar a ensayos adicionales también sería alto. Un ensayo de sangre de bajo coste con buena especificidad complementaría la CT para la detección precoz del cáncer. Otra estrategia para mejorar la utilidad de la CT implica el uso de un ensayo de sangre de alta sensibilidad para el cáncer de pulmón en estadio temprano. Tal ensayo se podría ofrecer a los pacientes como una alternativa a la CT o los rayos X. Si el ensayo fuera positivo, el paciente se exploraría por imagen; si el ensayo fuera negativo, el paciente no sería explorado por imagen, pero se

podría re-ensayar en el futuro. Si un ensayo de sangre ofrece una alta sensibilidad o una alta especificidad o, idealmente, ambas, tal ensayo sería útil en los protocolos actuales que se utilizan para detectar el cáncer de pulmón en estado temprano.

- 5 Adicionalmente, ha habido una reciente revisión de los marcadores tumorales y su utilidad cuando se combina en paneles para identificar individuos que tienen riesgo de cáncer de pulmón. Sin embargo, la falta de sensibilidad que era característica de los marcadores individuales evita que los paneles de marcadores tumorales sean útiles para la detección precoz del cáncer de pulmón. Los antígenos tumorales que se habían encontrado por métodos bioquímicos clásicos, incluyendo alfa fetoproteína (AFP), CA19-9, CA125, Antígeno Carcinoembrionario (CEA),
 10 Citoqueratina 8, Citoqueratina 1H, fragmentos de Citoqueratina 19 (CYFRA 21-1), Enolasa Neuronal específica (NSE), Péptido de liberación de pro-Gastrina (proGRP), y Antígeno del Carcinoma de Células Escamosas (SCC). Estos marcadores son útiles en la estadificación, clasificación, predicción de resultado, y seguimiento de pacientes de cáncer de pulmón después de que se ha hecho su diagnóstico; sin embargo, estos marcadores no han sido útiles, sea solos o en paneles, para la detección precoz de la enfermedad (véase, S. Ando, et al., *Anticancer Res.* 21: 3085-3092 (2001); U. S. Preventive Services Task Force, *Annals of Internal Medicine* 140:738-739 (2004)). Por el contrario un panel de inmunoensayo de marcadores conocidos, específicamente, CEA, CYFRA 21-1, SCC, NSE, y ProGRP, se sabe que es útil para hacer un diagnóstico histológico del cáncer de pulmón en circunstancias en las que es difícil obtener una muestra de biopsia (véase, C. Gruber et al., *Tumor Biology* 27 (Suplemento 1): 71 (2006) y P. Stieber et al., *Tumor Biology*, 27 (Suplemento 2):S5-4 (2006)).

20 Se han hecho intentos para descubrir marcadores tumorales mejores para el cáncer de pulmón que identifiquen primero componentes celulares que se expresen diferencialmente en el tejido tumoral de pulmón en comparación con el tejido pulmonar normal. Se ha utilizado la electroforesis en gel de poli(acrilamida) bidimensional para caracterizar las diferencias cuantitativas y cualitativas de la composición de polipéptidos (véase, T. Hirano et al., *Br. J. Cancer* 72:840-848 (1995); A. T. Endler et al., *J. Clin. Chem Clin. Biochem.* 24:981-992 (1986)). La sensibilidad de
 25 esta técnica, sin embargo, está limitada por el grado de resolución proteica de las dos etapas electroforéticas y por la etapa de detección que depende de proteínas de tinción en geles. También, la inestabilidad de los polipéptidos generará artefactos en el patrón bidimensional.

30 También se han hecho intentos para identificar biomarcadores y sus usos en la consecución del diagnóstico del cáncer de pulmón, tales como los que se describen en la Publicación Internacional N° WO 2005/098445 A2. Los biomarcadores tratados en el documento WO 2005/098445 se identificaron utilizando espectrometría de masas de desorción/ionización láser de superficie aumentada (SELDI). Se han desvelado varios marcadores, kits, métodos y métodos clínicos por árboles de decisión. Sin embargo, estos marcadores, kits y métodos no se han adoptado como
 35 de uso práctico rutinario ya que estos marcadores y métodos no se han duplicado en ningún laboratorio.

Se sabe desde hace mucho que el sistema inmunitario humano está implicado en el desarrollo y control del cáncer (véase, K. de Visser, et al., *Nature Reviews Cancer*, 6: 24-37 (2006)). Por lo tanto, no es sorprendente que se hayan explorado las proteínas que representan la respuesta inmunitaria innata y adaptativa como marcadores tumorales
 40 del cáncer de pulmón.

Las proteínas de fase aguda, miembros de la familia de proteínas de la inmunidad innata, incluyendo el amiloide sérico A (SAA), amiloide sérico P (SAP), y proteína reactiva C (CRP), alfa-1-antitripsina (alfa-1-AT), alfa-1-antitripsina (alfa-1-AT), alfa-2-macroglobulina (alfa-2-M), ceruloplasmina (Cp), haptoglobina (Hp), y transferrina (Tf)
 45 se han evaluado como biomarcadores en el cáncer de pulmón temprano, operable (véase, M. Kasprzyk, et al., *Przegl. Lek.*, 63: 936-940 (2006)). El fibrinógeno y CRP se han estudiado como biomarcadores en cáncer de pulmón operable (véase, J. Jones, et al., *Lung Cancer*, 53: 97-101 (2006)). No se ha comunicado la apolipoproteína CIII (apoCIII) como biomarcador del cáncer de pulmón, pero se ha comunicado que está asociado al cáncer pancreático (véase, J. Chen, et al., *J. Chromatogr. A*, (2007), doi: 10.1016/j.chroma.2007.03.096). La asociación de proteínas de fase aguda con el cáncer de pulmón se conoce desde hace más de 2 décadas (véase, P. Weinstein, et al., *Scand. J. Immunol.* 19: 193-198 (1984)); sin embargo, no se utiliza rutinariamente ninguna proteína de fase aguda en el diagnóstico del cáncer de pulmón en estadio temprano.

Hay proteínas que habitualmente no se consideraban como proteínas de fase aguda, pero que están elevadas en respuesta a un estrés del sujeto. Tales proteínas se pueden denominar proteínas de respuesta del huésped, una categoría más amplia que proteínas inmunitarias innatas o adaptativas. Una de tales proteínas es la timosina beta 4 (Tβ4), un péptido de 44 aminoácidos que está implicado en la curación de heridas. Se piensa que la expresión de timosina beta-4 en el tejido tumoral predice un fenotipo del tumor más agresivo y metastático. Los pacientes con
 55 cáncer de pulmón en estadio temprano que expresan ARNm de timosina beta-4 a altos niveles tienen peor supervivencia que los pacientes con baja expresión (véase, C. Muller-Tidow, et al., *Lung Cancer* 45: S145-150 (2004)).

Los autoanticuerpos, inmunoglobulinas que reaccionan con las propias proteínas del paciente (llamadas autoantígenos), comprenden las proteínas que implementan las acciones del sistema inmunitario adaptativo. Las publicaciones científicas sobre la presencia de autoanticuerpos en pacientes con cáncer de pulmón se remontan a
 65 40 años (véase, G. Levine, et al., *J. Lab. Clin. Med.* 69: 749-757 (1967)). De manera ocasional, particularmente en el

cáncer de pulmón de células pequeñas, los autoanticuerpos contra componentes del sistema nervioso dan como resultado que el paciente muestre síntomas neurológicos que se caracterizan por un síndrome paraneoplásico (véase, U. Seneviratne, et al., *Postgrad. Med. J.* 75: 516-520 (1999)). Aunque es dramático, muy pocos pacientes muestran un síndrome paraneoplásico, por tanto ni el síndrome ni los autoanticuerpos contra las células o proteínas neurológicas son útiles diagnósticamente para la mayoría de la población. La bibliografía anterior sobre autoanticuerpos tiene numerosas referencias a anticuerpos p53 circulantes. La P53 es una proteína supresora tumoral que se inactiva por mutación en muchos cánceres. Aunque los anticuerpos anti-p53 del paciente reaccionan contra ambas p53 nativa y mutada, se cree que la semivida tisular más larga de la p53 mutada da como resultado la aparición de este autoanticuerpo en aproximadamente el 30 % de los pacientes de cáncer de pulmón (véase, R. Lubin, et al., *Nat. Med.* 1: 701-702 (1995)). Hay informes de proteínas intracelulares implicadas en la regulación de la maquinaria celular, que dan lugar a autoanticuerpos en el cáncer. Los anticuerpos contra la proteína 40 de choque térmico humano (Hsp40) se ha informado que se encuentra en el suero de pacientes con cáncer de pulmón (véase, M. Oka, et al., *Jpn. J. Cancer Res.* 92: 316-320 (2001)). Esta proteína se conoce también como DnaJB1 o hDJ1. Aunque se han descrito muchas proteínas que se convierten en autoantígenos, dando lugar a autoanticuerpos circulantes en el cáncer de pulmón, debido al millón o más de proteínas que se cree que existen en un sujeto humano, quedan muchas más proteínas, tanto conocidas como desconocidas, que pueden demostrarse como la causa de autoanticuerpos diagnósticamente útiles en pacientes con cáncer de pulmón y otras enfermedades crónicas.

Si se utilizan múltiples autoanticuerpos para clasificar pacientes o hacer un diagnóstico, se necesita un protocolo de análisis de datos. J. Koziol, et al., (*Clin. Cancer Res.* 9:5120-5126 (2003)) utilizaban una partición recursiva para seleccionar los autoanticuerpos para el diagnóstico óptimo de varios tumores. Utilizaron siete antígenos asociados a tumores: c-myc, ciclina B1, IMP2, Koc, p53, p62, y survivina y consiguieron buenos resultados para un pequeño grupo (56) de pacientes de cáncer de pulmón.

Las estrategias sistemáticas para encontrar autoanticuerpos diagnósticamente útiles incluyen el uso de matrices proteicas para detectar pacientes y sueros libres de enfermedad para autoanticuerpos reactivos. Se han construido las matrices a partir de lisados de células tumorales (véase, J. Qiu, et al., *J. Proteome Res.* 3:261-267 (2004)) y se han utilizado para buscar nuevos autoantígenos que dan lugar a firmas de autoanticuerpos diagnósticamente útiles en los sueros de pacientes. Las matrices se pueden construir a partir de proteínas expresadas recombinantemente (véase, D. Mattoon, et al., *Expert. Rev. Proteomics* 2: 879-889 (2005)). Se han descrito las metodologías de cromatografía de afinidad que pueden ser útiles para encontrar autoanticuerpos diagnósticos (*J. Sep. Sci.* 30:352-358 (2007)). Hasta la fecha ninguna de estas estrategias ha producido nuevos autoanticuerpos útiles para la detección precoz del cáncer de pulmón.

Otra estrategia sistemática para descubrir autoanticuerpos diagnósticamente útiles es el método denominado SEREX. SEREX es por Análisis Serológico de Bibliotecas de Expresión de ADNc Recombinante y se ha utilizado para descubrir un nuevo antígeno testicular que se expresa de manera aberrante en una variedad de cánceres (véase, Y. Chen, et al., *Proc. Natl. Acad. Sci. U.S.A.* 94: 1914-1918 (1997)). NY-ESO-1 es una proteína de aproximadamente 18 kD y se ha estudiado extensamente como autoanticuerpo tumoral. Aproximadamente el 20 % de los pacientes con cáncer de pulmón de células no pequeñas tienen anticuerpos circulantes contra la proteína al medirse por un ELISA (véase, O. Tureci, et al., *Cancer Lett.* 236: 64-71 (2006)).

También se han hecho intentos para descubrir una respuesta inmunitaria específica para el cáncer de pulmón por medio de bibliotecas de péptidos de supervisión que se expresan en levaduras o bacterias con sueros de individuos enfermos y no enfermos. Las publicaciones del laboratorio de Hirschowitz (véase, L. Zhong et al., *Chest* 125:105-106 (2004), L. Zhong et al., *Am. J. Respir. Crit. Care Med.* 15:1308-1314 (2005)) han descrito el uso de bibliotecas de fagos para encontrar proteínas que con autoantígenos en pacientes con cáncer de pulmón. Los autores han comunicado de la identificación satisfactoria de pacientes de cáncer de pulmón sintomáticos y asintomáticos en estudios controlados. Sin embargo, el número de casos y controles es limitado (<200 sujetos totales) y el método tiene que validarse en una población mucho más grande.

Actualmente, la identificación de individuos en riesgo de cáncer de pulmón se basa mucho en la historia de fumador del individuo. Otras exposiciones ambientales tales como asbestos, partículas, etc., también pueden aumentar el riesgo de desarrollar cáncer de pulmón. Estos factores de riesgo conocidos se han combinado en uno o más algoritmos y están accesibles para los clínicos y el público para evaluar el riesgo de un individuo al cáncer de pulmón (véase, P. B. Bach et al., *J. Natl. Cancer Inst.* 95:470-478 (2003)). Desafortunadamente, este algoritmo no es suficientemente sensible ni específico para ser útil en la detección del cáncer de pulmón en estadio temprano. Además, basándose en el citado algoritmo, un individuo con una historia de fumador significativa tendrá un riesgo de 1/500 a 1/100 de desarrollar un cáncer de pulmón. Esto significa que incluso utilizando el método de Bach et al., hasta 499 de 500 exploraciones por CT no dan lugar al descubrimiento de un caso de cáncer de pulmón.

Por consiguiente, sigue habiendo una necesidad en la técnica de métodos y marcadores útiles para detectar el cáncer de pulmón que sea rápida, conveniente y económicamente eficaz para llevarlos a cabo. También sería ventajoso proporcionar métodos específicos y ciertas combinaciones de marcadores que se podrían utilizar para determinar si un paciente tiene riesgo de desarrollar un cáncer de pulmón o tiene cáncer de pulmón. Tales métodos

incluirían un método para ensayar uno o más marcadores en una muestra que indiquen el cáncer de pulmón y detectar tales marcadores. Tales métodos pueden incluir métodos mejorados para analizar espectros de masas de una muestra biológica para detectar marcadores o ensayar una muestra y entonces detectar marcadores como riesgo de desarrollar cáncer de pulmón o como indicación de cáncer de pulmón.

El documento WO 2007/076439 desvela métodos y combinaciones de marcadores para explorar la predisposición al cáncer de pulmón. Pollan et al., International Journal of Cancer (2003), 107:781-790 desvelan el valor clínico de p53, c-erbB-2, CEA y CA125 con respecto a la recaída, metástasis y muerte en cáncer de pulmón de células no pequeñas reseccionable.

Sumario de la invención

La invención se basa en parte en el descubrimiento de que se pueden basar métodos rápidos y sensibles que ayuden a la detección del cáncer de pulmón en un sujeto sospechoso de tener un cáncer de pulmón en ciertas combinaciones de biomarcadores y biomarcadores y parámetros biométricos.

El alcance de la invención y distintas realizaciones de la invención se definen por las reivindicaciones adjuntas. La presente invención proporciona en primer lugar un método para ayudar en el diagnóstico de un sujeto sospechoso de cáncer de pulmón, comprendiendo el método las etapas de:

- (a) cuantificar las cantidades de marcadores en una muestra de ensayo obtenida de un sujeto, donde los biomarcadores que se cuantifican son al menos anti-MAPKAPK3 anti-NY-ESO-1, CEA, CA125, citoqueratina 19, anti-p53, y anti-ciclina E2;
- (b) comparar la cantidad de cada biomarcador cuantificado con un punto de corte predeterminado de dicho biomarcador y asignar una puntuación para cada biomarcador basándose en dicha comparación;
- (c) combinar la puntuación asumida de cada biomarcador para obtener una puntuación total para dicho sujeto;
- (d) comparar la puntuación total determinada en la etapa (c) con una puntuación total predeterminada; y
- (e) determinar si dicho sujeto tiene riesgo de cáncer de pulmón basándose en la comparación de la puntuación total de la etapa (d).

La presente invención proporciona adicionalmente un kit que comprende:

- (a) reactivos que contienen antígenos para cuantificar anti-MAPKAPK3, anti-NY-ESO-1, anti-p53 y anti-ciclina E2;
- (b) reactivos para cuantificar CEA, CA125, y citoqueratina 19;
- (c) reactivos para cuantificar al menos uno o más de los biomarcadores que se seleccionan de entre el grupo que consiste en SCC, ProGRP, TPS, Pub11597, Pub4789, TFA2759, y Acn9459; y opcionalmente,
- (d) algoritmos para combinar y comparar la cantidad de cada antígeno y anticuerpo cuantificado en la muestra de ensayo frente a un valor de corte predeterminado y asignar una puntuación para cada antígeno y anticuerpo cuantificado para obtener una puntuación total, comparar la puntuación total con una puntuación total predeterminada y usar dicha comparación como una ayuda para determinar si el sujeto tiene cáncer de pulmón.

Además de la invención, la presente divulgación se refiere a otros aspectos distintos. La siguiente divulgación se refiere a la invención siempre que sea consistente con el alcance de las reivindicaciones como se ha definido anteriormente y en las reivindicaciones adjuntas.

En un aspecto, el método de la presente divulgación puede comprender las etapas de:

- a. cuantificar en una muestra de ensayo obtenida de un sujeto la cantidad de uno o más biomarcadores en un panel;
- b. comparar la cantidad de cada biomarcador en el panel con un valor de corte predeterminado para dicho biomarcador y asignar una puntuación para cada biomarcador basándose en dicha comparación;
- c. combinar la puntuación asignada para cada biomarcador determinada en la etapa b para obtener una puntuación total para dicho sujeto;
- d. comparar la puntuación total determinada en la etapa c con una puntuación total predeterminada; y
- e. determinar si dicho sujeto tiene riesgo de padecer cáncer de pulmón basándose en la comparación de la puntuación total en la etapa d.

En el método anterior, el DFI de los biomarcadores relacionados con el cáncer de pulmón preferiblemente es menor de aproximadamente 0,4.

Opcionalmente, el método anterior puede comprender además la etapa de obtener un valor para al menos un parámetro biométrico de un sujeto. Un ejemplo de un parámetro biométrico que se puede obtener es la historia de fumador del sujeto. Si el método anterior comprende además la etapa de obtener un valor para al menos un parámetro biométrico del sujeto, entonces el método puede comprender además la etapa de comparar el valor del al menos un parámetro biométrico con un punto de corte predeterminado para cada dicho parámetro biométrico y

asignar una puntuación para cada parámetro biométrico basándose en dicha comparación, combinando la puntuación asignada para cada parámetro biométrico con la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para cada sujeto de la etapa c, comparando la puntuación total con una puntuación total predeterminada en la etapa d y determinando si dicho sujeto tiene riesgo de cáncer de pulmón basándose en la puntuación total de la etapa e.

Ejemplos de biomarcadores que se pueden cuantificar en el método anterior son uno o más biomarcadores seleccionados de entre el grupo de anticuerpos, antígenos, regiones de interés o cualquier combinación de los mismos. Más específicamente, los biomarcadores que se pueden cuantificar incluyen, pero no se limitan a uno o más de: anti-p53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, proteína I tipo C1 anti-Niemann-Pick, dominio del péptido del extremo C (anti-NPC1L1C-dominio), anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSIN1, anti-RCV1, anti-MAPKAPK3, anti-Ciclina E2 (concretamente, al menos un anticuerpo contra la Ciclina E2 inmunorreactiva), citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide sérico A, alfa-1-anti-tripsina, apolipoproteína CIII y Tβ4, Acn6399, Acn9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3 959.

Opcionalmente, el panel que se utiliza en el método anterior puede comprender la cuantificación de la cantidad de dos o más biomarcadores, tres o más biomarcadores, cuatro o más biomarcadores, cinco o más biomarcadores, seis o más biomarcadores, siete o más biomarcadores, ocho o más biomarcadores, nueve o más biomarcadores, diez o más biomarcadores, once o más biomarcadores, doce o más biomarcadores, trece o más biomarcadores, catorce o más biomarcadores, quince o más biomarcadores, dieciséis o más biomarcadores, diecisiete o más biomarcadores, dieciocho o más biomarcadores, diecinueve o más biomarcadores o veinte biomarcadores o más, etc.

En otro aspecto, el método de la presente divulgación puede comprender las etapas de:

- a. obtener un valor para al menos un parámetro biométrico de un sujeto;
- b. comparar el valor del al menos un parámetro biométrico con un punto de corte predeterminado para cada dicho parámetro biométrico y asignar una puntuación para cada parámetro biométrico basándose en dicha comparación;
- c. cuantificar en una muestra de ensayo obtenida de un sujeto, la cantidad de dos o más biomarcadores de un panel, el panel comprende al menos un anticuerpo y al menos un antígeno;
- d. comparar la cantidad de cada biomarcador cuantificado en el panel para un punto de corte predeterminado para dicho biomarcador y asignar una puntuación para cada biomarcador basándose en dicha comparación;
- e. combinar la puntuación asignada para cada parámetro biométrico determinada en la etapa b con la puntuación asignada para cada biomarcador determinada en la etapa d para obtener una puntuación total para dicho sujeto;
- f. comparar la puntuación total determinada en la etapa e con una puntuación total predeterminada; y
- g. determinar si dicho sujeto tiene riesgo de cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa f.

En el método anterior, el DFI de los biomarcadores relacionados con el cáncer de pulmón es preferentemente menor de aproximadamente 0,4.

En el método anterior, el panel puede comprender al menos un anticuerpo que se selecciona de entre el grupo que consiste en: anti-p53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, anti-dominio-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSIN1, anti-RCV1, anti-MAPKAPK3 y anti-Ciclina E2 y al menos un antígeno que se selecciona de entre el grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA 19-9, proGRP, amiloide sérico A, alfa-1-anti-tripsina, apolipoproteína CIII y Tβ4.

En el método anterior. El parámetro biométrico que se obtiene del sujeto se selecciona de entre el grupo que consiste de la historia de fumador del sujeto, la edad, exposición a carcinógenos y género. Preferentemente, el parámetro biométrico es el de paquetes años como fumador del sujeto.

Opcionalmente, el método puede comprender además la cuantificación de al menos una región de interés en la muestra de ensayo. Si una región de interés se va a cuantificar en la muestra de ensayo, entonces el panel puede comprender además al menos una región de interés que se selecciona de entre el grupo que consiste en: Acn6399, Acn9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959.

Opcionalmente. El método anterior puede emplear también un Método de Puntuación Compensada, para determinar si un sujeto está en riesgo de desarrollar un cáncer de pulmón. Si el método anterior emplea tal Método de Puntuación Compensada, entonces en dicho método, la etapa b comprendía la comparación del valor de al menos un parámetro biométrico con un número de puntos de corte predeterminados para dicho parámetro biométrico y la asignación de una puntuación para cada parámetro biométrico basándose en dicha comparación, la etapa d comprende la comparación de la cantidad de cada biomarcador en el panel con un varios puntos de corte predeterminados para dicho biomarcador y la asignación de una puntuación para cada biomarcador basándose en dicha comparación, la etapa e comprende la combinación de la puntuación asignada a cada parámetro biométrico en

la etapa b con la puntuación asignada para cada biomarcador en la etapa d para llegar a una puntuación total para dicho sujeto, la etapa f comprende la comparación de la puntuación total determinada en la etapa e con una puntuación total predeterminada y la etapa g comprende determinar si dicho sujeto tiene cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa f.

5

En otro aspecto, el método de la presente divulgación puede comprender las etapas de:

- a. cuantificar en una muestra de ensayo obtenida de un sujeto, la cantidad de dos o más biomarcadores en un panel, el panel comprende al menos un anticuerpo y al menos un antígeno;
- 10 b. comparar la cantidad de cada biomarcador cuantificado en el panel con un punto de corte predeterminado para dicho biomarcador y asignar una puntuación para cada biomarcador basándose en dicha comparación;
- c. combinar la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto;
- d. comparar la puntuación total determinada en la etapa c con una puntuación total predeterminada; y
- 15 e. determinar si dicho sujeto tiene riesgo de cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa d.

En el método anterior, el DFI de los biomarcadores relacionados con cáncer de pulmón es preferentemente menor de aproximadamente 0,4.

20

En el método anterior, el panel puede comprender al menos un anticuerpo seleccionado de entre el grupo que consiste en: anti-p53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, anti-dominio-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PAC1SIN1, anti-RCV1, anti-MAPKAPK3 y anti-Ciclina E2. El panel puede comprender al menos un antígeno seleccionado de entre el grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide sérico A, alfa-1-anti-tripsina, apolipoproteína CIII y Tβ4.

25

Opcionalmente, el método puede comprender además la cuantificación de al menos una región de interés en la muestra de ensayo. Si se va a cuantificar una región de interés, entonces el panel puede comprender además al menos una región de interés que se selecciona de entre el grupo que consiste en: Acn6399, Acn9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959.

30

Opcionalmente, el método anterior puede emplear también un Método de Puntuación Ponderada, para determinar si un sujeto tiene riesgo de desarrollar cáncer de pulmón. Si el método anterior emplea dicho Método de Puntuación Ponderada, entonces en dicho método, la etapa b comprende la comparación de la cantidad de cada biomarcador del panel con varios puntos de corte predeterminados para dicho biomarcador y la asignación de una puntuación para cada biomarcador basándose en dicha comparación, la etapa c comprende la combinación de la puntuación asignada para cada biomarcador cuantificada en la etapa b para obtener una puntuación total para dicho sujeto, la etapa d comprende la comparación de la puntuación total determinada en la etapa c con una puntuación total predeterminada y la etapa e comprende determinar si dicho sujeto tiene cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa d.

35

40

En otro aspecto, el método de la presente divulgación puede comprender las etapas de:

- 45 a. cuantificar en una muestra de ensayo obtenida de un sujeto, una cantidad de al menos un biomarcador del panel, el panel comprende al menos un anti-Ciclina E2;
- b. comparar la cantidad de cada biomarcador cuantificado del panel con un punto de corte predeterminado para dicho biomarcador y asignar una puntuación para cada biomarcador basándose en dicha comparación;
- c. combinar la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto;
- 50 d. comparar la puntuación total determinada en la etapa c con una puntuación total predeterminada; y
- e. determinar si dicho sujeto tiene un riesgo de cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa d.

55 En el método anterior, el DFI de los biomarcadores relacionados con el cáncer de pulmón es preferentemente menor de aproximadamente 0,4.

Opcionalmente, el método anterior puede comprender además la cuantificación de al menos un antígeno en la muestra de ensayo, cuantificar al menos un anticuerpo en la muestra de ensayo, o cuantificar una combinación de al menos un antígeno y al menos un anticuerpo en la muestra de ensayo. Por consiguiente, si al menos un antígeno, al menos un anticuerpo o una combinación de al menos un antígeno y al menos un anticuerpo se van a cuantificar en la muestra de ensayo, entonces el panel puede comprender además al menos un antígeno que se selecciona de entre el grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide sérico A, alfa-1-anti-tripsina, apolipoproteína CIII y Tβ4, al menos un anticuerpo que se selecciona de entre el grupo que consiste en: anti-p53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, anti-dominio-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PAC1SIN1, anti-RCV1, anti-MAPKAPK3 o cualquier

60

65

combinación de los mismos.

Opcionalmente, el método puede comprender además la cuantificación de al menos una región de interés en la muestra de ensayo. Si se va a cuantificar una región de interés, entonces, el panel puede comprender además al menos una región de interés que se selecciona de entre el grupo que consiste en: Acn6399, Acn9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959.

Opcionalmente, el método anterior también puede emplear un Método de Puntuación Ponderada para determinar si un sujeto está en riesgo de desarrollar un cáncer de pulmón. Si el método anterior emplea tal Método de Puntuación Ponderada, entonces en dicho método, la etapa b comprende la comparación de la cantidad de cada biomarcador del panel con varios puntos de corte predeterminados para dicho biomarcador y la asignación de una puntuación para cada biomarcador basándose en dicha comparación, la etapa c comprende la combinación de la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto, la etapa d comprende la comparación de la puntuación total determinada en la etapa c con una puntuación total predeterminada y la etapa e comprende determinar si dicho sujeto tiene cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa d.

Opcionalmente, el método anterior puede comprender además la etapa de obtención de un valor para al menos un parámetro biométrico de un sujeto. Un parámetro biométrico que se puede obtener de un sujeto se puede seleccionar de entre el grupo que consiste en: la historia de fumador de un sujeto, edad, exposición a carcinógenos y género. Un parámetro biométrico preferido que se obtiene es el de paquetes años de fumador del sujeto. Si el método anterior comprende además la etapa de obtención de un valor para al menos un parámetro biométrico del sujeto, entonces el método puede comprender además la etapa de comparación del valor de al menos un parámetro biométrico con un punto de corte predeterminado para dicho parámetro biométrico y la asignación de una puntuación para cada parámetro biométrico basándose en dicha comparación, la combinación de la puntuación asignada para cada parámetro biométrico con la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto, comparando la puntuación total con una puntuación predeterminada total de la etapa c y determinar si dicho sujeto tiene un riesgo de cáncer de pulmón basándose en la comparación de la puntuación total en la etapa d.

En otro aspecto, el método de la presente divulgación puede comprender las etapas de:

- a. cuantificar en una muestra de ensayo obtenida de un sujeto al menos un biomarcador en un panel, el panel comprende al menos un biomarcador que se selecciona de entre el grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide sérico A, alfa-1-antitripsina, apolipoproteína CIII y Tβ4;
- b. comparar la cantidad de cada biomarcador cuantificado en el panel para un punto de corte predeterminado para dicho biomarcador y asignar una puntuación para cada biomarcador basándose en dicha comparación;
- c. combinar la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto;
- d. comparar la puntuación total cuantificada en la etapa c con una puntuación predeterminada total; y
- e. determinar si dicho sujeto tiene cáncer de pulmón basándose en la comparación de la puntuación total de la etapa d.

En el método anterior, el DFI de los biomarcadores relacionados con el cáncer de pulmón es preferentemente menor de aproximadamente 0,4.

Opcionalmente, el método anterior puede comprender además la cuantificación de al menos un anticuerpo en la muestra de ensayo. Por consiguiente, el panel puede comprender además al menos un anticuerpo que se selecciona de entre el grupo que consiste en: anti-p53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, anti-dominio-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSIN1, anti-RCV1, anti-MAPKAPK3 y anti-Ciclina E2 o cualquier combinación de los mismos.

Opcionalmente, el método anterior puede comprender además la cuantificación de al menos una región de interés en la muestra de ensayo. Si se va a cuantificar una región de interés, entonces el panel puede comprender además al menos una región de interés que se selecciona de entre el grupo que consiste en: Acn6399, Acn9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959.

Opcionalmente, el método anterior también puede emplear un Método de Puntuación Ponderada para determinar si un sujeto está en riesgo de desarrollar un cáncer de pulmón. Si el método anterior emplea tal Método de Puntuación Ponderada, entonces en dicho método, la etapa b comprende la comparación de la cantidad de cada biomarcador del panel con varios puntos de corte predeterminados para dicho biomarcador y la asignación de una puntuación para cada biomarcador basándose en dicha comparación, la etapa c comprende la combinación de la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto, la

etapa d comprende la comparación de la puntuación total determinada en la etapa c con una puntuación total predeterminada y la etapa e comprende determinar si dicho sujeto tiene cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa d.

- 5 Opcionalmente, el método anterior puede comprender además la etapa de obtención de un valor para al menos un parámetro biométrico de un sujeto. Un parámetro biométrico que se puede obtener de un sujeto se puede seleccionar de entre el grupo que consiste en: la historia de fumador de un sujeto, edad, exposición a carcinógenos y género. Un parámetro biométrico preferido que se obtiene es el de paquetes años de fumador del sujeto. Si el método anterior comprende además la etapa de obtención de un valor para al menos un parámetro biométrico del sujeto, entonces el método puede comprender la etapa de comparación del valor de al menos un parámetro biométrico con un punto de corte predeterminado para dicho parámetro biométrico y la asignación de una puntuación para cada parámetro biométrico basándose en dicha comparación, la combinación de la puntuación asignada para cada parámetro biométrico con la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto, comparando la puntuación total con una puntuación predeterminada total de la etapa c y determinar si dicho sujeto tiene un riesgo de cáncer de pulmón basándose en la comparación de la puntuación total en la etapa d.

En otro aspecto, el método de la presente divulgación puede comprender las etapas de:

- 20 a. cuantificar en una muestra de ensayo obtenida del sujeto, al menos un biomarcador en un panel, el panel comprende al menos un biomarcador, donde el biomarcador es una región de interés que se selecciona de entre el grupo que consiste en: Acn6399, Acn9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959.
 25 b. comparar la cantidad de cada biomarcador cuantificado en el panel con un punto de corte predeterminado para dicho biomarcador y asignar una puntuación para cada biomarcador basándose en dicha comparación;
 c. combinar la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto;
 d. comparar la puntuación total cuantificada en la etapa c con una puntuación total predeterminada; y
 30 e. determinar si dicho sujeto tiene un cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa d.

En el método anterior, el DFI de los biomarcadores relacionados con el cáncer de pulmón es preferentemente menor de aproximadamente 0,4.

- 35 Opcionalmente, el método anterior puede comprender además la cuantificación de al menos un antígeno en la muestra de ensayo, cuantificar al menos un anticuerpo en la muestra de ensayo, o cuantificar una combinación de al menos un antígeno y al menos un anticuerpo en la muestra de ensayo. Por consiguiente, si al menos un antígeno, al menos un anticuerpo o una combinación de al menos un antígeno o un anticuerpo se van a cuantificar en la muestra de ensayo, entonces el panel puede comprender además al menos un antígeno que se selecciona de entre el grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide sérico A, alfa-1-anti-tripsina, apolipoproteína CIII y Tβ4, al menos un anticuerpo que se selecciona de entre el grupo que consiste en: anti-p53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, anti-dominio-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSN1, anti-RCV1, anti-MAPKAPK3 o cualquier combinación de los mismos.
- 40
- 45 Opcionalmente, el método anterior también puede emplear un Método de Puntuación Ponderada para determinar si un sujeto está en riesgo de desarrollar un cáncer de pulmón. Si el método anterior emplea tal Método de Puntuación Ponderada, entonces en dicho método, la etapa b comprende la comparación de la cantidad de cada biomarcador del panel con varios puntos de corte predeterminados para dicho biomarcador y la asignación de una puntuación para cada biomarcador basándose en dicha comparación, la etapa c comprende la combinación de la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto, la etapa d comprende la comparación de la puntuación total determinada en la etapa c con una puntuación total predeterminada y la etapa e comprende determinar si dicho sujeto tiene cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa d.
- 50
- 55 Opcionalmente, el método anterior puede comprender además la etapa de obtención de un valor para al menos un parámetro biométrico de un sujeto. Un parámetro biométrico que se puede obtener de un sujeto se puede seleccionar de entre el grupo que consiste en: la historia de fumador de un sujeto, edad, exposición a carcinógenos y género. Un parámetro biométrico preferido que se obtiene es el de paquetes años de fumador del sujeto. Si el método anterior comprende además la etapa de obtención de un valor para al menos un parámetro biométrico del sujeto, entonces el método puede comprender la etapa de comparación del valor de al menos un parámetro biométrico con un punto de corte predeterminado para dicho parámetro biométrico y la asignación de una puntuación para cada parámetro biométrico basándose en dicha comparación, la combinación de la puntuación asignada para cada parámetro biométrico con la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto, comparando la puntuación total con una puntuación predeterminada total de la etapa c y determinar si dicho sujeto tiene un riesgo de cáncer de pulmón basándose en la comparación de la puntuación total en la etapa d.
- 60
- 65

En otro aspecto, el método de la presente divulgación puede comprender las etapas de:

- a. cuantificar en una muestra de ensayo obtenida de un sujeto, la cantidad de dos o más biomarcadores en un panel, el panel comprende dos o más de: citoqueratina 19, citoqueratina 18, CA 19-9, CEA, CA15-3, CA125, SCC, ProGRP, ACN9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 e Hic3959;
- b. comparar la cantidad de cada biomarcador en el panel con un punto de corte predeterminado para dicho biomarcador y asignar una puntuación para cada biomarcador basándose en dicha comparación;
- c. combinar la puntuación asignada para cada biomarcador determinada en la etapa b para obtener una puntuación total para dicho sujeto;
- d. comparar la puntuación total determinada en la etapa c con una puntuación total predeterminada; y
- e. determinar si dicho sujeto tiene riesgo de cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa d.

En el método anterior, el DFI de los biomarcadores relacionados con el cáncer de pulmón es preferentemente menor de aproximadamente 0,4.

Opcionalmente, el panel en el método anterior puede comprender: (1) citoqueratina 19, CEA, ACN9459, Pub11597, Pub4789 y TFA2759; (2) citoqueratina 19, CEA, ACN9459, Pub11597, Pub4789, TFA2759 y TFA9133; (3) citoqueratina 19, CA19-9, CEA, CA15-3, CA125, SCC, citoqueratina 18 y ProGRP; (4) Pub11597, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 e HIC3959; o (5) citoqueratina 19, CEA, CA125, SCC, citoqueratina 18, ProGRP, ACN9459, Pub11597, Pub4789, TFA2759, TFA9133.

Opcionalmente, el método anterior también puede emplear un Método de Puntuación Ponderada para determinar si un sujeto está en riesgo de desarrollar un cáncer de pulmón. Si el método anterior emplea tal Método de Puntuación Ponderada, entonces en dicho método, la etapa b comprende la comparación de la cantidad de cada biomarcador del panel con varios puntos de corte predeterminados para dicho biomarcador y la asignación de una puntuación para cada biomarcador basándose en dicha comparación, la etapa c comprende la combinación de la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto, la etapa d comprende la comparación de la puntuación total determinada en la etapa c con una puntuación total predeterminada y la etapa e comprende determinar si dicho sujeto tiene cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa d.

En otro aspecto, el método de la presente divulgación puede comprender las etapas de:

- a. cuantificar en una muestra de ensayo obtenida de un sujeto, la cantidad de dos o más biomarcadores en un panel, donde el panel comprende anti-p53, anti-NY-ESO-1, anti-MAPKAPK3, anti-Ciclina E2, citoqueratina 19, CEA, CA125 y ProGRP;
- b. comparar la cantidad de cada biomarcador en el panel con un punto de corte predeterminado para dicho biomarcador y asignar una puntuación para cada biomarcador basándose en dicha comparación;
- c. combinar la puntuación asignada para cada biomarcador determinada en la etapa b para obtener una puntuación total para dicho sujeto;
- d. comparar la puntuación total determinada en la etapa c con una puntuación total predeterminada; y
- e. determinar si dicho sujeto tiene riesgo de cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa d.

Opcionalmente, el panel puede comprender la cuantificación de la cantidad de tres o más biomarcadores, cuatro o más biomarcadores, cinco o más biomarcadores, seis o más biomarcadores, siete o más biomarcadores u ocho biomarcadores.

Opcionalmente, en el método anterior, la anti-Ciclina E2 puede ser un anticuerpo contra la SEQ ID NO: 3, SEQ ID NO: 4 o la SEQ ID NO: 5.

Opcionalmente, el método anterior puede comprender además la etapa de obtención de un valor para al menos un parámetro biométrico de un sujeto. Un parámetro biométrico que se puede obtener de un sujeto se puede seleccionar de entre el grupo que consiste en: la historia de fumador de un sujeto, edad, exposición a carcinógenos y género. Un parámetro biométrico preferido que se obtiene es el de paquetes años de fumador del sujeto. Si el método anterior comprende además la etapa de obtención de un valor para al menos un parámetro biométrico del sujeto, entonces el método puede comprender la etapa de comparación del valor de al menos un parámetro biométrico con un punto de corte predeterminado para dicho parámetro biométrico y la asignación de una puntuación para cada parámetro biométrico basándose en dicha comparación, la combinación de la puntuación asignada para cada parámetro biométrico con la puntuación asignada para cada biomarcador cuantificada en la etapa b para obtener una puntuación total para dicho sujeto, comparando la puntuación total con una puntuación predeterminada total de la etapa d y determinar si dicho sujeto tiene un riesgo de cáncer de pulmón basándose en la comparación de la puntuación total en la etapa c.

Opcionalmente, el método anterior también puede emplear un Método de Puntuación Ponderada para determinar si un sujeto está en riesgo de desarrollar un cáncer de pulmón. Si el método anterior emplea tal Método de Puntuación Ponderada, entonces en dicho método, la etapa b comprende la comparación de la cantidad de cada biomarcador del panel con varios puntos de corte predeterminados para dicho biomarcador y la asignación de una puntuación para cada biomarcador basándose en dicha comparación, la etapa c comprende la combinación de la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto, la etapa d comprende la comparación de la puntuación total determinada en la etapa c con una puntuación total predeterminada y la etapa e comprende determinar si dicho sujeto tiene cáncer de pulmón basándose en la comparación de la puntuación total determinada en la etapa d.

La presente divulgación también se refiere a una variedad de diferentes kits que se pueden utilizar en los métodos descritos anteriormente. En un aspecto, un kit puede comprender un péptido seleccionado de entre el grupo que consiste en: SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5 o cualquier combinación de los mismos. En otro aspecto, un kit puede comprender al menos un antígeno reactivo contra la Ciclina E2 inmunorreactiva o cualquier combinación de los mismos. En otro aspecto, un kit puede comprender al menos un antígeno reactivo contra la Ciclina E2 inmunorreactiva o cualquier combinación de los mismos. En un aspecto adicional, un kit puede comprender (a) reactivos que contienen al menos un anticuerpo para cuantificar uno o más antígenos de la muestra de ensayo, donde dichos antígenos son: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide sérico A, alfa-1-anti-tripsina, apolipoproteína CIII y Tβ4; (b) reactivos que contienen uno o más antígenos para cuantificar al menos un anticuerpo en una muestra de ensayo; donde dichos anticuerpos son: anti-p53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, anti-dominio-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACIN1, anti-RCV1, anti-MAPKAPK3 y anti-Ciclina E2; y (c) uno o más algoritmos para la combinación y comparación de la cantidad de cada antígeno y anticuerpo en la muestra de ensayo con un punto de corte predeterminado y asignar una puntuación para cada antígeno y anticuerpo basándose en dicha comparación, combinando la puntuación asignada para cada antígeno y anticuerpo para obtener una puntuación total, comparando la puntuación total con una puntuación total predeterminada y utilizando dicha comparación como ayuda para determinar si un sujeto tiene cáncer de pulmón. En un aspecto más, un kit puede comprender (a) reactivos que contienen al menos un anticuerpo para cuantificar uno o más antígenos en una muestra de ensayo, donde dichos antígenos son citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, seroamiloide A, alfa-1-anti-tripsina, apolipoproteína CIII y Tβ4; (b) reactivos que contienen uno o más antígenos para cuantificar al menos un anticuerpo en una muestra de ensayo; donde dichos anticuerpos son: anti-p53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, anti-dominio-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACIN1, anti-RCV1, anti-MAPKAPK3 y anti-Ciclina E2; (c) reactivos para cuantificar una o más regiones de interés seleccionadas de entre el grupo que consiste en: ACN9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 e Hic3959; y (d) uno o más algoritmos para la combinación y comparación de la cantidad de cada uno de los antígenos, anticuerpos y regiones de interés cuantificadas en la muestra de ensayo con un punto de corte predeterminado y la asignación de cada antígeno, anticuerpo y región de interés que se cuantifican basándose en dicha comparación, la combinación de la puntuación asignada para cada antígeno, anticuerpo y región de interés cuantificadas para obtener una puntuación total, la comparación de la puntuación total con una puntuación total predeterminada y la utilización de dicha comparación como una ayuda en la determinación de si un sujeto tiene cáncer de pulmón. En otro aspecto más, un kit puede comprender: (a) reactivos que contienen al menos un anticuerpo para cuantificar uno o más antígenos en la muestra de ensayo, donde dichos antígenos son citoqueratina 19, citoqueratina 18, CA19-9, CEA, CA125, SCC y proGRP; (b) reactivos para cuantificar una o más regiones de interés seleccionadas de entre el grupo que consiste en: ACN9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 e Hic3959; y (c) uno o más algoritmos para combinar y comparar la cantidad de cada antígeno y región de interés cuantificados en la muestra de ensayo con un punto de corte predeterminado, asignar una puntuación para cada antígeno y biomarcador cuantificados basándose en dicha comparación, combinando la puntuación asignada para cada antígeno y región de interés para obtener una puntuación total, comparando la puntuación total con una puntuación total predeterminada y utilizando dicha comparación como una ayuda para determinar si un sujeto tiene un cáncer de pulmón. Ejemplos de antígenos y regiones de interés que se pueden cuantificar son: (a) citoqueratina 19 y CEA y Acn9459, Pub11597, Pub4789 y Tfa2759; (b) citoqueratina 19 y CEA y Acn9459, Pub11597, Pub4789, Tfa2759 y Tfa9133; y (c) citoqueratina 19, CEA, CA125, SCC, citoqueratina 18, y ProGRP y ACN9459, Pub11597, Pub4789 y Tfa2759. En otro aspecto, un kit puede comprender (a) reactivos que contienen al menos un anticuerpo para cuantificar uno o más antígenos en una muestra de ensayo, donde dichos antígenos son citoqueratina 19, citoqueratina 18, CA 19-9, CEA, CA15-3, CA125, SCC y ProGRP; y (b) uno o más algoritmos para la combinación y comparación de la cantidad de cada antígeno cuantificado en la muestra de ensayo con un punto de corte predeterminado y asignar una puntuación para cada antígeno cuantificado basándose en dicha comparación, combinando la puntuación asignada para cada antígeno cuantificado para obtener una puntuación total, comparando la puntuación total con una puntuación total predeterminada y utilizando dicha comparación como ayuda para determinar si un sujeto tiene un cáncer de pulmón. Ejemplos de antígenos que se pueden cuantificar utilizando el kit son citoqueratina 19, citoqueratina 18, CA19-9, CEA, CA15-3, CA125, SCC y ProGRP. En otro aspecto, un kit puede comprender (a) reactivos para cuantificar uno o más biomarcadores, donde dichos biomarcadores son regiones de interés seleccionadas de entre el grupo que consiste en: ACN9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 e HIC3959; y (b) uno o más algoritmos para combinar y comparar la cantidad de cada biomarcador cuantificado en la muestra de ensayo con un punto de corte

predeterminado y asignar una puntuación para cada biomarcador cuantificado basándose en dicha comparación, combinando la puntuación asignada para cada biomarcador cuantificado para obtener una puntuación total, comparando la puntuación total con una puntuación total predeterminada y utilizando dicha comparación como una ayuda para determinar si un sujeto tiene cáncer de pulmón. Ejemplos de regiones de interés que pueden cuantificarse utilizando el kit se pueden seleccionar de entre el grupo que consiste en: Pub11597, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 e Hic3959. En otro aspecto más, un kit puede comprender (a) reactivos que contienen al menos un anticuerpo para cuantificar uno o más antígenos en una muestra de ensayo, donde dichos antígenos son: citoqueratina 19, CEA, CA125 y ProGRP; (b) reactivos que contienen uno o más antígenos para cuantificar al menos un anticuerpo en una muestra de ensayo; donde dichos anticuerpos son: anti-p53, anti-NY-ESO-1, anti-MAPKAPK3 y anti-Ciclina E2; y (c) uno o más algoritmos para combinar y comparar la cantidad de cada antígeno y anticuerpo cuantificado en la muestra de ensayo con un punto de corte predeterminado y asignar una puntuación para cada antígeno y anticuerpo cuantificados basándose en dicha comparación, combinando la puntuación asignada para cada antígeno y anticuerpo cuantificados para obtener una puntuación total, comparando la puntuación total con una puntuación total predeterminada y utilizando dicha comparación como ayuda para determinar si un sujeto tiene cáncer de pulmón. Opcionalmente, el kit puede comprender reactivos para cuantificar cada uno de citoqueratina 19, CEA, CA125 y ProGRP en una muestra de ensayo, reactivos para cuantificar cada uno de anti-p53, anti-NY-ESO-1, anti-MAPKAPK3 y anti-Ciclina E2 en una muestra de ensayo o reactivos para cuantificar cada uno de citoqueratina 19, CEA, CA125 ProGRP, anti-p53, anti-NY-ESO-1, anti-MAPKAPK3 y anti-Ciclina E2 en una muestra de ensayo.

La presente divulgación también se refiere a polipéptidos aislados o purificados. Los polipéptidos aislados o purificados contemplados en la presente invención son: (a) un polipéptido aislado o purificado que tiene (comprende) una secuencia de aminoácidos que se selecciona de entre el grupo que consiste en: la SEQ ID NO: 3 y un polipéptido que tiene un 60 % de homología con la secuencia de aminoácidos SEQ ID NO: 3; (b) un polipéptido aislado o purificado que consiste esencialmente en una secuencia de aminoácidos que se selecciona de entre el grupo que consiste en: la SEQ ID NO: 3 y un polipéptido que tiene un 60 % de homología con la secuencia de aminoácidos SEQ ID NO: 3; (c) un polipéptido aislado o purificado que consiste en una secuencia de aminoácidos de la SEQ ID NO: 3; (d) un polipéptido aislado o purificado que tiene una secuencia de aminoácidos que se selecciona de entre el grupo que consiste en: la SEQ ID NO: 4 y un polipéptido que tiene un 60 % de homología con la secuencia de aminoácidos SEQ ID NO: 4; (f) un polipéptido aislado o purificado que consiste en una secuencia de aminoácidos de la SEQ ID NO: 4; (g) un polipéptido aislado o purificado que tiene una secuencia de aminoácidos que se selecciona de entre el grupo que consiste en la SEQ ID NO: 5 y un polipéptido que tiene un 60 % de homología con la secuencia de aminoácido SEQ ID NO: 5; (h) un polipéptido aislado o purificado que consiste esencialmente en la secuencia de aminoácidos que se selecciona de entre el grupo que consiste en: la SEQ ID NO: 5 y un polipéptido que tiene un 60 % de homología con la secuencia de aminoácidos de la SEQ ID NO: 5; e (i) un polipéptido aislado o purificado que consiste en la secuencia de aminoácidos de la SEQ ID NO: 5.

La presente invención también se refiere a un Método de Puntuación Ponderada único. Este método se puede utilizar para puntuar uno o más marcadores obtenidos de un sujeto. Este método puede comprender las etapas de:

- a. cuantificar la cantidad de un marcador obtenido de una muestra de ensayo de un sujeto;
- b. comparar la cantidad de cada marcador cuantificado con varios puntos de corte predeterminados para dicho marcador y asignar una puntuación a cada marcador basándose en dicha comparación; y
- c. combinar la puntuación asignada para cada marcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto.

En el método anterior, los puntos de corte predeterminados se basan en curvas y la puntuación para cada marcador se calcula basándose en la especificidad del marcador. Adicionalmente, el marcador en el método anterior puede ser un biomarcador, un parámetro biométrico o una combinación de un biomarcador y un parámetro biométrico.

Adicionalmente, la presente divulgación proporciona un método para determinar el riesgo de un sujeto de desarrollar una afección médica utilizando el Método de Puntuación Ponderada. Este método puede comprender las etapas de:

- a. cuantificar en una muestra de ensayo obtenida de un sujeto la cantidad de al menos un marcador;
- b. comparar la cantidad de cada marcador cuantificado con varios puntos de corte predeterminados para dicho marcador y asignar una puntuación para cada marcador basándose en dichas comparaciones;
- c. combinar la puntuación asignada para cada marcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto;
- d. comparar la puntuación total determinada en la etapa c con una puntuación total predeterminada; y
- e. determinar si dicho sujeto tiene riesgo de desarrollar una afección médica basándose en la comparación de la puntuación total determinada en la etapa d.

En el método anterior, los puntos de corte predeterminados se basan en curvas y la puntuación de cada marcador se calcula basándose en la especificidad del marcador. Adicionalmente, el marcador del método anterior puede ser un biomarcador, un parámetro biométrico o una combinación de un marcador y un parámetro biométrico.

Breve descripción de las figuras

La **Figura 1** es un diagrama del flujo de trabajo bioinformático. Específicamente, se sometieron los datos de la MS e IA a varios métodos estadísticos. Se utilizó la regresión logística para generar curvas de Característica Operativa del Receptor (ROC) y para obtener el Área Bajo la Curva (AUC) para cada marcador. Los marcadores superiores con las AUC más altas se seleccionaron como marcadores candidatos. Los análisis multivariados (MVA) tales como Análisis Discriminatorio (DA), Análisis de Componente Principal (PCA) y Árboles de decisión (DT) identificaban los marcadores adicionales para introducirlos en el modelo. También se pueden incluir los parámetros biométricos. Los marcadores robustos que se producen en al menos el 50 % de los grupos de entrenamiento se identifican por el método/algoritmo de división y puntuación (SSM) y se seleccionan como supuestos marcadores. El proceso se repite n veces hasta que se obtiene un número de marcadores adecuado para el modelo predictivo final.

La **Figura 2** es un perfil de MS MALDI-TOF que muestra el biomarcador candidato Pub11597 a) tras la concentración de fracciones HPLC agrupadas y b) antes del proceso de concentración. La muestra es una mezcla compleja incluso tras el fraccionamiento HPLC.

La **Figura 3** es un gel teñido que muestra los componentes de las distintas muestras cargadas en el gel. Las calles a, f y g muestran una mezcla de proteínas de referencia de masas moleculares conocidas con fines de calibración. Adicionalmente, las calles b y e muestran una forma altamente purificada de la proteína sospechosa conocida como amiloide A sérico humano (HSAA), que se obtuvo comercialmente. Las calles c y d muestran las muestras fraccionadas que contienen el supuesto biomarcador. Hay un componente en la mezcla que migra a la misma distancia que el HSAA de referencia. Las bandas que tenían la misma distancia de migración que el HSAA se extirparon del gel y se sometieron a digestión en gel y análisis MS/MS para confirmar su identidad.

La **Figura 4** es una LS-MS/MS de la digestión trípica de Pub11597. Los paneles a-d muestran la MS/MS de 4 iones precursores principales. Los productos iónicos b e y se anotaron y se dio la secuencia de aminoácidos derivada de cada uno de los cuatro iones precursores. La búsqueda en la base de datos utilizando las masas moleculares de los iones generados, b e y, identificaron la fuente proteica como HSAA. La secuencia completa del fragmento observado (PM = 11526,51) se proporciona en la SEQ ID NO: 6.

La **Figura 5** da las curvas generadas del panel de 8 inmunoensayos de biomarcadores en muestras de 751 pacientes que se describe en el Ejemplo 1. Los rombos sólidos representan la curva generada por la puntuación total utilizando el Método de Puntuación Ponderada. Los cuadrados representan la curva generada por la puntuación total utilizando el método de puntuación binario utilizando puntos de división (puntos de corte) de una gran cohorte. Los triángulos sólidos representan la curva que se genera de la puntuación total utilizando el método de puntuación binario utilizando puntos de división (puntos de corte) de una pequeña cohorte.

La **Figura 6** da el AUC de un panel de 7 autoanticuerpos ensayados en cáncer de pulmón, benignos, y sujetos normales y los valores de todas las mediciones determinadas utilizando el método múltiple de medias (MoM) y el método de división y puntuación (SSM). En la Figura 6, -◇- es la puntuación SSM basándose en los puntos de división para dar la máxima precisión; -■- es la puntuación SSM basándose en los puntos de división para dar un 98 % de especificidad; y -●- es la puntuación MoM.

La **Figura 7** da el AUC para un panel de 4 inmunoensayos y 4 autoantígenos ensayado en sujetos con una historia de fumador empedernido (más de 20 paquetes años), en sujetos con una historia de fumador menos intensa o no fumador (menos de 20 paquetes años) y en sujetos en los que se desconoce su historia de fumador. En la Figura 7, - es menos de 20 paquetes años (packyrs) como fumador; -□- es mayor de 20 packyrs; y -●- es packyrs desconocido.

La **Figura 8** da el AUC para el panel de 4 inmunoensayos y 4 autoantígenos ensayados en sujetos que tienen un estadio temprano de cáncer de pulmón y enfermedad pulmonar benigna. Esta Figura 8 muestra que la capacidad de cada biomarcador individual ensayado para diferenciar el cáncer de pulmón en estado temprano y tardío de los sujetos con enfermedad pulmonar benigna y normales. En esta Figura 8, -●- muestra temprano frente a normal; -○- muestra temprano frente a benigno; - muestra tardío frente a normal; y ----- muestra tardío frente a benigno.

La **Figura 9** da el AUC de un panel de 4 inmunoensayos, 4 autoantígenos, 4 inmunoensayos y un marcador de respuesta inmunitaria (haptoglobina, amiloide sérico A, proteína C reactiva y apolipoproteína A), uno de 4 autoantígenos y uno de marcador de respuesta inmunitaria y uno de 4 inmunoensayos, 4 autoantígenos y uno de marcador de respuesta inmunitaria para la diferenciación de cáncer de pulmón en estadio temprano y tardío de la enfermedad pulmonar benigna. Para cada panel de biomarcadores, se representan cuatro barras. La primera barra representa el AUC para discriminar los pacientes de cáncer de pulmón de estadio temprano de la enfermedad pulmonar benigna; la segunda barra representa el AUC para discriminar el cáncer de pulmón en estadio temprano de los normales; la tercera barra representa el AUC para discriminar los pacientes de cáncer pulmonar en estado tardío de la enfermedad pulmonar benigna; y la cuarta barra representa el AUC para discriminar el cáncer de pulmón en estadio tardío de los normales.

La **Figura 10** da el AUC para el ensayo de Tβ4 realizado en un total de 118 muestras (44 de cáncer de pulmón y 74 no cancerosas) que contiene muestras normales, benignas y de pacientes de cáncer. El AUC = 0,8393.

Descripción detallada de la invención

DEFINICIONES

Como se utiliza en la presente solicitud, los siguientes términos tienen los siguientes significados. Todos los demás términos técnicos o científicos tienen el significado que comprenden comúnmente los expertos en la técnica.

El término “adsorbente” se refiere a cualquier material que es capaz de acumular (unirse) a una biomolécula. El adsorbente normalmente reviste una superficie biológicamente activa y está compuesto por un material simple o una pluralidad de diferentes materiales que sean capaces de unirse a una biomolécula o una variedad de biomoléculas basándose en sus características físicas. Tales materiales incluyen, pero no se limitan a, materiales de intercambio de aniones, materiales de intercambio de cationes, quelantes metálicos, polinucleótidos, oligonucleótidos, péptidos, anticuerpos, polímeros (sintéticos o naturales), papel, etc.

Como se utiliza en el presente documento, el término “anticuerpo” se refiere a una molécula de inmunoglobulina o una parte inmunológicamente activa de la misma, es decir, una parte de unión al antígeno. Ejemplos de partes inmunológicamente activas de moléculas de inmunoglobulina incluyen fragmentos F(ab), y F(ab')₂ que se pueden generar tratando un anticuerpo con una enzima, tal como la pepsina. Ejemplos de anticuerpos incluyen pero no se limitan a anticuerpos policlonales, anticuerpos monoclonales, anticuerpos quiméricos, anticuerpos humanos, anticuerpos humanizados, anticuerpos recombinantes, Fv de cadena sencilla (“scFv”), un anticuerpo de afinidad madura, anticuerpos de cadena sencilla, anticuerpos de dominio sencillo, fragmentos F(ab), fragmentos F(ab'), Fv unidos por disulfuro (“sdFv”), y anticuerpos antiidiotípicos (“anti-Id”) y fragmentos de unión al epítipo funcionalmente activos de cualquiera de los anteriores. Como se utiliza en el presente documento, el término “anticuerpo” también incluye autoanticuerpos (Los autoanticuerpos son anticuerpos que sintetiza un sujeto o paciente que se dirigen contra las proteínas propias normales (o antígenos propios) tales como, pero sin limitarse a, p53, calreticulina, alfa-enolasa, y HOXB7. Los autoanticuerpos contra un amplio intervalo de antígenos propios son bien conocidos por los expertos en la técnica y se han descrito en muchas enfermedades malignas incluyendo el cáncer de pulmón, cáncer de mama, cáncer de próstata, y cáncer de páncreas, entre otros). Un anticuerpo es un tipo de biomarcador.

Como se utiliza en el presente documento, el término “antígeno” se refiere a una molécula capaz de unirse a un anticuerpo y que además es capaz de inducir que un animal produzca anticuerpos capaces de unirse al menos a un epítipo de ese antígeno. Adicionalmente, una región de interés también puede ser un antígeno (en otras palabras, puede en último término determinarse que es un antígeno). Un antígeno es un tipo de biomarcador.

El término “AUC” se refiere al Área Bajo la Curva de una curva ROC. Se utiliza como una figura de la ventaja de un ensayo sobre una población de muestras determinada y da valores que fluctúan entre 1 para un ensayo perfecto a 0,5 en el que el ensayo da una respuesta completamente aleatoria en la clasificación de sujetos de ensayo. Como el intervalo del AUC es solo de 0,5 a 1, un pequeño cambio en el AUC tiene un mayor significado que un cambio en un sistema métrico que varía de 0 a 1 o de 0 al 100 %. Cuando se da el cambio del AUC en %, se calculará basándose en el hecho de que el intervalo completo del sistema métrico es de 0,5 a 1,0. El paquete estadístico JMP™ informa del AUC para cada curva generada. Las medidas del AUC son un medio valioso para comparar la precisión del algoritmo de clasificación a lo largo del intervalo de datos completo. Esos algoritmos de clasificación con un AUC mayor tienen por definición, una mayor capacidad para clasificar correctamente desconocidos entre los dos grupos de interés (enfermos y no enfermos). El algoritmo de clasificación puede ser tan simple como la medición de una única molécula o tan complejo como la medida e integración de múltiples moléculas.

La expresión “enfermedad pulmonar benigna” o “benigno” se refiere a una afección de una enfermedad asociada con el sistema pulmonar de un sujeto dado. En el contexto de la presente invención, una enfermedad pulmonar benigna incluye, pero no se limita a, trastorno pulmonar obstructivo crónico (COPD), inflamación aguda o crónica, nódulos benignos, neoplasias benignas, displasia, hiperplasia, atipia, bronquiectasia, histoplasmosis, sarcoidosis, fibrosis, granuloma, hematoma, enfisema, atelectasia, histiocitosis y otras enfermedades no cancerosas.

La expresión “superficie biológicamente activa” se refiere a cualquier extensión bi o tridimensional de un material con el que pueden interactuar o unirse biomoléculas, debido a las propiedades bioquímicas específicas de este material y las de las biomoléculas. Tales propiedades bioquímicas incluyen, pero no se limitan al carácter iónico (carga), hidrofobia, o hidrofilia.

Las expresiones “muestra biológica” y “muestra de ensayo” se refiere a todos los fluidos biológicos y excreciones aisladas de un sujeto determinado. En el contexto de la presente invención tales muestras incluyen pero no se limitan a la sangre, el suero sanguíneo, plasma sanguíneo, aspirado mamario, orina, semen, fluido seminal, plasma seminal, fluido prostático, excrementos, lágrimas, saliva, sudor, biopsias, ascitis, fluido cerebroespinal, leche, linfa, muestras de lavado bronquial y otros, o muestras de extractos tisulares. Normalmente se prefieren las muestras de ensayo de sangre, suero, plasma y lavado bronquial para su uso en el contexto de la presente invención.

El término “biomarcador” se refiere a una molécula biológica (o un fragmento de una molécula biológica) que se correlaciona con una condición física. Por ejemplo, los biomarcadores de la presente invención se correlacionan con el cáncer, preferentemente, con el cáncer de pulmón y se pueden utilizar como una ayuda en la detección de la presencia, ausencia, o riesgo de cáncer de pulmón. Tales biomarcadores incluyen, pero no se limita a las moléculas que comprenden nucleótidos, aminoácidos, azúcares, ácidos grasos, esteroides, metabolitos, polipéptidos, proteínas (tales como, pero sin limitarse a, antígenos y anticuerpos), carbohidratos, lípidos, hormonas, anticuerpos, regiones

de interés que sirven de sustitutos para moléculas biológicas, combinaciones de los mismos (por ejemplo, glucoproteínas, ribonucleoproteínas, lipoproteínas) y cualquier complejo que implica cualquiera de tales biomoléculas, tales como, pero sin limitarse a estos, un complejo formado entre un antígeno y un anticuerpo que se une a un epítipo disponible en dicho antígeno. El término “biomarcador” también se refiere a una parte de una

- 5 secuencia de un polipéptido (parental) que comprende al menos 5 restos de aminoácidos consecutivos, preferentemente al menos 10 restos de aminoácidos consecutivos, más preferentemente al menos 15 restos de aminoácidos consecutivos, y mantienen una actividad biológica y/o algunas características funcionales del polipéptido parental, por ejemplo, antigenicidad o características del dominio estructural.
- 10 La expresión “parámetro biométrico” se refiere a una o más cualidades intrínsecas físicas o de comportamiento que se utilizan para identificar únicamente pacientes como que pertenecen a un grupo bien definido o población. En el contexto de la presente invención, “parámetro biométrico” incluye pero no se limita a la altura del paciente, al peso del paciente, el género de un paciente, la historia de fumador, historia ocupacional, exposición a carcinógenos, exposición como fumador pasivo, historia familiar de cáncer de pulmón, y similares. La historia de fumador se
- 15 cuantifica habitualmente en términos de paquetes años (packyrs). Como se utiliza en el presente documento, la expresión “paquetes años” se refiere al número de años durante los que una persona ha fumado multiplicado por la media del número de paquetes fumados al día. Una persona que ha fumado, como media, 1 paquete de cigarrillos al día durante 35 años se designa que tiene 35 paquetes años de historia de fumador. La información del parámetro biométrico se puede obtener de un sujeto utilizando técnicas de rutina conocidas en la técnica, tales como a partir
- 20 del sujeto mismo por el uso de un cuestionario rutinario del paciente o un cuestionario de historia clínica, etc. De manera alternativa, el parámetro biométrico se puede obtener por una enfermera, un auxiliar, un asistente del médico o un médico a partir del sujeto.

- 25 Una “sustitución conservadora de aminoácidos” es en la que el resto de aminoácido se sustituye con un resto de aminoácido que tiene una cadena lateral similar. Las familias de restos de aminoácidos que tienen cadenas laterales similares están definidas en la técnica. Estas familias incluyen aminoácidos con cadenas laterales básicas (por ejemplo, lisina, arginina, histidina), cadenas laterales ácidas (por ejemplo, ácido aspártico, ácido glutámico), cadenas laterales sin carga polar (por ejemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales no polares (por ejemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano),
- 30 cadenas laterales con ramificación beta (por ejemplo, treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo, tirosina, fenilalanina, triptófano, histidina). Por lo tanto un resto de aminoácido previsto como no esencial se sustituye preferentemente con otro resto de aminoácido de la misma familia de cadenas laterales.

- 35 La frase “Análisis de Árbol de decisión” se refiere a la estrategia clásica en la que una serie de simples reglas dicotómicas (o síntomas) proporcionan una guía a lo largo de un árbol de decisión hasta un resultado de clasificación final o un nudo terminal del árbol. Su naturaleza inherentemente simple e intuitiva que produce una división recursiva muy flexible para un proceso diagnóstico.

- 40 El método necesita dos tipos de variables: variables factores (X) y variables respuestas (Y). Como se ha implementado, las variables X son continuas y las variables Y son categóricas (Nominales). En tales casos el paquete estadístico JMP utiliza un algoritmo que genera un valor de corte, que maximiza la pureza de los nodos. Las muestras se dividen en ramas o nodos basándose en los valores que están por encima o por debajo de este valor de corte.

- 45 Para la respuesta categórica variable, como en este caso, el valor ajustado convierte la probabilidad estimada para cada nivel de respuesta. En este caso la división se determina por la estadística chi cuadrado de la mayor probabilidad (G^2). Esto tiene el efecto de maximizar la diferencia en las respuestas entre las dos ramas de la división. Se puede encontrar una descripción más detallada del método y su implementación en la guía de estadísticas y Gráficos JMP.

- 50 Construir un árbol, sin embargo tiene sus propios problemas asociados. Un problema común es decidir el tamaño óptimo del árbol que proporcionará el mejor modelo predictivo sin sobre ajustar los datos. Con esto en mente, se desarrolló un método que utilizaba la información que se podía extraer en los distintos nodos del árbol para construir una curva ROC. Como se implementó, el método implica la construcción de un árbol de referencia con suficientes
- 55 nodos que sobre ajustarán con seguridad el grupo de datos que se va a modelar. Posteriormente, el árbol se poda, eliminando sucesivamente los nodos peores en cada etapa hasta que se alcanza el número mínimo de nodos (dos nodos terminales). Esto crea una serie o una familia de árboles de complejidad decreciente (menos nodos).

- 60 El programa de división recursiva intenta crear nodos terminales puros, es decir, solamente se incluyen los especímenes de un tipo de clasificación. Sin embargo, no siempre es posible. A veces los nodos terminales tienen poblaciones mezcladas. Por lo tanto, cada nodo terminal tendrá una probabilidad diferente en cuanto al cáncer. En un nodo puro terminal, la probabilidad de ser un espécimen de cáncer será del 100 %, y por el contrario, para un nodo puro terminal de no cáncer, la probabilidad de ser un espécimen de cáncer será del 0 %. La probabilidad de cáncer en cada nodo terminal se representa con (1-probabilidad de no cáncer) cada nodo.

- 65 Estos valores se representan para generar una curva que es representativa de ese árbol en particular. El AUC

calculada para cada árbol representa la “bondad” del árbol o modelo. Igual que en cualquier aplicación diagnóstica, cuanto mayor es el AUC, mejor es el ensayo, o en este caso el modelo. Un gráfico del AUC con el tamaño del árbol (número de nodos) tendrá como máximo el mejor modelo para el grupo de entrenamiento. Se lleva a cabo un procedimiento similar con un segundo subgrupo pero más pequeño de los datos para validar los resultados. Los modelos que tienen una actuación similar en los grupos de entrenamiento y validación se considera que son óptimos y por tanto se eligen para posteriores análisis y/o validación.

Las expresiones “grupo de datos de desarrollo” o “grupo de datos” se refiere a las características que incluyen el biomarcador completo o los datos del biomarcador y el parámetro biométrico recolectados para un grupo de muestras biológicas. Estas muestras mismas se extraen de los pacientes con resultados diagnosticados conocidos. Una característica o grupo de características se someten a un análisis estadístico apuntando hacia una clasificación de muestras en dos o más grupos de muestras diferentes (por ejemplo, cáncer y no cáncer) que se correlacionan con los resultados del paciente que se conocen. Cuando se utiliza la espectrometría de masas, entonces el espectro de masas del grupo puede ser diferente en sus intensidades, pero no en sus masas moleculares aparentes en la precisión de los aparatos.

El término “clasificador” se refiere a cualquier algoritmo que utiliza las características derivadas de un grupo de muestras para determinar la enfermedad asociada con la muestra. Un tipo de clasificador se crea por “entrenamiento” del algoritmo con datos del grupo de entrenamiento y cuya actuación se evalúa con los datos del grupo de ensayo. Ejemplos de clasificadores que se utilizan en conjunción con la invención son los análisis de discriminación, análisis de árboles de decisión, curvas de la característica operativa del receptor o análisis de división y puntuación.

La expresión “árbol de decisión” se refiere a un clasificador con una estructura de árbol tipo gráfico de flujo que se emplea para la clasificación. Los árboles de decisión consisten en divisiones repetidas de un grupo de datos en subgrupos. Cada división consiste en una simple regla que se aplica a una variable, por ejemplo, “si el valor de la ‘variable 1’ es mayor que el ‘umbral 1’; entonces ve a la izquierda, si no ve a la derecha”. En consecuencia, el espacio de determinada característica se divide en un grupo de rectángulos y cada rectángulo se asigna a una clase.

Las expresiones “ensayo diagnóstico” y “método diagnóstico” se refieren a la detección de la presencia o naturaleza de una afección patológica. Los ensayos diagnósticos se diferencian en cuanto a su sensibilidad y especificidad. Los sujetos que se prueban positivos para cáncer de pulmón y que están enfermos actualmente se consideran “positivos verdaderos”. En el contexto de la invención, la sensibilidad de un ensayo diagnóstico se define como el porcentaje de positivos verdaderos en una población enferma. Los sujetos que tienen cáncer de pulmón pero no se detectan en el ensayo diagnóstico se consideran “falsos negativos”. Los sujetos que no están enfermos y que se prueban negativos en el ensayo diagnóstico se consideran “negativos verdaderos”. El término especificidad de un ensayo diagnóstico, como se utiliza en el presente documento, se define como el porcentaje de los negativos verdaderos en la población no enferma.

La expresión “análisis discriminante” se refiere a un grupo de métodos estadísticos que se utilizan para seleccionar características que discriminan de manera óptima entre dos o más grupos que se producen naturalmente. La aplicación de análisis discriminantes a un grupo de datos permita al usuario enfocar las características más discriminativas para análisis posteriores.

La frase “Distancia del ideal” o “DFI” se refiere a un parámetro que se toma de la curva que es la distancia del ideal, que incorpora la sensibilidad y especificidad y que se define como $[(1-\text{sensibilidad}) + (1-\text{especificidad})^2]^{1/2}$. El DFI es 0 para un ensayo con una actuación del 100 % de sensibilidad y el 100 % de especificidad y aumenta a 1,414 para un ensayo con el 0 % de sensibilidad y el 0 % de especificidad. A diferencia del AUC que utiliza el intervalo completo de datos para su determinación, el DFI mide la actuación de un ensayo en un punto particular de la curva ROC. Los ensayos con valores DFI más bajos actúan mejor que los de valores DFI más altos. La DFI se trata en detalle en la Publicación de la Solicitud de Patente de EE. UU. Nº 2006/0211019A1.

Los términos “conjunto”, “árbol de conjunto” o “clasificador de conjunto” se pueden utilizar de manera intercambiable y se refieren a un clasificador que consiste en muchos clasificadores elementales más simples, por ejemplo, un conjunto de árboles de decisión es un clasificador que consiste en árboles de decisión. El resultado del clasificador de conjunto se obtiene combinando todos los resultados de sus clasificadores constituyentes, por ejemplo, por votación de la mayoría que sopesa todos los clasificadores constituyentes por igual. La votación de la mayoría es especialmente razonable cuando los clasificadores constituyentes se sopesan entonces naturalmente por la frecuencia con la que se generan.

El término “epítipo” quiere referirse a esa porción de un antígeno con la capacidad de unirse a un anticuerpo que también puede ser reconocido por ese anticuerpo. Los determinantes epitópicos habitualmente consisten en agrupaciones de moléculas de superficie químicamente activas tales como aminoácidos o cadenas laterales de azúcares y tienen características estructurales tridimensionales específicas así como características de carga específicas.

Los términos “característica” y “variable” se pueden utilizar de manera intercambiable y se refieren al valor de una medida de una característica de una muestra. Estas mediciones se pueden derivar de características físicas, químicas o biológicas de la muestra. Ejemplos de las mediciones incluyen pero no se limitan a un pico de espectro de masas, una señal de espectro de masas, una función de la intensidad de un ROL.

5 Los cálculos de la homología o identidad de secuencia entre secuencias (los términos se utilizan de manera intercambiable en el presente documento) se realizan de la siguiente manera.

10 Para determinar el porcentaje de identidad de dos secuencias de aminoácidos o de dos secuencias de ácidos nucleicos, las secuencias se alinean con fines de comparación óptima (por ejemplo se pueden introducir huecos en una o ambas de una primera y una segunda secuencia de aminoácidos o ácidos nucleicos para el alineamiento óptimo y se pueden ignorar las secuencias no homólogas para los fines de comparación). En una realización preferida, la longitud de una secuencia de referencia alineada para fines de comparación es al menos un 30 %, preferentemente al menos un 40 %, más preferentemente al menos un 50 %, incluso más preferentemente al menos un 60 %, e incluso más preferentemente al menos un 70 %, 80 %, 90 %, 95 %, 99 % o 100 % de la longitud de la secuencia de restos de aminoácidos de referencia que se alinean. Entonces se comparan los restos de aminoácidos o nucleótidos con las posiciones de aminoácidos o nucleótidos correspondientes. Cuando una posición de la primera secuencia está ocupada por el mismo resto de aminoácido o nucleótido en la posición correspondiente de la segunda secuencia, entonces las moléculas son idénticas en esa posición (como se utiliza en el presente documento “identidad” de aminoácidos o ácido nucleico es equivalente a “homología” de aminoácidos o ácido nucleico). El porcentaje de identidad entre las dos secuencias es una función el número de posiciones idénticas compartidas por las dos secuencias, tomando en cuenta el número de huecos, y la longitud de cada hueco, que es necesario introducir para el alineamiento óptimo de las dos secuencias.

25 La comparación de secuencias y la determinación del porcentaje de identidad entre dos secuencias se puede conseguir utilizando un algoritmo matemático. En una realización preferida, el porcentaje de identidad entre dos secuencias de aminoácidos se determina utilizando el algoritmo de Needleman y Wunsch (J. Mol. Biol. 48:444-453 (1970)) que se ha incorporado en el programa GAP en el paquete de software GCG, utilizando o una matriz Blossum 62 o una matriz PAM250, y un peso de huecos de 16, 14, 12, 10, 8, 6, o 4 y un peso de la longitud de 1, 2, 3, 4, 5, o 6. En otra realización preferida, el porcentaje de identidad entre dos secuencias de nucleótidos se determina utilizando el programa GAP en el paquete de software GCG, utilizando un NWSgapdna. La matriz CMP y un peso de huecos de 40, 50, 60, 70, o 80 y un peso de la longitud de 1, 2, 3, 4, 5, o 6. Se debería aplicar un grupo particularmente preferido de parámetros (y el que se debería utilizar si el médico no sabe sobre qué parámetros se deberían aplicar para determinar si una molécula está en una limitación de identidad u homología de la secuencia de la invención) utilizan una matriz de puntuación Blossum 62 con una falta de hueco abierto de 12, una falta de hueco extendido de 4, y una falta de hueco de desplazamiento de lectura de 5.

40 El porcentaje de identidad entre dos secuencias de aminoácidos o ácido nucleico se pueden determinar utilizando el algoritmo de E. Meyers y W. Miller (CABIOS, 4:11-17 (1989)) que se ha incorporado en el programa ALIGN (versión 2.0), utilizando una tabla de restos de peso PAM 120, una falta de longitud de huecos de 12, y una falta de huecos de 4.

45 Las secuencias de ácido nucleico y proteínas que se describen en el presente documento se pueden utilizar como una “secuencia de consulta” para llevar a cabo una búsqueda en las bases de datos públicas para, por ejemplo, identificar otros miembros de la familia o secuencias relacionadas. Tales búsquedas se pueden llevar a cabo utilizando los programas NBLAST y XBLAST (versión 2.0) de Altschul, et al., J. Mol. Biol. 215:403-10 (1990). Las búsquedas de proteínas BLAST se pueden llevar a cabo con el programa XBLAST, puntuación = 50, longitud de palabra = 3 para obtener secuencias de aminoácidos homólogas a una proteína Ciclina E2 inmunorreactiva de la presente invención. Para obtener los alineamientos con huecos con fines de comparación, se puede utilizar el BLAST con huecos como se describe en Altschul et al., Nucleic Acids Res. 25(17):3389-3402 (1997). Cuando se utilizan los programas BLAST y BLAST con huecos, se pueden utilizar los parámetros por defecto de los programas respectivos (por ejemplo, XBLAST y NBLAST).

55 Como se utiliza en el presente documento, la expresión “Ciclina E2 inmunorreactiva” se refiere a (1) un polipéptido que tiene una secuencia de aminoácidos de cualquiera de SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, o SEQ ID NO: 5; (2) cualquier combinación de cualquiera de las SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, o SEQ ID NO: 5; (3) un polipéptido que tiene una secuencia de aminoácidos que es al menos un 60 %, preferentemente al menos un 70 %, más preferentemente al menos un 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 % homóloga a la SEQ ID NO: 1, un polipéptido que tiene una secuencia de aminoácido que es al menos un 60 %, preferentemente al menos un 70 %, más preferentemente al menos un 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 % homóloga a la SEQ ID NO: 3, un polipéptido que tiene una secuencia de aminoácido que es al menos un 60 %, preferentemente al menos un 70 %, más preferentemente al menos un 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 % homóloga a la SEQ ID NO: 4, un polipéptido que tiene una secuencia de aminoácido que es al menos un 60 %, preferentemente al menos un 70 %, más preferentemente al menos un 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 % homóloga a la SEQ ID NO: 5 y cualquier combinación de las mismas; (4) un polipéptido Ciclina E2 que muestra una

inmunorreactividad similar a las SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5; y (5) un polipéptido que muestra una inmunorreactividad similar a las SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5.

Un polipéptido o proteína "aislada" o "purificada" está sustancialmente libre de material celular u otras proteínas contaminantes de la fuente celular o tisular de la que se deriva la proteína, o sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetizan químicamente. Cuando una proteína o parte biológicamente activa de la misma se produce recombinantemente, también está preferentemente sustancialmente libre de medio de cultivo, a saber, el medio de cultivo representa menos de aproximadamente el 20 %, más preferentemente menos de aproximadamente el 10 %, más preferentemente menos de aproximadamente el 5 % del volumen de la preparación proteica.

Como se utiliza en el presente documento, la frase "Análisis de Discriminación Lineal" se refiere a un tipo de análisis que proporciona una herramienta para identificar las variables o características que son las mejores para categorizar correctamente una muestra y la cual se puede implementar, por ejemplo, por el paquete estadístico JMP™. Utilizando la característica paso a paso del software, se pueden añadir variables a un modelo hasta que se clasifiquen todas las muestras correctamente. En general, el grupo de variables seleccionado de esta manera es sustancialmente menor que el número original de variables en el grupo de datos. Esta reducción en el número de características simplifica cualquiera de los siguientes análisis, por ejemplo, el desarrollo de un mecanismo de clasificación más general utilizando árboles de decisión, redes neurales artificiales, o similares.

La expresión "cáncer de pulmón" se refiere a un estado canceroso que se asocia con el sistema pulmonar de un determinado sujeto. En el contexto de la presente invención, los cánceres de pulmón incluyen, pero no se limitan a estos, adenocarcinoma, carcinoma epidermoide, carcinoma de células escamosas, carcinoma de células grandes, carcinoma de células pequeñas, carcinoma de células no pequeñas, carcinoma broncoalveolar. En el contexto de la presente invención, los cánceres pulmonares pueden estar en diferentes estadios, así como en varios grados de graduación. Los métodos para determinar el estadio de un cáncer de pulmón o su grado de graduación son bien conocidos por los expertos en la técnica.

La expresión "espectrometría de masas" se refiere al uso de una fuente de ionización para generar iones en fase gaseosa a partir de una muestra en una superficie y detectar los iones en fase gaseosa con un espectrómetro de masas. La expresión "espectrometría de masas de desorción láser" se refiere al uso de un láser como fuente de ionización para generar iones en fase gaseosa a partir de una muestra en una superficie y detectar los iones en fase gaseosa con un espectrofotómetro de masas. Un método preferido de espectrometría para biomoléculas es la espectrometría de masas de desorción/ionización láser asistida por matriz o MALDI. En MALDI, el analito se mezcla normalmente con un material de matriz que, tras el secado se co-cristaliza con el analito. El material de la matriz absorbe energía de la fuente de energía que de otra manera fragmentaría las moléculas o analitos lábiles. Otro método preferido es la espectrometría de masas de desorción/ionización por láser de superficie aumentada o SELDI. En el SELDI, la superficie en la que se aplica el analito tiene un papel activo en la captura y/o desorción del analito. En el contexto de la invención la muestra comprende una muestra biológica que puede someterse a cromatografía u otro procesamiento química y un sustrato de matriz adecuado.

En espectrometría de masas la "masa molecular aparente" se refiere a la masa molecular (en Daltons)-respecto al valor de la carga, m/z , de los iones detectados. El cómo se deriva la masa molecular aparente se deriva depende del tipo del espectrómetro de masas que se utilice. Con un espectrómetro de masas con tiempo de vuelo, la masa molecular aparente es función del tiempo desde la ionización a la detección.

El término "matriz" se refiere a una molécula que absorbe energía como fotones a partir de una fuente de luz apropiada, por ejemplo un UV/Vis o láser IR, en un espectrómetro de masas haciendo posible de esta manera la desorción de una biomolécula a partir de una superficie. Se utilizan frecuentemente derivados del ácido cinámico que incluyen el ácido α -ciano cinámico, ácido sinapínico y ácido dihidroxibenzoico como moléculas que absorben energía en la desorción láser de biomoléculas. Las moléculas que absorben energía se describen en la Patente de EE. UU. N° 5.719.060.

El término "normalización" y sus derivados, cuando se utilizan en conjunción con espectros de masas, se refiere a métodos matemáticos que se aplican a un grupo de espectros de masas para eliminar o minimizar las diferencias, debido primariamente a los parámetros del aparato, en las intensidades totales de los espectros.

El término "región de interés" o "ROI" se refiere a una adaptación estadística de un sujeto de un espectro de masas. Un ROI tiene una longitud mínima fija de señales consecutivas. Las señales consecutivas pueden contener huecos de longitud máxima fija dependiendo de cómo se elige el ROI. Las regiones de interés se refieren a los biomarcadores y pueden servir como sustitutos de los biomarcadores. Las regiones de interés pueden determinarse más tarde para representar una proteína, polipéptido, antígeno, anticuerpo, lípido, hormona, carbohidrato, etc.

La frase "Curva de Características Operativas del Receptor" o "curva ROC" se refiere a, en su aplicación más simple, un gráfico de la actuación de una característica particular (por ejemplo, un biomarcador o parámetro biométrico) para distinguir entre dos poblaciones (por ejemplo, casos (es decir, los sujetos que padecen un cáncer de pulmón) y los

controles (es decir, lo sujetos que son normales o tienen cáncer de pulmón benigno)). Los datos de característica a través de la población entera (es decir, los casos y controles), se ordenan en orden ascendente basándose en el valor de una única característica. Entonces, para cada valor de cada característica, se calculan las tasas de falsos positivos y falsos negativos de los datos. Se determina la tasa de positivos verdaderos por recuento del número de casos por encima del valor de cada característica que se considera y entonces se divide por el número total de casos. La tasa de falsos positivos se determina por recuento del número de controles por encima del valor de esa característica que se considera y luego dividiendo por el número total de controles. Aunque esta definición describe un escenario en el que una característica está elevada en casos en comparación con los controles, esta definición también engloba un escenario en el que una característica se suprime en los casos en comparación con los controles. En este escenario, se contarían las muestras por debajo del valor para la característica que se considera.

Las curvas ROC se pueden generar para una característica única así como para otros resultados únicos, por ejemplo, una combinación de dos o más características se combinan entre ellas matemáticamente (tales como, sumándolas, restándolas, multiplicándolas, etc.) para proporcionar un único valor de la suma, este único valor de la suma se puede representar en una curva ROC. Además, se puede representar cualquier combinación de múltiples características, de manera que la combinación dé lugar a un único valor de resultado. Estas combinaciones de características pueden comprender un ensayo. La curva ROC es la representación de la tasa de positivos verdaderos (sensibilidad) de un ensayo respecto a la tasa de falsos positivos (1-especificidad) del ensayo. El área bajo la curva ROC es una figura de cualidad para la característica de una determinada población de muestras y da valores que varían desde 1 para un ensayo perfecto y 0,5 en el que el ensayo da una respuesta completamente aleatoria en la clasificación de los sujetos de ensayo. Las curvas ROC proporcionan otros medios para explorar rápidamente un grupo de datos. Las características que parecen ser diagnósticas se pueden utilizar preferentemente para reducir el tamaño de grandes espacios característicos.

El término “selección” se refiere a una decisión diagnóstica con respecto a la disposición del paciente hacia el cáncer de pulmón. Un paciente se determina que está en riesgo de un cáncer de pulmón con un “ensayo de selección” positivo. Como resultado, el paciente puede someterse a ensayos adicionales, por ejemplo, por imagen, análisis del esputo, ensayos funcionales del pulmón, broncoescopia y/o procedimientos de biopsia y se hace un diagnóstico final.

El término “señal” se refiere a cualquier respuesta que se genera por una biomolécula bajo investigación. Por ejemplo, el término señal se refiere a la respuesta generada por una biomolécula que golpea el detector de un espectrómetro de masas. La intensidad de la señal se correlaciona con la cantidad o concentración de la biomolécula. La señal se define por dos valores: un valor aparente de la masa molecular y un valor de intensidad generado como se ha descrito. El valor de masa es una característica elemental de la biomolécula, mientras que el valor de la intensidad concuerda con una cierta cantidad o concentración de la biomolécula con el valor de masa molecular aparente correspondiente. Por lo tanto, la “señal” siempre se refiere a las propiedades de la biomolécula.

La frase “Método de división y puntuación” se refiere a un método adaptado de Mor et al., PNAS, 102(21):7677-7682 (2005). En este método, se toman múltiples mediciones de todas las muestras. Se determina un valor de corte de cada medición. Este valor de corte se puede fijar para maximizar la precisión de las clasificaciones correctas entre los grupos de interés (por ejemplo, enfermo y no enfermo) o se puede fijar para maximizar la sensibilidad o especificidad de un grupo. Para cada medición, se determina si el grupo de interés, por ejemplo, enfermo, está por encima o por debajo del valor de corte. Para cada medición, se asigna una puntuación para la muestra independientemente de si el valor de las mediciones está en el lado enfermo del valor de corte. Después de tomarse todas las mediciones en una muestra, se suman las puntuaciones para producir una puntuación total para el panel de mediciones. Es normal igualar el peso de todas las mediciones tal que un panel de 10 mediciones puede tener una puntuación máxima de 10 puntos (cada medición tienen una puntuación o de 1 o de 0) y una puntuación mínima de 0. Sin embargo, puede valorarse dar importancia a los valores desigualmente con una puntuación individual más alta para las mediciones más significativas.

Después de determinar las puntuaciones totales, se determina de nuevo un punto de corte para clasificar las muestras enfermas de las no enfermas basándose en el panel de mediciones. De nuevo en este caso, para un panel de mediciones con una puntuación máxima de 10 y una puntuación mínima de 0, se puede escoger un punto de corte para maximizar la sensibilidad (puntuación de 0 como punto de corte), o para maximizar la especificidad (puntuación de 10 como punto de corte), o para maximizar la precisión de la clasificación (puntuación de 0 – 10 como punto de corte).

Como se utiliza en el presente documento, la frase “Método de Puntuación Ponderado” se refiere a un método que implica la conversión de las mediciones de un biomarcador o un parámetro biométrico (designado colectivamente en el presente documento como “marcador(es)” que se identifica y cuantifica en una muestra de ensayo en una o más puntuaciones potenciales. Las puntuaciones se obtienen utilizando la siguiente ecuación:

$$\text{Puntuación} = \text{AUC} * \text{factor} / (1 - \text{especificidad})$$

donde el “factor” es un número entero (tal como 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,

21, 22, 23, 24, 25, etc.) y la “especificidad” es un valor escogido que es menor o igual a 1. La magnitud del “factor” aumenta para los marcadores que tienen una actuación clínica mejor, tal como, pero sin limitarse a, valores de AUC más altos, desviaciones estándar relativamente pequeñas, alta especificidad o sensibilidad o bajo DFI. Por consiguiente, la medición de un marcador se puede convertir en más o menos puntuaciones como se desee. Este método se basa en la curva de Características Operativas del Receptor que refleja el marcador/ensayo de actuación en la población de interés. La curva ROC es un gráfico de la tasa de positivos verdaderos (sensibilidad) o un ensayo contra la tasa de falsos positivos (1-especificidad) del ensayo. Cada punto de la curva representa un valor único de la característica/ensayo (marcador) que se está midiendo. Por lo tanto, algunos valores tendrán una tasa de falsos positivos baja en la población de interés (es decir, sujetos con riesgo de desarrollar un cáncer de pulmón) mientras que otros valores de la característica tendrán tasas altas de falsos positivos en esa población. Este método proporciona altas puntuaciones de los valores de la característica (a saber, biomarcadores o parámetros biométricos) que tienen tasas bajas de falsos positivos (teniendo de esta manera una alta especificidad) para la población de sujetos de interés. El método implica la elección de los niveles deseados de falsos positivos (1-especificidad) por debajo de la cual el ensayo da como resultado un aumento de la puntuación. En otras palabras, los marcadores que son altamente específicos dan una puntuación mayor o un mayor intervalo de puntuaciones que los marcadores que son menos específicos.

Como se utiliza en el presente documento, el término “sujeto” se refiere a un animal, preferentemente un mamífero, que incluye un ser humano o no humano. Los términos paciente y sujeto se puede utilizar de manera intercambiable en el presente documento.

La frase “Validación de diez veces de Modelos DT” se refiere al hecho de que las buenas prácticas analíticas necesitan que los modelos se validen contra una población nueva para evaluar su valor predictivo. En el lugar de una nueva población, los datos se pueden dividir en grupos de entrenamiento independientes y grupos de validación. Se generan diez subgrupos aleatorios para su uso como grupos de validación. Para cada grupo de validación, hay un correspondiente grupo de entrenamiento independiente que no tiene muestras en común. Se generan diez modelos DT a partir de los diez grupos de entrenamiento como se ha descrito anteriormente y se interrogan con los grupos de validación.

Las expresiones “grupo de ensayo” o “desconocido” o “grupo de validación” se refieren a un subgrupo del grupo de datos disponibles completo que consiste en las entradas no incluidas en el grupo de entrenamiento. Los datos de ensayo se aplican para evaluar la actuación del clasificador.

Las expresiones “grupo de entrenamiento” o “grupo conocido” o “grupo de referencia” se refieren a un subgrupo del grupo de datos disponibles completo. Este subgrupo se selecciona normalmente aleatoriamente, y solo se utiliza con el fin de la construcción del clasificador.

La expresión “Modelo de Regresión Logística Transformado” se refiere a un modelo, que también se implementa en el paquete estadístico JMP™, que proporciona un medio de combinación de varias características y permite un análisis de la curva ROC. Esta estrategia se aplica mejor para un grupo reducido de características ya que asume un modelo simplista para la relación de unas características con otras. Un resultado positivo sugiere que los métodos de clasificación más sofisticados deberían ser satisfactorios. Un resultado negativo aunque decepcionante no necesariamente implica el fallo de otros métodos.

POLIPÉPTIDOS DE CICLINA E2

En una realización, la presente divulgación se refiere a polipéptidos Ciclina E2 inmunorreactivos aislados o purificados o fragmentos biológicamente activos de los mismos que se pueden utilizar como inmunógenos o antígenos para dar lugar o ensayar (o más generalmente, para unirse) anticuerpos que se pueden utilizar en los métodos que se describen en el presente documento. Los polipéptidos Ciclina E2 inmunorreactivos de la presente divulgación se pueden aislar de fuentes celulares o tisulares utilizando técnicas de aislamiento de proteínas de referencia. De manera alternativa, los polipéptidos Ciclina E2 inmunorreactivos aislados o purificados y los fragmentos biológicamente activos de los mismos se pueden producir por técnicas de ADN recombinante o se sintetizan químicamente. Los polipéptidos Ciclina E2 inmunorreactivos aislados o purificados de la presente divulgación tienen las secuencias de aminoácidos que se muestran en las SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 y SEQ ID NO: 5. La SEQ ID NO: 1 es la secuencia de aminoácidos de una forma de ADNc expresado de la Ciclina E2 humana (Acceso de Genbank BC007015.1). La SEQ ID NO: 3 es una secuencia de 38 aminoácidos que comprende el extremo C de BC007015.1 más un aminoácido (cisteína) y que también se designa en el presente documento como “B2-1”. La SEQ ID NO: 4 es de 37 aminoácidos de longitud y es idéntica a la SEQ ID NO: 3 excepto que la SEQ ID NO: 4 no contiene, en su extremo amino, la cisteína muy primera de la SEQ ID NO: 3. La SEQ ID NO: 5 es una secuencia de 19 aminoácidos que comprende el extremo C de BC007015.1 y que se le designa como “E2-2”. Como se describe con más detalle en los Ejemplos, se comparó la inmunorreactividad de la SEQ ID NO: 1 con la inmunorreactividad de la SEQ ID NO: 2. La secuencia SEQ ID NO: 2 es otra forma de ADNc que se expresa de ciclina E2 humana (Acceso de Genbank BC020729.1). Se descubrió que la SEQ ID NO: 1 mostraba una fuerte inmunorreactividad con varios grupos de muestras de cáncer y mostraba mucha menos reactividad con grupos benignos y normales (no cancerosos). Por el contrario, la SEQ ID NO: 2 mostraba poca

reactividad con cualquier grupo de muestras de cáncer o no cancerosas. Se determinó la inmunorreactividad de la SEQ ID NO: 1 ser el resultado de los 37 primeros aminoácidos presentes en el extremo C de la SEQ ID NO: 1 que no está presente en la SEQ ID NO: 2. Las SEQ ID NO: 4 y SEQ ID NO: 5 o cualquier combinación de estas secuencias (tales como, anticuerpos generados contra la SEQ ID NO: 1 y SEQ ID NO: 3, anticuerpos generados contra la SEQ ID NO: 1 y SEQ ID NO: 4, anticuerpos generados contra la SEQ ID NO: 1 y SEQ ID NO: 5, anticuerpos generados contra la SEQ ID NO: 1, SEQ ID NO: 3 y SEQ ID NO: 4, anticuerpos generados contra la SEQ ID NO: 1, SEQ ID NO: 3 y SEQ ID NO: 5, anticuerpos generados contra la SEQ ID NO: 1, SEQ ID NO: 4 y SEQ ID NO: 5, anticuerpos generados contra la SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 y SEQ ID NO: 5, anticuerpos generados contra la SEQ ID NO: 3 y SEQ ID NO: 4, anticuerpos generados contra la SEQ ID NO: 3 y SEQ ID NO: 5, anticuerpos generados contra la SEQ ID NO: 3, SEQ ID NO: 4 y SEQ ID NO: 5, anticuerpos generados contra la SEQ ID NO: 4 y SEQ ID NO: 5 (a todos los que se designa en el presente documento como "anti-Ciclina E2")) se pueden utilizar en los métodos descritos en el presente documento. Por ejemplo, tales anticuerpos se pueden someter a los anticuerpos generados contra cualquiera de las SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 y SEQ ID NO: 5 o cualquier combinación de estas secuencias. Tales anticuerpos se pueden incluir en uno o más kits para su uso en los métodos de la presente divulgación descrita en el presente documento.

La presente divulgación también engloba polipéptidos que se diferencian de los polipéptidos descritos en el presente documento (a saber, SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 y SEQ ID NO: 5) por una o más sustituciones conservadoras de aminoácidos. Adicionalmente, la presente divulgación también engloba polipéptidos que tienen una similitud de secuencia total (identidad) u homología de al menos el 60 %, preferentemente al menos el 70 %, más preferentemente al menos el 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 % o más, con un polipéptido que tiene la secuencia de aminoácidos de SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 y SEQ ID NO: 5.

USO DE BIOMARCADORES Y PARÁMETROS BIOMÉTRICOS PARA DETECTAR LA PRESENCIA O RIESGO DE CÁNCER DE PULMÓN

En otra realización, la presente divulgación se refiere a métodos que ayudan eficazmente a la diferenciación entre sujetos normales y con cáncer o que están en riesgo de desarrollar una afección médica. La afección médica preferentemente es el cáncer, incluso más preferentemente el cáncer de pulmón. Los sujetos normales se considera que son los que no han sido diagnosticados con cualquier afección médica, tal como el cáncer, más preferentemente los que no han sido diagnosticados de cáncer de pulmón.

La presente divulgación proporciona ventajosamente métodos rápidos, sensibles y fáciles de utilizar para ayudar al diagnóstico de una afección médica, preferentemente, el cáncer, e incluso más preferentemente, cáncer de pulmón. Además, la presente divulgación se puede utilizar para identificar individuos en riesgo de desarrollar una condición médica, para seleccionar sujetos con riesgo de una condición médica y para controlar pacientes diagnosticados de o que se están tratando para una afección médica. La divulgación se puede utilizar para controlar la eficacia del tratamiento de un paciente que se está tratando para una afección médica. Preferentemente, la afección médica es un cáncer e incluso más preferentemente, el cáncer de pulmón. Como se utiliza en toda la presente identificación, particularmente en conexión con los métodos descritos en el presente documento, los sujetos que tienen una afección médica de interés (tal como los sujetos que tienen un cáncer de pulmón) se considera que son las mismas que las de los sujetos que tienen un riesgo de una afección médica de interés (tal como los sujetos que tienen alto riesgo de cáncer de pulmón).

En general, los métodos de la presente divulgación implican una muestra de ensayo obtenida de un sujeto. Normalmente se obtiene una muestra de ensayo de un sujeto y se procesa utilizando métodos de referencia que son conocidas por los expertos en la técnica. Para los especímenes de sangre y suero o plasma derivados de la misma, la muestra se puede obtener convenientemente de la vena antecubital por venopunción, y si se necesita un volumen menor, con un punzón en el dedo. En ambos casos, los elementos formados y coágulos se eliminan por centrifugación. La orina y heces se pueden recolectar directamente del paciente siempre que se procesen rápidamente o se estabilicen con conservantes si el procesamiento no se puede llevar a cabo inmediatamente. Las muestras más especializadas tales como lavados bronquiales o líquido pleural se pueden recolectar durante la broncoscopia o por biopsia transcutánea o abierta y se procesan de forma similar al suero o el plasma una vez que se han eliminado los materiales particulados por centrifugación.

Tras el procesamiento, la muestra de ensayo que se obtiene del paciente se interroga en cuanto a la presencia y cantidad de uno o más biomarcadores que se pueden correlacionar con un diagnóstico de cáncer de pulmón. Específicamente, los solicitantes han encontrado que la detección y cuantificación de uno o más biomarcadores o una combinación de biomarcadores y parámetros biométricos (tal como al menos 1 biomarcador, al menos 1 biomarcador y al menos 1 parámetro biométrico, al menos 2 biomarcadores, al menos 2 biomarcadores y 1 parámetro biométrico, al menos 1 biomarcador y al menos 2 parámetros biométricos, al menos 2 biomarcadores y al menos 2 parámetros biométricos, al menos 3 biomarcadores, etc.) son útiles como una ayuda en el diagnóstico del cáncer de pulmón en un paciente. El uno o más biomarcadores que se identifican y cuantifican en los métodos descritos en el presente documento pueden contenerse en uno o más paneles. El número de biomarcadores que comprenden un panel pueden incluir 1 biomarcador, 2 biomarcadores, 3 biomarcadores, 4 biomarcadores, 5

biomarcadores, 6 biomarcadores, 7 biomarcadores, 8 biomarcadores, 9 biomarcadores, 10 biomarcadores, 11 biomarcadores, 12 biomarcadores, 13 biomarcadores, 14 biomarcadores, 15 biomarcadores, 16 biomarcadores, 17 biomarcadores, 18 biomarcadores, 19 biomarcadores, 20 biomarcadores, etc. Sin embargo, cuando se eligen los biomarcadores que se incluyen en un panel, tales biomarcadores se deben seleccionar por complementariedad de forma que las combinaciones de biomarcadores que se seleccionan proporcionen una mejor actuación (en términos de AUC, sensibilidad y especificidad de DFI o cualquier combinación de AUC, DFI, sensibilidad y especificidad) que las combinaciones individuales.

Como se ha mencionado anteriormente, después de obtener una muestra de ensayo, los métodos de la presente divulgación implican la identificación de la presencia y luego la cuantificación de uno o más biomarcadores en una muestra de ensayo. Se puede cuantificar cualquier biomarcador que sea útil o que se crea útil para la ayuda en el diagnóstico de un paciente sospechoso de estar en riesgo de cáncer de pulmón, en los métodos descritos en el presente documento y que pueden estar contenidos en uno o más paneles. Por consiguiente, en un aspecto, el panel puede incluir uno o más biomarcadores. Ejemplos de biomarcadores que se pueden incluir en un panel, incluyen, pero no se limitan a estos, anti-53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, proteína 1 similar a C1 anti-Niemann-Pick, dominio del péptido del extremo C (anti-dominio-NPC1L1C), anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACIN1, anti-RCV1, anti-MAPKAPK3, anti-Ciclina E2 (a saber, al menos un anticuerpo inmunorreactivo contra la SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5 o cualquier combinación de las mismas), antígenos, tales como, pero sin limitarse a estos antígeno carcinoembrionario (CEA), antígeno de cáncer 125 (CA125), antígeno de cáncer 15-3 (CA 15-3), péptido liberador de progastrina (proGRP), antígeno de células escamosas (SCC), citoqueratina 8, péptidos o proteínas de la citoqueratina 19 (también designados solo como "CK-19, CYFRA 21-1, Cyfra" en el presente documento), y péptidos o proteínas de citoqueratina 18 (CK-18, TPS), antígenos carbohidratos, tales como el antígeno de cáncer 19-9 (CA 19-9), que es el grupo sanguíneo A de Lewis con restos de ácido siálico añadido, amiloide sérico A, alfa-1-anti-tripsina, apolipoproteína CIII y timosina β 4 ($\text{T}\beta$ 4), regiones de interés, tales como, pero sin limitarse a estas, Acn6399, Acn9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959 cualquier combinación de las mismas. Opcionalmente, el panel que se utiliza en el método anterior puede comprender la cuantificación de la cantidad de dos o más biomarcadores, tres o más biomarcadores, cuatro o más biomarcadores, cinco o más biomarcadores, seis o más biomarcadores, siete o más biomarcadores, ocho o más biomarcadores, nueve o más biomarcadores, diez o más biomarcadores, once o más biomarcadores, doce o más biomarcadores, trece o más biomarcadores, catorce o más biomarcadores, quince o más biomarcadores, dieciséis o más biomarcadores, diecisiete o más biomarcadores, dieciocho o más biomarcadores, diecinueve o más biomarcadores o veinte o más biomarcadores, etc.

En otro aspecto, el panel puede contener al menos un anticuerpo, al menos un antígeno, al menos una región de interés, al menos un antígeno y al menos un anticuerpo, al menos un antígeno y al menos una región de interés y al menos un anticuerpo y al menos una región de interés. Ejemplos de al menos un anticuerpo que se puede incluir en el panel, incluyen pero no se limitan a esos, anti-53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, anti-dominio-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACIN1, anti-RCV1, anti-MAPKAPK3 o anti-Ciclina E2. Ejemplos de al menos un antígeno que se puede incluir en el panel son, pero no se limitan a estos, citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA 125, SCC, CA 19-9, proGRP, amiloide sérico A, alfa-1-anti-tripsina, apolipoproteína CIII y $\text{T}\beta$ 4. Ejemplos de al menos una región de interés que se puede incluir en el panel incluyen, pero no se limitan a estas, Acn6399, Acn9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959. Además, se ha encontrado que ciertas regiones de interés se correlacionan altamente (lo que significa que estas regiones de interés tienen altos coeficientes de correlación unas con otras) con otras ciertas regiones de interés y por lo tanto son capaces de sustituirse unas por otras en el contexto de la presente invención. Específicamente, estas regiones de interés altamente correlacionadas se han unido en ciertas familias o "grupos" de correlación. La región de interés que están contenidas en estos "grupos" se pueden sustituir por otra en los métodos y kits de la presente invención. Estas familias o "grupos" de correlación de regiones de interés se describen a continuación:

Grupo A: Las regiones de interés: Pub3448 y Pub3493.

Grupo B: Las regiones de interés: Pub4487 y Pub4682.

Grupo C: Las regiones de interés: Pub8766, Pub8930, Pub9142, Pub9216, Pub9363, Pub9433, Pub9495, Pub9648 y Pub9722.

Grupo D: Las regiones de interés: Pub5036, Pub5139, Pub5264, Pub5357, Pub5483, Pub5573, Pub5593, Pub5615, Pub6702, Pub6718, Pub10759, Pub11066, Pub12193, Pub 13412, Acn10679 y Acn1877.

Grupo E: Las regiones de interés: Pub6391, Pub6533, Pub6587, Pub6798, Pub9317 y Pub13571.

Grupo F: Las regiones de interés: Pub7218, Pub7255, Pub7317, Pub7413, Pub7499, Pub7711, Pub14430 y Pub15599.

Grupo G: Las regiones de interés: Pub8496, Pub8546, Pub8606, Pub8662, Pub8734, Pub17121 y Pub17338.

Grupo H: Las regiones de interés: Pub6249, Pub12501 y Pub12717.

Grupo I: Las regiones de interés: Pub5662, Pub5777, Pub5898, Pub11597 y Acn11559.

Grupo J: Las regiones de interés: Pub7775, Pub7944, Pub7980, Pub8002 y Pub15895.

Grupo K: Las regiones de interés: Pub17858, Pub18422, Pub18766 y Pub18986.

Grupo L: Las regiones de interés: Pub3018, Pub3640, Pub3658, Pub3682, Pub3705, Pub3839, Hic2451,

Hic2646, Hic3035, Tfa3016, Tfa3635 y Tfa4321.

Grupo M: Las regiones de interés: Pub2331 y Tfa2331.

Grupo N: Las regiones de interés: Pub4557 y Pub4592.

Grupo O: Las regiones de interés: Acn4631, Acn5082, Acn5262, Acn5355, Acn5449 y Acn5455.

Grupo P: Las regiones de interés: Acn6399, Acn6592, Acn8871, Acn9080, Acn9371 y Acn9662.

Grupo Q: Las regiones de interés: Acn9459 y Acn9471.

Grupo R: Las regiones de interés: Hic2506, Hic2980, Hic3176 y Tfa2984.

Grupo S: Las regiones de interés: Hic2728 y Hit3276.

Grupo T: Las regiones de interés: Hic6381, Hic6387, Hic6450, Hic6649, Hic6816 e HIC6823.

Grupo U: Las regiones de interés: Hic8791 e HIC8897.

Grupo V: Las regiones de interés Tfa6453 y Tfa6652.

Grupo W: Las regiones de interés: Hic6005 e HIC5376.

Grupo X: Las regiones de interés: Pub4713, Pub4750 y Pub4861.

Los paneles preferidos que se pueden utilizar en los métodos de la presente invención, incluyen, pero no se limitan a:

1. Un panel que comprende al menos dos biomarcadores, donde dichos biomarcadores son al menos un anticuerpo y al menos un antígeno. Este panel puede además comprender biomarcadores adicionales tales como al menos una región de interés.

2. Un panel que comprende al menos un biomarcador, donde dicho biomarcador comprende anti-Ciclina E2. Adicionalmente, el panel puede también comprender opcionalmente biomarcadores adicionales, tales como, (a) al menos un antígeno; (b) al menos un anticuerpo, (c) al menos un antígeno y al menos un anticuerpo; (d) al menos una región de interés; (e) al menos un antígeno y al menos una región de interés; (f) al menos un anticuerpo y al menos una región de interés; o (g) al menos un antígeno, al menos un anticuerpo y al menos una región de interés.

3. Un panel que comprende al menos un biomarcador, donde el biomarcador se selecciona de entre el grupo que consiste en citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA 125, SCC, CA 19-9, proGRP, amiloide sérico A, alfa-1-anti-tripsina y apolipoproteína CIII. El panel puede comprender opcionalmente además biomarcadores adicionales, tales como, al menos un anticuerpo, al menos una región de interés y al menos una región de interés y al menos un anticuerpo de la muestra de ensayo.

4. Un panel que comprende al menos un biomarcador, donde el biomarcador es al menos una región de interés que se selecciona de entre el grupo que consiste en: Acn6399, Acn9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959. El panel puede opcionalmente comprender además biomarcadores adicionales, tales como, al menos un antígeno, al menos un anticuerpo y al menos un antígeno y al menos un anticuerpo de la muestra de ensayo.

5. Un panel que comprende al menos un biomarcador en un panel, donde el al menos un biomarcador se selecciona de entre el grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA 125, SCC, CA 19-9, proGRP, amiloide sérico A, alfa-1-anti-tripsina y apolipoproteína CIII, Acn6399, Acn9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959. El panel puede también opcionalmente comprender además biomarcadores adicionales tales como al menos un anticuerpo. Los paneles preferidos son paneles que comprenden: (a) citoqueratina 19, CEA, ACN9459, Pub11597, Pub4789 y TFA2759; (b) citoqueratina 19, CEA, ACN9459, Pub11597, Pub4789, TFA2759 y TFA9133; (c) citoqueratina 19, CA 19-9, CEA, CA 15-3, CA125, SCC, citoqueratina 18 y ProGRP; (d) Pub11597, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 e HIC3959; y (e) citoqueratina 19, CEA, CA125, SCC, citoqueratina 18, ProGRP, ACN9459, Pub115R7, Pub4789, TFA2759, TFA9133.

6. Un panel que comprende al menos dos biomarcadores en un panel, donde los al menos dos biomarcadores se seleccionan de entre el grupo que consiste en: anti-p53, anti-NY-ESO-1, anti-MAPKAPK3, uno o más anticuerpos contra la Ciclina E2 inmunorreactiva, citoqueratina 19, CEA, CA 125 y ProGRP. El panel puede comprender también al menos tres biomarcadores que se seleccionan de entre el grupo anterior, al menos cuatro biomarcadores que se seleccionan de entre el grupo anterior, al menos cinco biomarcadores que se seleccionan de entre el grupo anterior, al menos seis biomarcadores que se seleccionan de entre el grupo anterior, al menos siete biomarcadores que se seleccionan de entre el grupo anterior, los ocho biomarcadores que se han descrito anteriormente (a saber, cada uno de anti-p53, anti-NY-ESO-1, anti-MAPKAPK3, uno o más anticuerpos contra la Ciclina E2 inmunorreactiva, citoqueratina 19, CEA, CA 125 y ProGRP). Preferentemente, el panel comprende los ocho biomarcadores descritos anteriormente.

La presencia y cantidad de uno o más biomarcadores en la muestra de ensayo se puede obtener y cuantificar utilizando técnicas de rutina conocidas por los expertos en la técnica. Por ejemplo, los métodos para cuantificar antígenos o anticuerpos en muestras de ensayo son bien conocidas por los expertos en la técnica. Por ejemplo, la presencia y cuantificación de uno o más antígenos o anticuerpos en una muestra de ensayo se pueden determinar utilizando uno o más inmunoensayos que se conocen en la técnica. Los inmunoensayos comprenden normalmente: (a) proporcionar un anticuerpo (o antígeno) que se une específicamente a un biomarcador (a saber, un antígeno o un anticuerpo); (b) poner en contacto una muestra de ensayo con el anticuerpo o antígeno; y (c) detectar la

presencia de un complejo del anticuerpo unido al antígeno en la muestra de ensayo o un complejo del antígeno unido al anticuerpo en la muestra de ensayo.

Para preparar un anticuerpo que se una específicamente a un antígeno, se pueden utilizar antígenos purificados o sus secuencias de ácido nucleico. El ácido nucleico o las secuencias de aminoácidos se pueden obtener para posteriores caracterizaciones de estos antígenos. Por ejemplo, los antígenos pueden ser péptidos mapeados con varias enzimas (por ejemplo, tripsina, proteasa V8, etc.). Los pesos moleculares de los fragmentos de la digestión para cada antígeno se pueden utilizar para buscar en las bases de datos, tales como la base de datos SwissProt, secuencias que coincidirán con los pesos moleculares de los fragmentos de digestión generados por varias enzimas. Utilizando este método, se puede identificar el ácido nucleico y las secuencias de aminoácidos de otros antígenos si estos antígenos son proteínas conocidas en las bases de datos.

De manera alternativa, las proteínas se pueden secuenciar utilizando la secuenciación proteica en escalera, por ejemplo, fragmentando las moléculas y sometiendo los fragmentos a la digestión enzimática u otros métodos que retiran un único aminoácido secuencialmente del extremo del fragmento. Los métodos para la preparación de escaleras proteicas se describen por ejemplo en la Publicación Internacional WO 93/24834 y la Patente de EE. UU. Nº 5.792.664. La escalera se analiza luego por espectrometría de masas. La diferencia entre las masas de los fragmentos en escalera identifican el aminoácido retirado del extremo de la molécula.

Si los antígenos no son proteínas conocidas en la base de datos, el ácido nucleico y las secuencias de aminoácidos se pueden determinar por conocimiento de incluso una parte de la secuencia de aminoácidos del antígeno. Por ejemplo, se pueden hacer sondas de degeneración basándose en la secuencia de aminoácidos del extremo amino N del antígeno. Estas sondas se pueden utilizar entonces para explorar una biblioteca genómica o de ADNc que se crea a partir de una muestra de la que se detectó inicialmente un antígeno. Los clones positivos se pueden identificar, amplificar, y sus secuencias de ADN recombinante se pueden subclonar utilizando técnicas que son bien conocidas. Véase por ejemplo, *Current Protocols for Molecular Biology* (Ausubel et al., Green Publishing Assoc. and Wiley-Interscience 1989) y *Molecular Cloning: A Laboratory Manual*, 2ª Ed. (Sambrook et al., Cold Spring Harbor Laboratory, NY 1989).

Utilizando los antígenos purificados o sus secuencias de ácido nucleico, se pueden preparar los anticuerpos que se unen específicamente a un antígeno utilizando cualquiera de los métodos adecuados que se conocen en la técnica (véase, Coligan, *Current Protocols in Immunology* (1991); Harlow & Lane, *Antibodies: A Laboratory Manual* (1988); Goding, *Monoclonal Antibodies: Principles and Practice* (2ª ed. 1986); y Kohler & Milstein, *Nature* 256:495-497 (1975)). Tales técnicas incluyen, pero no se limitan a estas, preparación de anticuerpos por selección de anticuerpos a partir de bibliotecas de anticuerpos recombinantes en fagos o vectores similares, así como la preparación de anticuerpo policlonales o monoclonales inmunizando conejos o ratones (véase, por ejemplo, Huse et al., *Science* 246:1275-1281 (1989); Ward et al., *Nature* 341:544-546 (1989)).

Después de que se proporciona el anticuerpo, se puede detectar y/o cuantificar un antígeno utilizando cualquiera de los ensayos inmunológicos de unión bien reconocidos (véase por ejemplo, las Patentes de EE. UU. números 4.366.241, 4.376.110, 4.517.288, y 4.837.168). Los ensayos que se pueden utilizar en la presente invención incluyen, por ejemplo, un ensayo de inmunoabsorción ligado a enzimas (ELISA), que se conoce también como "ensayo sándwich", un inmunoensayo enzimático (EIA), un radioinmunoensayo (RIA), y fluoroinmunoensayo (FIA), un inmunoensayo quimioluminiscente (CLIA), un inmunoensayo de recuento (CIA), un inmunoensayo enzimático con medio de filtro (MEIA), un ensayo de inmunoabsorción ligado a fluorescencia (FLISA), inmunoensayos de aglutinación e inmunoensayos fluorescentes múltiplex (tales como el Luminex™ LabMAP), etc. Para una revisión de los inmunoensayos en general, véase también, *Methods in Cell Biology: Antibodies in Cell Biology*, volumen 37 (Asai, ed. 1993); *Basic and Clinical Immunology* (Stites & Terr, eds., 7ª ed. 1991).

En general, una muestra de ensayo obtenida de un sujeto se puede poner en contacto con el anticuerpo que se une específicamente con un antígeno. Opcionalmente, el anticuerpo se puede fijar en un soporte sólido antes de poner en contacto el anticuerpo con una muestra de ensayo para facilitar el lavado y posterior aislamiento del complejo. Los ejemplos de soportes sólidos incluyen cristal o plástico en forma de, por ejemplo, placa de microtitulación, portaobjetos de microscopio o cubreobjetos, un bastoncillo, una perla o una microperla. Los anticuerpos se pueden también adjuntar a un sustrato de sonda o sonda ProteinChip™ descrita anteriormente (véase, por ejemplo, Xiao et al., *Cancer Research* 62: 6029-6033 (2001)).

Después de incubar la muestra con anticuerpos, la mezcla se lava y se puede detectar el complejo antígeno-anticuerpo que se ha formado. Esto se puede conseguir incubando la mezcla lavada con un reactivo de detección. Este reactivo de detección puede ser, por ejemplo, un segundo anticuerpo que está marcado con un marcador detectable. En términos de marcador detectable se puede utilizar cualquier marcador detectable conocido en la técnica. Por ejemplo, el marcador detectable puede ser un marcador radioactivo (tales como, por ejemplo, ^3H , ^{125}I , ^{35}S , ^{14}C , ^{32}P , y ^{33}P), un marcador enzimático (tales como, por ejemplo, peroxidasa de rábano rusticano, fosfatasa alcalina, glucosa 6-fosfato deshidrogenasa, y similares), un marcador quimioluminiscente (tal como, por ejemplo, ésteres de acridinio, tioésteres de acridinio, sulfonamidas de acridinio, ésteres de fenantridinio, luminal, isoluminol y similares), un marcador fluorescente (tal como, por ejemplo, fluoresceína (por ejemplo, 5-fluoresceína, 6-

carboxifluoresceína, 3' 6-carboxifluoresceína, 5(6)-carboxifluoresceína, 6-hexacloro-fluoresceína, 6-tetraclorofluoresceína, isotiocianato de fluoresceína, y similares)), rodamina, ficobiliproteínas, R-ficoeritrina, puntos quantum (por ejemplo, selénido de cadmio cubierto por sulfuro de zinc), un marcador termométrico, o un marcador de la reacción en cadena de inmunopolimerasa. Una introducción a marcadores, procedimientos de marcado y
 5 detección de marcadores se encuentra en Polak y Van Noorden, *Introduction to Immunocytochemistry*, 2ª ed., Springer Verlag, N.Y. (1997) y en Haugland, *Handbook of Fluorescent Probes and Research Chemicals* (1996), que es un manual combinado y un catálogo publicado por Molecular Probes, Inc., Eugene, Oregón. De manera alternativa, el marcador se puede detectar en una muestra utilizando un ensayo indirecto, donde, por ejemplo, un
 10 segundo anticuerpo marcado se utiliza para detectar la unión a anticuerpos específicos del marcador, y/o en un ensayo de inhibición o competición donde, por ejemplo, se incuban simultáneamente un anticuerpo monoclonal que se une a un epitopo distintivo del antígeno con la mezcla.

En el transcurso de los ensayos, se pueden necesitar etapas de incubación y/o lavado tras cada combinación de reactivos. Las etapas de incubación pueden variar desde aproximadamente 5 segundos a varias horas,
 15 preferentemente desde aproximadamente 5 minutos a aproximadamente 24 horas. Sin embargo, el tiempo de incubación dependerá del formato del ensayo, el biomarcador (antígeno), volumen de la solución, concentraciones y similares. Habitualmente se llevarán a cabo los ensayos a temperatura ambiente, aunque se pueden realizar en un intervalo de temperaturas, tal como 10 °C a 40 °C.

Las técnicas de inmunoensayo se conocen bien en la técnica, y se puede encontrar una revisión general de la tecnología aplicable en Harlow & Lane, *supra*.

El inmunoensayo se puede utilizar para determinar una cantidad de ensayo de un antígeno en una muestra de un sujeto. Primero, una cantidad de ensayo de un antígeno en una muestra se puede detectar utilizando los métodos de
 25 inmunoensayo que se han descrito anteriormente. Si un antígeno está presente en la muestra, formará un complejo antígeno-anticuerpo con un anticuerpo que se une específicamente al antígeno en condiciones de incubación adecuadas descritas anteriormente. La cantidad de un complejo antígeno-anticuerpo se puede determinar comparando con una referencia. El AUC para el antígeno se puede entonces calcular utilizando técnicas conocidas, tales como, pero sin limitarse a, análisis ROC. De manera alternativa se puede calcular el DFI. Si el AUC es mayor
 30 de aproximadamente 0,5 o el DFI es menor de aproximadamente 0,5, el inmunoensayo se puede utilizar para discriminar sujetos con enfermedad o riesgo de enfermedad (tal como cáncer, preferentemente, cáncer de pulmón) de sujetos normales (o benignos).

Están disponibles comercialmente kits de inmunoensayo para varios antígenos. Por ejemplo, los kits para cuantificar
 35 Citoqueratina 19 están disponibles en F. Hoffmann-La Roche Ltd. (Basel, Switzerland) y Brahms Aktiengesellschaft (Hennigsdorf, Alemania), los kits para cuantificar Citoqueratina 18 están disponibles en IDL Biotech AD (Bromma, Suecia) y en Diagnostic Products Corporation (Los Angeles, CA), los kits para cuantificar CA125, CEA SCC y CA19-9 están disponibles cada uno en Abbott Diagnostics (Abbott Park, IL) y en F. Hoffman-La Roche Ltd., los kits para cuantificar amiloide sérico A y apolipoproteína CIII están disponibles en Linco Research, Inc. (St. Charles, MO), los
 40 kits para cuantificar ProGRP están disponibles en Advanced Life Science Institute, Inc. (Wako, Japón) y en IBL Immuno-Biological Laboratories (Hamburgo, Alemania), los kits para cuantificar alfa 1 antitripsina están disponibles en Autoimmune Diagnostica GMBH (Strassberg, Alemania) y GenWay Biotech, Inc. (San Diego, CA) y los kits para cuantificar Tβ4 están disponibles en ALPCO (Salem, NH).

La presencia y cuantificación de uno o más anticuerpos en una muestra de ensayo se puede determinar utilizando
 45 inmunoensayos similares a los descritos anteriormente. Tales inmunoensayos se llevan a cabo de manera similar que los inmunoensayos descritos anteriormente, excepto por el hecho de que se invierten los papeles de los anticuerpos y antígenos. Por ejemplo, se puede llevar a cabo un tipo de inmunoensayo que es un ensayo con perlas de autoanticuerpos. En este ensayo, un antígeno, tal como el antígeno p53 disponible comercialmente (que se
 50 puede adquirir en BioMol International L.P., Plymouth Landing, Pennsylvania), se puede fijar en un soporte sólido, por ejemplo, una perla, una microplaca de plástico, un portaobjetos o cubreobjetos de microscopio o una membrana hecha de un material tal como la nitrocelulosa que se une a los antígenos proteicos, utilizando técnicas de rutina que se conocen en la técnica o utilizando técnicas y métodos descritos en el Ejemplo 3 en el presente documento. De manera alternativa, si un antígeno no está disponible comercialmente, entonces el antígeno puede purificarse de
 55 líneas celulares cancerosas (preferentemente, líneas celulares de cáncer de pulmón) o los tejidos cancerosos propios de un sujeto (preferiblemente tejidos de cáncer de pulmón) (véase, S-H Hong, et al., *Cancer Research* 64: 5504-5510 (2004)) o se expresa a partir de un clon ADNc (véase, Y-L Lee, et al., *Clin. Chim. Acta* 349: 87-96 (2004)). Se pone entonces en contacto la perla que contiene el antígeno con la muestra de ensayo. Después de
 60 incubar la muestra de ensayo con la perla que contienen el antígeno unido, se lava la perla y se detecta cualquier complejo formado. Esta detección se puede llevar a cabo como se ha descrito anteriormente, a saber, incubando la perla lavada con un reactivo de detección. Este reactivo de detección puede ser por ejemplo, un segundo anticuerpo (tal como, pero sin limitarse a, inmunoglobulina G anti-humana (IgG), inmunoglobulina A anti-humana (IgA), inmunoglobulina M anti-humana (IgM), que está marcada con un marcador detectable. Después de la detección, se
 65 puede determinar la cantidad del complejo antígeno-anticuerpo comparando la señal con la generada por una referencia, como se ha descrito anteriormente. De manera alternativa, el complejo antígeno-anticuerpo se puede detectar aprovechando la naturaleza multivalente de las inmunoglobulinas. En vez de hacer reaccionar el complejo

antígeno-anticuerpo con un anticuerpo anti-humano, el complejo antígeno anticuerpo se puede exponer a un antígeno soluble que está marcado con un marcador detectable que contiene el mismo epítipo que el antígeno adjuntado a la fase sólida. Cualquiera de los sitios de unión del anticuerpo sin ocupar se unirá al antígeno soluble (que está marcado con un marcador detectable). Tras el lavado, se detecta el marcador detectable utilizando técnicas de rutina que se conocen por los expertos en la técnica. Cualquiera de los métodos que se han descrito anteriormente permiten la cuantificación de la sensibilidad y la especificidad de un anticuerpo específico en la muestra de ensayo. El AUC del anticuerpo (y por tanto, la utilidad del anticuerpo, tal como un autoanticuerpo, para detectar el cáncer de pulmón en un sujeto) se puede entonces calcular utilizando técnicas de rutina conocidas para los expertos en la técnica, tales como, pero sin limitarse a, análisis ROC. De manera alternativa se puede calcular el DFI. Si el AUC es mayor que aproximadamente 0,5 o el DFI es menor de aproximadamente 0,5, el inmunoensayo se puede utilizar para discriminar entre sujetos enfermos o con riesgo de enfermedad (tal como cáncer, preferentemente, cáncer de pulmón) de sujetos normales (o benignos).

La presencia y cantidad de regiones de interés se puede determinar utilizando técnicas de espectrometría de masas. Utilizando espectrometría de masas, los solicitantes han encontrado 212 regiones de interés que son útiles como ayuda al diagnóstico y detección de cáncer de pulmón en muestras de ensayo. Específicamente, cuando se utilizan técnicas de espectrometría de masas para detectar y cuantificar uno o más biomarcadores en una muestra de ensayo, la muestra de ensayo se tienen que preparar primero para el análisis de espectrometría de masas. La preparación de la muestra puede tener lugar de varias maneras, pero la que se utiliza más comúnmente implica poner en contacto la muestra con uno o más adsorbentes unidos a una fase sólida. Los adsorbentes pueden ser grupos aniónicos o catiónicos, grupos hidrófobos, grupos quelantes metálicos con o sin un ligando metálico, anticuerpos, sean policlonales o monoclonales, o antígenos adecuados para unirse a sus anticuerpos afines. La fase sólida puede ser una superficie plana fabricada en metal, cristal, o plástico. La fase sólida también puede ser de naturaleza microparticulada, sean micropérlas, particulados amorfos, o polímeros insolubles para aumentar el área de superficie. Además, los materiales microparticulados pueden ser magnéticos para el caso de manipulación. Los biomarcadores de interés se absorben en la fase sólida y las moléculas restantes se eliminan por lavado. Por análisis de masas, los biomarcadores de interés se eluyen de la fase sólida con un disolvente que reduce la afinidad del biomarcador por el adsorbente. Los biomarcadores se introducen luego en el espectrómetro de masas para su análisis. Preferentemente, los espectros periféricos se identifican y se omiten en la evaluación del espectro. Adicionalmente, se pueden utilizar también inmunoensayos tales como los que se han descrito anteriormente. Con la terminación de un inmunoensayo, el analito se puede eluir de la superficie inmunológica y se introduce en el espectrofotómetro para su análisis.

Una vez que se ha preparado la muestra de ensayo, se introduce en el analizador de masas. La desorción ionización por láser (por ejemplo, MALDI o SELDI) es una técnica común para las muestras que se presentan en forma sólida. En esta técnica, la muestra se co- cristaliza en una placa diana con una matriz eficaz para absorber y transferir la energía del láser a la muestra. Los iones se separan, se cuentan, y se calibran contra los iones de masa y carga conocida. Los datos de masa recolectados por cualquier muestra es un recuento iónico con una relación específica de masa/carga (m/z). Se anticipa que diferentes métodos de preparación de la muestra y diferentes técnicas de ionización darán como resultado diferentes espectros.

Los ensayos de cualificación para los datos del espectro de masas normalmente implican un proceso riguroso de análisis periférico con un mínimo pre-procesamiento de los datos originales. El proceso de identificar periféricos comienza con el cálculo de la corriente iónica total (TIC) del espectro bruto. No se aplican algoritmos de suavización o corrección de la línea base a los espectros brutos antes del cálculo de la TIC. La TIC se calcula sumando la intensidades de cada valor m/z a través del intervalo de masa (m/z) detectado. Esto muestra los fallos del instrumento, los problemas de moteado de la muestra, y otros defectos similares. Además de la TIC, se calcula el %CV medio (porcentaje del coeficiente de variación) a lo largo del espectro completo para cada muestra. Utilizando el número de mediciones replicadas para cada muestra, se calcula un %CV para cada valor m/z a lo largo del intervalo de masa detectado. Estos %CV se promedian juntos para obtener un %CV medio que es representativo para esa muestra en particular. El %CV medio puede o no utilizarse como una primera etapa de filtrado para identificar periféricos. En general, los replicados con altos %CV medios (mayores del 30 % o cualquier otro valor aceptable) indica baja reproducibilidad.

Como se ha descrito anteriormente, la TIC calculada y el %CV medio de cada espectro se podría utilizar como predictores para cualificar la reproducibilidad y la "bondad" de los espectros. Sin embargo, mientras estas mediciones proporcionan un buen descriptor para la propiedad en masa del espectro, no dan ninguna información sobre la reproducibilidad de las características destacadas de los espectros tales como las intensidades individuales en cada valor m/z . Este obstáculo se supera por una adaptación de los cálculos del Ángulo de Contraste Espectral (SCA) comunicados por Wan et. al. (J. Am. Soc. Mass Spectrom. 2002, 13, 85-88). En los cálculos del SCA, todo el espectro se trata como un vector cuyos componentes son los valores individuales de m/z . Con esta interpretación, el ángulo theta (θ) entre los dos vectores se da por la fórmula matemática de referencia

$$\cos(\theta) = \frac{v_1 \cdot v_2}{\sqrt{v_1 \cdot v_1} * \sqrt{v_2 \cdot v_2}}$$

Theta será pequeño, cerca de cero, para espectros similares.

En la práctica, el número total de cálculos y comparaciones se reducen calculando primero un espectro medio para cada replicado de muestra o para todas las muestras en un grupo en particular (por ejemplo, cánceres). Luego se calcula un SCA entre cada espectro y el espectro medio calculado. Los espectros que difieren drásticamente del espectro medio se consideran periféricos, siempre que, cumplan los criterios que se describen posteriormente.

El uso de más de un predictor para seleccionar periféricos es preferible porque un predictor no es suficiente para describir completamente un espectro de masa. Se puede llevar a cabo un análisis de periféricos multivariado utilizando múltiples predictores. Estos predictores podían ser, pero no se limitan a, la TIC, el %CV medio, y SCA. Utilizando el paquete estadístico JMP™ (SAS Institute Inc., Cary, NC), se calculan las distancias Mahalanobis para cada medición de cada replicado en el grupo (por ejemplo, cáncer). Un valor crítico (no un límite de confianza) se puede calcular tal que aproximadamente el 95 % de las observaciones caigan por debajo de este valor. El 5 % restante que cae por encima del valor crítico se consideran periféricos y se descartan para análisis posteriores.

Tras la cualificación de los datos de masa espectral, los espectros habitualmente se normalizan, se escalan las intensidades de forma que la TIC sea la misma para todos los espectros del grupo de datos o se escalan las intensidades con respecto a un pico en todos los espectros.

Tras la normalización, los espectros de masas se reducen a un grupo de características de intensidad. En otras aplicaciones, estos se reducen a una lista de intensidades espectrales a valores m/z asociados con las biomoléculas. Preferentemente, se utiliza otro tipo de característica, la región de interés o ROI.

Las regiones de interés son productos de una comparación entre dos o más grupos de datos de interés. Estos grupos de datos representan los grupos de interés (por ejemplo, enfermedad y no enfermedad). Se lleva a cabo un t-test de los valores de intensidad a través de todas las muestras por cada m/z. Se identifican estos valores m/z con valores de p del t-test menores de un umbral especificado para un operador. De los valores m/z identificados, los contiguos se agrupan juntos y se definen como una región de interés. El número mínimo de valores m/z contiguos necesarios para formar una ROI y cualquier hueco permitido en ese grupo contiguo, se pueden definir por el usuario. Otro calificador para la ROI es el valor absoluto del logaritmo de la relación de las medias de los dos grupos. Cuando este valor es mayor que un valor de corte del umbral, digamos 0,6 cuando se utilizan logaritmos de base 10, la localización de masa a carga se convierte en un candidato de inclusión en una ROI. La ventaja de utilizar el método ROI es que no solo marca diferencias en el patrón de altas intensidades entre los espectros de las dos clases sino que encuentra más diferencias sutiles como hombros e intensidades muy bajas que se perderían por los métodos de encontrar picos.

Una vez que se ha determinado la región de interés, la media o valor m/z medio del intervalo de la ROI a menudo se utiliza como un identificador de la región. Cada región es un marcador potencial que diferencia los grupos de datos. Se pueden extraer una variedad de parámetros (por ejemplo, intensidad total, intensidad máxima, intensidad media, o promedio de intensidad) de los datos de la muestra y asociarse con la ROI. Por lo tanto, cada espectro de una muestra se ha reducido desde muchos cientos de m/z, pares de intensidad para 212 ROI y su identificador, pares de función de intensidad. Estos descriptores se utilizan como variables de aportación para las técnicas de análisis de datos.

Opcionalmente, sea antes de la obtención de una muestra de ensayo o tras obtener la muestra de ensayo y antes de identificar y cuantificar uno o más biomarcadores en una muestra de ensayo o tras identificar y cuantificar uno o más biomarcadores en una muestra de ensayo, los métodos de la presente divulgación pueden incluir la etapa de obtención de un valor para al menos un parámetro biométrico de un sujeto. Por ejemplo, se pueden obtener 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc., factores biométricos de un sujeto. De manera alternativa, los métodos de la presente divulgación no tienen que incluir una etapa de obtención de un valor de cualquiera de los parámetros biométricos de un sujeto. El parámetro biométrico preferido que se obtiene de un sujeto es la historia de fumador del sujeto, específicamente, los paquetes-años de fumador del sujeto. Otros parámetros biométricos que se pueden obtener del sujeto incluyen, pero sin limitarse a estos, la edad, exposición a carcinógenos, género, historia familiar de fumadores, etc.

Como se ha mencionado anteriormente. En los métodos de la presente divulgación, la muestra de ensayo se analiza para determinar la presencia de uno o más biomarcadores contenidos en el panel. Si se determina que está presente un biomarcador en la muestra de ensayo, entonces se cuantifica la cantidad de cada tal biomarcador detectado (utilizando las técnicas descritas previamente en el presente documento). Una vez que la cantidad de cada biomarcador de la muestra de ensayo se cuantifica, entonces se compara la cantidad de cada biomarcador cuantificado con un punto de corte predeterminado (que es normalmente, un valor o un número, tal como un entero, y que se designa alternativamente en el presente documento como "punto de corte" o "punto de división") para esos biomarcador específico. El punto de corte predeterminado que se emplea en los métodos de la presente divulgación se puede determinar utilizando técnicas de rutina conocidas en la técnica, tales como, pero sin limitarse a estas, análisis multi-variado (véase la **Figura 1**), regresión logística transformada, un método de división y puntuación o cualquier combinación de las mismas. Por ejemplo, cuando se utiliza el Método de división y Puntuación, el valor o

número del punto de corte predeterminado dependerá del resultado deseado que se va a conseguir. Si el resultado deseado que se va a conseguir es para maximizar la precisión de calcificaciones correctas de cada marcador en un grupo de interés (a saber, identificación correcta de los sujetos en riesgo de desarrollar un cáncer de pulmón y los que no están en riesgo de desarrollar un cáncer de pulmón), entonces se elegirá un valor o número específico para un punto de corte predeterminado para ese biomarcador basándose en ese resultado deseado. Por el contrario, si el resultado deseado es maximizar la sensibilidad de cada marcador, entonces se debe elegir un valor o número diferente para el punto de corte predeterminado para ese biomarcador basándose en el resultado deseado. De la misma manera, si el resultado deseado es maximizar la especificidad de cada marcador, entonces se puede escoger un valor diferente del punto de corte predeterminado para ese biomarcador basándose en ese resultado deseado.

Una vez que se ha cuantificado la cantidad de cualquiera de los biomarcadores presentes en la muestra de ensayo, esta información se puede utilizar para generar curvas ROC, AUC y otra información que se puede utilizar por un experto en la técnica utilizando técnicas de rutina para determinar el punto de corte predeterminado adecuado para cada biomarcador dependiendo del resultado deseado. Después de que la cantidad de cada biomarcador se haya comparado con el punto de corte predeterminado, se le asigna entonces una puntuación (a saber, un número, que puede ser cualquier entero, tal como desde 0 a 100) a cada biomarcador basándose en la comparación. Además, si se obtienen además del uno o más biomarcadores, los valores de uno o más parámetros biométricos para un sujeto, entonces el valor de cada parámetro biométrico se compara con un punto de corte predeterminado para dicho parámetro biométrico. El punto de corte predeterminado para cualquier parámetro biométrico se puede determinar utilizando las mismas técnicas que se han descrito en el presente documento con respecto a la determinación de puntos de corte predeterminados para uno o más biomarcadores. Como con la comparación de biomarcadores, se asigna entonces una puntuación (a saber, un número, que puede ser un entero, tal como 1 a 100) a ese parámetro biométrico basándose en dicha comparación.

De manera alternativa, en vez de utilizar el método de puntuación descrito anteriormente, se puede utilizar un Método de Puntuación Ponderada. Si se utiliza un Método de Puntuación Ponderada, entonces una vez que se ha cuantificado la cantidad de cada biomarcador en una muestra de ensayo, la cantidad de cada biomarcador detectado en la muestra de ensayo se compara con un número de puntos de corte predeterminados para ese biomarcador específico. A partir de todos los puntos de corte predeterminados diferentes disponibles, se asigna entonces una única puntuación (a saber, un número, que puede ser cualquier entero, tal como de 0 a 100) a ese biomarcador. Este Método de Puntuación Ponderada también se puede utilizar con uno o más parámetros biométricos.

Una vez que se ha asignado una puntuación para cada uno de los biomarcadores cuantificados, y opcionalmente para cada uno de los parámetros biométricos obtenidos del sujeto, entonces la puntuación para cada biomarcador o cada biomarcador y cada parámetro biométrico se combinan para obtener una puntuación total para el sujeto. Esta puntuación total se compara entonces con una puntuación total predeterminada. Basándose en esta comparación, se puede hacer una determinación de si un sujeto está en riesgo de un cáncer de pulmón o no. La determinación de si un sujeto está o no en riesgo de un cáncer de pulmón se puede basar en si la puntuación es o no más alta o más baja que la puntuación total predeterminada. Por ejemplo, dependiendo del valor asignado respecto a la puntuación total predeterminada, un sujeto con una puntuación total que sea más alta que la puntuación total predeterminada se puede considerar que tiene un riesgo mayor y por lo tanto se puede remitir para ensayos posteriores o procedimientos de seguimiento. La puntuación total predeterminada (designada de manera alternativa en el presente documento como un "umbral") que se va a utilizar en el método se puede determinar utilizando las mismas técnicas que se han descrito anteriormente con respecto a las puntuaciones predeterminadas para los biomarcadores. Por ejemplo, la Figura 5 proporciona tres curvas ROC. Cada una de estas curvas ROC representa el resultado único de marcadores combinados, sin embargo un único marcador produciría una curva ROC similar. Las curvas ROC abarcan desde baja sensibilidad y la tasa baja de falsos positivos (1-especificidad) en un extremo hasta la sensibilidad más alta y la tasa alta de falsos positivos en el otro extremo. La forma de la curva entre estos dos extremos puede variar significativamente. Si se necesitara un método para tener al menos un 90 % de sensibilidad, entonces basándose en las curvas ROC que se muestran en la Figura 5, la tasa de falsos positivos sería del 60-70 % dependiendo de la curva escogida. Si el método necesitara tener al menos un 10 % de tasa de falsos positivos, entonces la sensibilidad sería del 40-55 % dependiendo de la curva escogida. Ambos métodos se derivan del mismo panel de marcadores, sin embargo, con el fin de proporcionar la actuación de características clínicas diferentes, el umbral (o puntuación total predeterminada) del panel se ha cambiado. Por medio del cálculo, por debajo de cada punto de la curva ROC hay un umbral (o puntuación total predeterminada) que se mueve de un extremo del intervalo de datos al otro extremo del intervalo de datos. Cuando el umbral (o puntuación total predeterminada) está en el extremo bajo del intervalo de datos, entonces todas las muestras son positivas y esto produce un punto en la curva ROC con alta sensibilidad y alta tasa de falsos positivos. Cuando el umbral (o puntuación total predeterminada) está en el extremo alto del intervalo de datos, entonces todas las muestras son negativas y esto produce un punto en la curva ROC con una baja sensibilidad y baja tasa de falsos positivos. A menudo se necesita un método para tener una característica clínica deseada, tal como un nivel mínimo de sensibilidad (es decir, el 90 %), un nivel mínimo de especificidad (es decir, un 90 %), o ambos. El cambio del umbral (o puntuación total predeterminada) de los marcadores puede ayudar a conseguir las características clínicas deseadas.

Las etapas descritas anteriormente de (a) comparar la cantidad de cada biomarcador de un panel con un punto de corte predeterminado (o varios puntos de corte predeterminados si se utiliza el Método de Puntuación Ponderada),

que asigna una puntuación (o una puntuación de una o varias puntuaciones posibles si se utiliza el Método de Puntuación Ponderada) para cada biomarcador basándose en la comparación, combinando la puntuación asignada a cada parámetro biométrico de un panel para obtener una puntuación total para el sujeto, comparando la puntuación total con una puntuación total predeterminada y determinando si un sujeto tiene riesgo de un cáncer de pulmón basándose en la comparación de la puntuación total con la puntuación predeterminada; o (b) comparar el valor de al menos un parámetro biométrico con un punto de corte predeterminado (o varios puntos de corte predeterminados si se utiliza el Método de Puntuación Ponderada) para cada parámetro biométrico basándose en dicha comparación, comparar la cantidad de cada biomarcador de un panel con un punto de corte predeterminado, asignar una puntuación para cada biomarcador basándose en la comparación, combinar la puntuación asignada para cada parámetro biométrico con la puntuación asignada para cada biomarcador cuantificado para obtener una puntuación total para el sujeto, comparando la puntuación total con una puntuación total predeterminada y determinar si un sujeto tiene un riesgo de cáncer de pulmón basándose en la comparación de la puntuación total con la puntuación predeterminada pueden llevarse a cabo manualmente por una persona, o se pueden llevar a cabo completa o parcialmente con un programa de computadora o algoritmo, junto con el hardware necesario, tal como una entrada, memoria, procesador, pantalla y dispositivos de salida.

Con fines solamente ilustrativos, un ejemplo de cómo se puede llevar a cabo el método de la presente divulgación se dará ahora. En este ejemplo, un paciente se ensaya para determinar la probabilidad del paciente de tener un cáncer de pulmón utilizando un panel que comprende 8 biomarcadores y el Método de División y Puntuación. Los biomarcadores del panel son citoqueratina 19, CEA, CA125, CA15-3, CA19-9, SCC, proGRP y citoqueratina 18. La puntuación total predeterminada (o umbral) para el panel es 3. Después de obtener la muestra de ensayo del paciente, se cuantifica la cantidad de cada uno de los 8 biomarcadores (citoqueratina 19, CEA, CA125, CA15-3, CA19-9, SCC, proGRP y citoqueratina 18) en la muestra del paciente. Para los fines de este ejemplo, la cantidad determinada de cada uno de los 8 biomarcadores de la muestra de ensayo era: citoqueratina 19: 1,95, CEA: 2,75, CA125: 15,26, CA15-3: 11,92, CA19-9: 9,24, SCC: 1,06, proGRP: 25,29 y citoqueratina 18: 61,13. La cantidad de cada uno de estos biomarcadores se comparan con el correspondiente punto de corte predeterminado (o punto de división). Los puntos de corte predeterminados de cada uno de los biomarcadores son: citoqueratina 19: 1,89, CEA: 4,82, CA125: 13,65, CA15-3: 13,07, CA19-9: 10,81, SCC: 0,92, proGRP: 14,62 y citoqueratina 18: 57,37. Para cada biomarcador que tiene una cantidad que es mayor de la del correspondiente punto de corte predeterminado (punto de división), se puede dar una puntuación de 1. Para cada biomarcador que tiene una cantidad que es menor o igual que su correspondiente punto de corte predeterminado, se puede dar una puntuación de 0. Por consiguiente, basándose en dicha comparación, se asignaría una puntuación a cada biomarcador como sigue: citoqueratina 19: 1, CEA: 0, CA125: 1, CA15-3: 0, CA19-9: 0, SCC: 1, proGRP: 1, y citoqueratina 18: 1. La puntuación para cada uno de los 8 biomarcadores se combinan entonces matemáticamente (es decir, sumando cada una de las puntuaciones de los biomarcadores juntos) para llegar a una puntuación total del paciente. La puntuación total del paciente es 5 (la puntuación total se calcula como sigue: $1 + 0 + 1 + 0 + 0 + 1 + 1 = 5$). La puntuación total del paciente se compara con la puntuación total predeterminada, que es 3. Una puntuación total mayor que la puntuación total predeterminada de 3 indicaría un resultado positivo para el paciente. Una puntuación total menor o igual a 3 indicaría un resultado negativo para el paciente. En este ejemplo, debido a que la puntuación total del paciente es mayor de 3, el paciente se consideraría que tiene un resultado positivo y por lo tanto se remitiría para posteriores ensayos por una indicación o sospecha de cáncer de pulmón. Por el contrario, si la puntuación total del paciente hubiera sido 2, el paciente se habría considerado que tenía un resultado negativo y no sería remitido para posteriores ensayos.

En un ejemplo más, el panel de 8 biomarcadores descrito anteriormente se utiliza de nuevo, sin embargo, en este ejemplo, se emplea el Método de Puntuación Ponderada. En este ejemplo, la puntuación total predeterminada (o umbral) para el panel es de 11,2 y las cantidades de biomarcadores cuantificados en la muestra de ensayo era la misma que se describe anteriormente. La cantidad de cada uno de los biomarcadores se compara entonces con 3 puntos de corte predeterminados diferentes para cada uno de los biomarcadores. Por ejemplo, los puntos de corte predeterminados para cada uno de los biomarcadores se proporcionan posteriormente en la Tabla A.

Tabla A

	CEA	Citoqueratina 18	ProGRP	CA15-3	CA125	SCC	Citoqueratina 19	CEA19-9
Punto de corte predeterminado @ 50 % especificidad	2,02	47,7	11,3	16,9	15,5	0,93	1,2	10,6
Punto de corte predeterminado @ 75 % especificidad	3,3	92,3	18,9	21,8	27	1,3	1,9	21,9
Punto de corte predeterminado @ 90 % especificidad	4,89	143,3	28,5	30,5	38,1	1,98	3,3	45,8

	CEA	Citoqueratina 18	ProGRP	CA15-3	CA125	SCC	Citoqueratina 19	CEA19-9
Puntuación por debajo del 50 % especificidad	0	0	0	0	0	0	0	0
Puntuación por encima del 50 % especificidad	<u>2,68</u>	<u>2,6</u>	2,48	1,16	2,68	<u>2,48</u>	4,2	1,1
Puntuación por encima del 75 % especificidad	5,36	5,2	<u>4,96</u>	2,32	5,36	4,96	<u>8,4</u>	2,2
Puntuación por encima del 90 % especificidad	13,4	13	12,4	5,8	13,4	12,4	21	5,5

Por lo tanto, se pueden dar 4 posibles puntuaciones para cada biomarcador. La cantidad de cada biomarcador cuantificado se compara con los puntos de corte predeterminados (puntos de división) que se proporcionan en la Tabla A anterior. Por ejemplo para CEA, como la cantidad de CEA cuantificada en la muestra de ensayo era de 2,74, que cae entre el punto de corte predeterminado de 2,02 para el 50 % de especificidad y el 3,3 para el 75 % de especificidad de la Tabla A. Por lo tanto, se asigna una puntuación para CEA de 2,68. Estos se repite para el resto de los biomarcadores que se evalúan de manera similar y se asigna a cada uno las siguientes puntuaciones: Citoqueratina 18: 2,6, proGRP: 4,96, CA15-3: 0, CA125: 0, SCC: 2,48, citoqueratina 19: 8,4 y CA19-9: 0. La puntuación de cada uno de los 8 biomarcadores se combinan entonces matemáticamente (es decir, sumando cada una de las puntuaciones de los biomarcadores juntas) para llegar a la puntuación total para el paciente. La puntuación total para el paciente es 21,12 (la puntuación total se calcula de la siguiente manera $2,68 + 2,6 + 4,96 + 0 + 0 + 2,48 + 8,4 + 0 = 21,12$). La puntuación total del paciente se compara con la puntuación total predeterminada que es 11,2. Una puntuación total mayor que la puntuación total predeterminada de 11,2 indicaría un resultado positivo para el paciente. Una puntuación total menor o igual de 11,2 indicaría un resultado negativo para el paciente. En este ejemplo, debido a que la puntuación total del paciente era mayor de 11,2, el paciente se consideraría que tiene un resultado positivo y por lo tanto se remitiría para posteriores ensayos por una indicación o sospecha de cáncer de pulmón.

Además, en otra realización, también se puede utilizar el Método de Puntuación Ponderada descrito en el presente documento para puntuar uno o más marcadores obtenidos del sujeto. Preferentemente, tales marcadores, si son uno o más biomarcadores, uno o más parámetros biométricos o una combinación de biomarcadores y parámetros biométricos se pueden utilizar como ayuda en el diagnóstico o la evaluación de si un sujeto está en riesgo de desarrollar una afección médica, tal como un cáncer o alguna otra enfermedad. El Método de Puntuación Ponderada se puede utilizar ampliamente en conexión con una afección médica en la que los marcadores se utilizan o se pueden utilizar para valorar si un sujeto está en riesgo de tal afección médica. Tal método comprende las etapas de:

- cuantificar en una muestra de ensayo obtenida de un sujeto, la cantidad de al menos un marcador;
- comparar la cantidad de cada marcador cuantificado con varios puntos de corte predeterminados para dicho marcador y asignar una puntuación para cada marcador basándose en dicha comparación; y
- combinar la puntuación asignada para cada marcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto.

Preferentemente, el método comprende las etapas de:

- cuantificar en una muestra de ensayo obtenida de un sujeto, la cantidad de al menos un marcador;
- comparar la cantidad de cada marcador cuantificado con varios puntos de corte predeterminados para dicho marcador y asignar una puntuación para cada marcador basándose en dicha comparación;
- combinar la puntuación asignada para cada marcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto;
- comparar la puntuación total determinada en la etapa c con una puntuación total predeterminada; y
- determinar si dicho sujeto tiene riesgo de desarrollar una afección médica basándose en la comparación de la puntuación total determinada en la etapa d.

DFI

Como se ha tratado anteriormente en el presente documento, los solicitantes han descubierto que la detección y cuantificación de uno o más biomarcadores y parámetros biométricos es útil como una ayuda en el diagnóstico del cáncer de pulmón en un paciente. Además, los solicitantes también han descubierto que el uno o más biomarcadores y las combinaciones del uno o más biomarcadores y uno o más parámetros biométricos que se describen en el presente documento tienen un DFI relacionado con el cáncer de pulmón menor de aproximadamente 0,5, preferentemente menor de aproximadamente 0,4, más preferentemente, menos de aproximadamente 0,3 e

incluso más preferentemente, menos de aproximadamente 0,2. Las Tablas 25-29 proporcionan ejemplos de paneles que contienen varios biomarcadores o combinaciones de biomarcadores y parámetros biométricos que muestran un DFI que es menor de aproximadamente 0,5, menos de aproximadamente 0,4, menos de aproximadamente 0,3 y menos de aproximadamente 0,2.

5

KITS

Uno o más biomarcadores, uno o más de los polipéptidos Ciclina E2 inmunorreactiva, parámetros biométricos o cualquier combinación de los mismos son adecuados para la formación de kits (tal como paneles) para su uso en la realización de los métodos de la presente invención. En un aspecto, el kit puede comprender un péptido que se selecciona de entre el grupo que consiste en las SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5 o combinaciones de las mismas.

En otro aspecto el kit puede combinar anti-Ciclina E2 o cualquier combinación de la misma.

15

En un aspecto más, el kit puede comprender (a) reactivos que contienen al menos un anticuerpo para cuantificar uno o más antígenos en la muestra de ensayo, donde dichos antígenos son: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide sérico A, alfa-1-antitripsina, apolipoproteína CIII y Tβ4; (b) reactivos que contiene uno o más antígenos para cuantificar al menos un anticuerpo en una muestra de ensayo; donde dichos anticuerpos son anti-p53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, anti-domino -NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACIN1, anti-RCV1, anti-MAPKAPK3 y anti-Ciclina E2; y opcionalmente (c) uno o más algoritmos o programas informáticos para llevar a cabo las etapas de combinación y comparación de la cantidad de cada antígeno y anticuerpo cuantificados en la muestra de ensayo con un punto de corte predeterminado (o con varios puntos de corte predeterminados) y asignar una puntuación para cada antígeno y anticuerpo (o una puntuación de una o varias puntuaciones posibles) cuantificados basándose en dicha comparación, combinar la puntuación asignada para cada antígeno y anticuerpo cuantificado para obtener una puntuación total, comparar la puntuación total con una puntuación total predeterminada y utilizando dicha comparación como una ayuda para determinar si un sujeto tiene cáncer de pulmón. De manera alternativa, en lugar de uno o más algoritmos o programas de computadora, se pueden proporcionar una o más instrucciones para que una persona lleve a cabo manualmente las etapas anteriores.

30

En otro aspecto más, el kit puede contener: (a) reactivos que contienen al menos un anticuerpo para cuantificar uno o más antígenos en una muestra de ensayo, donde dichos antígenos son citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide sérico A, alfa-1-anti-tripsina, apolipoproteína CIII y Tβ4; (b) reactivos que contienen uno o más antígenos para cuantificar al menos un anticuerpo en una muestra de ensayo; donde dichos anticuerpos son: anti-p53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, anti-domino -NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACIN1, anti-RCV1, anti-MAPKAPK3 y anti-Ciclina E2; (c) reactivos para cuantificar una o más regiones de interés que se seleccionan de entre el grupo que consiste en: ACN9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 e HIC3959; y opcionalmente, (d) uno o más algoritmos o programas de computadora para llevar a cabo las etapas de combinación y comparación de la cantidad de cada antígeno, anticuerpo y región de interés cuantificados en la muestra de ensayo con un punto de corte predeterminado (o con varios puntos de corte predeterminados) y asignar una puntuación para cada antígeno, anticuerpo y región de interés (o una puntuación de una de varias puntuaciones posibles cuantificados basándose en dicha comparación, combinar la puntuación asignada para cada antígeno, anticuerpo y región de interés cuantificados para obtener una puntuación total, comparar la puntuación total con una puntuación total predeterminada y utilizar dicha comparación como una ayuda para determinar si el sujeto tiene un cáncer de pulmón. De manera alternativa, en lugar de uno o más algoritmos o programas de computadora, se pueden proporcionar una o más instrucciones que una persona lleve a cabo las etapas anteriores manualmente. Los reactivos incluidos en el kit para cuantificar una o más regiones de interés pueden incluir un adsorbente que se una y retenga al menos una región de interés que esté contenida en un panel, soportes sólidos (tales como perlas) para usarse en conexión con dicho adsorbentes, uno o más marcadores detectables, etc. El adsorbente puede ser cualquiera de los muchos adsorbentes que se usan en química analítica e inmunoquímica, incluyendo quelantes metálicos, grupos catiónicos, grupos aniónicos, grupos hidrófobos, antígenos y anticuerpos. En otro aspecto más, el kit puede comprender: (a) reactivos que contienen al menos un anticuerpo para cuantificar uno o más antígenos en una muestra de ensayo, donde dichos antígenos son citoqueratina 19, citoqueratina 18, CA 19-9, CEA, CA15-3, CA125, SCC y ProGRP; (b) reactivos para cuantificar una o más regiones de interés seleccionadas de entre el grupo que consiste en ACN9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 e HIC3959; y opcionalmente, (c) uno o más algoritmos o programas de computadora para llevar a cabo las etapas de combinación y comparación de la cantidad de cada antígeno y región de interés cuantificados en la muestra de ensayo con un punto de corte predeterminado (o con varios puntos de corte predeterminados posibles) y asignar una puntuación para cada antígeno y región de interés (o una puntuación para una de varias puntuaciones posibles) que se cuantifican basándose en dicha comparación, combinando la puntuación asignada para cada antígeno y región de interés cuantificados para obtener una puntuación total, comparar la puntuación total con una puntuación total predeterminada y utilizar dicha comparación como una ayuda para determinar si un sujeto tiene cáncer de pulmón. De manera alternativa, el lugar de uno o más algoritmos o programas de computadora, se puede proporcionar una o más instrucciones para que una persona lleve a cabo

65

manualmente las etapas anteriores. Los reactivos incluidos en el kit para cuantificar una o más regiones de interés pueden incluir un adsorbente que se una y retenga al menos una región de interés contenida en un panel, soportes sólidos (tales como perlas) para usarse en conexión con dichos adsorbentes, uno o más marcadores detectables, etc. Preferentemente, el kit contiene los reactivos necesarios para cuantificar los siguientes antígenos y regiones de interés: (a) citoqueratina 19 y CEA y Acn9459, Pub11597, Pub4789 y Tfa2759; (b) citoqueratina 19 y CEA y Acn9459, Pub11597, Pub4789, Tfa2759 y Tfa9133; y (c) citoqueratina 19, CEA, CA125, SCC, citoqueratina 18, y ProGRP y ACN9459, Pub11597, Pub4789 y Tfa2759.

En otro aspecto, un kit puede comprender (a) reactivos que contienen al menos un anticuerpo para cuantificar uno o más antígenos en la muestra de ensayo, donde dichos antígenos son citoqueratina 19, citoqueratina 18, CA 19-9, CEA, CA15-3, CA125, SCC y ProGRP; y opcionalmente, (b) uno o más algoritmos o programas de computadora para llevar a cabo las etapas de combinación y comparación de la cantidad de cada antígeno cuantificado en la muestra de ensayo con un punto de corte predeterminado (o con varios puntos de corte predeterminados) y asignar una puntuación para cada antígeno (o una puntuación de una de varias puntuaciones posibles) cuantificado basándose en dicha comparación, combinar la puntuación asignada para cada antígeno cuantificado para obtener una puntuación total, comparar la puntuación total con una puntuación total predeterminada y utilizar dicha comparación como ayuda para determinar si un sujeto tiene cáncer de pulmón. De manera alternativa, en lugar de uno o más algoritmos o programas de computadora, se pueden proporcionar una o más instrucciones para que una persona lleve a cabo manualmente las etapas anteriores. El kit puede contener también uno o más marcadores detectables. Preferentemente, el kit contiene los reactivos necesarios para cuantificar los siguientes antígenos citoqueratina 19, citoqueratina 1H, CA 19-9, CEA, CA-15-3, CA125, SCC y ProGRP.

En otro aspecto un kit comprende (a) reactivos para cuantificar uno o más biomarcadores, donde dichos biomarcadores son regiones de interés que se seleccionan de entre el grupo que consiste en: ACN9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 e HIC3959; y opcionalmente, (b) uno o más algoritmos o programas de computadora para llevar a cabo las etapas de combinación y comparación de la cantidad de cada biomarcador cuantificado en la muestra de ensayo con un punto de corte predeterminado (o con varios puntos de corte predeterminados) y asignar una puntuación para cada biomarcador (o una puntuación de una de varias puntuaciones posibles) cuantificado basándose en dicha comparación, combinar la puntuación asignada para cada biomarcador cuantificado para obtener una puntuación total, comparar la puntuación total con una puntuación total predeterminada y utilizar dicha comparación como una ayuda para determinar si un sujeto tiene cáncer de pulmón. De manera alternativa, en lugar de uno o más algoritmos o programas de computadora se puede proporcionar una o más instrucciones para que una persona lleve a cabo manualmente las etapas anteriores. Preferentemente, las regiones de interés que se van a cuantificar en el kit se seleccionan de entre el grupo que consiste en: Pub11597, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 e HIC3959; Los reactivos incluidos en el kit para cuantificar una o más regiones de interés pueden incluir un adsorbente que se une y retiene al menos una región de interés contenida en un panel, soportes sólidos (tales como perlas) para usarse en conexión con dichos adsorbentes, uno o más marcadores detectables, etc.

En otro aspecto más, un kit puede comprender: (a) reactivos que contienen uno o más anticuerpos para cuantificar uno o más antígenos en una muestra de ensayo, donde dichos antígenos son citoqueratina 19, CEA, CA125 y proGRP; (b) reactivos que contienen uno o más antígenos para cuantificar al menos un anticuerpo en una muestra de ensayo; donde dichos anticuerpos son: anti-p53, anti-NY-ESO-1, anti-MAPKAPK3 y anti-Ciclina E2; y opcionalmente, (c) uno o más algoritmos o programas de computadora para llevar a cabo las etapas de combinación y comparación de la cantidad de cada antígeno, anticuerpo y región de interés cuantificados en la muestra de ensayo con un punto de corte predeterminado (o con varios puntos de corte predeterminados) y asignar una puntuación para cada antígeno, anticuerpo y región de interés (o una puntuación de una de varias puntuaciones posibles) cuantificados basándose en dicha comparación, combinar la puntuación asignada para cada antígeno, anticuerpo y región de interés cuantificados para obtener una puntuación total, comparar la puntuación total con una puntuación total predeterminada y utilizar dicha comparación como una ayuda para determinar si un sujeto tiene cáncer de pulmón. De manera alternativa, en lugar de uno o más algoritmos o programas de computadora, se pueden proporcionar una o más instrucciones para que una persona lleve a cabo manualmente las etapas anteriores. El kit puede contener un anticuerpo para cuantificar cada uno de citoqueratina 19, CEA, CA125 y proGRP, un antígeno para cuantificar cada uno de anti-p53, anti-NY-ESO-1, anti-MAPKAPK3 y anti-Ciclina E2 o un anticuerpo para cuantificar cada uno de citoqueratina 19, CEA, CA125 y proGRP y un antígeno para cuantificar cada uno de anti-p53, anti-NY-ESO-1, anti-MAPKAPK3 y anti-Ciclina E2.

IDENTIFICACIÓN DE BIOMARCADORES

Los biomarcadores de la divulgación se pueden aislar, purificar e identificar por técnicas bien conocidas por los expertos en la técnica. Estas incluyen técnicas cromatográficas, electroforéticas y de centrifugación. Estas técnicas se tratan en Current Protocols in Protein Science, J. Wiley and Sons, New York, NY, Coligan et al. (Eds) (2002) y Harris, E.L.V., S. Angal in Protein Purification Applications: A Practical Approach, Oxford University Press, New York, NY (1990) y otros sitios.

A modo de ejemplo, y sin limitación, se proporcionarán ahora ejemplos de la siguiente divulgación.

Ejemplos

Las muestras clínicas de suero sanguíneo de los pacientes se recolectaron (Ejemplo 1) y se analizaron por inmunoensayos de marcadores antigénicos (Ejemplo 2), por inmunoensayos de marcadores anticuerpos utilizando perlas (Ejemplo 3) o portaobjetos (Ejemplo 4), y por biomarcadores identificados por espectrometría de masas (Ejemplo 5). Los marcadores identificados se ordenaron y priorizaron utilizando una variedad de algoritmos (Ejemplo 6). Esos marcadores priorizados se combinaron utilizando un método de puntuación (Ejemplo 7) para identificar modelos predictivos (Ejemplo 8) para evaluar la utilidad clínica. Ejemplos del uso de los métodos que ayudan a detectar el cáncer de pulmón en pacientes sospechosos de tener un cáncer de pulmón se ilustra en el Ejemplo 9. Los biomarcadores identificados por espectrometría de masas como regiones de interés se analizaron para determinar su composición e identidad (Ejemplo 10). El Ejemplo 11 es un ejemplo profético que describe cómo se pueden detectar medir los biomarcadores identificados de acuerdo con la presente divulgación utilizando técnicas de inmunoensayo y técnicas inmunitarias espectrométricas de masas.

Ejemplo 1. Especímenes clínicos

Las muestras clínicas de suero del paciente se recolectaron bajo el protocolo aprobado por la Oficina de Revisión Institucional. Todos los sujetos que contribuyeron con un espécimen dieron su consentimiento informado para que se recolectaran los especímenes y se utilicen en este proyecto. Las muestras de suero se extrajeron en un tubo separador de suero y se permitieron coagular durante 15 minutos a temperatura ambiente. El coágulo se centrifugó y la muestra se vertió en alícuotas de 2 ml. A las 24 horas se congelaron las muestras a -80 °C y se mantuvieron a esa temperatura hasta que se sometieran a otros procesos. Las muestras se descongelaron en el momento inmediatamente antes del análisis. Al recibirlas, las muestras se descongelaron y se volvieron a hacer alícuotas en volúmenes más pequeños por conveniencia y se re congelaron. Las muestras se descongelaron entonces en el momento inmediatamente anterior a los análisis. Por lo tanto, cada muestra del grupo se congeló y descongeló dos veces antes del análisis.

Se recolectaron y analizaron un total de 751 especímenes. El grupo se compuso de 250 biopsias de pacientes confirmados de cáncer de pulmón, 274 biopsias de pacientes confirmados con enfermedad pulmonar benigna, y 227 de sujetos aparentemente normales. Los pacientes con cáncer y benignos se confirmaron en su diagnóstico por una biopsia definitiva. Los sujetos normales no se sometieron a ningún procedimiento diagnóstico definitivo y se valoraron "normales" por falta de una enfermedad maligna manifiesta. Tras este procedimiento diagnóstico definitivo, solo se seleccionaron los pacientes ≥ 50 años. Tras esta selección, quedaron 231 cánceres, 182 benignos, y 155 normales. Esta gran cohorte de cáncer, enfermedad pulmonar benigna, y sujetos aparentemente normales se denominará "cohorte grande". Un subgrupo de la gran cohorte se utilizó para enfocarla en la diferenciación entre enfermedad pulmonar benigna y cáncer de pulmón. Esta cohorte, denominada a partir de ahora "pequeña cohorte" consistía en 138 cánceres, 106 benignos, y 13 sujetos aparentemente normales. Tras retirar la "pequeña cohorte" en la "gran cohorte", permanecían 107 cánceres, 74 benignos, y 142 sujetos aparentemente normales. Esta cohorte, denominada a partir de ahora denominada "cohorte de validación" es independiente de la pequeña cohorte y se utilizó para validar los modelos predictivos generados. Las muestras clínicas que se prepararon como se ha descrito se utilizaron en los Ejemplos 2-10.

Ejemplo 2. Inmunoensayo de detección de biomarcadores

A. Ensayos Architect™ de Abbott Laboratories (Abbott Park, IL, a partir de aquí "Abbott")

Se adquirieron los kits Architect™ para los siguientes antígenos CEA, CA125, SCC, CA19-9 y CA15-3. Todos los ensayos se ejecutaron de acuerdo con las instrucciones del fabricante. Las concentraciones de los analitos en las muestras se proporcionaron por el instrumento Architect™. Estas concentraciones se utilizaron para generar los datos de la AUC que se muestran a continuación en la Tabla 1.

Tabla 1. Actuación clínica (AUC) de CA125, CEA, CA15-3, CA19-9, y SCC en las pequeña y gran cohortes. El nº obs se refiere al número total de individuos o muestras clínicas de cada grupo.

Marcador	gran cohorte		pequeña cohorte	
	nº obs	AUC	nº obs	AUC
Ca19-9	548	0,548	256	0,559
CEA	549	0,688	257	0,664
Ca15-3	549	0,604	257	0,569
Ca125	549	0,693	257	0,665
SCC	549	0,615	257	0,639

B. Ensayo Elecsys™ Roche

Se hicieron las mediciones Cyfra 21-1 (Citoqueratina 19, CK-19) con el sistema Elecsys™ 2010 (Roche Diagnostics GmbH, Mannheim, Alemania) de acuerdo con las instrucciones del fabricante. La concentración de Cyfra 21-1 se

proporcionó por el instrumento Elecsys™. Se generó una curva ROC con los datos y el AUC para las cohortes grande y pequeña se comunica en la Tabla 2.

Tabla 2. Actuación clínica (AUC) de la citoqueratina 19

	gran cohorte		pequeña cohorte	
Marcador	nº obs	AUC	nº obs	AUC
CK-19	537	0,68	248	0,718

5

C. Ensayos en placa de microtitulación

Se adquirieron los siguientes kits ELISA: ProGRP en Advanced Life Science Institute, Inc. (Japón), TPS (Citoqueratina 18, CK-18) en IDL Biotech AB (Bromma, Suecia) y Parainfluenza 1/2/3 IgG ELISA en IBL Immuno Biological Laboratories (Minneapolis, MN, EE. UU.). Se ejecutaron los análisis según las instrucciones del fabricante. Las concentraciones de los analitos se derivaron de los cálculos instruidos y proporcionados por el protocolo del fabricante. El AUC obtenida por los ensayos individuales se muestran a continuación en la Tabla 3.

15

Tabla 3. Actuación clínica (AUC) de Citoqueratina 18, proGRP, y parainfluenza 1/2/3

	Gran cohorte		Pequeña cohorte	
Marcador	nº obs	AUC	nº obs	AUC
CK-18	548	0,656	257	0,657
ProGRP	548	0,698	257	0,533
Parainfluenza 1/2/3	544	0,575	255	0,406

Ejemplo 3. Matriz de autoanticuerpos en perlas

A. Unión de proteínas humanas disponibles comercialmente (véase, la Tabla 4, posteriormente) a perlas Luminex™ SeroMap™ (Austin, Texas) y combinación de los grupos de perlas para preparar el reactivo. Se expusieron partes del reactivo a las muestras de suero humanas en condiciones que permitían que cualquier anticuerpo presente se uniera a las proteínas. El material que no se había unido se lavó y las perlas se expusieron entonces a un conjugado fluorescente de R-ficoeritrina unido a un anticuerpo que se une específicamente a la IgG humana. Tras el lavado, las perlas se pasaron a través de un instrumento Luminex™ 100, que identificó cada perla de acuerdo con su colorante interno, y midió la fluorescencia unida a la perla, que correspondía con la cantidad de anticuerpo unido a la perla. De esta manera, se evaluaron las respuestas de 772 muestras (251 de cáncer de pulmón, 244 normales, 277 benignos) contra 21 proteínas humanas, así como varias proteínas no humanas como controles (seroalbúmina bovina (BSA) y toxina tetánica).

Se obtuvieron los antígenos MUC-1 (Fujirebio Diagnostics INC, Malvern, PA), Citoqueratina 19 (Biodesign, Saco, ME), y CA-125 (Biodesign, Saco, ME) como fracciones de intercambio iónico de los cultivos celulares (Véase la Tabla 4, posteriormente). Estas preparaciones relativamente en crudo se sometieron a más fragmentaciones por peso molecular utilizando HPLC con una columna de exclusión por tamaño (BioRad SEC-250, Hercules, CA) con una fase móvil = PBS a 0,4 ml/minuto. Las fracciones se recolectaron comenzando a los 15 minutos con 1 minuto para cada fracción para un total de 23 fracciones para cada antígeno. Para el MUC-1, se inyectaron 250 ul; para la Citoqueratina 19 y CA-125, se inyectaron 150 ul. Las tres muestras mostraron señales que indicaban varias concentraciones de alto PM que se eluían de los 15-24 minutos, con señales demasiado altas para medirlas a tiempos mayores de 24 minutos, indicando altas concentraciones de materiales de bajo PM. Para el revestimiento de perlas se combinaron las siguientes fracciones: MUC-1-A fracciones 6,7; MUC-1-B fracciones 10,11; MUC-1-C fracciones 12,13; Citoqueratina 19-A fracciones 4, 5; Citoqueratina 19-B fracciones 8, 9; Citoqueratina 19-C fracciones 16, 17; CA125-A fracciones 5, 6; CA125-B fracciones 12, 13.

Tabla 4. Lista de proteínas

ID Perla	Antígeno	Fuente
1	MUC-1-A	Fujirebio Diagnostics INC
2	MUC-1-B	Fujirebio Diagnostics INC
3	MUC-1-C	Fujirebio Diagnostics INC
4	Citoqueratina 19-A	Biodesign, Saco, ME
5	Citoqueratina 19-B	Biodesign, Saco, ME
6	Citoqueratina 19-C	Biodesign, Saco, ME
7	CA125-A	Biodesign, Saco, ME

ID Perla	Antígeno	Fuente
8	CA125-B	Biodesign, Saco, ME
9	HSP27	US Biological, Swampscott, MA
10	HSP70	Alexis, San Diego, CA
11	HSP90	Alexis, San Diego, CA
12	Tétanos	Sigma, St. Louis, MO
13	HCG	Diosynth API, Des Plaines, IL
14	VEGF	Biodesign, Saco, ME
15	CEA	Biodesign, Saco, ME
16	NY-ESO-1	NeoMarkers, Fremont, CA
17	AFP	Cell Sciences, Canton, MA
18	ERB-B2	Invitrogen, Grand Island, NY
19	PSA	Fitzgerald, Concord, MA
20	P53	Lab Vision, Fremont, CA
21	JO-1	Biodesign, Saco, ME
22	Lactoferrina	Sigma, St. Louis, MO
23	HDJ1	Alexis, San Diego, CA
24	Queratina	Sigma, St. Louis, MO
25	RECAF62	BioCurex, Vancouver, BC Canadá
26	RECAF50	BioCurex, Vancouver, BC Canadá
27	Leche RECAF	BioCurex, Vancouver, BC Canadá
28	BSA	Sigma, St. Louis, MO

B. Revestimiento de perlas Luminex SeroMap™ con antígenos

- Se añadieron a los pocillos de una placa de ultrafiltración Omega 10K (Pall Corporation, Ann Arbor, Michigan) 50 ul de agua. Tras 10 minutos se colocó la placa al vacío. Cuando los pocillos estaban vacíos se añadieron 10 ul de agua para mantener la hidratación. Se añadieron a cada pocillo 50-100 ul de ácido morfolinoetanosulfónico (MES) 5 mM, pH 5,6, 50 ul de las perlas indicadas Luminex™ SeroMAP™ y el volumen apropiado correspondiente con 10-20 ug de cada antígeno indicado en la Tabla 4. Las perlas se suspendieron con la pipeta. A las perlas se añadieron 10 ul de EDAC (2,0 mg en 1,0 ml de MES 5 mM pH 5,6). La placa se cubrió y se colocó en un agitador en oscuridad. Tras 14 horas, la placa se succionó por vacío, se lavó con agua, y finalmente se resuspendieron las perlas en 50 ul de trietanolamina (TEA) 20 mM pH 5,6. La placa se agitó con el agitador en oscuridad. Se añadió a cada pocillo unos segundos 10 ul de EDAC (2,0 mg en 1,0 ml de MES 5 mM pH 5,6), y la placa se colocó en un agitador durante una hora en oscuridad. Tras el lavado, se añadieron a cada pocillo 200 ul de tampón PBS que contenía un 1 % de BSA y un 0,08 % de azida sódica (PBN), seguido por sonicación con sonda, y se colocó en oscuridad.

C. Ensayo de las muestras de suero con perlas revestidas

- Se prepararon las muestras en microplacas en una dilución de 1:20 en PBN, con 80 muestras por microplacas. A 50 ul del grupo de perlas descrito anteriormente se añadieron a 5 ul de suero de conejo (de un conejo inmunizado con un antígeno sin relación con los que se ensayan en el presente ensayo). El grupo de perlas se removió y se colocó a 37 °C. Tras 35 minutos, se añadió 1 ml de PBN que contenía un 5 % de suero de conejo y un 1 % de CHAPS (BRC). El grupo de perlas se removió, se centrifugó y se resuspendió en 1,05 ml de BRC. Se lavaron los pocillos de una placa filtro de 1,2 u Supor (Pall Corporation) con 100 ul de PBN. A cada pocillo se le añadieron 50 ul de BRC, 10 ul de la dilución 1:20 de muestra de suero y 10 ul de las perlas resuspendidas. La placa se agitó a temperatura ambiente en oscuridad durante 1 hora, se filtró y luego se lavó 3 veces durante 10 minutos con 100 ul de BRC. Se añadió el conjugado de detección 50 ul de (20 ul RPE anti-IgG humana en 5,0 ml de BRC) y se agitó la placa en oscuridad durante 30 minutos después de que se resuspendieran las perlas con una pipeta. Se añadieron entonces 100 ul de BRC, se agitaron las perlas con una pipeta y se analizaron las muestras en un instrumento Luminex™ 100.
- Los resultados (intensidad media de las perlas para cada muestra y antígeno) se evaluaron por análisis ROC con los siguientes resultados para las cohortes grande y pequeña que se muestra en la Tabla 5:

Tabla 5. Actuación clínica de la matriz de autoanticuerpos en perlas que contenían las proteínas de la Tabla 4 en las cohortes grande y pequeña

Biomarcador	Gran cohorte		Pequeña cohorte	
	nº obs	AUC	nº obs	AUC
MUC-1-A	579	0,53	253	0,56
MUC-1-B	579	0,55	253	0,59
MUC-1-C	579	0,57	253	0,61
Citoqueratina 19-A	579	0,57	253	0,58
Citoqueratina 19-B	579	0,53	253	0,49

Biomarcador	Gran cohorte		Pequeña cohorte	
	nº obs	AUC	nº obs	AUC
Citoqueratina 19-C	579	0,62	253	0,65
CA125-A	579	0,53	253	0,5
CA125-B	579	0,62	253	0,59
HSP27	579	0,56	253	0,56
HSP70	579	0,49	253	0,51
HSP90	579	0,54	253	0,53
Tétanos	579	0,57	253	0,56
HCG	579	0,54	253	0,5
VEGF	579	0,53	253	0,51
CEA	579	0,57	253	0,55
NY-ESO-1	579	0,58	253	0,58
AFP	579	0,51	253	0,55
ERB-B2	579	0,61	253	0,57
PSA	579	0,6	253	0,57
P53	579	0,6	253	0,54
JO-1	579	0,57	253	0,54
Lactoferrina	579	0,49	253	0,49
HDJ1	579	0,62	253	0,63
Queratina	579	0,58	253	0,55
RECAP62	579	0,54	253	0,53
RECAF50	579	0,53	253	0,53
Leche RECAF	579	0,54	253	0,62
BSA	579	0,57	253	0,59

Ejemplo 4. Matriz de autoanticuerpos en portaobjetos

A. Preparación de antígenos

Se prepararon aproximadamente 5000 proteínas derivadas de la Ultimate ORF Collection™ de Invitrogen (Invitrogen, Grand Island, NY) como fusiones recombinantes de la secuencia de la glutatión-S-transferasa (GST) con una proteína humana de longitud completa. El marcador GST permite la evaluación de la cantidad de proteína que se une a la matriz independientemente de otras características de la proteína.

B. Revestimiento de portaobjetos con antígeno

La ProtoArray consiste en una superficie de cristal (portaobjetos) revestido con nitrocelulosa punteada con las aproximadamente 5000 proteínas mencionadas anteriormente, así como numerosas características de control.

C. Ensayo de las muestras de suero con portaobjetos revestidos

La matriz se bloqueó primero con PBS/1 % BSA/0,1 % de Tween 20 durante 1 hora a 4 °C. Entonces se expuso a la muestra de suero diluida 1:120 en Tampón de Perfil (el "tampón de Perfil" que se trata en el presente documento contenía PBS, 5 mM de MgCl₂, 0,5 mM de ditiotretitol, 0,05 % Triton-X-100, 5 % de glicerol, 1 % BSA) durante 90 minutos a 4 °C. La matriz se lavó entonces tres veces con el Tampón de Perfil durante 8 minutos por lavado. La matriz se expuso entonces al conjugado AlexaFluor anti-IgG humana a 0,5 ul/ml en Tampón de Perfil durante 90 minutos a 44 °C. La matriz se lavó entonces tres veces con el Tampón de Perfil durante 9 minutos por lavado. Tras el secado en una centrifuga se exploró utilizando un escáner de micromatriz fluorescente Axon GenePix 4000B (Molecular Devices, Sunnyvale, CA).

D. Selección del biomarcador

Comparando la distribución de las señales positivas del suero de los pacientes de cáncer con las de los pacientes normales se determinaron las identidades de las proteínas que producían autoanticuerpos característicos de los pacientes de cáncer. Para aumentar la probabilidad de encontrar autoanticuerpos específicos de cáncer con un número limitado de matrices, se utilizaron los siguientes grupos de muestras: 10 grupos que contenían cada uno el suero de 4 o 5 pacientes de cáncer de pulmón, 10 grupos que contenían cada uno el suero de 4 o 5 pacientes con enfermedad de pulmón benigna. Estos grupos se enviaron a Invitrogen para su procesamiento como se ha descrito anteriormente. Las intensidades de la fluorescencia que corresponden con cada proteína para cada grupo se presentaron en una hoja de cálculo. Cada proteína se representó dos veces, que corresponde con puntos duplicados en la matriz.

El fabricante (Invitrogen) suministró en un algoritmo para evaluar la especificidad para el cáncer de la respuesta

inmunitaria de una proteína particular, un valor de punto de corte que distinguía mejor las intensidades de señal de las muestras de cáncer de las muestras no cancerosas. Se determinó el número de muestras de cada grupo con intensidades por encima de este punto de corte (recuento de cáncer y recuento no canceroso respectivamente) y se colocó en la hoja de cálculo como parámetros. Adicionalmente se calculó el valor de p, que representaba la probabilidad de que no había aumento de señal en un grupo comparado con el otro. Los datos se ordenaron entonces para poner en lo más alto las proteínas con menos positivos en el grupo no canceroso y más positivos en el grupo de cáncer, y se ordenaron además por el valor de p de menor a mayor. Ordenando por esta fórmula se proporcionó la siguiente información proporcionada a continuación en la Tabla 7.

Tabla 7. Lista de ID de Antígenos

Identificación del Antígeno	Recuento de cáncer	Recuento no canceroso	Valor de P
proteína vesícula acrosómica 1 (ACRV1)	6	0	0,0021
forkhead box A3 (FOXA3)	6	0	0,0072
factor general transcripción IIA	6	0	0,5539
WW dominio que contiene E3 ubiquitina proteína ligasa 2	5	0	0,0018
PDZ dominio que contiene 1 (PDZ1)	5	0	0,0018
ciclina E2	5	0	0,0018
ciclina E2	5	0	0,0018
dominio que contiene ácido fosfatídico fosfatasa tipo 2 3 (PPAPDC3)	5	0	0,0088
dominio que contiene motivo alfa repetido y estéril de anquirina 3	5	0	0,0563
dedos de zinc	5	0	0,0563
cisteinil-tARN sintetasa	4	0	0,0077
cisteinil-tARN sintetasa	4	0	0,0077
factor transcripción que se une al potenciador IGHM 3 (TFE3)	4	0	0,0077
WW dominio que contiene E3 ubiquitina proteína ligasa 2	4	0	0,0077
Fase de lectura abierta 7 del Cromosoma 21	4	0	0,0077
Fase de lectura abierta 7 del Cromosoma 21	4	0	0,0077
IQ motivo que contiene F1 (IQCF1)	4	0	0,0077
proteína 1 citosólica de linfocitos (L-plastina) (LCP1)	4	0	0,0077
proteína vesícula acrosómica 1 (ACRV1)	4	0	0,0077
homólogo DnaJ (Hsp40)	4	0	0,0077
homólogo DnaJ (Hsp40)	4	0	0,0077
factor 2 de unión al receptor nuclear	4	0	0,0077
factor 2 de unión al receptor nuclear	4	0	0,0077
dominio que contiene PDZ 1 (PDZK1)	4	0	0,0077
Sustrato de proteína quinasa C y caseína quinasa in neuronas 2	4	0	0,0077
LIM dominio quinasa 2	4	0	0,0077
polipéptido D de polimerasa (ARN) III (ADN dirigido)	4	0	0,0077
Proteína del motivo de unión ARN	4	0	0,0077
Asociado al ciclo de división celular 4 (CDCA4)	4	0	0,0312
factor de intercambio de nucleótido guanina Rho (GEF) 1	4	0	0,076
LUC7-tipo 2 (S. cerevisiae)	4	0	0,2302
similar a RIKEN ADNc 2310008M10 (LOC202459)	4	0	0,2302
ribulosa-5-fosfato-3-epimerasa	3	0	0,0296
ribulosa-5-fosfato-3-epimerasa	3	0	0,0296
proteína 1 de unión a heme (HEBP1)	3	0	0,0296
proteína 1 de unión a heme (HEBP1)	3	0	0,0296
subfamilia C receptor de células citolíticas tipo lectina	3	0	0,0296
subfamilia C receptor de células citolíticas tipo lectina	3	0	0,0296
LATS	3	0	0,0296
N-acilesfingosina amidohidrolasa (ceramidasa ácida) 1 (ASH1)	3	0	0,0296
N-acilesfingosina amidohidrolasa (ceramidasa ácida) 1 (ASH1)	3	0	0,0296
Paralemmina	3	0	0,0296
Paralemmina	3	0	0,0296
proteína 1 que interactúa con PIN2	3	0	0,0296
proteína quinasa S6 ribosómica	3	0	0,0296
proteína quinasa S6 ribosómica	3	0	0,0296
dominio que contiene SH3 y PX 3 (SH3PX3)	3	0	0,0296
dominio que contiene SH3 y PX 3 (SH3PX3)	3	0	0,0296
Pareja de fusión TCF3 (E2A) (en Leucemia infantil) (TFPT)	3	0	0,0296
Pareja de fusión TCF3 (E2A) (en Leucemia infantil) (TFPT)	3	0	0,0296

Identificación del Antígeno	Recuento de cáncer	Recuento no canceroso	Valor de P
Factor de transcripción que se une al potenciador IGHM 3 (TFE3)	3	0	0,0296
Fase de lectura abierta 117 del cromosoma 1	3	0	0,0296
Fase de lectura abierta 117 del cromosoma 1	3	0	0,0296
Proteína sobre-expresada asociada a la resistencia al cisplatino	3	0	0,0296
proteína de interacción con hsp70	3	0	0,0296
proteína hipotética FLJ22795	3	0	0,0296
proteína hipotética FLJ22795	3	0	0,0296
Proteína transmembrana inducida por interferón 1 (9-27)	3	0	0,0296
Proteína transmembrana inducida por interferón 1 (9-27)	3	0	0,0296
IQ motivo que contiene F1 (IQCF1)	3	0	0,0296
motivo que contiene repeticiones ricas en leucina y IQ 2 (LRRIQ2)	3	0	0,0296
motivo que contiene repeticiones ricas en leucina y IQ 2 (LRRIQ2)	3	0	0,0296
paralemina 2	3	0	0,0296
paralemina 2	3	0	0,0296
dominio que contiene RWD 1	3	0	0,0296
familia 7 vehículo de solutos	3	0	0,0296
familia 7 vehículo de solutos	3	0	0,0296
tropomiosina 1 (alfa)	3	0	0,0296
tropomiosina 1 (alfa)	3	0	0,0296
Candidato subtransferible supresor de tumores 4	3	0	0,0296
ubiquitina-tipo 4A	3	0	0,0296
vestigial tipo 4 (Drosophila) (VGLL4)	3	0	0,0296
dominio 16 repetido WD	3	0	0,0296
dominio 16 repetido WD	3	0	0,0296
proteína quinasa activada por Mitógeno - proteína quinasa activada 3	3	0	0,0296
proteína quinasa activada por Mitógeno - proteína quinasa activada 3	3	0	0,0296
proteína quinasa asociada a muerte 1 (DAK1)	3	0	0,0296
dimetilarginina dimetilaminohidrolasa 2 (DDAH2)	3	0	0,0296
dimetilarginina dimetilaminohidrolasa 2 (DDAH2)	3	0	0,0296
proteína 2 de choque térmico de 70 kDa	3	0	0,0296
familia H antígeno melanoma	3	0	0,0296
proteína quinasa activada por Mitógeno - proteína quinasa activada 3 (MAPKAPK3)	3	0	0,0296
nei tipo 2 (E. coli) (NEIL2)	3	0	0,0296
Sustrato de proteína quinasa C y caseína quinasa en neuronas 2	3	0	0,0296
SMAD	3	0	0,0296
SMAD	3	0	0,0296
proteína de unión a ARN asociada a gránulos citotóxicos TIA1	3	0	0,0296
factor trefoil 2 (proteína espasmolítica 1) (TFF2)	3	0	0,0296
uroporfinógeno III sintasa (porfiria eritropoyética congénita) (UROS)	3	0	0,0296
Proteína de 29 kDa inducida por citoquina (CIP29)	3	0	0,0296
proteína transmembrana 106C (TMEM106C)	3	0	0,0296
Fase de lectura abierta 11 del Cromosoma 9	3	0	0,0296
O-6-metilguanina-ADN metiltransferasa (MGMT)	3	0	0,0296
proteína 1 asociada a PDGFA (PDAP1)	3	0	0,0296
proteína 1 asociada a PDGFA (PDAP1)	3	0	0,0296
polipéptido D de polimerasa (ARN) III (dirigido a ADN)	3	0	0,0296
Rho-asociado	3	0	0,0296
Rho- asociado	3	0	0,0296
Motivo de proteína de unión a ARN	3	0	0,0296
tetraspanina 17	3	0	0,0296

Un segundo algoritmo calculaba la especificidad de cáncer de la respuesta inmunitaria para una proteína como la diferencia entre la señal media para las muestras de cáncer y la señal media para no cancerosas dividida por la desviación estándar de las intensidades de señal de las muestras no cancerosas. Esto tiene la ventaja de que las respuestas inmunitarias fuertes afectan al resultado más que las débiles. Los datos se ordenan entonces para poner en la parte superior las proteínas con los valores más altos. Los 100 más altos que se listaron identificados por este orden se muestra a continuación en la Tabla 8:

Tabla 8: Listado de ID de antígenos ordenados para poner en la parte superior las proteínas con la relación S/N más

altas. El S/N se calculó dividiendo la diferencia de la intensidad media de la señal de los dos grupos (media de cáncer-media de no canceroso) por la desviación estándar del grupo no canceroso (DE de no canceroso)

Identificación del Antígeno	Dif. de las medias / DE (no canceroso)
Pareja de fusión TCF3 (E2A) (en Leucemia infantil) (TFPT)	21,4
proteasa específica de ubiquitina 45 (USP45)	16,1
proteasa específica de ubiquitina 45 (USP45)	15,6
enzima de conjugación con ubiquitina E2O	15,1
proteasa específica de ubiquitina 45 (USP45)	13,9
enzima de conjugación con ubiquitina E2O	12,3
Superenrollamiento rico en prolina 1 (PRRC1)	11,5
Superenrollamiento rico en prolina 1 (PRRC1)	10
CLL de células B/linfoma 10	9,8
familia 7 vehículo de solutos	8,8
CLL de células B/linfoma 10	8,7
homólogo DnaJ (Hsp40)	8,2
homólogo DnaJ (Hsp40)	8
familia 7 vehículo de solutos	7,9
vestigial tipo 4 (Drosophila) (VGLL4)	6,5
dominio que contiene SH3 y PX 3 (SH3PX3)	6,3
ciclina E2	6,1
dominio que contiene SH3 y PX 3 (SH3PX3)	6,1
ciclina E2	6
Clon de ADNc IMAGE:3941306	5,9
Paralemmína	5,8
Proteína transmembrana inducida por interferón 1 (9-27)	5,6
Paralemmína	5,4
ribulosa-5-fosfato-3-epimerasa	5,4
motivo que contiene repeticiones ricas en leucina y IQ 2 (LRRIQ2)	5,3
ribulosa-5-fosfato-3-epimerasa	5,3
Asociado al ciclo de división celular 4 (CDCA4)	5,2
Proteína transmembrana inducida por interferón 1 (9-27)	4,8
motivo que contiene repeticiones ricas en leucina y IQ 2 (LRRIQ2)	4,7
proteína quinasa activada por Mitógeno - proteína quinasa activada 3	4,5
proteína quinasa dependiente del calcio/calmodulina I (CAMK1)	4,4
Proteína que interactúa con RAB3A tipo (rabin3) 1 (RAB3IL1)	4,3
dimetilarginina dimetilaminohidrolasa 2 (DDAH2)	4,2
proteína de interacción con hsp70	4,1
Fase de lectura abierta 11 del Cromosoma 9	4,1
proteína quinasa activada por Mitógeno - proteína quinasa activada 3	4,1
proteína vesícula acrosómica 1 (ACRV1)	4,1
triosafosfato isomerasa 1	4
triosafosfato isomerasa 1	3,8
uroporfirinógeno III sintasa (porfiria eritropoyética congénita) (UROS)	3,7
subfamilia C receptor de células citolíticas tipo lectina	3,7
receptor alfa relacionado con estrógeno (ESRRA)	3,6
proteína vesícula acrosómica 1 (ACRV1)	3,6
Asociado al ciclo de división celular 4 (CDCA4)	3,6
Proteína que interactúa con RAB3A tipo (rabin3) 1 (RAB3IL1)	3,5
proteína quinasa asociada a muerte 1 (DAPK1)	3,5
Sustrato de proteína quinasa C y caseína quinasa in neuronas 2	3,5
Tropomodulina 1	3,4
Tropomodulina 1	3,4
Fase de lectura abierta 117 del Cromosoma 1	3,4
dimetilarginina dimetilaminohidrolasa 2 (DDAH2)	3,4
receptor alfa relacionado con estrógeno (ESRRA)	3,2
Dominio que contiene homología a la pleckstrina	3,1

Identificación del Antígeno	Dif. de las medias / DE (no canceroso)
uroporfirinógeno III sintasa (porfiria eritropoyética congénita) (UROS)	3,1
proteína hipotética FLJ22795	3,1
oncogén FYN relacionado con SRC	3,1
proteína quinasa activada por Mitógeno - proteína quinasa activada 3 (MAPKAPK3)	3,1
Homólogo 37 tipo 1 (S. cerevisiae) del ciclo de división celular CDC37	3
Candidato subtransferible de supresión tumoral 4	3
dominio que contiene RWD 1	3
proteína hipotética FLJ22795	3
Homólogo 37 tipo 1 (S. cerevisiae) del ciclo de división celular CDC37	2,9
WW dominio que contiene E3 ubiquitina proteína ligasa 2	2,9
Dominio que contiene PDZ 1 (PDZK1)	2,9
proteína quinasa activada por Mitógeno - proteína quinasa activada 3 (MAPKAPK3)	2,9
factor de transcripción que se une a potenciador IGHM 3 (TFE3)	2,9
forkhead box A3 (FOXA3)	2,8
Fase de lectura abierta 117 del Cromosoma 1	2,8
dominio que contiene motivo alfa repetido y estéril de anquirina 3	2,8
Dominio que contiene OCIA 1 (OCIAD1)	2,8
polimerasa (DNA dirigida)	2,8
SMAD	2,8
KIAA0157 (KIAA0157)	2,8
CLL de células B/linfoma 7C (BCL7C)	2,8
proteína quinasa ribosómica S6	2,8
Fase de lectura abierta 11 del Cromosoma 9	2,7
proteína quinasa ribosómica S6	2,7
proteína inducida por citoquina de 29 kDa (CIP29)	2,7
Factor de unión al receptor nuclear 2	2,7
regulador del factor de célula huésped C1 1 (XPO1 dependiente) (HCFC1R1)	2,7
quinasa tipo STE20 (levadura) (SLK)	2,7
Dominio que contiene OCIA 1 (OCIAD1)	2,6
Sustrato de proteína quinasa C y caseína quinasa in neuronas 2	2,6
homólogo quaking	2,6
clasificación nexina 16 (SNX16)	2,6
proteína citosólica de linfocitos 1 (L-plastina) (LCP1)	2,6
Fase de lectura abierta 7 del Cromosoma 21	2,5
quinasa tipo STE20 (levadura) (SLK)	2,5
regulador del factor de célula huésped C1 1 (XPO1 dependiente) (HCFC1R1)	2,5
Proteína que interactúa con hsp70	2,5
homología quaking	2,5
factor transcripción que se une al potenciador IGHM 3 (TFE3)	2,5
SMAD	2,4
WW dominio que contiene E3 ubiquitina proteína ligasa 2	2,4
Fase de lectura abierta 7 del Cromosoma 21	2,4
Dominio que contiene PDZ 1 (PDZK1)	2,4
acetilserotonina tipo O-metiltransferasa	2,4
CLL de células B/linfoma 7C (BCL7C)	2,3
proteína ribosómica S19 (RPS19)	2,3
O-6-metilguanina-DNA metiltransferasa (MGMT)	2,3

Comparando los tipos de resultados de las Tablas 7 y 8 y examinando las señales generadas por las muestras de cáncer y no cancerosas para cada proteína, se seleccionaron las siguientes 25 proteínas que se muestran a continuación en la Tabla 9 para investigación posterior.

Tabla 9. Mejores 25 proteínas seleccionadas para investigación posterior

Clon	Identificación del Antígeno
BC007015.1	ciclina E2
NM_002614.2	Dominio que contiene PDZ 1 (PDZK1)
NM_001612.3	proteína vesícula acrosómica 1 (ACRV1)
NM_006145.1	homólogo DnaJ (Hsp40)
BC011707.1	factor 2 de unión al receptor nuclear
BC008567.1	Fase de lectura abierta 7 del Cromosoma 21
BC000108.1	WW dominio que contiene E3 ubiquitina proteína ligasa 2
BC001662.1	proteína quinasa activada por Mitógeno - proteína quinasa activada 3
BC008037.2	Sustrato de proteína quinasa C y caseína quinasa in neuronas 2
NM_005900.1	SMAD
NM_013974.1	dimetilarginina dimetilaminohidrolasa 2 (DDAH2)
NM_000375.1	uroporfirinógeno III sintasa (porfiria eritropoyética congénita) (UROS)
NM_145701.1	Asociado al ciclo de división celular 4 (CDCA4)
BC016848.1	Fase de lectura abierta 117 del Cromosoma 1
BC014307.1	Fase de lectura abierta 11 del Cromosoma 9
BC000897.1	Proteína transmembrana inducida por interferón 1 (9-27)
NM_024548.2	motivo que contiene repeticiones ricas en leucina y IQ 2 (LRRIQ2)
BC013778.1	familia 7 vehículo de solutos
BC032449.1	Paralemmína
NM_153271.1	dominio que contiene SH3 y PX 3 (SH3PX3)
NM_013342.1	Pareja de fusión TCF3 (E2A) (en Leucemia infantil) (TFPT)
NM_006521.3	factor transcripción que se une al potenciador IGHM 3 (TFE3)
BC016764.1	ribulosa-5-fosfato-3-epimerasa
BC014133.1	Homólogo 37 tipo 1 (S. cerevisiae) del ciclo de división celular CDC37
BC053545.1	tropomiosina 1 (alfa)

E. Ciclina E2

- 5 Se encontraron dos formas de Ciclina E2 presentes en la ProtoArray™. La forma identificada como el registro en Genbank BC007015.1 (SEQ ID NO: 1) que muestra una fuerte inmunorreactividad con varios de los grupos de muestras de cáncer y una reactividad mucho más baja con los grupos benigno y normal (no canceroso). Por el contrario, la forma identificada como Acceso de Genbank BC020729.1 (SEQ ID NO: 2) mostraba poca reactividad con cualquiera de las muestras de cáncer y no cancerosas. Como se muestra posteriormente, un alineamiento de
- 10 secuencia de las dos formas mostraban una identidad por encima de 259 aminoácidos, con diferencias tanto en las regiones del extremo N como del extremo C. El B020729.1 tiene 110 aminoácidos en el extremo N y 7 aminoácidos en el extremo C que no están presentes en BC007015.1. El BC007015.1. tiene 37 aminoácidos en el extremo C que no están presentes en B020729.1. Debido a que solo la forma BC007015.1. muestra inmunorreactividad, esto se atribuye a la parte de 37 aminoácidos del extremo C.
- 15

Se sintetizaron dos péptidos del extremo C de BC007015.1: E2-1 (SEQ ID NO: 3) que contienen los 37 aminoácidos del extremo C de BC007015.1. E2-2 (SEQ ID NO: 5) que contiene el extremo C de 18 aminoácidos de BC007015.1. Ambos péptidos se sintetizaron para incluir una cisteína en el extremo N para proporcionar un sitio reactivo para un enlace covalente para una proteína portadora o de superficie.

```

BC007015.1 1      M
BC020729.1 1      MSRRSSRLQAKQQPQPSQTESPQEAIIQAKKRKTQDVKKRREEVTKKHQYEIRNCWPP
*

BC007015.1
BC020729.1 61      VLSGGISPCIIETPHKEIGTSDFSRFTNYRFPKNLFINPSPLPDLSWGC

BC007015.1 2      SKEVWLNMLKKESRYVHDKHFEVLHSDLEPQMRISILLDWLLEVCEVYTLHRETFYLAQDF
BC020729.1 110     SKEVWLNMLKKESRYVHDKHFEVLHSDLEPQMRISILLDWLLEVCEVYTLHRETFYLAQDF
*****

BC007015.1 62      FDRFMLTQKDINKNMLQLIGITSIFIASKLEEIYAPKLQEFAYVTDGACSEEDILRMELI
BC020729.1 170     FDRFMLTQKDINKNMLQLIGITSIFIASKLEEIYAPKLQEFAYVTDGACSEEDILRMELI
*****

BC007015.1 122     ILKALKWELCPVTIISWLNLFLOVDALKDAPKVLPLQYSQETFIQIAQLLDLCILAIDSL
BC020729.1 230     ILKALKWELCPVTIISWLNLFLOVDALKDAPKVLPLQYSQETFIQIAQLLDLCILAIDSL
*****

BC007015.1 182     EFQYRILTAALCHFTSIEVVKASGLEWDSISECVDMVFPVNVVKSTSPVKLKTFFKKI
BC020729.1 290     EFQYRILTAALCHFTSIEVVKASGLEWDSISECVDMVFPVNVVKSTSPVKLKTFFKKI
*****

BC007015.1 242     PMEDRHNIQHTNYLAMEEVNYINTFRKGGQLSPVCNGGIMTPPKSTKPPGKH
BC020729.1 350     PMEDRHNIQHTNYLAMELVNYINTFRKGGQLSPVCNGGIMTPPKSTKPPGKH
*****

```

Alineamiento de secuencias de BC007015.1 (SEQ ID NO: 1) y BC020729.1 (SEQ ID NO: 2)

E2-1 : CEEVNYINTFRKGGQLSPVCNGGIMTPPKSTEKPPGKH (SEQ ID NO: 3)

5 E2-2 : CNGGIMTPPKSTEKPPGKH (SEQ ID NO: 5)

Péptidos derivados de BC007015.1

10 Los péptidos E2-1 y E2-2 se unieron cada uno a BSA activando el BSA con maleimida seguido por acoplamiento del péptido. El BSA activado se preparó siguiendo el siguiente protocolo: A 0,8 mg de BSA en 200 µl de PBS se le añadió 1 mg de GMBS (N-(gamma-maleimido-butil-oxi) succinimida, Pierce, Rockford IL) en 20 µl de DMF y 10 µl de trietanolamina 1 M a pH 8,4. Tras 60 minutos, la mezcla se pasó a través de una columna Sephadex G50 con fracciones de 400 µl de tampón de recolección PBS. Al BSA-Mal (100 µl) activado se le añadieron 0 2,5 mg de péptido E2-1 o 3,2 mg de péptido E2-2. En ambos casos, la mezcla se removió y se colocó en hielo durante 15 minutos, tras lo cual la mezcla se movió a temperatura ambiente durante 25 minutos. Los productos acoplados BSA-Mal-E2-1 1 (BM-E2-1) y BSA-Mal-E2-2 (BM-E2-2), se pasaron a través de una columna Sephadex G50 para su limpieza.

20 Las proteínas y péptidos se acoplaron a microesferas Luminex™ utilizando dos métodos. El primer método se describe en el Ejemplo 10C y se denomina "método directo". El segundo método se denomina "método pre-activo" y usa el siguiente protocolo: A los pocillos de una placa de ultrafiltración Omega 10 k se añadieron 100 µl de agua, tras 10 minutos colocada al vacío. Cuando los pocillos estaban vacíos, se añadieron 20 µl de MES (100 mM) pH 5,6 y 50 µl de cada grupo de perlas Luminex™ SeroMap™ como se muestra en la Tabla 10, a continuación. A los pocillos de la columna 1 filas A, B, C, y D y los pocillos de la columna 2 filas A, B, C, D, y E se añadieron 10 µl de NHS (20 mg/ml) en MES y 10 µl de EDAC (10 mg/ml) en MES. Tras 45 minutos de agitación en oscuridad, la placa se colocó al vacío para succión del tampón y los reactivos que no reaccionaron. Cuando los pocillos estaban vacíos se añadieron 100 µl de MES y se les permitió pasar a través de las membranas. Se retiró la placa del vacío y se añadieron 20 µl de MES y 50 µl de agua. A los pocillos indicados en la Tabla 10 se añadieron 4 µl de cada proteína o péptido (excepto de DNAJB1, que se añadieron 2 µl) y se agitaron con pipetas para dispersar las perlas. La placa se agitó durante 30 minutos en un agitador, luego se añadieron 5 µl de EDAC 10 mg/ml en MES a la columna 1, filas E F G H (por acoplamiento directo), y se agitó la placa en un agitador durante 30 minutos, luego se colocó al vacío para retirar el tampón y los reactivos que no reaccionaron. Cuando los pocillos estaban vacíos se añadieron 50 µl de PBS y se agitaron las mezclas y la placa se puso al vacío. Cuando los pocillos estaban vacíos se añadieron 50 µl de PBS y se agitaron las mezclas con pipetas para dispersar las perlas, y se incubaron 60 minutos en el agitador. Para 35 parar la reacción 200 µl de PBN se añadieron y se sonicaron las mezclas.

La Tabla 10 posterior resume las diferentes presentaciones de los péptidos E2 y proteínas de los diferentes grupos de perlas. Los péptidos E2-1 y E2-2, se acoplaron a BSA que se acopló a las perlas utilizando el método preactivado (Id de perlas 25 y 26) o el método directo (id de perlas 30 y 31). Los péptidos E2-1 y E2-2, también se acoplaron a las perlas sin BSA utilizando el método preactivado (id de perlas 28 y 29) o el método directo (id de perlas 33 y 34) 40 Las perlas 35, 37, 38, 39 y 40 se revistieron con proteína utilizando el método preactivado.

Tabla 10. Resumen de las diferentes presentaciones de péptidos ciclina E2 y proteínas en las diferentes perlas

Columna	Fila	ID de perlas	antígeno	Fuente	Método de acoplamiento
1	A	25	BM-E2-1	3,9 mg/ml	Preactivado
1	B	26	BM-E2-2	2,4 mg/ml	Preactivado
1	C	28	E2-1	21 mg/ml	Preactivado
1	D	29	E2-2	40 mg/ml	Preactivado
1	E	30	BM-E2-1	3,9 mg/ml	Directo
1	F	31	BM-E2-2	2,4 mg/ml	Directo
1	G	33	E2-1	21 mg/ml	Directo
1	H	34	E2-2	40 mg/ml	Directo
2	A	35	CCNE2	(GenWay, San Diego, CA) 0,6 mg/ml	Preactivado
2	B	37	MAPKAPK3	(GenWay, San Diego, CA) 05 mg/ml	Preactivado
2	C	38	p53	(Biomol, Plymouth Meeting, PA) 0-25 mg/ml	Preactivado
2	D	39	TMOD1	(GenWay, San Diego, CA) 0,8 mg/ml	Preactivado
2	E	40	DNAJB1	(Axxora, San Diego, CA) 1 mg/ml	Preactivado

45 Las perlas se ensayaron con sueros de pacientes de la siguiente manera: a 1 ml de PBN se añadieron 5 µl de cada preparación de perlas. La mezcla se sonicó y centrifugó, y las perlas aglomeradas se lavaron con 1 ml de BSA 1 % en PBS, y se resuspendieron en 1 ml del mismo tampón. Se añadieron a una placa filtro Supor de 1,2 u (Pall Corporation, East Hills, NY) 100 µl de PBN/Tween (1 % BSA en PBS que contiene un 20 % de Tween 20). Tras 10 minutos se filtró la placa, y se añadieron 50 µl de PBN 0,2 % Tween (1 % BSA en PBS que contiene un 0,2 % de

Tween 20). A cada pocillo se añadieron 20 ul de la mezcla de perlas y 20 ul de suero (1:50) como se muestra en la Tabla 11. El suero era o suero de pacientes humanos o de suero de conejo anti-GST. La placa se colocó en un agitador en oscuridad. Tras 1 hora, la placa se filtró y se lavó con 100 ul de PBN/Tween tres veces. Se añadieron 50 ul de RPE-anti-IgG humana (1:400) (Sigma, St. Louis, MO) para detectar anticuerpos humanos mientras que 50 ul de RPE-anti-IgG de conejo (1:200) se añadió para detectar los anticuerpos anti-GST de conejo. La placa se colocó en un agitador en oscuridad durante 30 minutos tras lo cual se filtraron las perlas, se lavaron y se procesaron en Luminex™.

Los resultados de seis muestras de suero y anti-GST de conejo se muestran en la Tabla 11 a continuación.

Tabla 11. Resultados del Luminex para las perlas revestidas con péptidos ciclina F2 y proteínas, expuestos a sueros de pacientes.

	ID de perla								
	25	26	28	29	35	30	31	33	34
	Preactivado					Directo			
ID del Suero	BM-E2-1	BM-E2-2	E2-1	E2-2	CCNE2	BM-E2-1	BM-E2-2	E2-1	E2-2
A2	18	12	7	4	17	16	13	9	5
A4	4	4	3	3	4	2	5	4	3
B2	9	16	5	4	12	8	10	9	5
B4	4380	172	1985	11	358	4833	132	2298	18
C4	227	44	66	9	50	243	40	87	7
D4	406	15	64	7	19	440	13	107	8
F4	3721	156	1592	8	299	4034	140	1997	19
conej-antiGST	13	14	40	21	1358	10	13	56	22

Es evidente en la Tabla 11 que las perlas 25 y 30, que contenían el péptido E2-1 unido a BSA y se acoplaban directamente (utilizando el método directo) o por medio de preactivación (o el método preactivado) de perlas, respectivamente, daban las señales más fuertes. El péptido E2-1 acoplado sin el portador BSA también daba señales fuertes, aunque solo aproximadamente la mitad de las que daba con el portador BSA. La proteína de longitud completa CCNE2 (que contiene el marcador de fusión GST en el extremo N) mostraba señales bastante por encima de cualquier forma de péptido E2-2, pero aún mucho menor que el del péptido E2-1, sugiriendo que contenía la parte inmunoreactiva de las secuencias, pero con una densidad más baja en las perlas. Su señal con el anti-GST de conejo muestra que esta proteína de fusión GST estaba acoplada satisfactoriamente a la microesfera.

Se revistieron las perlas Luminex SeroMap™ con las proteínas que se muestran en la Tabla 12, posteriormente, por método de preactivación y directo como se ha descrito anteriormente, y por revestimiento pasivo. Para el revestimiento pasivo, se añadieron 5 ug de la proteína, en solución tal cual se recibe del vendedor, a 200 ul de perlas SeroMap™, la mezcla se removió, y se incubó 5 horas a temperatura ambiente, a 18 horas a 4 °C, entonces se centrifugó para sedimentarlo, y el aglomerado se lavó y se resuspendió en PBN.

Tabla 12. Proteínas que revisten perlas Luminex SeroMap™ por métodos de preactivación y directo.

Revestimiento	Proteína	Perla	Fuente
Preactivado	TMP21-ECD	1	Abbott, North Chicago, IL
Preactivado	domino NPC1L1C	5	Abbott, North Chicago, IL
Preactivado	PSN2(1-86aa)	14	Abbott, North Chicago, IL
Preactivado	IgG humana	22	Abbott, North Chicago, IL
Preactivado	BM-E2-2	26	Abbott, North Chicago, IL
Directo	BM-E2-1	30	Abbott, North Chicago, IL
Preactivado	TMOD1	39	Genway, San Diego, CA
Preactivado	DNAJB1	40	Axxora, San Diego, CA
Preactivado	PSMA4	41	Abnova, Taipei City, Taiwan
Preactivado	RPE	42	Abnova, Taipei City, Taiwan
Preactivado	CCNE2	43	Abnova, Taipei City, Taiwan
Preactivado	PDZK1	46	Abnova, Taipei City, Taiwan
Directo	CCNE2	48	Genway, San Diego, CA
Preactivado	Paxilina	53	BioLegend, San Diego, CA

Revestimiento	Proteína	Perla	Fuente
Directo	AMPHIPHYSIN	54	LabVision, Fremont, CA
Preactivado	CAMK1	55	Upstate, Charlottesville, VA
Pasivo	DNAJB11	67	Abnova, Taipei City, Taiwan
Pasivo	RGS1	68	Abnova, Taipei City, Taiwan
Pasivo	PACSIN1	70	Abnova, Taipei City, Taiwan
Pasivo	SMAD1	71	Abnova, Taipei City, Taiwan
Pasivo	p53	72	Biomol, Plymouth Meeting, PA
Pasivo	RCV1	75	Genway, San Diego, CA
Pasivo	MAPKAPK3	79	Genway, San Diego, CA

Se ensayaron muestras de suero de 234 pacientes (87 cánceres, 70 benignos, y 77 normales). Los resultados de estos ensayos se analizaron por curvas ROC. El AUC calculado para cada antígeno se muestra en la Tabla 13 a continuación.

5

Tabla 13. AUC calculada para antígenos derivados de muestras de suero.

Proteína	AUC
péptido ciclina E2 1	0,81
proteína ciclina E2 (Genway)	0,74
péptido ciclina E2 2	0,71
TMP21-ECD	0,66
Dominio NPC1L1C	0,65
PACSIN1	0,65
p53	0,63
proteína quinasa activada por mitógeno proteína quinasa activada (MAPKAPK3)	0,62
Tropomodulina 1 (TMOD1)	0,61
PSEN2(1-86aa)	0,60
proteína de unión DNA J 1(DNAJB1)	0,60
proteína de unión DNA J 11(DNAJB11)	0,58
RCV1	0,58
(proteína quinasa dependiente de calcio/calmodulina - 1 CAMK1)	0,57
SMAD1	0,57
AMPHIPHYSIN Lab Vision	0,55
RGS1	0,55
PSMA4	0,51
ribulosa-5-fosfato-3-epimerasa (RPE)	0,51
Paxilina	0,51
proteína ciclina E2 (Abnova)	0,49
proteína que contiene el dominio PDZ 1(PDZK1)	0,47

Ejemplo 5. Espectrometría de masas

10 A. Preparación de muestras por Elución Secuencial de una perla magnética mixta (MMB)

Las muestras de suero se descongelaron y se mezclaron con un volumen igual de tampón Sol B de Invitrogen. Se removió la mezcla y se filtró a través de un filtro de 0,8 µm (Sartorius, Goettingen, Alemania) para clarificar y eliminar los residuos antes de más procesamientos. Se llevó a cabo la preparación automática de la muestra en una placa KingFisher® de 96 pocillos (Thermo Fisher, Scientific, Inc., Waltham, MA) utilizando una mezcla de perlas magnéticas Dynal® (Invitrogen) de intercambio aniónico fuerte y Abbott Laboratories (Abbott, Abbott Park, IL) de intercambio catiónico débil. Normalmente las perlas de intercambio aniónico tienen hidrocarburos basados en aminas- aminas cuaternarias o grupos dietil aminas – como grupos terminales funcionales y las perlas de intercambio catiónico débil normalmente tienen grupos funcionales basados en ácido sulfónico (ácido carboxílico).

Las perlas de intercambio catiónico Abbott (perlas CX) estaban en una concentración de 2,5 % (masa/volumen) y las perlas de intercambio aniónico fuerte Dynal® (perlas AX) estaban a 10 mg/ml de concentración. Justo antes de la preparación de las muestras, las perlas de intercambio catiónico se lavaron una vez con 20 mM Tris. HCl, pH 7,5, 0,1 % de Triton X100 reducido (tampón Tris-Triton). Otros reactivos, Tris. HCl 20 mM, pH 7,5 (tampón Tris), 0,5 % ácido trifluoroacético (a partir de aquí “solución TFA”) y un 50 % de acetonitrilo (a partir de aquí “solución de Acetonitrilo”), que se utilizan en esta preparación de la muestra se prepararon en el laboratorio. Los reactivos y muestras se dispusieron en las placas de 96 pocillos de la siguiente manera:

La fila A contenía una mezcla de 30 ul de perlas AX, 20 ul de perlas CX y 50 ul de tampón Tris.
La fila B contenía 100 ul de tampón Tris.

La fila C contenía 120 µl de tampón Tris y 30 µl de muestra.

La fila D contenía 100 µl de tampón Tris.

La fila E contenía 100 µl de Agua desionizada.

La fila F contenía 50 µl de solución TFA.

5 La fila G contenía 50 µl de solución Acetonitrilo.

La fila H estaba vacía.

10 Las perlas y el tampón en la fila A se premezclaron y las perlas se recolectaron con un recuento de recolección de 3 (parámetro del aparato que indica las veces que la sonda magnética se introduce en la solución para recolectar las perlas magnéticas) y se transfirieron a la fila B para lavarlas en tampón Tris – con ajuste de liberación “rápido”, ajuste de lavado – medio, y un tiempo de lavado de 20 segundos. Al final de la etapa del lavado de perlas, las perlas se recolectaron con recuento de recolección de 3 y se transfirieron a la fila C para unirse a la muestra. El ajuste de liberación de la perlas era rápido. La muestra unida se lleva a cabo con el ajuste en “lento”, con un tiempo de unión de 5 minutos. Al final de la etapa de unión, se recolectaron las perlas con un recuento de recolección de 3. Las perlas recolectadas se transfirieron a la fila E para la segunda etapa de lavado – ajuste de liberación “rápida”, ajuste de elución – medio, con un tiempo de lavado de 20 segundos. Al final de la segunda etapa de lavado, las perlas se recolectaron con un recuento de recolección de 3. Las perlas recolectadas se transfirieron a la fila F para la elución en solución TFA – con el ajuste de liberación “rápido”, ajuste de elución – rápido y tiempo de elución de 2 minutos. Al final de la etapa de elución TFA, se recolectaron las perlas con un recuento de Recolección de 3. Este eluyente TFA se recolectó y analizó por espectrometría de masas. Las perlas recolectadas se transfirieron a la fila G para la elución en solución de Acetonitrilo – con ajuste de liberación “rápido”, ajuste de elución – rápido y tiempo de elución 2 minutos. Tras la elución, las perlas se retiraron con un recuento de recolección de 3 y se pusieron en la Fila A. El eluyente de Acetonitrilo (AcN) se recolectó y analizó por espectrometría de masas.

25 Todas las muestras se procesaron por duplicado, pero no en la misma placa para evitar errores sistemáticos. Las muestras eluidas se aspiraron manualmente y se colocaron en placas de 96 pocillos para la preparación de muestra diana MALDI. Por tanto, cada muestra proporcionaba dos eluyentes para el análisis de espectrometría de masas.

30 Se utilizó un robot CLINPROT (Bruker Daltonics Inc., Billerica, MA) para preparar las dianas MALDI antes de la interrogación MS. En resumen, el proceso implica la carga de la placa de muestras que contiene las muestras de suero eluido y los viales que contenían la solución matriz MALDI (10 mg/ml ácido sinapínico en Acetonitrilo al 70 %) en las posiciones designadas del robot. Se cargó y se inició un archivo que contenía el procedimiento de aplicación puntual desde la computadora que controla el robot. En este caso, la aplicación puntual implicaba la aspiración de 5 µl de solución matriz y la dispersión en la placa matriz seguido por 5 µl de muestra. La premezcla de la muestra y la matriz se conseguía aspirando 5 µl de la mezcla y dispersándola varias veces en la placa matriz. Tras la premezcla, se aspiraron 5 µl de la mezcla y se depositaron 0,5 µl en cuatro puntos contiguos en el ancla del chip diana (Bruker Daltonics Inc., Billerica, MA). Los restantes 3 µl de solución se depositaron en el contenedor de desechos. El aspirado de más muestra de la necesaria minimizaba la formación de burbujas de aire en las puntas desechables que puede llevar a puntos perdidos durante el depósito de la muestra en el ancla del chip diana.

40

B. Preparación de muestras por Cromatografía Hidrofóbica de Perlas Magnéticas C8 (C8 MB-HIC)

45 Las muestras de suero se mezclaron con tampón SOLB y se clarificaron con filtros como se describe en el Ejemplo 5A. La preparación automática de las muestras se llevó a cabo en palcas KingFisher® de 96 pocillos utilizando los kits de purificación CLINPROT conocido como 100 MB-HIC 8 (Bruker Daltonics Inc., Billerica, MA). El kit incluye perlas magnéticas C8, solución de unión, y solución de lavado. Todos los demás reactivos se adquirieron en Sigma Chem. Co., si no se establece otra cosa. Los reactivos y muestras se dispusieron en la placa de 96 pocillos de la siguiente manera:

50 La fila A contenía una mezcla de 20 µl de perlas magnéticas C8 de Bruker y 80 µl de agua DI.

La fila B contenía una mezcla de 10 µl de muestra de suero y 40 µl de solución de unión.

Las filas C-E contenían 100 µl de solución de lavado.

La fila F contenía 50 µl de acetonitrilo al 70 % (añadido justo antes de la etapa de elución para minimizar la evaporación del disolvente orgánico).

55 La fila G contenía 100 µl de agua DI.

La fila H estaba vacía.

60 Se premezclaron las perlas en la fila A y se recolectaron con un “recuento de recolección” de 3 y se transfirieron a la fila B para unirse a la muestra. El ajuste de liberación se ajustó a “rápido” con un tiempo de liberación de 10 segundos. La unión a la muestra se llevó a cabo con el ajuste en “lento” durante 5 minutos. Al final de la etapa de unión, se recolectaron las perlas con un “recuento de recolección” de 3 y se transfirieron a la fila C para la primera etapa de lavado (ajuste de liberación = rápido con un tiempo = 10 segundos, ajuste de lavado = medio con un tiempo = 20 segundos). Al final de la primera etapa de lavado, las perlas se recolectaron con “recuento de recolección” de 3 y se transfirieron a la fila D para la segunda etapa de lavado con los mismos parámetros de la primera etapa de lavado. Al final de la segunda etapa de lavado, las perlas se recolectaron una vez más y se transfirieron a la fila E para la tercera y última etapa de lavado como se ha descrito anteriormente. Al final de la tercera etapa de lavado, se

pausó el KingFisher® durante la etapa de transferencia de la Fila E a la Fila F y se añadieron 50 µl de acetonitrilo al 70 % a la Fila F. Tras la adición de acetonitrilo, se reasumió el proceso. Las perlas recolectadas de la Fila se transfirieron a la Fila F para la etapa de elución (ajuste de liberación = rápido con un tiempo = 10 segundos, ajuste de elución = rápido con un tiempo = 2 minutos). Tras la etapa de elución, las perlas se retiraron y se dispusieron en la Fila G. Todas las muestras se procesaron por duplicado, como se ha descrito anteriormente en el Ejemplo 5A.

Se utilizó un robot CLINPROT (Bruker Daltonics Inc., Billerica, MA) para preparar las dianas MALDI antes de la interrogación por MS como se ha descrito en la sección anterior con solo mínimas modificaciones en la matriz MALDI que se utilizó. En este caso, en vez de SA se utilizó HCCA (1 mg/ml de HCCA en 40 % de ACN /50 % de MeOH /10 % de agua, v/v/v). Todos los demás parámetros permanecieron igual.

C. Preparación de muestras utilizando el chip SELDI

Se utilizaron los siguientes reactivos:

1. 100 mM de tampón fosfato, pH 7,0, preparado mezclando 250 ml de agua desionizada con 152,5 ml de solución de fosfato disódico 200 mM y 97,5 ml de solución de fosfato monosódico 200 mM.
2. 10 mg/ml de solución de ácido sinapínico, preparado disolviendo una cantidad pesada de ácido sinapínico en una suficiente cantidad de una solución preparada mezclando volúmenes iguales de acetonitrilo y un 0,4 % de ácido trifluoroacético acuoso (v/v) para dar una concentración final de 10 mg de ácido sinapínico por ml de solución.
3. El agua desionizada, el ácido sinapínico y ácido trifluoroacético eran de Fluka Chemicals. El acetonitrilo era de Burdick y Jackson.

Las matrices Q10 ProteinChip en la configuración de ocho puntos y los Bioprocesadores utilizados para mantener las matrices en una matriz de 12 x 8 con una huella idéntica a las microplacas de referencia se adquirieron en CIPHERGEN. La superficie activa Q10 es un intercambiador aniónico fuerte de amina cuaternaria. Se utilizó un Sistema CIPHERGEN ProteinChip, un espectrómetro de masas de desorción ionización por láser asistido por matriz con tiempo de vuelo (MALDI) Serie 4000 para analizar los péptidos unidos a la superficie del chip. Todos los productos CIPHERGEN se adquirieron en CIPHERGEN Biosystems, Dumbarton, California.

Todas las transferencias de líquidos, diluciones, y lavados se llevaron a cabo mediante un pipeteador robótico STAR Hamilton Microlab de la Hamilton Company, Reno, Nevada.

Las muestras de suero se descongelaron a temperatura ambiente y se mezclaron removiendo suavemente. Los viales que contenían la muestra se cargaron en las 24 posiciones de porta muestras en el pipeteador Hamilton; se cargaron cuatro porta-muestras con un total de 96 muestras. Se colocaron en la plataforma del pipeteador Hamilton dos bioprocesadores que tenían los chips Q10 (192 puntos en total). Se cargaron los contenedores con tampón fosfato 100 mM y agua desionizada en el pipeteador Hamilton. Las puntas de pipeta desechables se colocaron también en la plataforma del instrumento.

Todo el proceso de las muestras era totalmente automático. Cada muestra se diluyó 1 a 10 en dos alícuotas separadas mezclando 5 microlitros de suero con 45 microlitros de tampón fosfato en dos pocillos separados de una microplaca en la plataforma del pipeteador Hamilton. Los chips Q10 se activaron exponiendo cada punto a dos alícuotas de 150 microlitros de tampón fosfato. Se permitió que el tampón activara la superficie durante 5 minutos a continuación de cada adición. Después de que se aspirara la segunda alícuota de cada punto, se añadieron 25 microlitros de suero diluido en cada punto y se incubó durante 30 minutos a temperatura ambiente. Cada muestra se diluyó dos veces con una alícuota simple de cada dilución situada en un punto de un chip Q10. Después de la aspiración del suero diluido, cada punto se lavó cuatro veces con 150 microlitros de tampón fosfato y finalmente con 150 microlitros de agua desionizada. Los chips procesados se secaron con aire y se trataron con ácido sinapínico, la matriz utilizada capacitaba el proceso MALDI en el CIPHERGEN 4000. La solución matriz de ácido sinapínico se cargó en el pipeteador Hamilton colocando una microplaca de 96 pocillos, con cada pocillo lleno de solución de ácido sinapínico, en la plataforma del instrumento. Se utilizó un pipeteador de 96 cabezas para añadir 1 microlitro de matriz de ácido sinapínico a cada punto en un bioprocesador simultáneamente. Tras un periodo de secado de 15 minutos, se añadió una segunda alícuota de 1 microlitro a cada punto y se dejó secar.

D. Adquisición de datos AutoFlex MALDI-TOF de la preparación de muestras con perlas mezcladas

El intervalo de adquisición del instrumento se fijó en 400 a 100.000 m/z. El instrumento se calibró externamente en modo lineal utilizando las referencias de calibración de Bruker que cubrían un intervalo de masas de 2-17 kDa. Con el fin de recolectar espectros de alta calidad, las adquisiciones se automatizaron completamente con el control difuso encendido, excepto para el láser. El control difuso del láser se apagó de forma que la potencia del láser permanecía constante durante la duración del experimento. Como el instrumento está calibrado en general con una potencia de láser fija, la precisión se beneficia del mantenimiento de una potencia del láser constante. Los otros parámetros de control difusos controlaban la resolución y el S/N de los picos en el intervalo de masas de 2-10 kDa. Estos valores se optimizaron antes de cada adquisición y se eligieron para maximizar la calidad del espectro mientras se minimiza el

número de adquisiciones fallidas de la muestra a muestra o puto a punto. También se encendió el deflector para desviar los iones de baja masa molecular (< 400 m/z) para evitar la saturación del detector con iones de matriz y maximizar la señal que proviene de la muestra. Además, antes de cada adquisición, se dispararon 5 disparos de calentamiento (LP ca. 5-10 % por encima del umbral) para eliminar cualquier exceso de matriz cuando el rayo láser se rasteriza a través de la superficie de la muestra. Para cada espectro de masa, se co-añaden 600 disparos láser juntos, solo si cumplían los criterios de resolución y S/N fijados anteriormente. Todos los demás espectros de calidad inferior se ignoraron y descartaron y no se utilizaron correcciones de la línea base ni algoritmos de suavización durante la adquisición del espectro en bruto.

Los datos se archivaron, se transformaron en un eje común m/z para facilitar la comparación y se exportó en un formato portátil ASCII que podría analizarse por varios paquetes de software estadístico. La transformación en un eje m/z común se conseguía utilizando un algoritmo de interpolación desarrollado en el laboratorio.

E. Adquisición de los datos AutoFlex MALDI-TOF de C8 MB-HIC

El intervalo de adquisición del instrumento se fijó de 1000 a 20.000 m/z y se optimizó la sensibilidad y resolución. Todos los demás parámetros de adquisición y métodos de adquisición se fijaron como se han descrito en el Ejemplo 5D anterior, con la excepción de que se co-añadieron 400 disparos láser a cada espectro de masa.

F. Adquisición de datos con Ciphergen 4000 SELDI-TOF del chip Q-10.

Se cargaron los bioprocesadores en el espectrómetro de masas Ciphergen 4000 SELDI con tiempo de vuelo utilizando los parámetros optimizados para el intervalo de masas entre 0-50.000 Da. Los datos se digitalizaron y promediaron sobre las 530 adquisiciones por punto para obtener un solo espectro de corriente iónica vs. Masa/carga (m/z). Cada espectro se exportó a un servidor y se recuperó posteriormente como un archivo ASCII para el análisis post-adquisición.

G. Análisis de los datos de las regiones de interés de la espectrometría de masas

Los datos espectrométricos de masas consisten en los valores de masa/carga de 0- 50.000 y sus correspondientes valores de intensidad. Se construyeron los grupos de datos de cáncer y no cancerosos. El grupo de datos de cáncer consiste en el espectro de masa de todas las muestras de cáncer, mientras que los grupos de datos no cancerosos consisten en los espectros de masa de cada muestra no cancerosa, incluyendo los sujetos normales y los pacientes con enfermedad pulmonar benigna. Los grupos de datos de cáncer y no cancerosos se cargaron por separado en un programa de software que actúa de la siguiente manera:

- Se determina el t-test de Student de cada valor masa/carga registrado para dar un valor de p.
- Los espectros de cáncer y no canceroso se promedian a uno representativo de cada grupo.
- Se determina la relación logarítmica (Relación Log) de intensidad de los espectros medios de cáncer y espectros medios no cancerosos.

Se especificaron los ROI para tener diez o más valores de masa consecutivos con un valor de p de menos de 0,01 y una Relación Log absoluta mayor de 0,1. Se encontraron 18, 36, y 26 ROI en los grupos de datos MMB-TFA, MMB-AcN, y MB-HIC, respectivamente (Tablas 14a- 14c). Además, se encontraron 124 ROI (< 20 kDa) en los datos SELDI como se muestra en la Tabla 14d. Las tablas 14a a 14d enumeran los ROI de la presente invención, ordenados por el valor de masa medio de menor a mayor. El ROI proporcionado en la tabla es el valor de masa medio para el intervalo calculado (media del valor de masa de inicio y final para el intervalo dado). El promedio de masa ROI se denominará simplemente el ROI de aquí en adelante. Las intensidades de cada ROI para cada muestra se sometieron al análisis ROC. El AUC para cada marcador también se muestran en las tablas 14a- 14d posteriormente. En las Tablas 14a-14c a continuación, se muestran los ROI calculados que se obtuvieron del análisis de los perfiles MS de grupos enfermos y no enfermos. Las muestras individuales se procesaron utilizando tres métodos diferentes: cromatografía de intercambio aniónico/catiónico de perlas magnéticas mezcladas eluidas con a) TFA (tfa) y eluidas secuencialmente con b) acetonitrilo (acn), c) utilizando cromatografía de interacción hidrofóbica (hic). Cada método de preparación de muestras se analizó independientemente con el fin de obtener el ROI. Todos los espectros se recolectaron con el espectrómetro de masas Bruker AutoFlex MALDI-TOF. En la Tabla 14d posteriormente, el ROI calculado, obtenido a partir del análisis de los perfiles de MS de grupos enfermos y no enfermos. Todas las muestras se procesaron utilizando un chip Q-10. Todos los espectros se recolectaron utilizando un espectrofotómetro de masas Ciphergen 4000 SELDI-TOF.

Tabla 14a.

ROI	ROI	Promedio	ROI	Gran cohorte		Pequeña cohorte	
comienzo m/z	final m/z	ROI	Nombre	nº obs	AUC	nº obs	AUC
2322,911	2339,104	2331	tfa2331	538	0,66	236	0,52

ROI	ROI	Promedio	ROI	Gran cohorte		Pequeña cohorte	
comienzo m/z	final m/z	ROI	Nombre	nº obs	AUC	nº obs	AUC
2394,584	2401,701	2398	tfa2398	538	0,68	236	0,55
2756,748	2761,25	2759	tfa2759	538	0,65	236	0,60
2977,207	2990,847	2984	tfa2984	538	0,69	236	0,52
3010,649	3021,701	3016	tfa3016	538	0,63	236	0,48
3631,513	3639,602	3636	tfa3635	538	0,61	236	0,54
4188,583	4198,961	4194	tfa4193	538	0,60	236	0,56
4317,636	4324,986	4321	tfa4321	538	0,61	236	0,51
5000,703	5015,736	5008	tfa5008	538	0,70	236	0,57
5984,935	5990,126	5988	tfa5987	538	0,70	236	0,49
6446,144	6459,616	6453	tfa6453	538	0,74	236	0,65
6646,05	6658,513	6652	tfa6652	538	0,72	236	0,71
6787,156	6837,294	6812	tfa6815	538	0,71	236	0,53
8141,621	8155,751	8149	tfa8148	538	0,62	236	0,64
8533,613	8626,127	8580	tfa8579	538	0,71	236	0,58
8797,964	8953,501	8876	tfa8872	538	0,68	236	0,52
9129,621	9143,87	9137	tfa9133	538	0,63	236	0,60
12066,33	12093,36	12080	tfa12079	538	0,66	236	0,63

Tabla 14b.

ROI	ROI	Promedio	ROI	Gran cohorte		Pequeña cohorte	
comienzo m/z	final m/z	ROI	Nombre	nº obs	AUC	nº obs	AUC
3022,726	3026,825	3025	acn3024	519	0,63	244	0,51
3144,614	3182,554	3164	acn3163	519	0,70	244	0,60
3183,395	3188,023	3186	acn3186	519	0,63	244	0,54
4128,262	4135,209	4132	acn4132	519	0,61	244	0,59
4152,962	4161,372	4157	acn4157	519	0,65	244	0,65
4183,519	4194,373	4189	acn4189	519	0,52	244	0,55
4627,389	4635,759	4632	acn4631	519	0,74	244	0,68
5049,048	5114,402	5082	acn5082	519	0,68	244	0,62
5229,648	5296,428	5263	acn5262	519	0,68	244	0,61
5338,006	5374,554	5356	acn5355	519	0,64	244	0,52
5375,101	5383,848	5379	acn5378	519	0,67	244	0,62
5446,925	5457,382	5452	acn5455	519	0,68	244	0,54
5971,68	5981,476	5977	acn5976	519	0,64	244	0,58
6150,986	6166,194	6159	acn6158	519	0,63	244	0,54
6314,273	6338,877	6327	acn6326	519	0,62	244	0,58
6391,206	6406,112	6399	acn6399	519	0,67	244	0,60
6455,723	6461,713	6459	acn6458	519	0,56	244	0,65
6574,845	6607,218	6591	acn6592	519	0,68	244	0,58
6672,509	6689,568	6681	acn6681	519	0,53	244	0,70
8759,205	8791,323	8775	acn8775	519	0,64	244	0,58
8850,827	8888,382	8870	acn8871	519	0,69	244	0,55
9067,056	9095,468	9081	acn9080	519	0,65	244	0,57
9224,586	9277,996	9251	acn9251	519	0,64	244	0,59
9358,22	9384,195	9371	acn9371	519	0,65	244	0,55
9453,639	9467,414	9461	acn9459	519	0,66	244	0,76
9470,315	9473,579	9472	acn9471	519	0,70	244	0,71
9651,055	9674,867	9663	acn9662	519	0,66	244	0,52
10008,34	10022,51	10015	acn10015	519	0,63	244	0,56
10217,84	10221,98	10220	acn10216	519	0,64	244	0,55
10669,51	10689,53	10680	acn10679	519	0,61	244	0,52
10866,73	10886,56	10877	acn10877	519	0,63	244	0,50
11371,68	11745,49	11559	acn11559	519	0,63	244	0,68
14293,87	14346,94	14320	acn14319	519	0,62	244	0,58
22764,38	22771,69	22768	acn22768	519	0,68	244	0,62
22778,44	22788	22783	acn22783	519	0,68	244	0,63

ROI	ROI	Promedio	ROI	Gran cohorte		Pequeña cohorte	
comienzo m/z	final m/z	ROI	Nombre	nº obs	AUC	nº obs	AUC
22791,38	23147,21	22969	acn22969	519	0,70	244	0,63

Tabla 14c.

ROI	ROI	Promedio	ROI	Gran cohorte		Pequeña cohorte	
comienzo m/z	final m/z	ROI	Nombre	nº obs	AUC	nº obs	AUC
2016,283	2033,22	2025	hic2025	529	0,65	245	0,53
2304,447	2308,026	2306	hic2306	529	0,64	245	0,66
2444,629	2457,914	2451	hic2451	529	0,60	245	0,50
2504,042	2507,867	2506	hic2506	529	0,65	245	0,53
2642,509	2650,082	2646	hic2646	529	0,54	245	0,45
2722,417	2733,317	2728	hic2728	529	0,61	245	0,56
2971,414	2989,522	2980	hic2980	529	0,64	245	0,53
3031,235	3037,804	3035	hic3035	529	0,54	245	0,45
3161,146	3191,075	3176	hic3176	529	0,70	245	0,61
3270,723	3280,641	3276	hic3276	529	0,64	245	0,57
3789,504	3797,883	3794	hic3794	529	0,64	245	0,57
3942,315	3975,73	3959	hic3959	529	0,74	245	0,59
4999,913	5006,107	5003	hic5003	529	0,66	245	0,56
5367,59	5384,395	5376	hic5376	529	0,68	245	0,48
6002,824	6006,289	6005	hic6005	529	0,69	245	0,51
6181,86	6195,934	6189	hic6189	529	0,72	245	0,51
6380,634	6382,272	6381	hic6381	529	0,70	245	0,55
6382,569	6392,1	6387	hic6387	529	0,71	245	0,54
6438,218	6461,563	6450	hic6450	529	0,66	245	0,57
6640,279	6658,057	6649	hic6649	529	0,62	245	0,59
6815,125	6816,816	6816	hic6816	529	0,72	245	0,56
6821,279	6823,896	6823	hic6823	529	0,71	245	0,58
8788,878	8793,595	8791	hic8791	529	0,58	245	0,47
8892,247	8901,211	8897	hic8897	529	0,61	245	0,52
8908,948	8921,088	8915	hic8915	529	0,64	245	0,55
9298,469	9318,065	9308	hic9308	529	0,68	245	0,59

Tabla 14d.

ROI	ROI	Promedio	ROI	Gran cohorte		Pequeña cohorte	
comienzo m/z	final m/z	ROI	Nombre	nº obs	AUC	nº obs	AUC
2327	2336	2331	Pub2331	513	0,65	250	0,62
2368	2371	2369	Pub2369	513	0,64	250	0,60
2384	2389	2387	Pub2386	513	0,67	250	0,62
2410	2415	2413	Pub2412	513	0,67	250	0,63
2431	2435	2433	Pub2433	513	0,72	250	0,72
2453	2464	2459	Pub2458	513	0,70	250	0,62
2672	2682	2677	Pub2676	513	0,73	250	0,68
2947	2955	2951	Pub2951	513	0,72	250	0,64
2973	2979	2976	Pub2976	513	0,63	250	0,58
3016	3020	3018	Pub3018	513	0,50	250	0,51
3168	3209	3189	Pub3188	513	0,69	250	0,59
3347	3355	3351	Pub3351	513	0,70	250	0,67
3409	3414	3412	Pub3411	513	0,60	250	0,57
3441	3456	3449	Pub3448	513	0,72	250	0,58
3484	3503	3494	Pub3493	513	0,72	250	0,67
3525	3531	3528	Pub3527	513	0,62	250	0,55
3548	3552	3550	Pub3550	513	0,62	250	0,62
3632	3650	3641	Pub3640	513	0,63	250	0,57
3656	3662	3659	Pub3658	513	0,51	250	0,49

ROI	ROI	Promedio	ROI	Gran cohorte		Pequeña cohorte	
comienzo m/z	final m/z	ROI	Nombre	nº obs	AUC	nº obs	AUC
3678	3688	3683	Pub3682	513	0,72	250	0,69
3702	3709	3706	Pub3705	513	0,57	250	0,55
3737	3750	3744	Pub3743	513	0,69	250	0,67
3833	3845	3839	Pub3839	513	0,62	250	0,59
3934	3955	3944	Pub3944	513	0,65	250	0,57
4210	4217	4214	Pub4213	513	0,62	250	0,56
4299	4353	4326	Pub4326	513	0,69	250	0,59
4442	4448	4445	Pub4444	513	0,61	250	0,52
4458	4518	4488	Pub4487	513	0,75	250	0,69
4535	4579	4557	Pub4557	513	0,73	250	0,68
4590	4595	4592	Pub4592	513	0,70	250	0,66
4611	4647	4629	Pub4628	513	0,77	250	0,66
4677	4687	4682	Pub4682	513	0,72	250	0,69
4698	4730	4714	Pub4713	513	0,73	250	0,70
4742	4759	4751	Pub4750	513	0,76	250	0,73
4779	4801	4790	Pub4789	513	0,70	250	0,72
4857	4865	4861	Pub4861	513	0,72	250	0,75
4987	4996	4992	Pub4991	513	0,67	250	0,57
5016	5056	5036	Pub5036	513	0,65	250	0,54
5084	5194	5139	Pub5139	513	0,61	250	0,51
5208	5220	5214	Pub5213	513	0,57	250	0,52
5246	5283	5265	Pub5264	513	0,59	250	0,56
5295	5420	5357	Pub5357	513	0,64	250	0,54
5430	5537	5484	Pub5483	513	0,62	250	0,54
5570	5576	5573	Pub5573	513	0,59	250	0,57
5590	5595	5593	Pub5592	513	0,60	250	0,54
5612	5619	5615	Pub5615	513	0,55	250	0,53
5639	5648	5644	Pub5643	513	0,68	250	0,63
5679	5690	5685	Pub5684	513	0,66	250	0,59
5752	5804	5778	Pub5777	513	0,71	250	0,63
5839	5886	5862	Pub5862	513	0,73	250	0,67
5888	5909	5898	Pub5898	513	0,63	250	0,56
6008	6018	6013	Pub6013	513	0,61	250	0,57
6047	6058	6053	Pub6052	513	0,64	250	0,63
6087	6103	6095	Pub6094	513	0,59	250	0,54
6111	6124	6118	Pub6117	513	0,70	250	0,67
6153	6160	6156	Pub6156	513	0,57	250	0,51
6179	6188	6183	Pub6183	513	0,65	250	0,60
6192	6198	6195	Pub6194	513	0,57	250	0,49
6226	6272	6249	Pub6249	513	0,66	250	0,63
6277	6286	6281	Pub6281	513	0,62	250	0,65
6297	6307	6302	Pub6302	513	0,71	250	0,67
6352	6432	6392	Pub6391	513	0,65	250	0,56
6497	6570	6534	Pub6533	513	0,63	250	0,59
6572	6603	6587	Pub6587	513	0,60	250	0,55
6698	6707	6702	Pub6702	513	0,57	250	0,52
6715	6723	6719	Pub6718	513	0,64	250	0,57
6748	6849	6799	Pub6798	513	0,77	250	0,69
7197	7240	7219	Pub7218	513	0,73	250	0,65
7250	7262	7256	Pub7255	513	0,72	250	0,65
7310	7326	7318	Pub7317	513	0,71	250	0,65
7401	7427	7414	Pub7413	513	0,73	250	0,69
7435	7564	7499	Pub7499	513	0,76	250	0,73
7611	7616	7614	Pub7613	513	0,67	250	0,60
7634	7668	7651	Pub7651	513	0,70	250	0,63
7699	7723	7711	Pub7711	513	0,72	250	0,66
7736	7748	7742	Pub7742	513	0,69	250	0,65

ROI	ROI	Promedio	ROI	Gran cohorte		Pequeña cohorte	
comienzo m/z	final m/z	ROI	Nombre	nº obs	AUC	nº obs	AUC
7768	7782	7775	Pub7775	513	0,63	250	0,57
7935	7954	7945	Pub7944	513	0,64	250	0,61
7976	7985	7981	Pub7980	513	0,62	250	0,59
7999	8006	8003	Pub8002	513	0,58	250	0,60
8134	8239	8186	Pub8186	513	0,73	250	0,62
8286	8308	8297	Pub8297	513	0,69	250	0,62
8448	8461	8455	Pub8454	513	0,61	250	0,59
8476	8516	8496	Pub8496	513	0,69	250	0,64
8526	8567	8547	Pub8546	513	0,73	250	0,66
8579	8634	8606	Pub8606	513	0,80	250	0,70
8640	8684	8662	Pub8662	513	0,80	250	0,71
8710	8758	8734	Pub8734	513	0,74	250	0,67
8771	8781	8776	Pub8776	513	0,56	250	0,59
8913	8947	8930	Pub8930	513	0,68	250	0,64
8961	8977	8969	Pub8969	513	0,65	250	0,57
9122	9162	9142	Pub9142	513	0,66	250	0,66
9199	9233	9216	Pub9216	513	0,59	250	0,62
9311	9323	9317	Pub9317	513	0,57	250	0,60
9357	9370	9364	Pub9363	513	0,58	250	0,63
9409	9458	9434	Pub9433	513	0,67	250	0,65
9478	9512	9495	Pub9495	513	0,61	250	0,63
9629	9667	9648	Pub9648	513	0,62	250	0,64
9696	9749	9722	Pub9722	513	0,70	250	0,67
9977	10281	10129	pub10128	513	0,66	236	0,48
10291	10346	10318	pub10318	513	0,66	236	0,56
10692	10826	10759	pub10759	513	0,62	236	0,51
10867	11265	11066	pub11066	513	0,61	236	0,55
11339	11856	11597	pub11597	513	0,75	236	0,77
12080	12121	12100	pub12100	513	0,63	236	0,54
12159	12228	12194	pub12193	513	0,59	236	0,49
12422	12582	12502	pub12501	513	0,66	236	0,64
12620	12814	12717	pub12717	513	0,73	236	0,60
12839	12854	12846	pub12846	513	0,72	236	0,56
13135	13230	13182	pub13182	513	0,69	250	0,53
13386	13438	13412	pub13412	513	0,54	250	0,56
13539	13604	13572	pub13571	513	0,71	250	0,64
14402	14459	14430	pub14430	513	0,74	250	0,67
15247	15321	15284	pub15284	513	0,69	250	0,60
15414	15785	15600	pub15599	513	0,76	250	0,71
15872	15919	15896	pub15895	513	0,58	250	0,57
16366	16487	16427	pub16426	513	0,66	250	0,60
16682	16862	16772	pub16771	513	0,69	250	0,61
16984	17260	17122	pub17121	513	0,68	250	0,60
17288	17389	17339	pub17338	513	0,81	250	0,72
17431	18285	17858	pub17858	513	0,81	250	0,68
18321	18523	18422	pub18422	513	0,73	250	0,59
18728	18804	18766	pub18766	513	0,65	250	0,52
18921	19052	18987	pub18986	513	0,69	250	0,55

H. Identificación de familias de ROI: Se utilizó la función de análisis multivariado del programa del paquete estadístico JMP™ (SAS Institute Inc., Cary, NC) para identificar los ROI que estaban altamente correlacionados. Se extrajo del programa JMP un coeficiente de la matriz de correlación de dos dimensiones y luego se analizaron por Microsoft Excel. Para cada ROI, se identificó un grupo de ROI cuyo coeficiente de correlación superaba 0,8. Estos ROI juntos se convertían en una familia de ROI correlacionados. La Tabla 15 muestra las familias correlacionadas, sus correspondientes ROI miembros, el valor del AUC para el miembro ROI en la gran cohorte, y el promedio de coeficientes de correlación con otros miembros de la familia. Por lo tanto, se puede ver que los ROI que tienen masas de 3449 y 3494 están altamente correlacionados y se pueden sustituir unos con otros en el contexto de la presente invención.

Grupo A (n=2)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub3448	3449	0,72	0,81
Pub3493	3494	0,72	0,81
Grupo B (n=2)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub4487	4488	0,75	0,8
Pub4682	4682	0,72	0,8
Grupo C (n=9)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub8776	8776	0,56	0,8
Pub8930	8930	0,68	0,83
Pub9142	9142	0,66	0,92
Pub9216	9216	0,59	0,91
Pub9363	9363	0,58	0,88
Pub9433	9434	0,67	0,94
Pub9495	9495	0,61	0,94
Pub9648	9648	0,62	0,93
Pub9722	9722	0,7	0,89
Grupo D (n=15)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub5036	5036	0,65	0,71
Pub5139	5139	0,61	0,81
Pub5264	5265	0,59	0,79
Pub5357	5357	0,64	0,85
Pub5483	5484	0,62	0,87
Pub5573	5573	0,59	0,8
Pub5593	5593	0,6	0,78
Pub5615	5615	0,55	0,77
Pub6702	6702	0,57	0,79
Pub6718	6718	0,64	0,73
Pub10759	10759	0,62	0,77
Pub1066	11066	0,61	0,84
Pub12193	12194	0,59	0,79
Pub13412	13412	0,54	0,78
acn10679	acn10679	0,61	0,73
acn10877	acn10877	0,62	0,77
Grupo E (n=6)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub6391	6392	0,65	0,9
Pub6533	6534	0,63	0,9
Pub6587	6587	0,6	0,87
Pub6798	6799	0,76	0,85
Pub9317	9317	0,57	0,7
Pub13571	13571	0,71	0,67
Grupo F (n=8)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub7218	7219	0,73	0,82
Pub7255	7255	0,72	0,73
Pub7317	7318	0,71	0,88
Pub7413	7414	0,73	0,81
Pub7499	7499	0,76	0,84
Pub7711	7711	0,72	0,76
Pub14430	14430	0,74	0,77
Pub15599	15600	0,76	0,82
Grupo G (n=7)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub8496	8496	0,69	0,78
Pub8546	8547	0,73	0,88
Pub8606	8606	0,8	0,84
Pub8662	8662	0,79	0,77
Pub8734	8734	0,74	0,45
Pub17121	17122	0,68	0,78
Pub17338	17339	0,81	0,54

Grupo H (n=3)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub6249	6249	0,66	0,82
Pub12501	12502	0,73	0,87
Pub12717	12717	0,73	0,87
Grupo I (n=5)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub5662	5662	0,73	0,93
Pub5777	5777	0,71	0,92
Pub5898	5898	0,63	0,89
Pub11597	11597	0,75	0,93
acn11559	acn11559	0,63	0,84
Grupo J (n=5)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub7775	7775	0,63	0,39
Pub7944	7944	0,64	0,83
Pub7980	7980	0,62	0,72
Pub8002	8002	0,58	0,77
Pub15895	15895	0,58	0,75
Grupo K (n=4)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub17858	17858	0,81	0,84
Pub18422	18422	0,13	0,92
Pub18766	18766	0,69	0,89
Pub18986	18986	0,65	0,91
Grupo L (n=12)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub3018	3018	0,5	0,78
Pub3640	3640	0,62	0,82
Pub3658	3658	0,51	0,81
Pub3682	3682	0,72	0,77
Pub3705	3705	0,57	0,79
Pub3839	3839	0,62	0,75
hic2451	hic2451	0,6	0,78
hic2646	hic2646	0,54	0,7
hic3035	hic3035	0,54	0,72
tfa3016	tfa3016	0,63	0,78
tfa3635	tfa3635	0,61	0,78
tfa4321	tfa4321	0,61	0,74
Grupo M (n=2)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub2331	2331	0,65	0,9
tfa2331	tfa2331	0,66	0,9
Grupo N (n=2)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub4557	4557	0,73	0,81
Pub4592	4592	0,71	0,81
Grupo O (n=6)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
acn4631	acn4631	0,74	0,81
acn5082	acn5082	0,68	0,85
acn5262	acn5262	0,68	0,9
acn5355	acn5355	0,64	0,87
acn5449	acn5449	0,7	0,88
acn5455	acn5455	0,68	0,88
Grupo P (n=6)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
acn6399	acn6399	0,67	0,78
acn6592	acn6592	0,68	0,8
acn8871	acn8871	0,69	0,79
acn9080	acn9080	0,65	0,84
acn9371	acn9371	0,65	0,83
acn9662	acn9662	0,66	0,79
Grupo Q (n=2)			

Nombre de ROI	Miembros	AUC	Coef. Corr.
acn9459	acn9459	0,66	0,91
acn9471	acn9471	0,7	0,91
Grupo R (n=4)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
hic2506	hic2506	0,65	0,82
hic2980	hic2980	0,64	0,87
hic3176	hic3176	0,69	0,8
tfa2984	tfa2984	0,69	0,78
Grupo S (n=2)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
hic2728	hic2728	0,61	0,81
hic3276	hic3276	0,64	0,81
Grupo T (n=6)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
hic6381	hic6381	0,7	0,83
hic6387	hic6387	0,71	0,84
hic6450	hic6450	0,66	0,81
hic6649	hic6649	0,62	0,73
hic6816	hic6816	0,72	0,81
hic6823	hic6823	0,71	0,79
Grupo U (n=2)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
hic8791	hic8791	0,58	0,8
hic8897	hic8897	0,61	0,8
Grupo V (n=2)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
tfa6453	tfa6453	0,74	0,84
tfa6652	tfa6652	0,72	0,84
Grupo W (n=2)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
hic6005	hic6005	0,69	0,74
hic5376	hic5376	0,68	0,74
Grupo X (n=3)			
Nombre de ROI	Miembros	AUC	Coef. Corr.
Pub4713	4714	0,73	0,83
Pub4750	4751	0,76	0,66
Pub4861	4861	0,72	0,65

Tabla 15. Familias de Regiones de interés correlacionadas

Ejemplo 6. Análisis multivariado de biomarcadores utilizando análisis discriminatorio, análisis de árboles de decisión y análisis de componente principal

- 5 Se llevaron a cabo los análisis multivariados en los biomarcadores del inmunoensayo y las Regiones de Interés. Todos los diferentes análisis se llevaron a cabo utilizando el paquete estadístico JMP. Para simplificar, los análisis discriminatorios (DA), el análisis de componente principal (PCA) y árboles de decisión (DT) se denominan en general en el presente documento métodos multivariados (MVM). Es digno de mención que en el PCA solo se extrajeron los primeros 15 componentes principales, que supone más del 90 % de la variabilidad total en los datos. Los factores de
- 10 carga y/o comunitarios se utilizaron para extraer solamente el único factor (biomarcador) que contribuía a la mayoría de cada componente principal. Como el cuadrado de cargas del factor reflejan la contribución relativa de cada factor en cada componente principal, se utilizaron estos valores como base para seleccionar el marcador que contribuía con la mayoría de cada componente principal. Así, se extrajeron 15 factores (biomarcadores) que contribuían a la mayoría de los 15 componentes principales. En el DA, el proceso de marcadores de selección se llevó a cabo hasta
- 15 que la adición de más marcadores no tenía efecto en el resultado de la clasificación. En general, el DA se utilizó entre 5 y 8 marcadores. En el caso de los DT, se construyeron y evaluaron árboles de 6 nodos con aproximadamente 5 biomarcadores.

- Los biomarcadores se evaluaron utilizando métodos de validación bien establecidos de bootstrapping y dejar uno
- 20 fuera (Richard O. Duda et al. In Pattern Classification , 2ª Edición, pp. 485, Wiley-Interscience (2000)). Se utilizó un proceso de entrenamiento de 10 veces para identificar los biomarcadores robustos que se muestran regularmente. Los biomarcadores robustos se definieron como los marcadores que emergen en al menos un 50 % de los grupos de entrenamiento. Por tanto, los biomarcadores con una frecuencia mayor o igual a 5 en nuestro proceso de entrenamiento de diez veces se seleccionaron para posterior evaluación. La Tabla 16 a continuación resume los
- 25 biomarcadores que mostraban regularidad en cada método en cada cohorte.

La estrategia de descubrimiento de biomarcadores que utiliza varios métodos estadísticos ofrece una ventaja distintiva proporcionando un repertorio más amplio de biomarcadores candidatos (Figura 1). Mientras que algunos métodos tales como DA y PCA funcionan bien con datos distribuidos normalmente, otros métodos no paramétricos tales como la regresión logística y los árboles de decisión funcionan mejor con datos que están separados, distribuidos no uniformemente o tienen variaciones extremas. Tal estrategia es ideal cuando los marcadores (tales como biomarcadores y parámetros biométricos) de diversas fuentes (espectrometría de masas, inmunoensayos, historia clínica, etc.) se van a combinar en un panel único aunque los marcadores pueden estar o no distribuidos normalmente en la población.

- 10 Tabla 16. Marcadores identificados utilizando análisis multivariado (MVM). Solamente los marcadores que se muestran al menos un 50 % de las veces se seleccionaron para consideración posterior. En la Tabla, no hay diferencias entre "x" y "X".

			pequeña cohorte						gran cohorte		
	AUC	Marcadores superiores	DA	PCA	DT		AUC	Marcadores superiores	DA	PCA	DT
1	0,76	acn9459	x			1	0,81	pub17858	x		x
2	0,75	pub4861	x		x	2	0,81	pub17338			x
3	0,66	CEA			x	3	0,8	pub8606	x		
4	0,65	pub9433		x		4	0,72	pub4861	x		x
5	0,64	pub9648			x	5	0,69	pub3743	x		x
6	0,64	pub2951	x			6	0,67	acn6399		x	
7	0,63	pub6052	x			7	0,66	tfa2331		x	
8	0,6	tfa2759	x			8	0,65	pub9433		x	
9	0,6	tfa9133	x			9	0,58	acn6592		x	
10	0,59	acn4132		x		10	0,56	pub4213		x	
11	0,58	acn6592		x		11	0,55	acn9371		x	
12	0,57	pub7775		x				Total	4	6	4
13	0,56	pub4213		x							
14	0,55	acn9371		x							
		Total	6	6	3						

Ejemplo 7. Método de división y puntuación (en adelante "SSM")

- 15 A. Método de división y puntuación (SSM) mejorado

El Software interactivo que implementa el punto de división (punto de corte) del método de puntuación descrito por Mor et al. (véase, PNAS, 102(21):7677 (2005)) se ha escrito para ejecutarse bajo Microsoft® Windows. Este software lee las hojas de cálculo de Microsoft® Excel que son los vehículos naturales para guardar los resultados del análisis del marcador (biomarcadores y parámetros biométricos) para un grupo de muestras. Los datos se pueden guardar en una única hoja de datos con un campo para designar la enfermedad de la muestra, se puede guardar en dos hojas de cálculo, una para muestras enfermas y la otra para muestras no enfermas, o en cuatro hojas de cálculo, un par para muestras de entrenamiento, enfermas y no enfermas y el otro par para las muestras de ensayo, enfermas y no enfermas. En los primeros dos casos, el usuario puede utilizar el software para generar automáticamente aleatoriamente pares de entrenamiento y de ensayo de la entrada. En el caso final, se pueden leer múltiples archivos Excel una vez y analizarse en una sola ejecución.

El software presenta una lista de todos los marcadores recolectados en los datos. El usuario selecciona un grupo de marcadores de esta lista para utilizarlo en los análisis. El software calcula automáticamente puntos de división (puntos de corte) para cada marcador de los grupos de datos de entrenamiento enfermos y no enfermos así como determina si el grupo de enfermos está aumentado o disminuido con respecto al de no enfermos. El punto de división (punto de corte) se elige para maximizar la precisión de cada marcador único. Los puntos de corte (o puntos de división) también se agrupan y ajustan manualmente.

En todos los análisis, se calculan la precisión, especificidad y sensibilidad de cada valor de umbral utilizando el grupo seleccionado de marcadores para tanto los grupos de entrenamiento como de ensayo. En los análisis que producen resultados múltiples estos resultados se ordenan por las precisiones del grupo de entrenamiento.

Están disponibles tres modos de análisis. El modo más simple calcula los resultados de referencia utilizando solamente los marcadores seleccionados. Un segundo modo determina el menor marcador valorable en la lista seleccionada. Se llevan a cabo múltiples cálculos, uno por cada subgrupo posible de marcadores formado por la eliminación de un solo marcador. El subgrupo con la mayor precisión sugiere que el marcador eliminado para crear el subgrupo tiene la menor contribución en el grupo completo. Los resultados de estos dos primeros modos son esencialmente inmediatos. El cálculo más implicado explora todas las combinaciones posibles de los marcadores seleccionados. Se muestran los veinte mejores resultados. Esta opción final puede implicar un gran número de

candidatos. Por tanto, es bastante intensivo computacionalmente y puede tomar algún tiempo en completarse. Cada marcador adicional que se utiliza dobla el tiempo de ejecución.

Parra aproximadamente 20 marcadores, a menudo se ha encontrado que hay normalmente 6 a 10 marcadores que aparecen en los 20 resultados mejores. Estos se cotejan con otros 2 a 4 marcadores del grupo. Esto sugiere que puede haber alguna flexibilidad en la selección de marcadores para un panel diagnóstico. Los veinte mejores resultados superiores generalmente son similares en precisión pero pueden diferenciarse significativamente en sensibilidad y especificidad. Viendo todas las posibles combinaciones de marcadores de esta forma se proporciona una visión de las combinaciones que pueden ser las más útiles clínicamente.

B. Método de Puntuación Ponderada

Como se ha tratado anteriormente en el presente documento, este método es un método de puntuación ponderada que implica la conversión de la medición de un marcador en uno de muchos marcadores potenciales. Estas puntuaciones se derivan utilizando la ecuación:

$$\text{Puntuación} = \text{AUC} \times \text{factor} / (1 - \text{especificidad})$$

Se puede utilizar el marcador citoqueratina 19 como un ejemplo ilustrativo. Los niveles de citoqueratina 19 varían desde 0,4 a 89,2 ng/ml en la pequeña cohorte. Utilizando el software Analyze-it, se generó una curva ROC con los datos de Citoqueratina 19 tal que los cánceres eran positivos. La tasa de falsos positivos (1-especificidad) se representó en el eje X y la tasa de positivos verdaderos (sensibilidad) se representó en el eje Y y se generó una hoja de cálculo con el valor de citoqueratina 19 correspondiente para cada punto de la curva. A un punto de corte de 3,3 ng/ml, la especificidad era del 90 % y la tasa de falsos positivos era del 10 %. Un factor de tres se dio arbitrariamente para este marcador mientras que su AUC era mayor de 0,7 y menor de 0,8 (véase, la Tabla 2). Sin embargo, cualquier número integral se puede utilizar como un factor. En este caso, los números en aumento se utilizan con biomarcadores que tienen un AUC mayor que indica una mejor actuación clínica. La puntuación para un individuo con un valor de citoqueratina 19 mayor o igual a 3,3 ng/ml se calculó de esta manera.

$$\text{Puntuación} = \text{AUC} \times \text{factor} / (1 - \text{especificidad})$$

$$\text{Puntuación} = 0,70 \times 3 / (1 - 0,90)$$

$$\text{Puntuación} = 21$$

Para cada valor de citoqueratina 19 mayor de 3,3 ng/ml, se daba de esta manera una puntuación de 21. Para cualquier valor de citoqueratina 19 mayor de 1,9 pero menor de 3,3, se daba una puntuación de 8,4 y así.

Tabla 17a. Las cuatro posibles puntuaciones determinadas para Citoqueratina 19

Punto de corte	CITOQUERATINA 19	
	AUC	0,70
	Especificidad	Puntuación
3,3	0,90	21
1,9	0,75	8,4
1,2	0,50	4,2
0	0	0,0

La puntuación aumenta de valor según aumenta el nivel de especificidad. Los valores escogidos para la especificidad se pueden ajustar a medida a cualquier marcador. El número del nivel de especificidad escogido para cualquier marcador se puede ajustar a medida. Este método permite que la especificidad mejore la contribución de un biomarcador a un panel.

Se hizo una comparación del método de puntuación ponderada con el método de puntuación binario descrito en el Ejemplo 7A anterior. En este ejemplo, el panel estaba constituido por ocho inmunoensayos de biomarcadores: CEA, Citoqueratina 19, Citoqueratina 18, CA125, CA15-3, CA19-9, proGRP, y SCC. Los AUC, factores, niveles de especificidad escogidos, y puntuaciones a cada uno de estos niveles de especificidad se tabulan para cada uno de los marcadores en la Tabla 17b a continuación. Utilizando estos puntos de corte y puntuaciones individuales, cada muestra estaba tabulada para los ocho biomarcadores. La puntuación total de cada muestra se sumó y representó en una curva ROC. Esta curva ROC se comparó con las curvas ROC generados utilizando el método de puntuación binaria con los puntos de corte de la pequeña cohorte (puntos de división) o puntos de corte de la gran cohorte (puntos de división) proporcionados en la Tabla 18 (véase, el Ejemplo 8A). Los valores del AUC para el método de puntuación ponderada, los puntos de corte (puntos de división) para la gran cohorte en el método de puntuación binaria, y los puntos de corte (puntos de división) para la pequeña cohorte en el método de puntuación binaria eran 0,78, 0,76, y 0,73, respectivamente. Además de la actuación total mejorada del panel como se indica por el valor del AUC, el método de puntuación ponderada proporciona un gran número de valores de puntuación posibles para el

panel. Una ventaja del número grande de puntuaciones posibles en el panel es que hay más opciones para fijar el punto de corte para un ensayo positivo (véase la **Figura 5**). El método de puntuación binaria aplicada a un panel de 8 biomarcadores puede tener valores de resultado de un panel que varía de 0 a 8 con incrementos de 1 (véase la **Figura 5**).

5

Tabla 17b.

	CEA	CK-18	proGRP	CA15-3	CA125	SCC	CK-19	CA19-9
AUC	0,67	0,65	0,62	0,58	0,67	0,62	0,7	0,55
Factor	2	2	2	1	2	2	3	1
valor @ 50 % de especificidad*	2,02	47,7	11,3	16,9	15,5	0,93	1,2	10,6
valor @ 75 % de especificidad*	3,3	92,3	18,9	21,8	27	1,3	1,9	21,9
valor @ 90 % de especificidad*	4,89	143,3	28,5	30,5	38,1	1,98	3,3	45,8
puntuación por debajo del 50 % de especificidad	0	0	0	0	0	0	0	0
puntuación por encima del 50 % de especificidad	2,68	2,6	2,48	1,16	2,68	2,48	4,2	1,1
puntuación por encima del 75 % de especificidad	5,36	5,2	4,96	2,32	5,36	4,96	8,4	2,2
puntuación por encima del 90 % de especificidad	13,4	13	12,4	5,8	13,4	12,4	21	5,5

* Cada uno de esos valores representan un punto de corte (punto de división).

La solicitud de patente titulada "Métodos de Puntuación Ponderada y el uso de los mismos en la Selección", presentada electrónicamente el 29 de junio de 2007 como Número de Expediente 8612.US.P1, describe entre otras cosas, el Método de Puntuación Ponderada.

10

Ejemplo 8. Modelos predictivos para el cáncer de pulmón utilizando el método de división y puntuación (SSM)

15 A. SSM de inmunoensayos de biomarcadores

Como se ha tratado en el Ejemplo 2, se detectaron algunos biomarcadores por ensayos inmunológicos. Estos incluían la Citoqueratina 19, CEA, CA125, SCC, proGRP, Citoqueratina 18, CA19-9, y CA15-3. Estos datos se evaluaron utilizando el SSM. Estos biomarcadores juntos mostraban una utilidad clínica limitada. En la pequeña cohorte, que representan la enfermedad pulmonar benigna y el cáncer de pulmón, la precisión del panel de 8 biomarcadores con un umbral de 4 o más como resultado positivo, alcanzaba un promedio de precisión del 64,8 % (AUC 0,69) en los 10 grupos de ensayo de la pequeña cohorte. En la gran cohorte, que representa los normales así como enfermedad pulmonar benigna y cáncer de pulmón, la precisión del panel de 8 biomarcadores con un umbral de 4 o mayor como resultado positivo, alcanzaba un promedio del 77,4 % (AUC 0,79) en los 10 grupos de ensayo de la gran cohorte.

20

25

Incluyendo el parámetro biométrico de paquetes-años mejoraba la precisión predictiva de estos biomarcadores en casi el 5 %. Por tanto, la precisión del panel de los 8 biomarcadores y 1 parámetro biométrico con un umbral de 4 o mayor como resultado positivo, alcanzaba un promedio del 69,6 % (AUC 0,75) en los 10 grupos de ensayo de la pequeña cohorte.

30

Tabla 18. Puntos de división (puntos de corte) calculados para cada inmunoensayo de marcador individual utilizando el algoritmo SSM

	pequeña cohorte		gran cohorte		grupo control
	Prom. punto de división (punto de corte predeterminado)	Dev.est.	Prom. punto de división (punto de corte predeterminado)	Dev.est.	
CEA	4,82	0	9,2	0	norm <= punto de división
CK 19	1,89	0,45	2,9	0,3	norm <= punto de división
CA125	13,65	8,96	26	2,6	norm <= punto de división
CA15-3	13,07	3,39	20,1	2,6	norm <= punto de división

	pequeña cohorte		gran cohorte		
	Prom. punto de división (punto de corte predeterminado)	Dev.est.	Prom. punto de división (punto de corte predeterminado)	Dev.est.	grupo control
CA19-9	10,81	11,25	41,1	18,5	norm <= punto de división
SCC	0,92	0,11	1,1	0,1	norm <= punto de división
proGRP	14,62	8,53	17,6	0	norm <= punto de división
CK-18	57,37	2,24	67,2	9,5	norm <= punto de división
parainfluenza	103,53	32,64	79,2	9,8	norm >= punto de división
Pack-yrs	30		30		Norm <= punto de división

B. SSM de biomarcadores y parámetros biométricos seleccionados por ROC/AUC

Al contrario que en el Ejemplo 6, en el que se identificaban supuestos marcadores utilizando métodos estadísticos multivariados, se utilizó en este caso un método no paramétrico que implicaba análisis ROC/AUC para identificar los supuestos biomarcadores. Aplicando este método, se escogieron marcadores individuales con actuación clínica aceptable ($AUC > 0,6$) para análisis posteriores. Solamente los 15 biomarcadores superiores y el parámetro biométrico (paquetes años) se seleccionaron y se llamó a los grupos los grupos 16AUC (pequeño y grande) de aquí en adelante. Estos marcadores se enumeran en la Tabla 19, a continuación.

Tabla 19. Los 15 biomarcadores superiores y un parámetro biométrico (paquetes años)

gran cohorte			pequeña cohorte		
Marcador	nº obs	AUC	Marcador	nº obs	AUC
pub17338	513	0,813	pub11597	236	0,766
pub17858	513	0,812	acn9459	244	0,761
pub8606	513	0,798	pub4861	250	0,75
pub8662	513	0,796	pack-yrs	257	0,739
pub4628	513	0,773	pub4750	250	0,729
pub6798	513	0,765	pub7499	250	0,725
pub7499	513	0,762	pub2433	250	0,719
pub4750	513	0,76	CK 19	248	0,718
pub15599	513	0,757	pub4789	250	0,718
pub11597	513	0,751	pub17338	250	0,718
pub4487	513	0,747	pub8662	250	0,713
tfa6453	538	0,744	acn9471	244	0,712
Paquetes años	249	0,741	pub15599	250	0,711
pub8734	513	0,741	tfa6652	236	0,71
pub14430	513	0,741	pub8606	250	0,703
hic3959	529	0,741	acn6681	244	0,703

Las combinaciones optimizadas (paneles) de la pequeña cohorte de 16AUC se determinaron utilizando el SSM de cada uno de los 10 subgrupos de entrenamiento. Este proceso se hizo tanto en ausencia (Tabla 20a) como en presencia (Tabla 20b) del parámetro biométrico de historia de fumador (paquetes años) utilizando el SSM. Así, se introdujeron las variables de los 15 biomarcadores (excluyendo el parámetro biométrico pack-yrs) o los 15 biomarcadores y el 1 parámetro biométrico (paquetes años) (el 16 AUC) para el método de división y puntuación. Se determinó el panel óptimo de cada uno de los 10 grupos de entrenamiento basándose en la precisión total. Cada panel se ensayó contra las muestras remanentes sin ensayar y se registró las características estadísticas. Se compararon entonces los 10 paneles y se anotó la frecuencia de cada biomarcador. El proceso se ejecutó dos veces, incluyendo y excluyendo el biométrico paquetes años. Los resultados de estos dos procesos se presentan en las Tablas 20a y 20b a continuación. Una vez más, los marcadores robustos con una frecuencia mayor o igual a 5 se seleccionaron para posteriores consideraciones. El proceso se repitió en la gran cohorte y los resultados se presentan en la Tabla 20c. Las Tablas 20a y 20b contenían una lista parcial de los resultados SSM de la pequeña cohorte, mostrando la frecuencia de los marcadores para a) los biomarcadores 15 AUC solos y b) los biomarcadores 15 AUC y el parámetro biométrico packyrs. Se señala que en la primera tabla (20a) solo 5 marcadores tienen frecuencias mayores o iguales a 5. En la tabla 20b, 7 marcadores cumplían ese criterio. La Tabla 20c contiene una lista parcial de los resultados SSM de la gran cohorte que muestra la frecuencia de los marcadores de los marcadores 15 AUC. Se señala que 11 marcadores tienen frecuencias mayores o iguales a 5.

Tabla 20a.

Grupo de entrenamiento n°	CK 19	pub4789	acn9459	Pub11597	tfa6652	pub2433	pub4713
1	x	x	x	x			
2	x	x	x	x			x
3	x	x	x		x		
4	x	x		x		x	
5	x	x	x		x		
6	x	x	x		x		
7	x	x	x	x			x
8	x	x	x	x	x	x	
9	x	x	x		x		x
10	x	x	x	x		x	
Frecuencia	10	10	9	6	5	3	3

Tabla 20b.

Grupo de entrenamiento n°	acn 9459	CK 19	packyrs	Pub115 97	pub 4789	pub 2433	pub 4861	tfa 6652	acn 9471
1	x	x	x			x	x		
2	x	x	x	x	x	x			
3	x	x	x	x	x			x	
4	x	x	x	x	x	x		x	x
5	x	x	x	x			x	x	
6	x		x	x	x				x
7	x	x	x	x	x	x	x		
8	x	x	x		x		x		x
9	x	x	x	x			x	x	
10	x	x		x	x	x			x
Frecuencia	10	9	9	8	7	5	5	4	4

5

Tabla 20c.

Grupo de entrenamiento n°	Pub 115 97	pub 4487	pub 17338	pub 8606	Pub 6798	Tfa 6453	pub 4750	hic 3959	pub 8662	pub 4628	pub 17858
1	x	x	x	x	x	x		x			
2	x	x	x	x	x	x	x	x	x		
3	x	x	x	x			x	x	x		x
4	x		x		x	x		x			
5	x	x	x	x		x	x		x	x	x
6	x	x		x	x	x	x	x	x	x	
7	x	x			x	x	x	x	x	x	x
8	x	x	x		x				x	x	
9	x	x		x			x			x	x
10	x	x	x	x	x	x	x	x		x	x
Frecuencia	10	9	7	7	7	7	7	7	6	6	5

C. SSM de biomarcadores seleccionados por MVM

- Un ejemplo de un método multivariado es el análisis en árbol de decisión. Los biomarcadores utilizando el análisis en árbol de decisión solo se tomaron en conjunto y se utilizaron en el SSM. Este grupo de biomarcadores demostraron una utilidad clínica similar al grupo de biomarcadores llamado 16 AUC. Como ejemplo, el grupo de ensayo 1 (de 10) tenía una AUC de 0,90 (ensayo) sin el parámetro biométrico paquetes años, y 0,91 (ensayo) con el parámetro biométrico paquetes años.
- Los biomarcadores DT se combinaron con marcadores identificando que utilizaban PCA y DA para generar el grupo MVM. El grupo 14MVM se evaluó con y sin el parámetro biométrico de historia de fumador (paquetes años) utilizando el SSM. Una vez más, los marcadores robustos con una frecuencia mayor o igual a 5 se seleccionaron para posterior consideración (resultados no mostrados). Como se puede ver en las tablas anteriores, los paquetes años (historia de fumador) tiene un efecto en el número y tipo de biomarcadores que emergen como biomarcadores robustos. Esto no es totalmente inesperado ya que algunos biomarcadores pueden tener efectos sinérgicos o perjudiciales sobre otros biomarcadores. Un aspecto de la invención implica el hallazgo de los marcadores que

funcionan juntos como un panel mejorando la capacidad predictiva del modelo. Siguiendo la misma línea, los biomarcadores que se identificaron que funcionaban sinérgicamente con el parámetro biométrico paquetes años en ambos métodos (AUC y MVM) se combinaron en un esfuerzo para identificar un panel superior de marcadores (véase el Ejemplo 8D).

- 5 Los marcadores multivariados identificados en la gran cohorte se evaluaron con el SSM. Una vez más, solo se seleccionaron los marcadores con frecuencias mayores o iguales a 5 para posterior consideración. La Tabla 21 a continuación resume los resultados SSM para la gran cohorte.

10 Tabla 21. Lista parcial de los resultados SSM de la gran cohorte que muestra la frecuencia de los marcadores de los marcadores 11MVM. Se señala que 7 marcadores tienen frecuencias mayores o iguales a 5.

Grupo de entrenamiento n°	pub 3743	pub 4861	pub 8606	Pub 17338	pub 17858	acn 6399	tfa 2331
1	x	x	x	x	x	x	
2	x	x	x	x	x		x
3	x	x	x	x		x	
4	x	x		x	x	x	
5	x	x	x	x			x
6	x	x	x		x	x	
7	x	x	x	x	x	x	x
8	x	x	x	x			
9	x	x	x	x		x	x
10	x		x	x	x		x
Frecuencia	10	9	9	8	6	6	5

D. SSM de marcadores combinados (AUC + MVM + paquetes años)

- 15 En una etapa posterior, todos los marcadores (biomarcadores y parámetros biométricos) con frecuencias mayores o iguales a 5 (en los 10 grupos de entrenamiento) se combinaron para producir una segunda lista de marcadores que contenían marcadores de ambos grupos AUC y MVM para ambas cohortes. De los resultados SSM, se seleccionaron 16 únicos marcadores únicos de la pequeña cohorte y 15 marcadores únicos de la gran cohorte con frecuencias mayores o iguales a cinco. La tabla 22 siguiente resume los marcadores que se seleccionaron.

20

Tabla 22. Marcadores combinados para ambos grupos AUC y MVM

	pequeña cohorte					gran cohorte			
	AUC	Marcadores	16AUC	14MVM		AUC	Marcadores	15AUC	11MVM
1	0,77	Pub11597	x		1	0,813	Pub17338	x	x
2	0,76	Acn9459	x	x	2	0,812	pub17858	x	x
3	0,75	Pub4861	x	x	3	0,798	pub8606	x	x
4	0,74	packyrs	x	x	4	0,796	pub8662	x	
5	0,72	Pub2433	x		5	0,773	pub4628	x	
6	0,72	CK 19	x		6	0,765	pub6798	x	
7	0,72	Pub4789	x		7	0,76	pub4750	x	
8	0,71	Tfa6652	x		8	0,751	pub11597	x	
9	0,66	cea		x	9	0,747	pub4487	x	
10	0,64	Pub2951		x	10	0,744	tfa6453	x	
11	0,63	Pub6052		x	11	0,741	hic3959	x	
12	0,6	Tfa2759		x	12	0,72	pub4861		x
13	0,6	Tfa9133		x	13	0,69	pub3743		x
14	0,59	Acn4132		x	14	0,67	acn6399		x
15	0,58	Acn6592		x	15	0,66	tfa2331		x
16	0,57	Pub7775		x			Total	11	7
		Total	8	11					

Las listas anteriores de marcadores se tomaron para un ciclo de evaluación final con el SSM. Como se estableció anteriormente, la combinaciones de marcadores se optimizaron para los 10 subgrupos de entrenamiento y se determinó la frecuencia de cada biomarcador y parámetro biométrico. Aplicando el criterio de selección de que un marcador esté presente al menos en el 50 % de los grupos de entrenamiento, se seleccionaron 13 de los 16 marcadores para la pequeña cohorte y 9 de los 15 marcadores de la gran cohorte.

25

Tabla 23a. Lista de marcadores con frecuencias mayores o iguales a 5.

pequeña cohorte					gran cohorte			
	AUC	Marcadores	Frecuencia			AUC	Marcadores	Frecuencia
1	0,718	CK 19	9		1	0,67	acn6399	10
2	0,761	acn9459	8		2	0,69	pub3743	8
3	0,74	packyrs	8		3	0,798	pub8606	7
4	0,664	cea	8		4	0,751	pub11597	7
5	0,603	tfa2759	8		5	0,744	tfa6453	7
6	0,766	pub11597	7		6	0,747	pub4487	6
7	0,718	pub4789	7		7	0,72	pub4861	6
8	0,6	tfa9133	7		8	0,765	pub6798	5
9	0,75	pub4861	6		9	0,741	hic3959	5
11	0,719	pub2433	6					
10	0,589	acn4132	6					
12	0,57	Pub7775	6					
13	0,635	pub2951	5					

Para cada marcador, se determinó un punto de división (punto de corte) evaluando cada grupo de datos de entrenamiento para la precisión más alta sobre la clasificación al optimizar el nivel del marcador. Los puntos de división (puntos de corte) para los ocho marcadores más frecuentes utilizados en la pequeña cohorte se enumeran a continuación.

Tabla 23b.

	Marcadores	Grupo Control	Prom.	Desv. Est.
1	CK 19	Norm <= SP	1,89	0,45
2	acn9459	Norm >= SP	287,3	23,67
3	packyrs	Norm <= SP	30,64	4,21
4	cea	Norm <= SP	4,82	0
5	tfa2759	Norm >= SP	575,6	109,7
6	pub11597	Norm <= SP	34,4	2,52
7	pub4789	Norm <= SP	193,5	18,43
8	tfa9133	Norm >= SP	203,6	46,38

La tabla 23b muestra la lista de los 8 marcadores más frecuentes con su promedio (prom.) de puntos de división (cada uno es un punto de corte predeterminado). También se incluyen las desviaciones estándar de cada punto de división (punto de corte) (Desv. Est.). Se da en la segunda columna por la izquierda la posición del grupo de control respecto al punto de división (punto de corte). Como un ejemplo, en la Citoqueratina 19, el grupo normal o el grupo control (no canceroso) es menor o igual que el valor 1,89 del punto de división /punto de corte).

Ejemplo 9. Validación de los modelos predictivos

Los subgrupos de la lista de 13 biomarcadores y parámetros biométricos de la pequeña cohorte (véase la Tabla 23a anterior) proporciona una buena utilidad clínica. Por ejemplo, los 8 biomarcadores y parámetros biométricos más frecuentes usados juntos en un panel en el método de división y puntuación tienen una AUC de 0,90 para el subgrupo de ensayo 1 (véase la Tabla 23b anterior).

Los modelos predictivos comprendían un panel de 7 marcadores (marcadores 1-7, Tabla 23b) y un panel de 8 marcadores (marcadores 1-8, Tabla 23b) se validaron utilizando 10 grupos de ensayo aleatorios. Las Tablas 24a y 24b a continuación resumen los resultados de los dos modelos. Todas las condiciones y parámetros de cálculo eran idénticos en ambos casos con la excepción del número de marcadores en cada modelo.

Tabla 24a.

Grupo de entrenamiento nº	AUC	Precisión (%)	Sensibilidad (%)	Especificidad (%)	Nº de Marcadores	Umbral
1	0,91	85	80,7	90,7	7	3
2	0,92	85	78,2	93,3	7	3

Grupo de entrenamiento nº	AUC	Precisión (%)	Sensibilidad (%)	Especificidad (%)	Nº de Marcadores	Umbral
3	0,89	80	78,8	82,4	7	3
4	0,89	82	78,0	86,0	7	3
5	0,90	85	78,7	90,6	7	3
6	0,89	83	76,9	89,6	7	3
7	0,92	86	78,4	93,9	7	3
8	0,89	83	79,6	87,0	7	3
9	0,91	84	79,6	89,1	7	3
10	0,92	86	81,8	91,1	7	3
Prom.	0,90	83,9	79,1	89,4		
Desv. Est.	0,01	1,9	1,4	3,5		

La Tabla 24b muestra la actuación clínica del panel de 7 marcadores con diez grupos de ensayo aleatorios. Los 7 marcadores y los puntos de división (puntos de corte) que se usaron en los cálculos se dieron en la Tabla 16b. Se utilizó un valor umbral de 3 para separar el grupo enfermo del grupo no enfermo. El promedio del AUC para el modelo es 0,90, que se corresponde con un promedio de precisión del 83,9 % y una sensibilidad y especificidad del 79,1 % y 89,4 % respectivamente.

Tabla 24b.

Grupo de entrenamiento nº	AUC	Precisión (%)	Sensibilidad (%)	Especificidad (%)	Nº de Marcadores	Umbral
1	0,90	81	91,2	67,4	8	3
2	0,91	86	92,7	77,8	8	3
3	0,89	83	90,9	67,6	8	3
4	0,89	83	90,0	76,0	8	3
5	0,91	83	91,5	75,5	8	3
6	0,90	83	88,5	77,1	8	3
7	0,92	88	92,2	83,7	8	3
8	0,90	85	92,6	76,1	8	3
9	0,93	84	92,6	73,9	8	3
10	0,92	85	92,7	75,6	8	3
Prom.	0,91	84,1	91,5	75,1		
Desv. Est.	0,01	1,8	1,4	4,7		

La Tabla 24b muestra la actuación clínica del panel de 8 marcadores con 10 grupos de ensayo aleatorios. Los 8 marcadores y la media de puntos de división (puntos de corte) utilizados en el cálculo se dieron en la Tabla 16b. Un valor del umbral de 3 (una puntuación total predeterminada) se utilizó para separar el grupo enfermo del grupo no enfermo. El AUC promedio para el modelo es 0,91, que corresponde con una precisión promedio del 84,1 % y una sensibilidad y especificidad del 91,5 % y 71,5 % respectivamente.

Una comparación de las Tablas 24a y 24b muestra que ambos modelos son comparables en términos de AUC y precisión y se diferencian solo en la sensibilidad y especificidad. Como se puede ver en la Tabla 24a, el panel de 7 marcadores muestra una especificidad mayor (89,4 % vs. 75,1 %). Por el contrario, el panel de 8 marcadores muestra mejor sensibilidad (91,5 % frente a 79,1 %) como se valora por sus valores promedio (Prom.). Se debería señalar que se escogió el umbral (o puntuación total predeterminada) que maximizaba la precisión de la clasificación que es semejante a maximizar el AUC de la curva ROC. Por tanto, el umbral escogido de 3 (una puntuación predeterminada) no solo maximiza la precisión sino que también ofrece el mejor compromiso entre la sensibilidad y la especificidad del modelo. En la práctica, esto lo que significa es que un individuo normal se considera que esta en bajo "riesgo" de desarrollar cáncer de pulmón si dichos ensayos individuales son positivos para menos o igual a 3 de los 7 marcadores posibles en este modelo (o menos o igual a 3 de los 8 para el segundo modelo). Los individuos con puntuaciones mayores (puntuación total) del umbral fijado (o puntuación total predeterminada) se considera que tiene un riesgo más alto y se convierten en candidatos para posteriores ensayos o procedimientos de seguimiento. Se debería señalar que el umbral del modelo (a saber, la puntuación total predeterminada) puede aumentarse o disminuirse con el fin de maximizar la sensibilidad o la especificidad de dicho modelo (a expensas de la precisión). Esta flexibilidad es ventajosa ya que permite al modelo ajustarse para abordar diferentes cuestiones diagnósticas y/o poblaciones en riesgo, por ejemplo, diferenciando individuos normales de individuos sintomáticos y/o asintomáticos.

Se resumen varios modelos predictivos en las Tablas 25a y 25b a continuación. Para cada modelo predictivo, se indican los biomarcadores y los parámetros biométricos que constituyen el modelo, como son el umbral (a saber, la puntuación total predeterminada), el AUC promedio, la sensibilidad, y especificidad con sus correspondientes desviaciones estándar (entre paréntesis) a través de los 10 grupos de ensayo. El panel de 8 marcadores descrito

- 5 anteriormente es el Modelo Mezclado 2 y el panel de 7 marcadores descrito anteriormente es el Modelo Mezclado 3. El Modelo Mezclado 1A y el Modelo Mezclado 1B contienen los mismos marcadores. La única diferencia entre el Modelo Mezclado 1A y el Modelo Mezclado 1B es el umbral (a saber, la puntuación total predeterminada). De manera similar, el Modelo Mezclado 10A y el Modelo Mezclado 10B contiene los mismos marcadores. La única diferencia entre el Modelo Mezclado 10A y el Modelo Mezclado 10B es en el umbral (a saber, la puntuación total predeterminada).

Tabla 25a.

	pequeña cohorte										
Marcadores	Modelo 8/A	Modelo 9/A	Modelo 1A- pkys	Modelo MS	Modelo MS pkys	Modelo Mezclado 1A	Modelo Mezclado 1B	Modelo Mezclado 2	Modelo Mezclado 3	Modelo Mezclado 4	Modelo Mezclado 5
CK 19	X	X	X			X	X	X	X		
CA 19-9	X	X	X								
CEA	X	X	X			X	X	X	X	X	X
CA15-3	X	X	X								
CA125	X	X	X								
SCC	X	X	X								
CK 18	X	X	X								
ProGRP	X	X	X								
Parainflu		X	X								
Pkys			X		X			X	X	X	
Acn9459				X	X	X	X	X	X	X	X
Pub11597				X	X	X	X	X	X	X	X
Pub4789				X	X	X	X	X	X	X	X
TFA2759				X	X	X	X	X	X	X	X
TFA9133				X	X	X	X	X		X	X
pub3743											
pub8606											
pub4487											
pub4861											
pub6798											
tfa6453											
hic3959											
Umbral*	1/8	4/9	4/10	3/5	3/6	2/7	3/7	3/8	3/7	3/7	3/6

pequeña cohorte											
Marcadores	Modelo 8IA	Modelo 9IA	Modelo IA-pkys	Modelo MS	Modelo MS pkys	Modelo Mezclado 1A	Modelo Mezclado 1B	Modelo Mezclado 2	Modelo Mezclado 3	Modelo Mezclado 4	Modelo Mezclado 5
AUC		0,73 (0,04)	0,80(0,03)	0,83 (0,02)	0,86(0,02)		0,87(0,02)	0,91 (0,01)	0,90(0,01)	0,89(0,01)	0,86 (0,02)
Precisión		66,0 (4,1)	70,0(2,4)	77,0 (3,7)	80,0(2,1)		78,8 (2,0)	84,1 (2,0)	83,9(1,9)	83,0(1,9)	79,4(3,6)
Sensibilidad	90,2 (3,1)	69,5 (8,5)	85,0(5,0)	63,4 (4,6)	72,0(3,5)	91,3(2,0)	81,6(2,3)	91,5(1,4)	79,1 (1,4)	81,3(1,8)	70,9(4,3)
Especificidad	30 (4,7)	62,0 (6,8)	52,3(3,9)	93,3 (2,5)	89,0(2,6)	42,7(3,6)	75,5(3,1)	75,1 (3,1)	89,4(3,5)	84,8 (4,7)	89,6(3,0)
DFI	0,71	0,49	0,50	0,37	0,30	0,58	0,31	0,26	0,23	0,24	0,31
*Puntuación Total predeterminada, En la tabla anterior, no hay diferencia entre "x" y "X".											

Tabla 25b.

pequeña cohorte						
Marcadores	Modelo mezclado 6	Modelo mezclado 7	Modelo mezclado 8	Modelo mezclado 9	Modelo mezclado 10A	Modelo mezclado 10B
CK 19		x		x	x	x
CA 19-9						
CEA	x	x		x	x	x
CA15-3						
CA125				x	x	x
SCC				x	x	x
CK 18	x			x	x	x
ProGRP				x	x	x
Parainflu						
Packyrs	x				x	x
Acn9459	x	x		x	x	x
Pub11597	x	x	x	x	x	x
Pub4789	x	x		x	x	x
TFA2759	x	x		x	x	x
TFA9133	x					
pub3743			x			
pub8606			x			
pub4487			x			
pub4861			x			
pub6798			x			
tfa6453			x			
hic3959			x			
Umbral*	3/8	2/6	3/8	3/10	3/11	4/11
AUC	0,90 (0,01)					
Precisión	80,2 (1,7)					
Sensibilidad	92,6 (2,0)	87,8 (2,3)	88,2 (3,3)	89,1 (3,4)	94,3 (1,2)	86,6 (4,40)
Especificidad	65,5 (2,7)	63,7 (4,9)	64,2 (3,7)	52,3 (3,9)	47,6 (4,9)	63,9 (4,0)
DFI	0,35	0,38	0,38	0,49	0,53	0,39
*Puntuación total predeterminada.						

Las Tablas 25a y b. Resumen varios modelos predictivos.

- 5 De manera similar, se pueden optimizar varios modelos predictivos para la gran cohorte en cuenta a la precisión, sensibilidad, o especificidad. Se resumen cuatro modelos potenciales en la Tabla 26 a continuación.

Tabla 26. Cuatro modelos potenciales

gran cohorte				
Marcadores	Modelo 1 MS	Modelo 2 MS	Modelo 3 MS	Modelo MS 4
acn6399	x	x	x	x
pub3743	x	x	x	x
pub8606	x	x	x	x
pub11597	x	x	x	x
tfa6453	x	x	x	x
pub4487	x	x	x	x
pub4861	x	x	x	
pub6798	x	x		
hic3959	x			
Umbral*	3/9	3/8	3/7	2/6
AUC				
Precisión	75,7 (2,6)	80,0 (2,0)	84,2 (1,7)	78,9 (2,6)
Sensibilidad	95,1 (2,0)	89,7 (2,6)	80,7 (4,4)	88,5 (4,0)
Especificidad	67,7 (3,1)	78,0 (2,2)	85,7 (1,4)	74,9 (2,7)
DFI	0,33	0,26	0,24	0,28
*Puntuación Total predeterminada. En la Tabla anterior, no hay diferencia entre "x" y "X".				

De manera similar, los modelos predictivos de la cohorte de ciclina (subgrupo de individuos con anticuerpos medidos anti-proteína ciclina E2 y anticuerpos anti-péptido ciclina E2) se resumen en las Tablas 27a y 27b a continuación.

Tabla 27a

Cohorte de Ciclina (234 muestras)											
Marcadores	Modelo A	Modelo B	Modelo C	Modelo D	Modelo E	Modelo F	Modelo G	Modelo H	Modelo I	Modelo J	Modelo
CK19									X	X	
CA19-9											
CEA											
CA15-3											
CA125						X	X	X	X		
SCC								X	X		
CK18							X	X	X		
ProGRP					X	X	X	X	X		
Parainflu											
Pkys				X	X	X	X	X	X		
Acn9459											
Pub11597										X	X
Pub4789											
TFA2759											
TFA9133											
Pub6453											X
Pub2951										X	
Pub4861										X	
Pub2433										X	
Pub3743											
Pub17338											
TFA6652											
Pép. Ciclina E2-1	X			X	X	X	X	X	X	X	X
Proteína Ciclina E2		X									

Cohorte de Ciclina (234 muestras)											
Marcadores	Modelo A	Modelo B	Modelo C	Modelo D	Modelo E	Modelo F	Modelo G	Modelo H	Modelo I	Modelo J	Modelo
Pép. Ciclina E2-2			X								
Umbral*	0/1	0/1	0/1	0/2	0/3	0/4	0/5	0/6	0/7	2/6	1/3
Precisión	79,0	75,4	67,4	84,1	86,2	85,2	83,5	81,2	80,4	88,4	88,4
Sensibilidad	61,2	44,7	31,8	93,2	87	91,8	95,3	95,3	95,5	80,0	74,1
Especificidad	89,9	94,2	89,2	72,9	85,6	81,3	76,2	72,7	71,4	93,5	97,1
DFI	0,40	0,56	0,69	0,28	0,19	0,20	0,24	0,28	0,29	0,21	0,26
* Puntuación Total predeterminada.											

Tabla 27b. Las Tablas 27a y 27b proporcionan modelos predictivos para la cohorte de ciclina.

Marcadores	Modelo L	Modelo M	Modelo N	Modelo O	Modelo P	Modelo Q	Modelo R	Modelo S	Modelo T	Modelo U	Modelo V
CK19		X	X	X	X						
CA19-9											
CEA			X	X	X	X	X				
CA15-3											
CA125					X						
SCC					X						
CK18					X	X					
ProGRP					X	X	X	X	X	X	
Parainflu											
Pkys											
Acn9459											
Pub11597	X			X							
Pub4789											
TFA2759											
TFA9133											
Pub6453	X										
Pub2951											
Pub4861									X	X	
Pub2433									X		
Pub3743									X	X	X
Pub17338									X	X	X
TFA6652									X		
pép. Ciclina E2-1		X	X	X	X	X	X	X	X	X	
proteína Ciclina E2	X										
pép. Ciclina E2-2											

Marcadores	Modelo L	Modelo M	Modelo N	Modelo O	Modelo P	Modelo Q	Modelo R	Modelo S	Modelo T	Modelo U	Modelo V
Umbral*	1/3	0/2	0/3	1/4	1/7	0/4	0/3	0/2	2/8	1/5	0/2
Precisión	84,4	80,3	80,8	82,6	63,8	82,1	83,0	82,1	93,8	92,9	85,2
Sensibilidad	64,7	80,0	81,1	58,8	94,1	80	75,3	72,9	90,6	89,4	85,9
Especificidad	96,4	80,6	80,6	97,1	45,3	83,4	87,8	87,8	95,7	95	84,9
DFI	0,35	0,28	0,27	0,41	0,55	0,26	0,28	0,30	0,10	0,12	0,21
*Puntuación Total predeterminada.											

De manera similar, los modelos predictivos que utilizan anticuerpos se resumen en la Tabla 28 a continuación.

Tabla 28. Modelos predictivos utilizando ensayos de autoAc.

Marcadores	Modelo Aab1	Modelo Aab2
TMP21	x	x
Dominio-NPC1L1C	x	x
CCNE2BM-E2-1	x	x
TMOD1	x	x
CAMK1	x	x
RGS1	x	x
PACSIN1	x	x
p53	x	x
RCV1	x	
MAPKAPK3	x	x
Umbral*	1/10	1/9
Precisión	82	82,9
Sensibilidad	74,7	73,5
Especificidad	86,4	88,4
DFI	0,29	0,29
*Puntuación Total predeterminada.		

- 5 Cinco de estos modelos se utilizaron contra la cohorte de validación. La Tabla 29 a continuación resume la actuación clínica para cada uno de los modelos predictivos para las cohortes independientes, pequeña cohorte y cohorte de validación.

Tabla 29.

	Modelo Mezclado 71	Modelo Mezclado 1	Modelo 8 IA	Modelo MS 5	Modelo Mezclado 9
CK 19	x	x	x		x
CEA	x	x	x		x
CA19-9			x		
CA15-3			x		
CA125			x		x
SCC			x		x
CK 18			x		x
proGRP			x		x
parainfluenza					
acn9459	x	x			x
pub11597	x	x		x	x
pub4789	x	x			x
tfa2759	x	x			x
tfa9133		x			
pub3743				x	
pub8606				x	
pub4487				x	
pub4861				x	
pub6798				x	
tfa6453				x	
hic3959				x	
pack-ys					
Umbral*	2/6	2/7	1/8	3/8	3/10
Pequeña Cohorte					
AUC					
Precisión					
Sensibilidad	87,8	91,3	90,2	88,2	89,1
Especificidad	63,7	42,7	30,0	64,2	52,3
DFI	0,38	0,58	0,71	0,38	0,49
Cohorte de Validación					
AUC					
Precisión					
Sensibilidad	75,6	87,2	94,2	82,5	88,4

	Modelo Mezclado 71	Modelo Mezclado 1	Modelo 8 IA	Modelo MS 5	Modelo Mezclado 9
Especificidad	62,9	55,7	35,2	86,0	58,6
DFI	0,44	0,46	0,65	0,22	0,43
*Puntuación Total predeterminada. En la Tabla anterior, no hay diferencia entre "x" y "X".					

Ejemplo 10. Identificación de biomarcadores

A. Fraccionamiento HPLC

Con el fin de obtener la identidad de los candidatos del biomarcador MS de la Tabla 22, era necesario fraccionar primero las muestras de suero agrupadas y/o individuales por HPLC de fase inversa utilizando protocolos de referencia. La obtención de material suficiente para la electroforesis en gel y para los análisis MS se necesitan varios ciclos de fragmentación. Las fracciones individuales se perfilaron por MS MALDI-TOF y las fracciones que contenían los picos de interés se agruparon juntos y se concentraron en un speedvac. Todos los otros biomarcadores candidatos se procesaron como se ha descrito anteriormente.

La Figura 2 muestra un supuesto marcador (Pub11597) antes y después de la concentración. Se señala que el biomarcador candidato de 11 kDa en la muestra de inicio está muy diluido. Tras la concentración la intensidad es más alta pero la muestra no es suficientemente pura para el análisis y se necesita una posterior separación por SDS-PAGE con el fin de aislar el biomarcador de interés.

B. Digestión en gel y análisis LC-MS/MS

Tras la concentración, las fracciones que contienen los biomarcadores candidatos se sometieron a SDS-PAGE para aislar el péptido/proteína deseado que tiene una masa molecular que corresponde al biomarcador candidato. Se llevó a cabo la electroforesis en gel (SDS-PAGE) utilizando la metodología de referencia proporcionada por el fabricante (Invitrogen, Inc.). En resumen, el procedimiento implica la carga de muestras que contienen los biomarcadores candidatos y proteínas de referencia de masa molecular conocida en diferentes pocillos en el mismo gel como se muestra en la Figura 3. Comparando las distancias de migración de las proteínas de referencia con las de la muestra "desconocida", se identifica la banda con masa molecular deseada y se escinde del gel.

La banda de gel escindida se sometió entonces a digestión automática triptica en gel utilizando una estación Waters MassPREP™. Posteriormente, la muestra digerida se extrajo del gel y se sometió a ESI-LC-MS/MS de fase inversa en línea. El espectro del producto iónico se utilizó entonces para la búsqueda en la base de datos. Si era posible, la proteína identificada se obtenía comercialmente y se sometía a SDS-PAGE y digestión en gel como se ha descrito previamente. La coincidencia en la electroforesis en gel, los resultados MS/MS y la búsqueda en la base de datos entre las dos muestras daba más pruebas de que el biomarcador se había identificado correctamente. Como se puede ver en la **Figura 3**, hay una buena coincidencia entre el amiloide sérico A humano (HSAA) disponible comercialmente y el supuesto biomarcador en la muestra fraccionada a 11,5 kDa. El análisis MS/MS y la búsqueda en la base de datos confirmaron que ambas muestras eran la misma proteína. La **Figura 4** mostró que los espectros MS/MS del biomarcador candidato Pub11597. La secuencia de aminoácidos derivada de los iones b e y se anotaron en la parte superior de cada panel. El biomarcador candidato se identificó como un fragmento de la proteína del amiloide sérico humano (HSAA).

Los biomarcadores candidatos pequeños que no son adecuados para la digestión se sometieron a ESI-q-TOF y/o fragmentación MALDI-TOF-TOF seguida de la secuenciación de novo y la búsqueda en la base de datos (BLAST) para obtener información de la secuencia y el ID de la proteína.

C. Búsqueda en la base de datos e ID de la proteína

Con el fin de caracterizar completamente los biomarcadores candidatos era imperativo identificar las proteínas de las que se derivaron. La identificación de proteínas desconocidas implicaba la digestión en gel seguido por espectrometría de masas en tándem seguido por la secuenciación de novo y la búsqueda en la base de datos de los fragmentos tripticos. Los productos iónicos resultantes del proceso MS/MS se buscaron en la base de datos de proteínas SwissProt para identificar la proteína fuente. Para los biomarcadores candidatos que tenían bajas masas moleculares, la espectrometría de masas en tándem seguida por la secuenciación de novo y la búsqueda en la base de datos era el método de elección para identificar la proteína fuente. Las búsquedas se consideraban solo del genoma de Homo sapiens y con precisiones de masa de $\pm 1,2$ Da para iones precursores y $\pm 0,8$ Da para los productos iónicos (MS/MS). Solamente se permitió una escisión omitida por tripsina. Las únicas dos modificaciones variables permitidas para las búsquedas en la base de datos eran las carbamidometilación (C) y la oxidación (M). Se asignó un ID final de proteína tras el reconocimiento con la maquinaria de búsqueda Mascot resultante y la interpretación manual de los espectros de MS y MS/MS relacionados. La precisión de los resultados se verificó mediante mediciones replicadas.

Tabla 30

Marcador candidato	Nº de registro	Nombre de la Proteína	Secuencia peptídica observada	Prom. de PM (Da)
Pub11597	Q6FG67	Proteína A Amiloide Humana	SFFSFLGRAFDGARDMWRAYSDMREA NYIGSDKYFHARGNYDAAKRGPGGA WAAEIVSDARENIQRFFGHGAEDSLAD QAANEWG RSGKDPNHFRPAGLPEKY (SEQ ID NO: 7)	11526,51
ACN9459	P02656	ApoCIII ₁	SEAEDASLLSFMQGYMKIATKTAKDA LSSVQRSQVAQQARGWVTDGFSSLKD YWSTV KDKFSEFWDLDPVVRP *(T) SAVAA (SEQ ID NO: 8) *(sitio Glicosilado)	9421,22
TFA9133	P02656	ApoCIII ₁	ApoCIII ₁ tras la pérdida de ácido siálico	9129,95
Pub4789	P01009	alfa-1 antitripsina	LEAIPMSIPPEVKFN *(E) PFVFLMTDQ NTKSPLFMGKVVNPTQK (SEQ ID NO: 8) *(posible sustitución de K por E)	4776,69
TFA2759	Q56G89	Péptido Albúmina Humana	DAHKSEVAHRFKDLGEENFKAL VL (SEQ ID NO: 10)	2754,10

La Tabla 30 anterior da la proteína fuente de los varios biomarcadores candidatos con su ID proteico. Los marcadores se identificaron por digestión en gel y LC-MS/MS y/o secuenciación de novo. Se señala que solo las secuencias de aminoácidos de los fragmentos observados se muestran y el promedio de PM incluye el PTM donde se indica. Los números de registro se obtuvieron de la base de datos SwissProt y se dan solamente como referencia. Es interesante señalar que ACN9459 y TFA9133 eran los mismos fragmentos proteicos con la excepción de que el último había perdido un ácido siálico (-291,3 Da) del resto glucosilado. Tanto ACN9459 y TFA9133 se identificaron como una variante de la apolipoproteína CIII. Los hallazgos de los investigadores concuerdan con la secuencia conocida publicada y la masa molecular de esta proteína (Bondarcnko et. al, J. Lipid Research, 40:543-555 (1999)). El Pub4789 se identificó como la proteína alfa-1-antitripsina. El examen minucioso del espectro del producto iónico sugiere que podía haber una sustitución de K por E en el sitio indicado en la Tabla 30. La incertidumbre en la precisión de la masa excluye la asignación. Adicionalmente, los biomarcadores de la Tabla 30 anterior se conocen como proteínas de fase aguda (a las cuales también se denominan “respondedores de fase aguda (APR)) o “biomarcadores de respuesta inmunitaria”). Como se ha tratado anteriormente en el presente documento, las proteínas de fase aguda son miembros de la familia de proteína de inmunidad innata.

Ejemplo 11. Detección de Cáncer de pulmón

A. Inmunoensayo para péptidos o proteínas: Los biomarcadores descritos en el Ejemplo 9 anterior se pueden detectar y medir por técnicas de inmunoensayo. Por ejemplo, el sistema de inmunoensayo Architect™ de Abbott Diagnostics se utilizó para el ensayo automático de un desconocido en una muestra sospechosa de contener un biomarcador de la presente invención. Como se sabe en la técnica, el sistema utiliza micropartículas magnéticas revestidas con anticuerpos, que son capaces de unirse al biomarcador de interés. Bajo el control del instrumento, se mezcló una alícuota de la muestra con un volumen igual de micropartículas magnéticas revestidas de anticuerpos y dos veces el volumen del diluyente del espécimen, que contenía tampones, sales, tensioactivos, y proteínas solubles. Tras la incubación se lavaron las micropartículas con un tampón de lavado que comprendía tampón, sal, tensioactivo y conservante. Se añadió una alícuota de conjugado marcado con acridinio junto con un volumen igual el diluyente del espécimen y se redispersaron las partículas. La mezcla se incubó y luego se lavó con tampón de lavado. Las partículas lavadas se redispersaron en un predesencadenante ácido que contenía ácido nítrico y peróxido de hidrógeno para disociar el conjugado de acridinio de las partículas. Se añadió entonces una solución de NaOH para desencadenar la reacción quimioluminiscente. Se midió la luz con un fotomultiplicador y el resultado desconocido se cuantificó por comparación con la luz emitida por una serie de muestras que contenían cantidades conocidas del péptido biomarcador utilizado para construir una curva de referencia. La curva de referencia se utilizó entonces para estimar la concentración del biomarcador en una muestra clínica que se procesó de manera idéntica. El resultado se puede utilizar por sí mismo o en combinación con otros marcadores como se describe posteriormente.

B. Inmunoensayo multiplexado para péptidos o proteínas: Cuando se necesita la detección de múltiples biomarcadores en una muestra única, puede ser más económico y conveniente llevar a cabo un ensayo multiplexado. Para cada analito en cuestión, se necesitan un par de anticuerpos específicos y una micropartícula únicamente teñida para su uso en un analizador Luminex 100™. Cada anticuerpo de captura del par reviste individualmente una única micropartícula. El otro anticuerpo del par se conjuga con un fluoróforo tal como R-Ficoeritrina. Las micropartículas se agrupan y se diluyen a una concentración de aproximadamente 1000 partículas únicas por microlitro que corresponde a aproximadamente 0,01 % p/v. El diluyente contiene tampón, sal, y tensioactivo. Si hay 10 marcadores en el panel, los sólidos totales serían aproximadamente 10.000 partículas por microlitro o aproximadamente 0,1 % de sólidos p/v. Los conjugados se agruparon y se ajustaron a

una concentración final de aproximadamente 1 a 10 nM de cada en el diluyente de micropartículas. Para llevar a cabo el ensayo, una alícuota de la muestra sospechosa de contener uno o más de los analitos se colocó en un pocillo de incubación seguido por un volumen a la mitad de micropartículas agrupadas. La suspensión se incubó durante 30 minutos seguida por la adición de una mitad de volumen de solución de conjugado agrupado. Tras una incubación adicional de 30 minutos, se diluyó la reacción por adición de dos volúmenes de solución tamponada que contenía una sal y un tensioactivo. La suspensión se mezcla y se aspira un volumen de aproximadamente el doble que el de la muestra por el instrumento Luminex 100™ para el análisis. Opcionalmente, las micropartículas se pueden lavar tras cada incubación y luego resuspenderse para análisis. Se midió la fluorescencia de cada partícula individual con 3 longitudes de onda; dos se utilizaron para identificar la partícula y su analito asociado y la tercera se utiliza para cuantificar la cantidad de analito unido a la partícula. Se midieron 100 micropartículas de cada tipo y se calculó la fluorescencia media para cada analito. Opcionalmente, 50 partículas o menos de cada tipo se pueden medir con una pequeña pérdida de precisión y sensibilidad concomitante. La cantidad de analito en la muestra se calcula por comparación con una curva de referencia que se genera llevando a cabo el mismo análisis en una serie de muestras que contenían cantidades conocidas del péptido o proteína y representando la fluorescencia media de las muestras conocidas contra la concentración conocida. Una muestra desconocida se clasificaba como que era de cáncer o no cancerosa basándose en la concentración de analito (si se elevaba o deprimía) con respecto a especímenes conocidos de cáncer o no cancerosos utilizando modelos tales como el método de división y puntuación o división y método de puntuación ponderada como se describe en el Ejemplo 7. Opcionalmente, se pueden utilizar otros métodos como el Múltiple y Medio, descrito en el Ejemplo 12, como método de puntuación para un panel de analitos.

Por ejemplo, un paciente se puede ensayar para determinar la probabilidad del paciente detener un cáncer de pulmón utilizando el panel de 8 inmunoensayos (IA) de la Tabla 18 y el método de división y puntuación. Después de obtener una muestra de ensayo del paciente, se cuantifica la cantidad de cada uno de los 8 biomarcadores en la muestra del paciente (es decir, el suero) y la cantidad de cada uno de los biomarcadores se compara entonces con el correspondiente punto de división predeterminado (punto de corte predeterminado) para el biomarcador, tal como los enumerados en la Tabla 18 (es decir, el punto de corte predeterminado que se puede utilizar para la Citoqueratina 19 es 1,89 o 2,9). Para cada biomarcador que tenga una cantidad que sea mayor que su correspondiente punto de división predeterminado (punto de corte predeterminado), se puede dar una puntuación de 1. Para cada biomarcador que tenga una cantidad que sea menor o igual a su correspondiente punto de división predeterminado (punto de corte predeterminado), se puede dar una puntuación de 0. La puntuación para cada uno de los 8 biomarcadores se combinaron entonces matemáticamente (es decir, sumando cada uno de las puntuaciones de los biomarcadores en conjunto) para llegar a una puntuación total para el paciente. Esta puntuación total se convierte en la puntuación del panel. La puntuación del panel se compara con el umbral predeterminado (puntuación total predeterminada) del modelo 8 IA de la Tabla 25a, a saber 1. Una puntuación del panel mayor de 1 sería un resultado positivo para el paciente. Una puntuación del panel menor o igual a 1 sería un resultado negativo para el paciente. En la población de estudio descrito en el Ejemplo 9, este panel ha demostrado una especificidad del 30 %, una tasa de falsos positivos del 70 % y una sensibilidad del 90 %. Un resultado positivo del panel para el paciente tiene un 70 % de oportunidades de ser un falso positivo. Además, el 90 % de los pacientes con cáncer de pulmón tendrán un resultado del panel positivo. Por lo tanto, el paciente que tiene un resultado del panel positivo puede remitirse para posteriores ensayos de una indicación o sospecha de cáncer de pulmón.

Por medio de un ejemplo más, utilizando de nuevo el panel 8 IA y el Método de Puntuación Ponderada, tras la obtención de una muestra de ensayo de un paciente. La cantidad de cada uno de los 8 biomarcadores en la muestra de ensayo del paciente (es decir, suero) se cuantifica y la cantidad de cada uno de los biomarcadores se compara entonces con puntos de división predeterminados (puntos de corte predeterminados) tales como los puntos de división (puntos de corte) enumerados en la Tabla 17b (es decir, los puntos de corte predeterminados que se pueden utilizar para la Citoqueratina 19 son 1,2, 1,9 y 3,3). En este ejemplo, cada biomarcador tiene tres puntos de división predeterminados (puntos de corte predeterminados). Por lo tanto, hay 4 puntuaciones posibles que se pueden dar para cada biomarcador. La puntuación para cada uno de los 8 biomarcadores se combinan entonces matemáticamente (es decir, sumando cada una de las puntuaciones de los biomarcadores en conjunto) para llegar a la puntuación total del paciente. La puntuación total se convierte entonces en la puntuación del panel. La puntuación del panel se puede comparar con el umbral predeterminado (o puntuación total predeterminada) para el modelo 8 IA, que se había calculado que era 11,2. Una puntuación del panel del paciente mayor de 11,2 sería un resultado positivo. Una puntuación del panel del paciente menor o igual a 11,2 sería un resultado negativo. En la población de estudio descrita en el Ejemplo 9, este panel demostró una especificidad del 34 %, una tasa de falsos positivos del 66 % y una sensibilidad del 90 %. El resultado positivo del panel tenía una probabilidad del 66 % de ser un falso positivo. Además, el 90 % de los pacientes con cáncer de pulmón tienen un resultado positivo del panel. Por tanto, un paciente que tiene un resultado del panel positivo se puede remitir para posteriores ensayos para una indicación o sospecha de cáncer de pulmón.

C. Inmunoensayo multiplexado para autoanticuerpos para el paciente: Si el biomarcador o biomarcadores que se utilizan para hacer un diagnóstico clínico son autoanticuerpos circulantes, se puede utilizar una estrategia similar a la descrita anteriormente en los Ejemplos 11A y 11B. Si se va a utilizar un único autoanticuerpo o un número pequeño de autoanticuerpos, se puede llevar a cabo el ensayo en un analizador automático tal como el sistema de inmunoensayo Architect™ disponible en Abbott Diagnostics, Abbott Park, Illinois. Se determina la positividad o

negatividad del suero del paciente para el biomarcador por comparación de la señal generada por el suero de ensayo con un suero de referencia o control y el paciente se determina positivo si el valor excede un punto de corte. Si se necesitan múltiples autoanticuerpos para hacer una determinación, se utilizan el analizador Luminex 100™, o instrumentos relacionados, disponible en Luminex Corporation, Austin, Texas. Las micropartículas agrupadas revestidas con los autoantígenos elegidos se describen en el Ejemplo 12, posteriormente, se exponen a una alícuota diluida del suero del paciente (o plasma). Tras la primera incubación, las partículas se lavaron y se expusieron al reactivo de detección que comprende un anticuerpo contra IgG humana marcado con un marcador detectable. Después de la segunda incubación, las partículas se lavan de nuevo y el reactivo detectable se detecta con el uso de un marcador fluorescente; si el marcador detectable es biotina. El marcador detectable puede ser o antibiotina, avidina o estreptavidina unida a R-ficoeritrina. A continuación de la tercera incubación y lavado (o aspiración), las partículas se resuspendieron en un tampón y se introduce en el analizador Luminex 100™. Las partículas de control se pueden llevar a través de un proceso para ensayar la unión no específica (control negativo) y la unión específica (control positivo). De la misma manera, se puede incluir un control de procedimiento, por ejemplo una partícula revestida con anti-IgG humana. Por comparación de los valores obtenidos de la muestra del paciente con muestras negativas, se puede determinar si un biomarcador es positivo o negativo. Cuando se miden múltiples autoanticuerpos, cada biomarcador se mide por turno y los resultados se combinan para llegar a una evaluación total de positividad o negatividad del paciente. Los resultados múltiples se pueden combinar óptimamente utilizando una herramienta de análisis de datos como el método de división y puntuación del Ejemplo 11B, anteriormente, o el Múltiple o Medio (MOM) descrito en el Ejemplo 12.

Si se desea que se procesen ambos antígenos y autoanticuerpos en cada paciente y las concentraciones de los antígenos son lo suficientemente altas para permitir una dilución comparable a la que se utiliza en la medición de autoanticuerpos, todas las micropartículas se pueden combinar en un reactivo y el ensayo se ejecuta en una alícuota del suero o plasma del paciente. La única modificación con el protocolo de autoanticuerpos es que además de del reactivo anti-IgG humana, se deben incluir anticuerpos marcados detectablemente contra cada analito antígeno en el segundo reactivo. Todos se pueden detectar con un reactivo de r-ficoeritrina en la última etapa. Los ensayos de antígeno pueden necesitar una alta sensibilidad, y por lo tanto, la muestra no se puede diluir apropiadamente para la detección de autoanticuerpos. En este caso, los antígenos se miden en una muestra de suero diluida mínimamente y los autoanticuerpos se miden en una muestra de suero altamente diluida. Aunque se dobla el número de determinaciones, ambos tipos de mediciones se pueden llevar a cabo fácilmente en la familia de instrumentos Luminex. De manera alternativa, las mediciones se pueden hacer en otros citómetros de flujo, tales como los disponibles en BD Biosciences, San Jose, California. El cambio del formato a una matriz, similar a la ProtoArray de Invitrogen, Carlsbad, California, elimina la conveniencia de preparar reactivos de micropartículas individualmente, pero es técnicamente factible.

D. Análisis inmuno espectrométrico de masas: La preparación de las muestras para la espectrometría de masas puede utilizar también métodos inmunológicos así como métodos cromatográficos o electroforéticos. Se ajustan las micropartículas superparamagnéticas revestidas con anticuerpos específicos para un péptido biomarcador a una concentración de aproximadamente 0,1 % p/v en una solución tampón que contenía sal. Una alícuota de muestra de suero del paciente se mezcla con un volumen igual de micropartículas revestidas con anticuerpo y dos veces ese volumen de diluyente. Después de una incubación, las micropartículas se lavan con tampón de lavado que contenía una sal tamponada y, opcionalmente, sal y tensioactivos. Las micropartículas se lavaron entonces con agua desionizada. El analito inmunopurificado se eluyó de las micropartículas añadiendo un volumen de acetonitrilo acuoso que contenía ácido trifluoroacético. La muestra se mezcla entonces con un volumen igual de solución matriz de ácido sinapínico y un pequeño volumen (aproximadamente 1 a 3 microlitros) se aplica a un análisis de masas MALDI diana para tiempo de vuelo. La corriente iónica a la m/z deseada se compara con la corriente iónica derivada de una muestra que contenía una cantidad conocida del péptido biomarcador que se ha procesado de manera idéntica.

Se debería señalar que la corriente iónica se relaciona directamente con la concentración y la corriente iónica (o intensidad) a un valor m/z particular (o ROI) se puede convertir a concentración si se desea. Tales concentraciones o intensidades se pueden utilizar entonces como entrada en cualquiera de los modelos que construyen los algoritmos descritos en el Ejemplo 7.

E. Espectrometría de masas para ROI: Se obtiene una muestra de sangre de un paciente y se deja coagular para formar una muestra de suero. La muestra se prepara para el análisis espectrométrico de masas SELDI y se cargó en un Chip proteico en un Bioprocesador y se trata como se proporciona en el Ejemplo 2. El Chip Proteico se carga en un espectrómetro de masas MALDI con tiempo de vuelo Ciphergen 4000 y se analiza como en el Ejemplo 3. Cada espectro se ensaya en cuanto a su aceptación utilizando un análisis multivariado. Por ejemplo, se calculan la corriente total y el ángulo de contraste espectral (entre la muestra desconocida y una población de referencia conocida). Se determina entonces la distancia Mahalanobis. Para el espectro en el que la distancia Mahalanobis es menor que el valor crítico establecido, el espectro se cualifica. Para el espectro en el que la distancia de Mahalanobis mayor que el valor crítico establecido, el espectro se excluye de posteriores análisis y la muestra se re-procesaría. Tras la cualificación, el espectro de masa se normaliza.

El espectro de masa resultante se evaluó para medir la corriente iónica en regiones de interés apropiadas para el modelo de análisis de datos que se elija. Basándose en el resultado del análisis, el paciente se valora de estar en riesgo de o tiene una alta probabilidad de tener un cáncer de pulmón y debería someterse procedimientos diagnósticos adicionales.

Para el uso del Método de División y Puntuación, las intensidades en los ROI a los valores m/z dados en la Tabla 5 se miden para el paciente. El resultado del paciente se puntúa para señalar si los valores del paciente son en el lado del cáncer o en el lado del no canceroso de los valores del punto de división promedio que se dan en la Tabla 7. Una puntuación de 1 se da para cada valor de ROI que se encuentra en el lado del cáncer del punto de división. Las puntuaciones de 3 y mayores indican que el paciente tiene un riesgo elevado de cáncer y debería remitirse para procedimientos diagnósticos adicionales.

Ejemplo 12. Método del Múltiplo de Mediana

El método del múltiplo de mediana (MoM) de integración de datos permite puntuar un panel de biomarcadores. En el método del múltiplo de mediana, se determina el valor de la mediana de cada biomarcador en una cohorte normal. El valor de la mediana se utiliza para normalizar todas las mediciones de ese biomarcador específico. El MoM se ha utilizado para analizar tanto anticuerpos del suero como niveles de antígeno en el suero. W. Kutteh, et al. (Obstet. Gynecol. 84:811-815 (1994)) utilizaba el MoM para comunicar los niveles de anticuerpos anti-fosfolípidos en pacientes. Adicionalmente, G. Palomaki et al., (Clin. Chem. Lab. Med.) 39:1137-1145 (2001)) ha utilizado el MoM para analizar los niveles de proteínas en el suero. Así, se divide cualquier nivel medido de un biomarcador por la mediana del grupo normal, resultando en un múltiplo de la mediana. Esos múltiplos de los valores medios se combinan (a saber, se suman o añaden) para cada biomarcador en el panel que resulta en un panel de valores MoM o puntuación MoM para cada muestra. Con fines clínicos, se puede determinar un punto de corte para el panel del valor MoM que indica la presencia o riesgo de enfermedad.

Una característica del método MoM es la retención de la distribución de los datos de la población. Como ejemplo, se midieron siete anticuerpos en 1059 individuos (291 de cáncer de pulmón, 387 de enfermedad pulmonar benigna, y 381 normales). Los valores medios de las mediciones de los 7 autoanticuerpos se determinaron para la cohorte normal y todas las mediciones se convirtieron a un múltiplo del valor de la mediana. Los siete múltiplos de los valores de la mediana se sumaron para los 1059 individuos. Estos 1059 valores del panel MoM se sometieron a una curva de la característica operativa del receptor (curva ROC). La curva resultante se muestra en la **Figura 6**. También se representa en la **Figura 6** las curvas resultantes del método de división y puntuación binario utilizando puntos de división determinados para las necesidades de máxima precisión o para el 98 % de especificidad.

Ejemplo 13. Una matriz de autoanticuerpos en perlas

A. Se fijaron proteínas humanas disponibles en el mercado (véase, Tabla 31, posteriormente) a perlas Luminex™ SeroMap™ (Austin, Texas) y los grupos de perlas individuales se combinaron para preparar el reactivo. Partes del reactivo se expusieron a muestras de suero humanas en condiciones que permitían a los anticuerpos presentes unirse a las proteínas. El material no unido se lavó y las perlas se expusieron a un conjugado fluorescente de R-ficoeritrina unido a un anticuerpo que se une específicamente a la IgG humana. Tras el lavado, las perlas se pasaron a través de un instrumento Luminex™ 100, que identificaba cada perla según sus colorantes internos, y medía la fluorescencia unida a la perla, que corresponde con la cantidad de anticuerpo unido a la perla. De esta manera, se evaluaron las respuestas inmunitarias de 772 muestras (251 de cáncer de pulmón, 244 normales, 277 benignas) contra 21 proteínas humanas, así como varias proteínas no humanas como controles (seroalbúmina bovina (BSA) y toxina tetánica).

B. Revestimiento de perlas Luminex SeroMap™ con antígenos

Las perlas para este estudio se revistieron como se ha indicado para el método directo descrito en el Ejemplo 3B, el método de preactivación descrito en el Ejemplo 4E, o por adsorción pasiva descrita aquí:

A 100 ul de perlas Luminex SeroMap se les añadió un volumen apropiado correspondiente con 2,5 ug de la proteína para revestirlas, y la mezcla se removió, y se dejó a temperatura ambiente en oscuridad. Tras 90 min las mezclas se colocaron a 37 °C durante 60 minutos, luego se añadió 1 ml de PBN/Tween y las mezclas se removieron y se centrifugaron. Los aglomerados se resuspendieron en 1 ml de PBN/Tween y se sonicaron las muestras. Las perlas se centrifugaron y los aglomerados se resuspendieron en 500 ul de PBN/Tween.

C. Ensayo de muestras de suero con perlas revestidas

Se prepararon las muestras de suero en microplacas a una dilución 1:30 en PBN, con 80 muestras por microplaca. Protocolo: Se añadieron 100 ul de PBN/Tween a los pocillos de una placa filtro 1,2 u Supor. Después de aproximadamente 5 minutos, se aplicó vacío. Cuando los pocillos estaban vacíos, se secó el fondo de la placa con papel y se añadieron 50 ul de 1 x Master Mix por pocillo. A cada pocillo se añadieron 20 ul de la muestra 1:30 de la placa de muestra y luego se incubó durante 60 minutos en un agitador a máxima agitación. Se aplicó vacío para retirar todo el líquido. Se añadieron 100 ul de PBN/Tween y se aplicó vacío para retirar todo el líquido. Se añadieron 100 ul de PBN/Tween, se colocó en un agitador durante 10 minutos y se aplicó vacío para retirar todo el líquido. Se añadieron 100 ul de conjugado 1:500 (20,8 ul de RPE anti-IgG humana en 10,4 ml de PBN/Tween) a cada pocillo, se agitó con pipetas, se colocó en el agitador durante 30 minutos y se aplicó vacío para retirar todo el líquido. Se añadieron 100 ul de PBN/Tween, se agitó con pipetas, se transfirió a una microplaca y se procesó en el Luminex.

D. Análisis de paneles de autoanticuerpos

a. Análisis ROC

- 5 El resultado bruto registrado para cada autoantígeno ensayado con muestras individuales era el valor medio. Estos resultados se sometieron a un análisis de la Características Operativa del Receptor (ROC) y se calculó el Área Bajo la Curva (AUC) para discriminar los grupos de cáncer de los no cancerosos. La Tabla 31 a continuación, presenta los valores de AUC para los autoantígenos individuales.

10

Tabla 31.

Perla	Prep	Revestimiento	Nombre	Fuente	AUC
43	10/03/06	Preactivado	CCNE2	Abnova	0,60
1	10/04/06	Preactivado	TMP21-ECD	Abbott	0,64
5	10/04/06	Preactivado	Domino NPC1L1C	Abbott	0,64
14	10/04/06	Preactivado	PSEN2 (1-86aa) his	Abbott	0,60
55	10/05/06	Preactivado	CAMK-1	Invitrogen	0,61
49	10/18/06	Directo	CCNE2	Genway	0,70
61	12/05/06	Preactivado	tropomiosina 1	Invitrogen	0,60
62	12/05/06	Preactivado,	PDZK1	Invitrogen	0,57
63	12/05/06	Preactivado	ACRV1	Invitrogen	0,59
64	12/05/06	Preactivado	DNAJB1	Invitrogen	0,58
65	12/05/06	Preactivado	NRBF2	Invitrogen	0,59
66	12/05/06	Preactivado	C21orf7	Invitrogen	0,56
67	12/05/06	Preactivado	MAPKAPK3	Invitrogen	0,58
68	12/05/06	Preactivado	PKCKS	Invitrogen	0,58
69	12/05/06	Preactivado	SMAD1	Invitrogen	0,57
70	12/05/06	Preactivado	DDAH2	Invitrogen	0,55
71	12/05/06	Preactivado	UROS	Invitrogen	0,59
72	12/05/06	Preactivado	CDC44	Invitrogen	0,54
73	12/05/06	Preactivado	IITP	Invitrogen	0,54
74	12/05/06	Preactivado	LRRIQ2	Invitrogen	0,56
75	12/05/06	Preactivado	SCF7m6os	Invitrogen	0,60
76	12/05/06	Preactivado	SH3PX3	Invitrogen	0,57
77	12/05/06	Preactivado	TFPT	Invitrogen	0,58
79	12/05/06	Preactivado	RPE	Invitrogen	0,55
56	12/11/06	Preactivado	WW-E3	Invitrogen	0,60
57	12/11/06	Preactivado	C1orf117	Invitrogen	0,62
58	12/11/06	Preactivado	C9orf11	Invitrogen	0,60
59	12/11/06	Preactivado	paralemmi	Invitrogen	0,61
60	12/11/06	Preactivado	CDC37-1	Invitrogen	0,60
78	12/11/06	Preactivado	TFE3	Invitrogen	0,56
80	12/11/06	Preactivado	CCNE2	Invitrogen	0,58
26	02/13/07	Pasivo	p53	Biomol	0,69
27	02/13/07	Pasivo	MAPKAPK3	Genway	0,70
30	02/13/07	Directo	BM-E2-1	Abbott	0,73
31	02/13/07	Pasivo	Aldolasa	Sigma	0,62
36	02/14/07	Preactivado	TMOD-1	Lab Vision	0,67
32	02/15/07	Directo	NY-ESO-1	NeoMarkers	0,66
33	02/15/07	Directo	HDJ1	Alexis	0,64
23	02/19/07	Directo	IgG	Abbott	0,53
38	02/19/07	Directo	BSA	Sigma	0,60
40	02/19/07	Directo	H. Pylori	Fitzgerald	0,81

b. Método de puntuación

- 15 Se utilizó un método de puntuación binario para evaluar paneles múltiples de biomarcadores en la discriminación de grupos de cáncer de no cancerosos. La Tabla 32 a continuación, enumera los puntos de división (puntos de corte) utilizado para cada biomarcador. La Tabla 33 posteriormente identifica los marcadores individuales en cada panel y el AUC para cada panel.

Tabla 32.

Biomarcador	Punto de división (máx. precisión)	Por encima del punto de división
p53	84	Cáncer
NY-ESO-1	100	Cáncer
cíclico c2-1 pcp	653	Cáncer
Mapkapk	54,5	Cáncer
Cyfra	1,89	Cáncer
Cea	4,82	Cáncer
ca125	22,4	Cáncer
Scc	1,1	Cáncer
Tps	55,5	Cáncer
Progrp	11,8	Cáncer
4789	193,4	Cáncer
11597	34,4	Cáncer
2759	575,5	no canceroso
acn9459	287,3	no canceroso

c. Múltiplo de la Mediana

- 5 Los datos brutos para cada biomarcador se convirtieron en un múltiplo del valor de la mediana dividido por el valor de la mediana del grupo "normal". Para cada valor, los múltiplos de los valores de la mediana se sumaban para proporcionar un panel de múltiplos del valor de la mediana para cada muestra. Estos paneles de múltiplos de los valores de la mediana se sometieron a un análisis de Característica Operativa del Receptor y el AUC se expone en las tablas 33a y 33b a continuación.

10

		Aac/IA/MS														
	Marcadores	4/6/4	4/4/2	4/4/1	4/2/1	4/0/4	4/4/0	4/6/0	4/2/0	0/4/1	0/6/0	0/4/0	4/0/0	8/0/0	0/0/4	
8 auto antígenos	ciclina e2-l	x	x	x	x	x	x	x	x				x	x		
	mapkapk3	x	x	x	x	x	x	x	x				x	x		
	p53	x	x	x	x	x	x	x	x				x	x		
	tmod-1													x		
	nyeso	x	x	x	x	x	x	x	x				x	x		
	hdjl													x		
	npelllc													x		
	tmp21													x		
6 inmunoensayos	ck19	x	x	x	x		x	x	x	x	x	x				
	oca	x	x	x	x		x	x	x	x	x	x				
	cal25	x	x	x			x	x		x	x	x				
	scc	x						x			x					
	progrp	x	x	x			x	x		x	x	x				
	tps	x						x			x					
4 espec. de masas	pub11597	x	x	x	x	x				x					x	
	pub4789	x				x									x	
	lfa2759	x				x									x	
	acn9459	x	x			x									x	
AUC	Puntuación binaria	0,88	0,88	0,87	0,85	0,85	0,84	0,83	0,81	0,81	0,77	0,77	0,77	0,75	0,78	
	MoM	0,88	0,88	0,88	0,86	0,82	0,88	0,88	0,85	0,83	0,81	0,82	0,81	0,81	0,76	

Tabla 33a.

Análisis de dejar uno fuera en un panel de 14 marcadores utilizando datos MoM													
		Aac/IA/MS											
Marcadores		2/1/0	2/2/0	3/2/0	3/3/0	3/4/0	4/4/0	4/4/1	4/4/2	4/4/3	4/5/3	4/5/4	4/6/4
8 autoantígenos	Ciclina e2-1	x	x	x	x	x	x	x	x	x	x	x	x
	mapkapk3						x	x	x	x	x	x	x
	p53	x	x	x	x	x	x	x	x	x	x	x	x
	tmod-1												
	NY-ESO-1			x	x	x	x	x	x	x	x	x	x
	hdjl												
	npclllc												
	Imp21												
6 Inmunoensayos	ckl9					x	x	x	x	x	x	x	x
	Cea		x	x	x	x	x	x	x	x	x	x	x
	Ca125				x	x	x	x	x	x	x	x	x
	Sccl												x
	Progrp	x	x	x	x	x	x	x	x	x	x	x	x
	Tps										x	x	x
Espec. masas	Pub11597											x	x
	pub4789								x	x	x	x	x
	tra2759							x	x	x	x	x	x
	acn9459									x	x	x	x
AUC	Puntuación binaria	0,76	0,79	0,82	0,83	0,83	0,84	0,86	0,87	0,88	0,88	0,88	0,88
	MoM	0,82	0,84	0,86	0,87	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88

Tabla 33b.

Como se puede ver al comparar los paneles de las Tablas 33a y 33b anteriores con solo los autoantígenos como biomarcadores, el panel 4/0/0 que incluye NY-ESO-1, p53, mapkapk3, y péptido ciclina E2-1 era tan bueno como el panel 8/0/0 que incluye cuatro autoantígenos más. Además, los valores del AUC más altos se obtenían cuando se incluían al menos tres autoantígenos en el panel, a saber, NY-ESO-1, p53, y péptido ciclina E2-1.

En el análisis dejar uno fuera (Tabla 33b anterior), se alcanzaba el mismo AUC tanto en los métodos del Múltiplo de la Media (MoM) y el de puntuación binaria. Sin embargo, un panel mayor de biomarcadores era necesario con el método de puntuación binaria al compararse con el método del análisis MoM. Este método Múltiplo de la Media determinó que el panel 3/4/0 que incluye autoanticuerpos contra NY-ESO-1, p53, y ciclina E2-1 y antígenos tumorales CK19, CEA, CA125 y proGRP proporcionaba el AUC más alta que se conseguía con menor biomarcadores. Añadiendo biomarcadores adicionales, al menos en este Ejemplo, no cambiaba el AUC. Sin embargo, utilizando el método de puntuación binario, el panel con el AUC más alto y menos biomarcadores era el 4/4/3 que incluye autoanticuerpos contra NY-ESO-1, p53, ciclina E2-1, y mapkapk3 y antígenos tumorales CK-19, CEA, CA125, y proGRP y biomarcadores de espectrometría de masas PUB4789, Tfa2759, y ACN9459.

El panel 4/4/0 (Tablas 33a y 33b anteriores) se ensayó en el grupo de fumadores empedernidos (≥ 20 packyrs), poco fumadores y grupo de no fumadores (<20 packyrs) y el grupo con historia de packyrs desconocida. Los pacientes individuales se representaron solo en un grupo (Tabla 34). El panel actuaba de forma similar en la distinción cánceres de no cancerosos tanto si todos los individuos eran fumadores empedernidos, si todos los individuos eran poco o no fumadores o si las historias de fumadores eran desconocidas (véase la Figura 7). El ensayo actuaba igualmente bien en los tres grupos demostrando que el resultado era independiente de la historia de fumador. La capacidad de distinguir los pacientes de cáncer de los no cancerosos en una cohorte de fumadores empedernidos era de 0,84, en una cohorte de poco o no fumadores era del 0,85 y en una cohorte de historia de fumador desconocida era de 0,88.

Tabla 34

	>20 py	<20 py	py desconocido
cáncer	127	45	92
Temprano	85	25	46
Tardío	42	20	46
no canceroso	118	89	305
benigno	109	85	76

Los pacientes individuales de cáncer se agruparon según el estadio del cáncer de pulmón al diagnóstico. El estadio temprano de cáncer de pulmón se definió con los estadios I y II. El cáncer de pulmón en estadio tardío se definió como los estadios III y IV. El panel 4/4/0 se ensayó por su capacidad para diferenciar el cáncer del no canceroso utilizando o una cohorte de enfermedad pulmonar benigna como control o la cohorte normal como control. Los pacientes individuales se representaron en un solo grupo. La Tabla 35, a continuación identifica el número de pacientes en cada grupo y el AUC para cada comparación.

			Cáncer	
			Temprano	Tardío
			(156)	(108)
no canceroso	normal	(241)	0,869	0,927
	benigno	(270)	0,814	0,894

Tabla 35.

Se ha hecho una mejora significativa en la diferenciación de cáncer de pulmón en estadio temprano y enfermedad pulmonar benigna utilizando el panel 4/4/0 y otros compuestos de manera similar. En la Figura 8 y la Tabla 36a, cada biomarcador individual se ensayó en cuanto a su capacidad para diferenciar las cohortes de cáncer de pulmón en estadio temprano y tardío de las normales y de enfermedad pulmonar benigna. Está claro que los antígenos tumorales (CK 19, CA19-9, CEA, CA 15-3, CA 125, SCC, CK18, y proGRP) demostraban todos una dependencia significativa del estadio, en que la enfermedad en estadio tardío se diferencia de normales y benignos con alguna competencia (AUC ~ 0,7 – 0,8), sin embargo, la diferenciación de cáncer de pulmón en estado temprano de los normales y benignos es menos competente (AUC ~ 0,5 – 0,7).

Los marcadores de respuesta inmunitaria incluyendo la haptoglobina, amiloide sérico A, proteína reactiva C, y apolipoproteína CIII demostraron una dependencia del estadio similar a los antígenos tumorales (Figura 8).

Los autoanticuerpos como marcadores individuales son mucho menos dependientes del estadio. Los cánceres de pulmón en estadio temprano y tardío se diferencian de los normales y benignos con casi una competencia equivalente. La mejor actuación de los autoantígenos incluyen a los p53, mapkapk3, ciclina E2-1, NY-ESO-1, HDJ1 y tmod, que tienen AUC que varían de 0,6 – 0,75.

Los marcadores MS (4789, 11597, 2759, y ACN9459, cada uno de los cuales es un APR) se comportan de manera similar a los autoanticuerpos en que son menos dependientes del estadio en su comportamiento.

El poder de estas combinaciones queda claro en la **Figura 9** y la Tabla 36b. Combinando los antígenos tumorales con los marcadores de respuesta inmunitaria se proporciona un panel con una dependencia del estadio mejorada y con una persistente incapacidad de diferenciar el estadio temprano del cáncer de pulmón de la enfermedad pulmonar benigna (AUC ~0,65-0,68). El cáncer de pulmón en estadio temprano se diferencia de la cohorte normal con un AUC ~ 0,75 - 0,8.

Combinando los biomarcadores autoantígenos con los marcadores de respuesta inmunitaria se proporcionaba un panel con una mayor consistencia total en la diferenciación del cáncer del no canceroso incluyendo cáncer de pulmón de estadio temprano y tardío de la normalidad y de la enfermedad pulmonar benigna (Figura 9). El cáncer de pulmón en estadio temprano se diferencia competentemente de los normales y enfermedad pulmonar benigna con un AUC de ~ 0,75 - 0,82. El AUC para la enfermedad en estado tardío y normal y enfermedad pulmonar benigna es similar (0,75 - 0,85) y no tan buena como cuando se utilizan antígenos tumorales.

La combinación de antígenos tumorales y autoanticuerpos proporciona un buen panel. El panel 4/4/0 descrito como 4 IA + 4 Aac de la Figura 9 diferencia cáncer de pulmón en estadio temprano de enfermedad pulmonar benigna con un AUC ~ 0,82. Además, combinando marcadores de respuesta inmunitaria (proteína reactiva C, amiloide sérico A, o haptoglobina) tiene una poca mejora. Además, la adición de 3 marcadores de respuesta inmunitaria al panel 4/4/0 en realidad reduce la actuación del panel.

5

	Grupo de caso	Cáncer de pulmón en estadio temprano		Cáncer de pulmón en estadio tardío	
	<i>Grupo control</i>	<i>benigno</i>	<i>Normal</i>	<i>benigno</i>	<i>normal</i>
Biomarcadores					
Biomarcadores de respuesta inmunitaria	haptoglobina	0,56	0,65	0,69	0,76
	SAA	0,47	0,63	0,72	0,74
	CRP	0,56	0,64	0,73	0,8
	apoCIII	0,6	0,52	0,67	0,58

Biomarcadores de espect. de masas	4789	0,65	0,65	0,75	0,75
	11597	0,75	0,66	0,83	0,76
	2759	0,66	0,63	0,66	0,64
	acn9459	0,64	0,69	0,64	0,69
Autoanticuerpos	Imp21	0,6	0,64	0,59	0,63
	npc11lc	0,59	0,58	0,61	0,6
	p53	0,63	0,67	0,69	0,73
	mapkapk	0,68	0,7	0,71	0,74
	E2-1	0,69	0,72	0,72	0,75
	NYESO	0,62	0,62	0,68	0,68
	hdjl	0,65	0,61	0,68	0,65
	tmod	0,61	0,64	0,71	0,73
	E2 Abnova	0,58	0,58	0,61	0,61
	E2 Genway	0,65	0,65	0,72	0,74
antígenos tumorales	CYFRA	0,6	0,63	0,76	0,79
	19-9	0,56	0,51	0,65	0,69
	CEA	0,61	0,7	0,72	0,78
	15-3	0,56	0,62	0,63	0,69
	CA125	0,61	0,68	0,78	0,83
	SCC	0,55	0,57	0,61	0,63
	TPS	0,6	0,64	0,72	0,75
	proGRP	0,55	0,69	0,59	0,69

Tabla 36a.

Tabla 36b.

Grupo de caso	Cáncer de pulmón en estadio temprano		Cáncer de pulmón en estadio tardío		Todos los cánceres de pulmón	
<i>Grupo control</i>	<i>benigno</i>	<i>Normal</i>	<i>benigno</i>	<i>normal</i>	<i>benigno + normal</i>	
	AUC	AUC	AUC	AUC	DFI AUC	
4IA	0,668	0,8	0,886	0,942	0,39	0,8
4IA+CRP	0,66	0,792	0,875	0,938	0,38	0,8
4IA+SAA	0,681	0,769	0,859	0,9		0,81
4IA+hapto	0,672	0,81	0,894	0,955	0,37	0,78
4IA+3APR	0,666	0,772	0,855	0,904		0,78

Grupo de caso	Cáncer de pulmón en estadio temprano		Cáncer de pulmón en estadio tardío		Todos los cánceres de pulmón	
<i>Grupo control</i>	<i>benigno</i>	<i>Normal</i>	<i>benigno</i>	<i>normal</i>	<i>benigno + normal</i>	
	AUC	AUC	AUC	AUC	DFI AUC	
4Aac	0,793	0,83	0,781	0,808	0,36	0,8
	AUC	AUC	AUC	AUC	DFI	AUC
4Aab+CRP	0,779	0,833	0,812	0,856	0,34	0,81
4Aab+SAA	0,767	0,8	0,806	0,832		0,81
4Aab+hapto	0,798	0,836	0,789	0,824		0,8
4Aab+3APR	0,767	0,817	0,815	0,85	0,37	0,81
4IA+4Aac	0,816	0,87	0,893	0,93	0,28	0,87
4IA+4Aah+CRP	0,803	0,87	0,885	0,923	0,3	0,86
4IA+4Aab+SAA	0,786	0,832	0,866	0,895	0,28	0,87
4IA+4Aab+hapto	0,818	0,877	0,892	0,929	0,33	0,84
4IA+4Aab+3APR	0,782	0,841	0,865	0,9	0,34	0,84

(Las Tablas 36a y 36b tienen los valores del AUC tabulados que se muestran en las Figuras 8 y 9)

Ejemplo 14. Timosina $\beta 4$ EIA (T $\beta 4$)

El kit timosina $\beta 4$ EIA se adquirió en ALPCO Diagnostics, (Salem, NH). Los ensayos se ejecutaron según las instrucciones del fabricante. Las muestras de suero se utilizaron directamente sin diluirlas. Se procesó el grupo seleccionado de 118 muestras que contenían muestras normales, benignas y de pacientes de cáncer. Las concentraciones de los analitos se derivaron del ajuste de la curva logística de 4 parámetros (4PLC) de concentraciones de referencia conocidas como instruye el protocolo del fabricante. Todos los cálculos se llevaron a cabo utilizando el algoritmo de ajuste de la curva no lineal proporcionada en el paquete estadístico JMP 5.0. El AUC que se obtiene por el ensayo T $\beta 4$ se muestra en la **Figura 10**.

Como se ha mostrado anteriormente, la T $\beta 4$ por sí misma era capaz de discriminar entre los grupos enfermos y no enfermos relativamente bien. Sin embargo, a alta especificidad, su sensibilidad es relativamente poca como se muestra en la Tabla 37, posteriormente (Especificidad = 90 %, Sensibilidad = 53 %). Aquí, se evaluó la mejora de actuación de algunos paneles multimarcaador preferidos cuando estaba incluida la T $\beta 4$ en los paneles. Se utilizaron tres grupos diferentes de paneles de marcadores (4IA (CK19, CEA, CA125 y ProGRP), 4Aab (Ciclina E2-1, MAPKAPK3, P53 y NY-ESO-1) y 3APR (Amiloide sérico A (Pub11597), proteína reactiva C, y haptoglobina (sospechoso de ser el 17858 de masas)). La tabla 37 a continuación compara la actuación de los paneles con y sin T $\beta 4$. También se compararon el AUC y las sensibilidades con dos especificidades diferentes (Hspec y Lspec). La Hspec se fijó al 90 % y Lspec al 70 % con el fin de comparar la actuación de los paneles a alta y baja especificidad respectivamente.

Tabla 37. Actuación de paneles seleccionados con y sin T $\beta 4$. La Hspec se fijó al 90 % y Lspec al 70 %

Paneles	AUC	Sens@Hspec	Sens@Lspec
Tp4	0,84	0,53	0,83
4Aac	0,9	0,8	0,9
4Aab + T $\beta 4$	0,9	0,75	0,93
4IA	0,86	0,72	0,78
4IA + T $\beta 4$	0,92	0,83	0,93
4APR	0,64	0,38	0,5
4APR+T $\beta 4$	0,74	0,35	0,7

Como se puede ver en la Tabla 37 anterior, el panel mejor sin T $\beta 4$ es 4Aac con un AUC de 0,9. La sensibilidad de este panel es aproximadamente 80 % y 90 % a alta y baja especificidad respectivamente. Sin embargo, cuando se añadía T $\beta 4$ a este panel, la actuación se deterioraba ligeramente a alta especificidad (Hspec) ya que la sensibilidad cae del 80 % al 75 %. Como el AUC permanece igual, sugiere que el panel es capaz de compensar la pérdida, por ganancia de alguna especificidad. Esto sugiere que la T $\beta 4$ tiene un efecto sinérgico en la actuación de este panel. Como se muestra en la Tabla 37 anterior, el AUC mejora de 0,86 a 0,92 y la sensibilidad a alta especificidad mejora aproximadamente un 10 %. A baja especificidad, la mejora es todavía más alta (ca. 15 %). Sin embargo, aunque el panel 3APR muestra algo de mejora cuando se incluye T $\beta 4$, puede no ser suficientemente robusto para utilizarse

como ensayo clínico para discriminar entre estados de enfermedad y no enfermedad. La sensibilidad del panel es relativamente poca a especificidades tanto alta como baja.

Por tanto, los resultados anteriores sugieren que la Tβ4 se puede utilizar con el panel 4IA ya que mejora el poder de discriminación del panel a especificidades tanto alta como baja. Dicho panel combinado sería entonces más útil que el que se usa tal cual. Las inspecciones más detenidas de los parámetros característicos sugieren que la actuación del panel combinado es adecuada para un ensayo clínico relativamente robusto (AUC = 0,92, a una especificidad = 90 %, Sensibilidad = 83 %). Además, la correlación de análisis entre Tβ4 y los miembros del panel no muestran ninguna correlación, confirmando la independencia de los marcadores.

Un experto en la técnica apreciaría rápidamente que la presente invención está bien adaptada para llevar a cabo los objetivos y obtener los fines y ventajas mencionadas, así como las inherentes al presente documento. Las composiciones, formulaciones, métodos, procedimientos, tratamientos, moléculas, compuestos específicos descritos en el presente documento son actualmente representativas de realizaciones preferidas, son ejemplares, y no se pretenden como limitaciones de la invención. Será fácilmente evidente para un experto en la materia que pueden efectuarse diversas sustituciones y modificaciones a la invención divulgada en el presente documento.

Todas las patentes y publicaciones mencionadas en la especificación son indicativas de los niveles de los expertos en la técnica a la que pertenece la invención.

La presente divulgación incluye los siguientes artículos:

1. Un método para ayudar en el diagnóstico de un sujeto sospechoso de padecer cáncer de pulmón, comprendiendo el método las etapas de:

- a. cuantificar en una muestra de ensayo obtenida de un sujeto, la cantidad de dos o más biomarcadores en un panel, comprendiendo el panel al menos, un anticuerpo y al menos un antígeno;
- b. comparar la cantidad de cada biomarcador cuantificado en el panel con un valor de corte predeterminado para dicho biomarcador y determinar una puntuación para cada biomarcador basándose en dicha comparación;
- c. combinar la puntuación asignada para cada biomarcador cuantificada en la etapa b para obtener una puntuación total para dicho sujeto;
- d. comparar la puntuación total determinada en la etapa c con una puntuación total predeterminada; y
- e. determinar si dicho sujeto tiene riesgo de padecer cáncer de pulmón basándose en la comparación de la puntuación total en la etapa d.

2. El método del artículo 1, donde el panel comprende al menos un anticuerpo que se selecciona de entre el grupo que consiste en: anti-p53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, dominio anti-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSLN1, anti-RCV1, anti-MAPKAPK3 y anti-ciclina E2.

3. El método del artículo 1, donde el panel comprende al menos un antígeno que se selecciona de entre grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA19-9, CA15-3, SCC, proGRP; seroamiloides A, alfa-1-antitripsina, apolipoproteína CIII y timosina β4.

4. El método del artículo 1, donde el panel comprende además al menos una región de interés que se selecciona de entre el grupo que consiste en: Acn6399, Acn9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959.

5. El método del artículo 1, donde el DFI de los biomarcadores con respecto al cáncer de pulmón es menor de aproximadamente 0,4.

6. El método del artículo 1, donde la etapa b comprende la comparación de la cantidad de cada biomarcador en el panel con respecto a varios puntos de corte predeterminados para dicho biomarcador y asignando una puntuación para cada biomarcador basándose en dicha comparación, la etapa c comprende la combinación de la puntuación asignada para cada biomarcador cuantificada en la etapa b para obtener una puntuación total para dicho sujeto, la etapa d comprende la comparación de la puntuación total determinada en la etapa c con una puntuación total predeterminada y la etapa e comprende la determinación de si el sujeto tiene cáncer de pulmón basándose en la comparación de la puntuación total de la etapa d.

7. Un método de ayuda al diagnóstico de un sujeto sospechoso de cáncer de pulmón, comprendiendo el método las etapas de:

- a. obtener un valor para al menos un parámetro biométrico de un sujeto;
- b. comparar el valor de al menos un parámetro biométrico contra un punto de corte predeterminado para cada dicho parámetro biométrico y asignar una puntuación para cada parámetro biométrico basándose en dicha comparación;

- c. cuantificar en una muestra de ensayo obtenida de un sujeto, la cantidad de dos o más biomarcadores en un panel, comprendiendo el panel al menos un anticuerpo y al menos un antígeno;
- d. comparar la cantidad de cada biomarcador cuantificado en el panel con respecto a un punto de corte predeterminado para dicho biomarcador y asignar una puntuación para cada biomarcador basándose en dicha comparación;
- e. combinar la puntuación asignada para cada parámetro biométrico determinado en la etapa b con la puntuación asignada para cada biomarcador cuantificado en la etapa d para obtener una puntuación total para dicho sujeto.
- f. comparar la puntuación total determinada en la etapa 3 con una puntuación predeterminada total; y
- g. determinar si dicho sujeto tiene riesgo de cáncer de pulmón basándose en la comparación de la puntuación total de la etapa f.
8. El método del artículo 7, donde el panel comprende al menos un anticuerpo que se selecciona de entre el grupo que consiste en: anti-p53, anti-TMP21, anti-NY-ESO-1, anti-HDJ1, dominio-anti-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSLN1, anti-RCV1, anti-MAPKAPK3 y anti-Ciclina E2.
9. El método del artículo 7, donde el panel comprende al menos un antígeno que se selecciona de entre el grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, proGRP, CA19-9, seroamiloides A, alfa-1-anti-tripsina, apolipoproteína CIII y timosina [beta]4.
10. El método del artículo 7, donde el panel comprende además al menos una región de interés seleccionada de entre el grupo que consiste en: Acn6399, Acn949, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959.
11. El método del artículo 7, donde el parámetro biométrico se selecciona de entre el grupo que consiste en:
- la historia de fumador del sujeto, edad, exposición a carcinógenos y género.
12. El método del artículo 11, donde el parámetro biométrico es los paquetes-años de fumador.
13. El método del artículo 7, donde la etapa b comprende la comparación del valor de cada parámetro biométrico contra varios puntos de corte predeterminados para cada parámetro biométrico y la asignación de uno de varios puntuaciones posibles a cada dicho parámetro biométrico basándose en dicha comparación, la etapa d comprende la comparación de la cantidad de cada biomarcador en el panel de varios puntos de corte predeterminados para dicho marcador y la asignación de una puntuación para cada biomarcador basándose en dicha comparación, la etapa 3 comprende la combinación de la puntuación asignada para cada biomarcador cuantificado en la etapa d con la puntuación asignada para el parámetro biométrico en la etapa b para obtener una puntuación total para dicho sujeto, la etapa f comprende la comparación de la puntuación total determinada en la etapa e con una puntuación total predeterminada y la etapa g comprende la determinación de si dicho sujeto tiene cáncer de pulmón basándose en la comparación de la puntuación total y en la etapa f.
14. Un método para ayudar el diagnóstico de un sujeto sospechoso de cáncer de pulmón, comprendiendo el método las etapas de:
- a. cuantificar en una muestra de ensayo obtenida de un sujeto, la cantidad de dos o más biomarcadores en un panel, donde el panel comprende anti-p53, anti-NY-ESO-1, anti-MAPKAPK3, anti-Ciclina E2, citoqueratina 19, CEA, CA125 y ProGRP;
- b. comparar la cantidad de cada biomarcador cuantificado en el panel con un punto de corte predeterminado para dicho biomarcador y asignar una puntuación para cada biomarcador basándose en dicha comparación;
- c. combinar la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto;
- d. comparar la puntuación total determinado en la etapa c con una puntuación total predeterminada; y
- e. determinar si dicho sujeto tiene riesgo de cáncer de pulmón basándose en la comparación de la puntuación total de la etapa d.
15. El método del artículo 14, donde la etapa b comprende comparar la cantidad de cada biomarcador en el panel con una serie de valores de corte predeterminados para dicho biomarcador y asignar una puntuación para cada biomarcador basándose en dicha comparación; la etapa c comprende combinar la puntuación asignada para cada biomarcador cuantificado en la etapa b para obtener una puntuación total para dicho sujeto; la etapa d comprende comparar la puntuación total determinada en la etapa c con una puntuación total predeterminada y la etapa e comprende determinar si dicho sujeto tiene cáncer de pulmón basándose en la puntuación total determinada en la etapa d.
16. El método del artículo 14, donde el método comprende además obtener un valor para al menos un parámetro biométrico de un sujeto y comparar el valor del al menos un parámetro biométrico frente a un valor de corte predeterminado para cada uno de dichos parámetros biométricos y asignar una puntuación para cada parámetro biométrico basándose en dicha comparación.

17. El método del artículo 16, donde el parámetro biométrico es paquetes años.
18. El método del artículo 14, donde el método implica la cuantificación de la cantidad de tres o más biomarcadores en el panel.
- 5 19. El método del artículo 14, donde el método implica la cuantificación de la cantidad de cuatro o más biomarcadores en el panel.
- 10 20. El método del artículo 14, donde el método implica la cuantificación de la cantidad de siete o más biomarcadores en el panel.
21. El método del artículo 14, donde el método implica la cuantificación de la cantidad de ocho biomarcadores en el panel.
- 15 22. El método del artículo 14, donde el anti-ciclina E2 es un anticuerpo contra la SEQ ID NO: 3, la SEQ ID NO: 4 o la SEQ ID NO: 5.
23. Un kit que comprende: un péptido que se selecciona de entre el grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5 o combinaciones de los mismos.
- 20 24. Un kit que comprende:

al menos un antígeno contra anti-ciclina E2.
- 25 25. Un kit que comprende: a. reactivos que contienen al menos un anticuerpo para cuantificar uno o más antígenos en una muestra de ensayo, en donde dichos antígenos son: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA 15-3, SCC, CA19-9, proGRP, seroamiloide A, alfa-1-antitripsina, apolipoproteína CIII o timosina β [beta]4;
- 30 b. reactivos que contienen uno o más antígenos para cuantificar al menos un anticuerpo en una muestra de ensayo; donde dichos anticuerpos son; anti-TMP21, dominio-anti-NPCILIC, anti-TMOD1, anti-p53, anti-NY-ESO-1, anti-MAPKAPK3 y anti-ciclina E2; y opcionalmente,
- 35 c. algoritmos para combinar y comparar la cantidad de cada antígeno y cuantificar los anticuerpos en la muestra de ensayo contra un punto de corte predeterminado y asignar una puntuación a cada antígeno y anticuerpo cuantificados basándose en dicha comparación, combinar la puntuación asignada para cada antígeno y anticuerpo cuantificados para obtener una puntuación total, comparar la puntuación total con una puntuación total predeterminada y utilizar dicha comparación como una ayuda para determinar si un sujeto tiene cáncer de pulmón.
- 40 26. El kit del artículo 25, que comprende además reactivos para cuantificar una o más regiones de interés que se seleccionan de entre el grupo que consiste en: ACN9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 e Hic3959.
27. El kit del artículo 25, que comprende además: a. reactivos que contienen al menos un anticuerpo para cuantificar uno o más antígenos en una muestra de ensayo, donde dichos antígenos son: citoqueratina 19, CEA, CA125 y ProGRP;
- 45 b. reactivos que contienen uno o más antígenos para cuantificar al menos un anticuerpo en la muestra de ensayo;
- 50 donde dichos anticuerpos son: anti-p53, anti-NY-ESO-1, anti-MAPKAPK3 y anti-Ciclina E2; y opcionalmente,
- 55 c. algoritmos para combinar y comparar la cantidad de cada antígeno y anticuerpo cuantificado en la muestra de ensayo frente a un valor de corte predeterminado y asignar una puntuación para cada antígeno y anticuerpo cuantificado basándose en dicha comparación, combinar la puntuación asignada para cada antígeno y anticuerpo cuantificado para obtener una puntuación total, comparar la puntuación total con una puntuación total predeterminada usando dicha comparación y como una ayuda para determinar si dicho sujeto padece cáncer de pulmón.
- 60 28. El kit del artículo 27, donde dicho kit contiene reactivos para cuantificar cada uno de citoqueratina 19, CEA, CA125 y ProGRP en una muestra de ensayo.
29. El kit del artículo 27, donde dicho kit contiene reactivos para cuantificar cada uno de anti-p53, anti-NY-ESO-1, anti-MAPKAPK3 y al menos un anticuerpo contra Ciclina E2 inmunorreactiva en una muestra de ensayo.
- 65 30. El kit del artículo 27, donde dicho kit contiene reactivos para cuantificar cada uno de citoqueratina 19, CEA, CA125 ProGRP, anti-p53, anti-NY-ESO-1, anti-MAPKAPK3 y anti-Ciclina E2 en una muestra de ensayo.

31. Un polipéptido aislado que tiene una secuencia de aminoácidos que se selecciona de entre el grupo que consiste en:

5 SEQ ID NO: 3 y un polipéptido que tiene un 60 % de homología con la secuencia de aminoácidos de SEQ ID NO: 3.

32. Un polipéptido aislado que tiene una secuencia de aminoácidos que se selecciona de entre el grupo que consiste en:

10 SEQ ID NO: 4 y un polipéptido que tiene un 60 % de homología con la secuencia de aminoácidos de SEQ ID NO: 4.

33. Un polipéptido aislado que tiene una secuencia de aminoácidos que se selecciona de entre el grupo que consiste en:

15 SEQ ID NO: 5 y un polipéptido que tiene un 60 % de homología con la secuencia de aminoácidos de SEQ ID NO: 5.

LISTADO DE SECUENCIAS

20 <100> ABBOTT LABORATORIES

<120> Métodos y combinaciones de marcadores para detectar la predisposición al cáncer de pulmón

25 <130> 065483-1246

<150> 11/771.727

<151> 29-06-2007

30 <160> 10

<170> PatentIn versión 3.2

<210> 1

35 <211> 296

<212> PRT

<213> *Homo sapiens*

40 <400> 1

ES 2 633 596 T3

Met	Ser	Lys	Glu	Val	Trp	Leu	Asn	Met	Leu	Lys	Lys	Glu	Ser	Arg	Tyr	1	5	10	15
Val	His	Asp	Lys	His	Phe	Glu	Val	Leu	His	Ser	Asp	Leu	Glu	Pro	Gln	20	25	30	
Met	Arg	Ser	Ile	Leu	Leu	Asp	Trp	Leu	Leu	Glu	Val	Cys	Glu	Val	Tyr	35	40	45	
Thr	Leu	His	Arg	Glu	Thr	Phe	Tyr	Leu	Ala	Gln	Asp	Phe	Phe	Asp	Arg	50	55	60	
Phe	Met	Leu	Thr	Gln	Lys	Asp	Ile	Asn	Lys	Asn	Met	Leu	Gln	Leu	Ile	65	70	75	80
Gly	Ile	Thr	Ser	Leu	Phe	Ile	Ala	Ser	Lys	Leu	Glu	Glu	Ile	Tyr	Ala	85	90	95	
Pro	Lys	Leu	Gln	Glu	Phe	Ala	Tyr	Val	Thr	Asp	Gly	Ala	Cys	Ser	Glu	100	105	110	
Glu	Asp	Ile	Leu	Arg	Met	Glu	Leu	Ile	Ile	Leu	Lys	Ala	Leu	Lys	Trp	115	120	125	
Glu	Leu	Cys	Pro	Val	Thr	Ile	Ile	Ser	Trp	Leu	Asn	Leu	Phe	Leu	Gln	130	135	140	
Val	Asp	Ala	Leu	Lys	Asp	Ala	Pro	Lys	Val	Leu	Leu	Pro	Gln	Tyr	Ser	145	150	155	160

Gln Glu Thr Phe Ile Gln Ile Ala Gln Leu Leu Asp Leu Cys Ile Leu
165 170 175

Ala Ile Asp Ser Leu Glu Phe Gln Tyr Arg Ile Leu Thr Ala Ala Ala
180 185 190

Leu Cys His Phe Thr Ser Ile Glu Val Val Lys Lys Ala Ser Gly Leu
195 200 205

Glu Trp Asp Ser Ile Ser Glu Cys Val Asp Trp Met Val Pro Phe Val
210 215 220

Asn Val Val Lys Ser Thr Ser Pro Val Lys Leu Lys Thr Phe Lys Lys
225 230 235 240

Ile Pro Met Glu Asp Arg His Asn Ile Gln Thr His Thr Asn Tyr Leu
245 250 255

Ala Met Leu Glu Glu Val Asn Tyr Ile Asn Thr Phe Arg Lys Gly Gly
260 265 270

Gln Leu Ser Pro Val Cys Asn Gly Gly Ile Met Thr Pro Pro Lys Ser
275 280 285

Thr Glu Lys Pro Pro Gly Lys His
290 295

<210> 2
<211> 374
<212> PRT
<213> *Homo sapiens*

5

<400> 2

Met Ser Arg Arg Ser Ser Arg Leu Gln Ala Lys Gln Gln Pro Gln Pro
1 5 10 15

Ser Gln Thr Glu Ser Pro Gln Glu Ala Gln Ile Ile Gln Ala Lys Lys
20 25 30

Arg Lys Thr Thr Gln Asp Val Lys Lys Arg Arg Glu Glu Val Thr Lys
35 40 45

Lys His Gln Tyr Glu Ile Arg Asn Cys Trp Pro Pro Val Leu Ser Gly
50 55 60

Gly Ile Ser Pro Cys Ile Ile Ile Glu Thr Pro His Lys Glu Ile Gly
65 70 75 80

ES 2 633 596 T3

Thr	Ser	Asp	Phe	Ser	Arg	Phe	Thr	Asn	Tyr	Arg	Phe	Lys	Asn	Leu	Phe
				85					90					95	
Ile	Asn	Pro	Ser	Pro	Leu	Pro	Asp	Leu	Ser	Trp	Gly	Cys	Ser	Lys	Glu
			100					105					110		
Val	Trp	Leu	Asn	Met	Leu	Lys	Lys	Glu	Ser	Arg	Tyr	Val	His	Asp	Lys
		115					120					125			
His	Phe	Glu	Val	Leu	His	Ser	Asp	Leu	Glu	Pro	Gln	Met	Arg	Ser	Ile
	130					135					140				
Leu	Leu	Asp	Trp	Leu	Leu	Glu	Val	Cys	Glu	Val	Tyr	Thr	Leu	His	Arg
145					150					155					160
Glu	Thr	Phe	Tyr	Leu	Ala	Gln	Asp	Phe	Phe	Asp	Arg	Phe	Met	Leu	Thr
				165					170					175	
Gln	Lys	Asp	Ile	Asn	Lys	Asn	Met	Leu	Gln	Leu	Ile	Gly	Ile	Thr	Ser
			180					185					190		
Leu	Phe	Ile	Ala	Ser	Lys	Leu	Glu	Glu	Ile	Tyr	Ala	Pro	Lys	Leu	Gln
		195					200					205			
Glu	Phe	Ala	Tyr	Val	Thr	Asp	Gly	Ala	Cys	Ser	Glu	Glu	Asp	Ile	Leu
	210					215					220				
Arg	Met	Glu	Leu	Ile	Ile	Leu	Lys	Ala	Leu	Lys	Trp	Glu	Leu	Cys	Pro
225					230					235					240
Val	Thr	Ile	Ile	Ser	Trp	Leu	Asn	Leu	Phe	Leu	Gln	Val	Asp	Ala	Leu
				245					250					255	
Lys	Asp	Ala	Pro	Lys	Val	Leu	Leu	Pro	Gln	Tyr	Ser	Gln	Glu	Thr	Phe
			260					265					270		
Ile	Gln	Ile	Ala	Gln	Leu	Leu	Asp	Leu	Cys	Ile	Leu	Ala	Ile	Asp	Ser
		275					280					285			
Leu	Glu	Phe	Gln	Tyr	Arg	Ile	Leu	Thr	Ala	Ala	Ala	Leu	Cys	His	Phe
	290					295					300				
Thr	Ser	Ile	Glu	Val	Val	Lys	Lys	Ala	Ser	Gly	Leu	Glu	Trp	Asp	Ser
305					310					315					320
Ile	Ser	Glu	Cys	Val	Asp	Trp	Met	Val	Pro	Phe	Val	Asn	Val	Val	Lys
				325					330					335	

ES 2 633 596 T3

Ser Thr Ser Pro Val Lys Leu Lys Thr Phe Lys Lys Ile Pro Met Glu
340 345 350

Asp Arg His Asn Ile Gln Thr His Thr Asn Tyr Leu Ala Met Leu Cys
355 360 365

Met Ile Ser Ser His Val
370

<210> 3
<211> 38
<212> PRT
<213> *Homo sapiens*

<400> 3

Cys Glu Glu Val Asn Tyr Ile Asn Thr Phe Arg Lys Gly Gly Gln Leu
1 5 10 15

Ser Pro Val Cys Asn Gly Gly Ile Met Thr Pro Pro Lys Ser Thr Glu
20 25 30

Lys Pro Pro Gly Lys His
35

<210> 4
<211> 37
<212> PRT
<213> *Homo sapiens*

<400> 4

Glu Glu Val Asn Tyr Ile Asn Thr Phe Arg Lys Gly Gly Gln Leu Ser
1 5 10 15

Pro Val Cys Asn Gly Gly Ile Met Thr Pro Pro Lys Ser Thr Glu Lys
20 25 30

Pro Pro Gly Lys His
35

<210> 5
<211> 19
<212> PRT
<213> *Homo sapiens*

<400> 5

Cys Asn Gly Gly Ile Met Thr Pro Pro Lys Ser Thr Glu Lys Pro Pro
1 5 10 15

Gly Lys His

<210> 6
<211> 103

ES 2 633 596 T3

<212> PRT

<213> *Homo sapiens*

<400> 6

5

```

Ser Phe Phe Ser Phe Leu Gly Glu Ala Phe Asp Gly Ala Arg Asp Met
1           5           10           15

Trp Arg Ala Tyr Ser Asp Met Arg Glu Ala Asn Tyr Ile Gly Ser Asp
          20           25           30

Lys Tyr Phe His Ala Arg Gly Asn Tyr Asp Ala Ala Lys Arg Gly Pro
          35           40           45

Gly Gly Ala Trp Ala Ala Glu Val Ile Ser Asp Ala Arg Glu Asn Ile
50           55           60

Gln Arg Phe Phe Gly His Asp Ala Glu Asp Ser Leu Ala Asp Gln Ala
65           70           75           80

Ala Asn Glu Trp Gly Arg Ser Gly Lys Asp Pro Asn His Phe Arg Pro
          85           90           95

Ala Gly Leu Pro Glu Lys Tyr
          100

```

<210> 7

<211> 103

<212> PRT

<213> *Homo sapiens*

<400> 7

10

ES 2 633 596 T3

Ser Phe Phe Ser Phe Leu Gly Glu Ala Phe Asp Gly Ala Arg Asp Met
1 5 10 15

Trp Arg Ala Tyr Ser Asp Met Arg Glu Ala Asn Tyr Ile Gly Ser Asp
20 25 30

Lys Tyr Phe His Ala Arg Gly Asn Tyr Asp Ala Ala Lys Arg Gly Pro
35 40 45

Gly Gly Ala Trp Ala Ala Glu Val Ile Ser Asp Ala Arg Glu Asn Ile
50 55 60

Gln Arg Phe Phe Gly His Gly Ala Glu Asp Ser Leu Ala Asp Gln Ala
65 70 75 80

Ala Asn Glu Trp Gly Arg Ser Gly Lys Asp Pro Asn His Phe Arg Pro
85 90 95

Ala Gly Leu Pro Glu Lys Tyr
100

<210> 8
<211> 79
<212> PRT
<213> *Homo sapiens*

<400> 8

Ser Glu Ala Glu Asp Ala Ser Leu Leu Ser Phe Met Gln Gly Tyr Met
1 5 10 15

Lys His Ala Thr Lys Thr Ala Lys Asp Ala Leu Ser Ser Val Gln Glu
20 25 30

Ser Gln Val Ala Gln Gln Ala Arg Gly Trp Val Thr Asp Gly Phe Ser
35 40 45

Ser Leu Lys Asp Tyr Trp Ser Thr Val Lys Asp Lys Phe Ser Glu Phe
50 55 60

Trp Asp Leu Asp Pro Glu Val Arg Pro Thr Ser Ala Val Ala Ala
65 70 75

<210> 9
<211> 42
<212> PRT
<213> *Homo sapiens*

<400> 9

ES 2 633 596 T3

Leu Glu Ala Ile Pro Met Ser Ile Pro Pro Glu Val Lys Phe Asn Glu
1 5 10 15

Pro Phe Val Phe Leu Met Ile Asp Gln Asn Thr Lys Ser Pro Leu Phe
20 25 30

Met Gly Lys Val Val Asn Pro Thr Gln Lys
35 40

<210> 10

<211> 24

5 <212> PRT

<213> *Homo sapiens*

<400> 10

Asp Ala His Lys Ser Glu Val Ala His Arg Phe Lys Asp Leu Gly Glu
1 5 10 15

Glu Asn Phe Lys Ala Leu Val Leu
20

REIVINDICACIONES

1. Un método para ayudar al diagnóstico de un sujeto sospechoso de cáncer de pulmón, comprendiendo el método las etapas de:

- (a) cuantificar las cantidades de biomarcadores en una muestra de ensayo obtenida de un sujeto, donde los biomarcadores cuantificados son al menos anti-MAPKAPK3, anti-NY-ESO-1, CEA, CA125, citoqueratina 19, anti-p53, y anti-Ciclina E2;
- (b) comparar la cantidad de cada uno de los biomarcadores cuantificados con un punto de corte predeterminado para dicho biomarcador y asignar una puntuación para cada uno de los biomarcadores basándose en dicha comparación;
- (c) combinar la puntuación asignada para cada uno de los biomarcadores para obtener una puntuación total para dicho sujeto;
- (d) comparar la puntuación total determinada en la etapa (c) con una puntuación total predeterminada; y
- (e) determinar si dicho sujeto tiene un riesgo de cáncer de pulmón basándose en la comparación de la puntuación total de la etapa (d).

2. El método de la reivindicación 1, donde los biomarcadores cuantificados comprenden además uno o más biomarcadores que se seleccionan de entre el grupo que consiste en SCC, ProGRP, TPS, Pub11597, Pub4789, TFA2759, y Acn9459.

3. El método de la reivindicación 2, donde los biomarcadores cuantificados comprenden además ProGRP, Pub11597, y Acn9459.

4. El método de la reivindicación 2, donde

- (i) los biomarcadores cuantificados comprenden además al menos un anticuerpo que se selecciona de entre el grupo que consiste en: anti-TMP21, anti-HDJ1, anti-dominio-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSN1, y anti-RCV1; o
- (ii) los biomarcadores cuantificados comprenden además al menos un antígeno que se selecciona de entre el grupo que consiste en: citoqueratina 8, citoqueratina 18, CA19-9, CA15-3, seroamiloide A, alfa-1-anti-tripsina, timosina β 4, y apolipoproteína CIII; o
- (iii) los biomarcadores cuantificados comprenden además al menos una región de interés que se selecciona de entre el grupo que consiste en: Acn6399, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959.

5. El método de la reivindicación 1, donde la etapa (b) comprende la comparación de la cantidad de cada uno de los biomarcadores del panel con varios puntos de corte predeterminados para dicho biomarcador y la asignación de una puntuación para cada biomarcador basándose en dicha comparación, la etapa (c) comprende la combinación de la puntuación asignada para cada uno de los biomarcadores para obtener una puntuación total para dicho sujeto, la etapa (d) comprende la comparación de la puntuación total determinada en la etapa (c) con una puntuación total predeterminada y la etapa (e) comprende la determinación de si el sujeto tiene un cáncer de pulmón basándose en la comparación de la puntuación total de la etapa (d).

6. El método de la reivindicación 1, donde el método comprende además la obtención de un valor para al menos un parámetro biométrico de un sujeto y la comparación del valor del al menos un parámetro biométrico con un punto de corte predeterminado para cada uno de los parámetros biométricos y la asignación de una puntuación para cada uno de los parámetros biométricos basándose en dicha comparación.

7. El método de la reivindicación 6, donde el parámetro biométrico se selecciona de entre el grupo que consiste en: la historia de fumador del sujeto, edad, exposición a carcinógenos y género.

8. El método de la reivindicación 6, donde el método comprende además la comparación del valor de cada uno de los parámetros biométricos con varios puntos de corte predeterminados para dicho parámetro biométrico y la asignación de varias posibles puntuaciones para cada uno de dichos parámetros biométricos basándose en dicha comparación, y donde la etapa (b) comprende la comparación de la cantidad de cada biomarcador en el panel con varios puntos de corte predeterminados para dicho biomarcador y la asignación de una puntuación para cada uno de los biomarcadores basándose en dicha comparación, la etapa (c) comprende la combinación de la puntuación asignada para cada biomarcador con la puntuación asignada para el parámetro biométrico para obtener una puntuación total para dicho sujeto, la etapa (d) comprende la comparación de la puntuación total determinada en la etapa (c) con una puntuación total predeterminada y la etapa (e) comprende la determinación de si dicho sujeto tiene cáncer de pulmón basándose en la comparación de la puntuación total de la etapa (d).

9. Un kit que comprende:

- (a) reactivos que contienen antígenos para cuantificar anti-MAPKAPK3, anti-NY-ESO-1, anti-p53, y anti Ciclina

E2;

(b) reactivos para cuantificar CEA, CA125, y citoqueratina 19;

(c) reactivos para cuantificar al menos uno o más biomarcadores que se seleccionan de entre el grupo que consiste en SCC, ProGRP, TPS, Pub11597, Pub4789, TFA2759, y Acn9459; y opcionalmente,

5 (d) algoritmos para la combinación y comparación de la cantidad de cada antígeno y anticuerpo cuantificados en la muestra de ensayo con un punto de corte predeterminado y la asignación de una puntuación para cada antígeno y anticuerpo cuantificados basándose en dicha comparación, la combinación de la puntuación asignada para cada antígeno y anticuerpo cuantificados para obtener una puntuación total, la comparación de la puntuación total con una puntuación total predeterminada y la utilización de dicha comparación como una ayuda
10 para determinar si un sujeto tiene cáncer de pulmón.

10. El kit de la reivindicación 9, que comprende además reactivos para cuantificar uno o más biomarcadores seleccionados de entre el grupo que consiste en: anti-TMP21, anti-HDJ1, anti-dominio NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSIN1, anti-RCV1, citoqueratina 8, citoqueratina 18, CA19-9, CA15-3, seroamiloide A,
15 alfa-1-anti-tripsina, timosina β 4, apolipoproteína CIII, Acn6399, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 e HIC3959.

11. El kit de la reivindicación 9, que comprende además un péptido seleccionado de entre el grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, y combinaciones de los mismos.
20

12. El método de la reivindicación 1, donde los biomarcadores cuantificados comprenden además proGRP.

Fig. 1

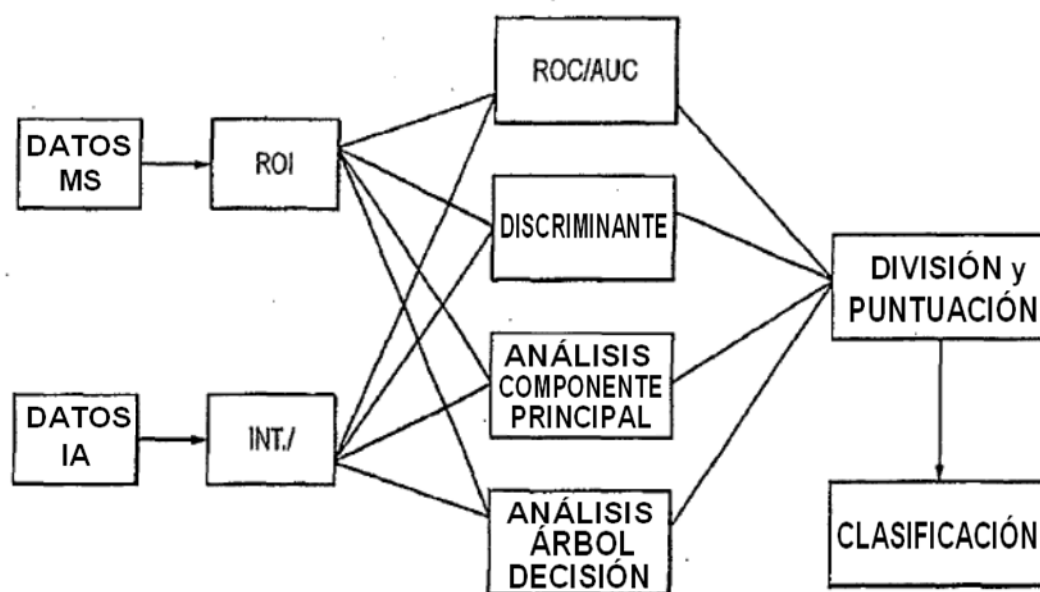
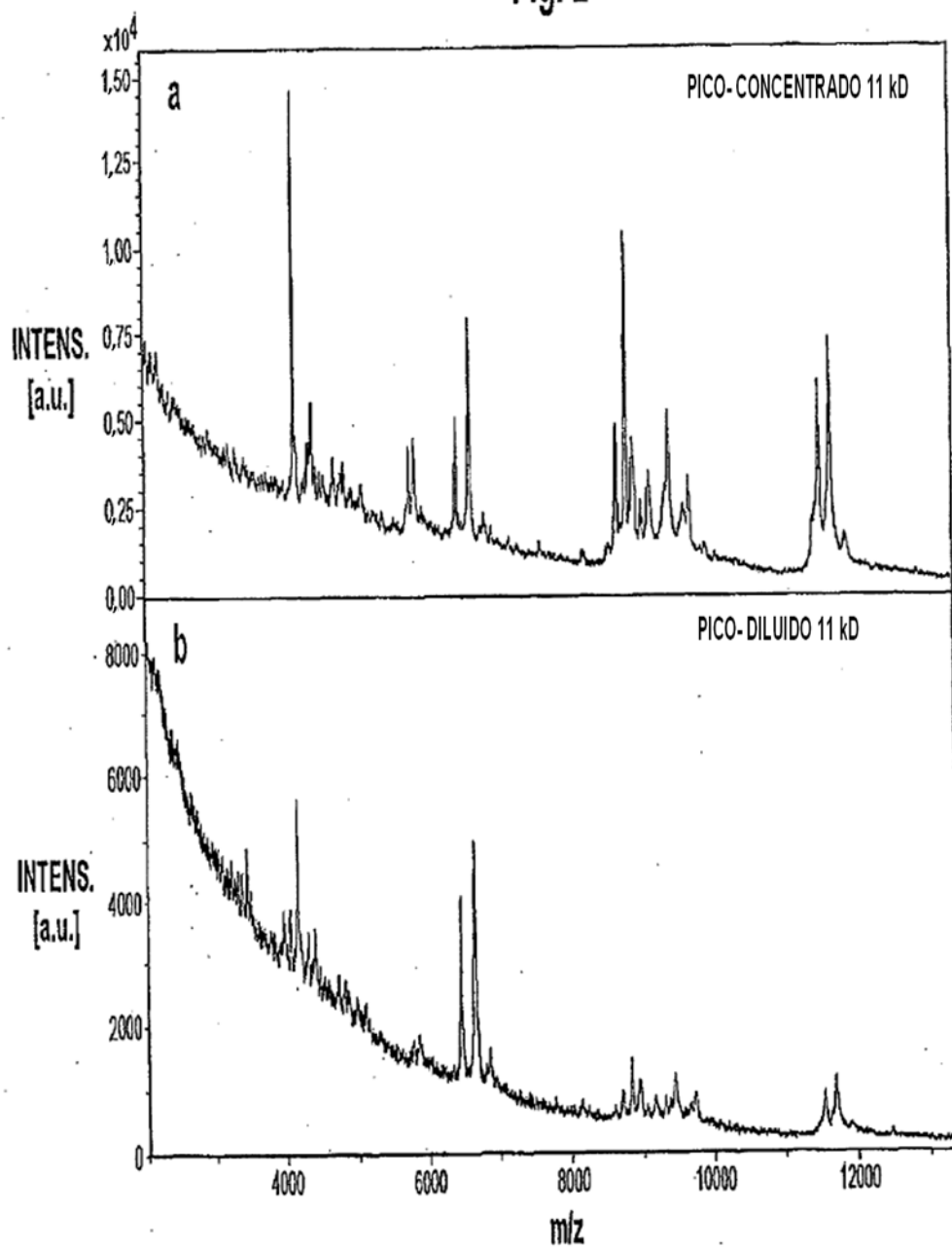


Fig. 2



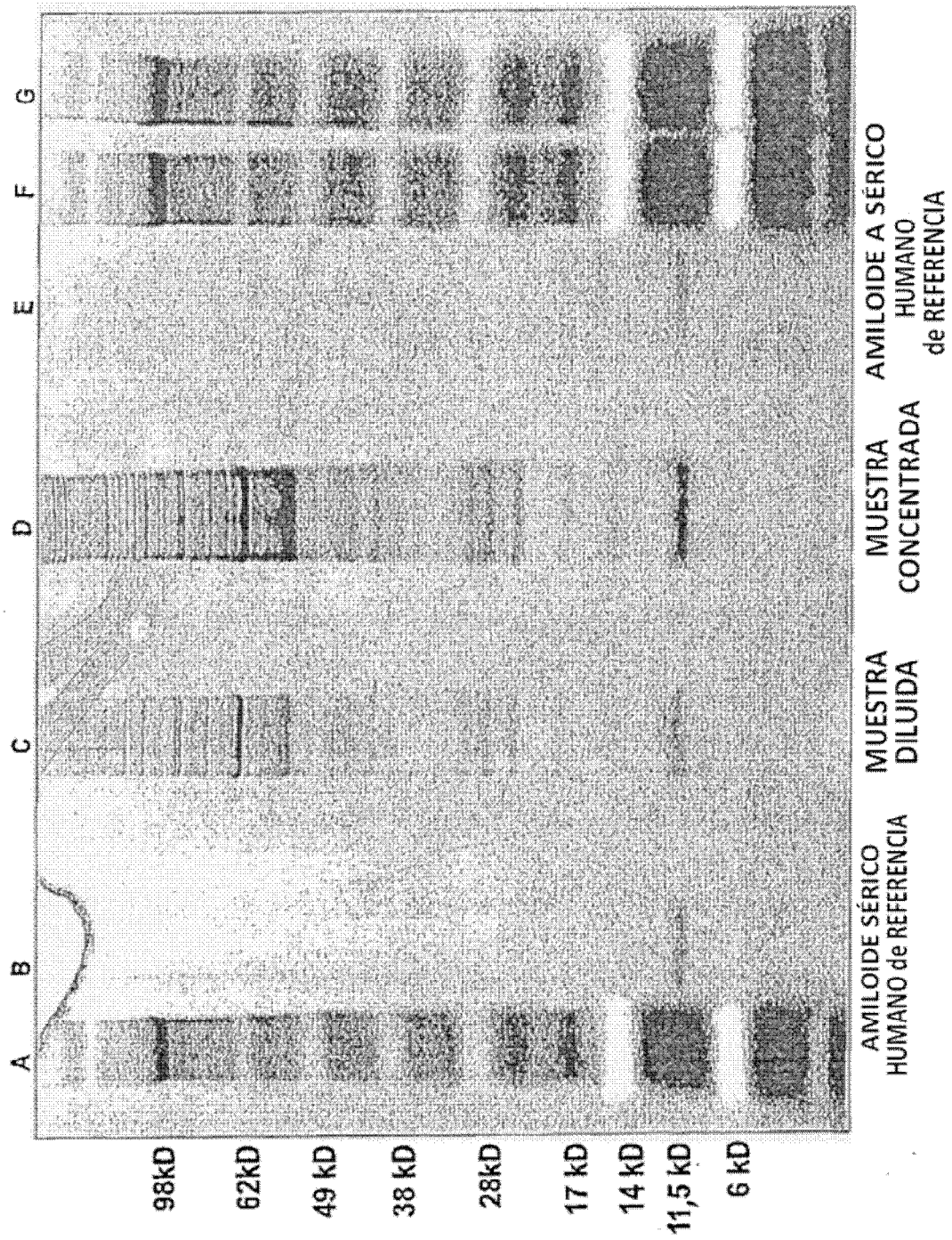


Fig. 3

Fig. 4a

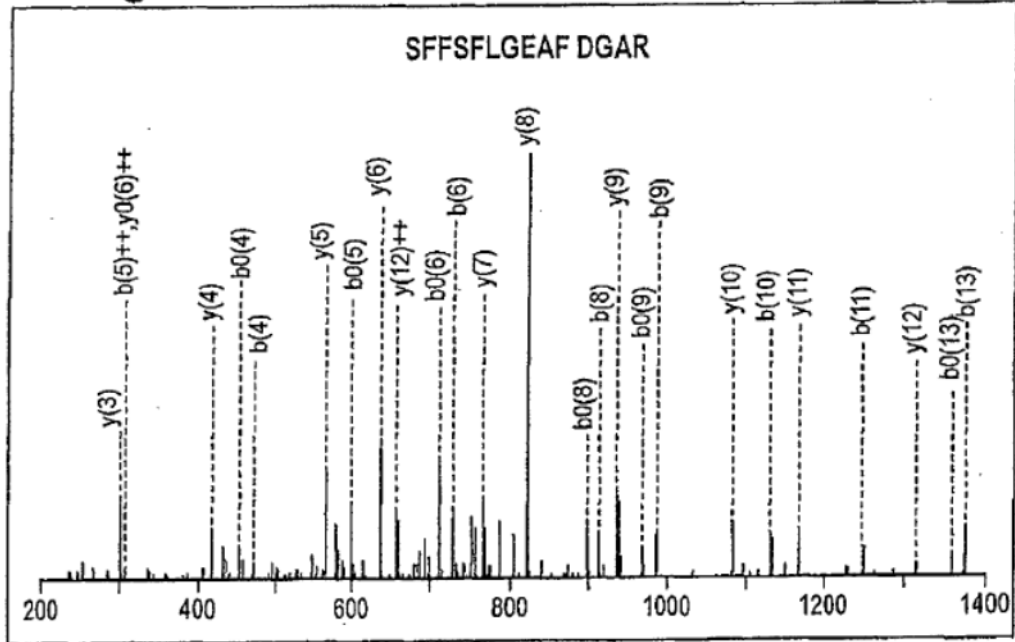


Fig. 4b

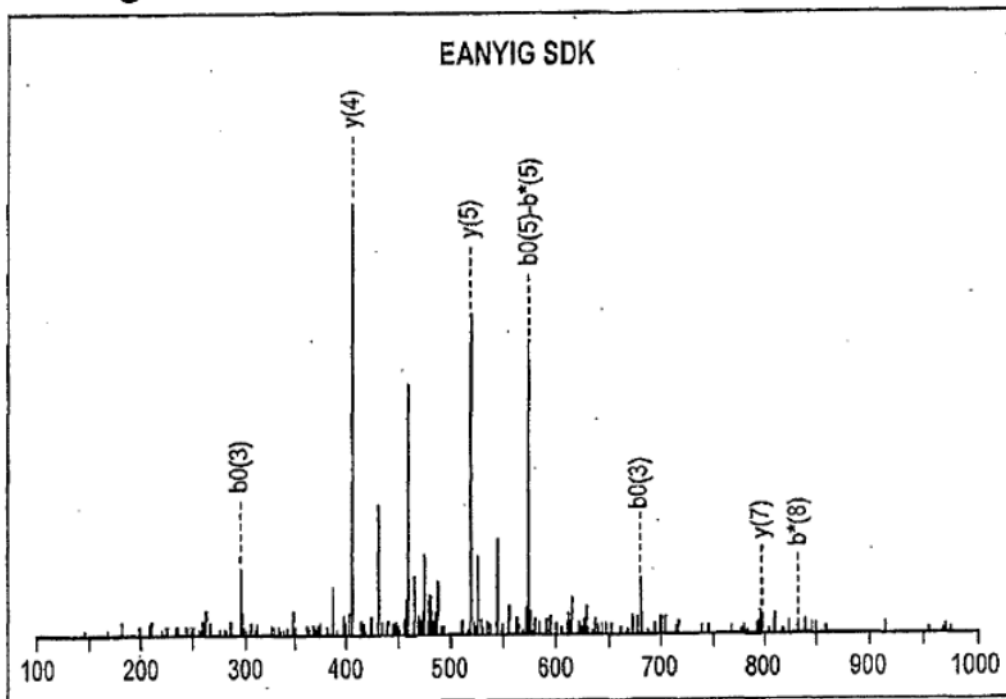


Fig. 4c

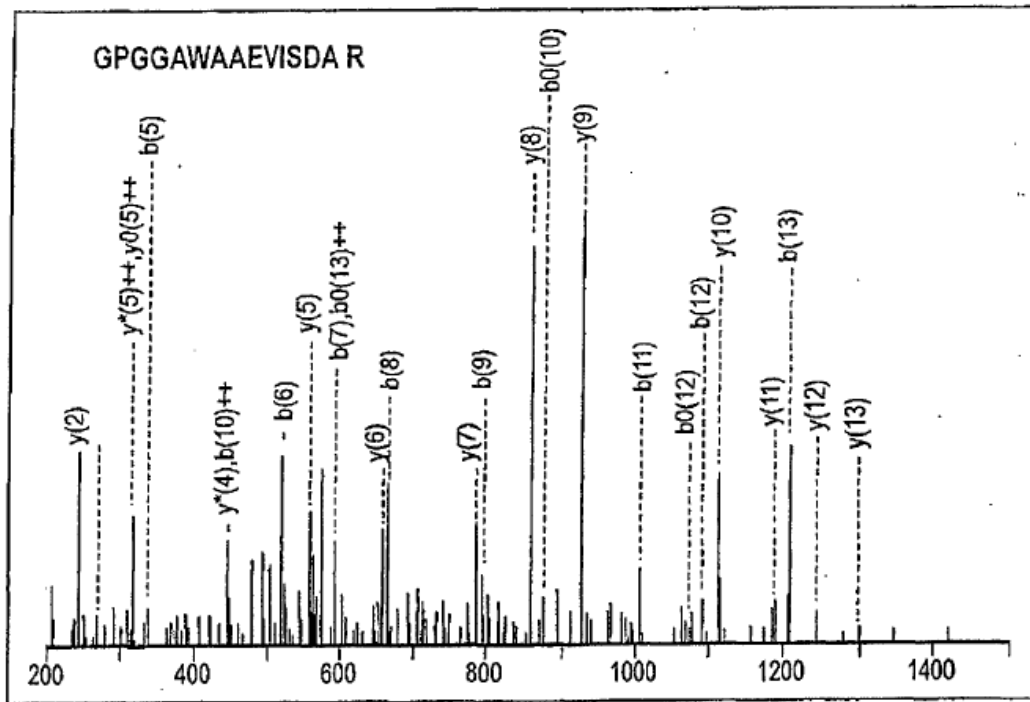


Fig. 4d

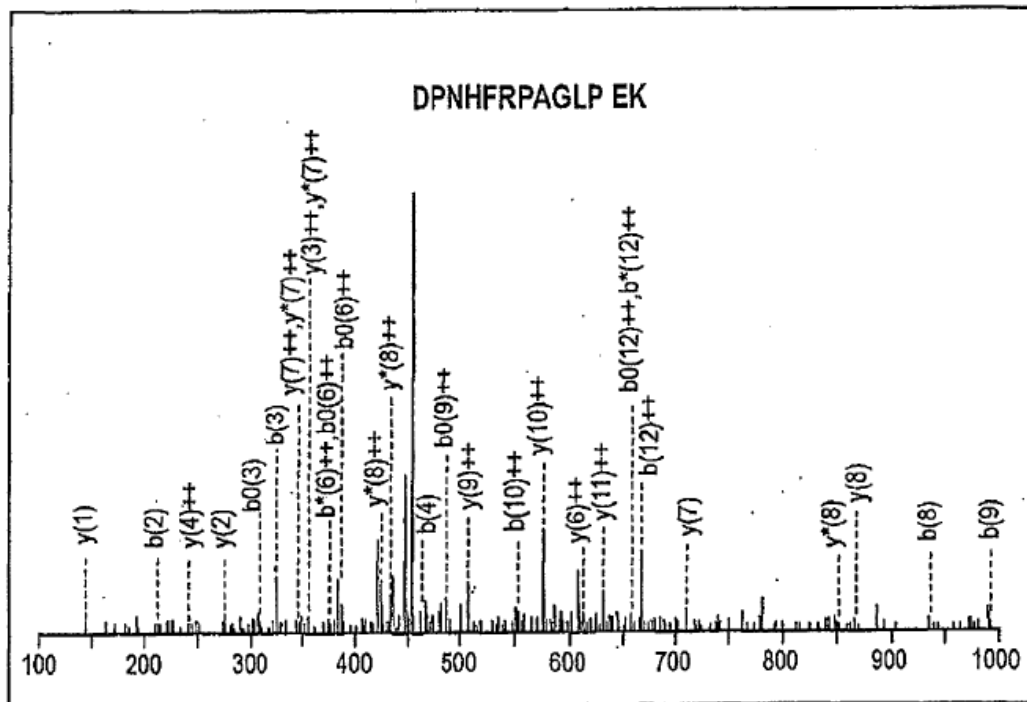


Fig. 5

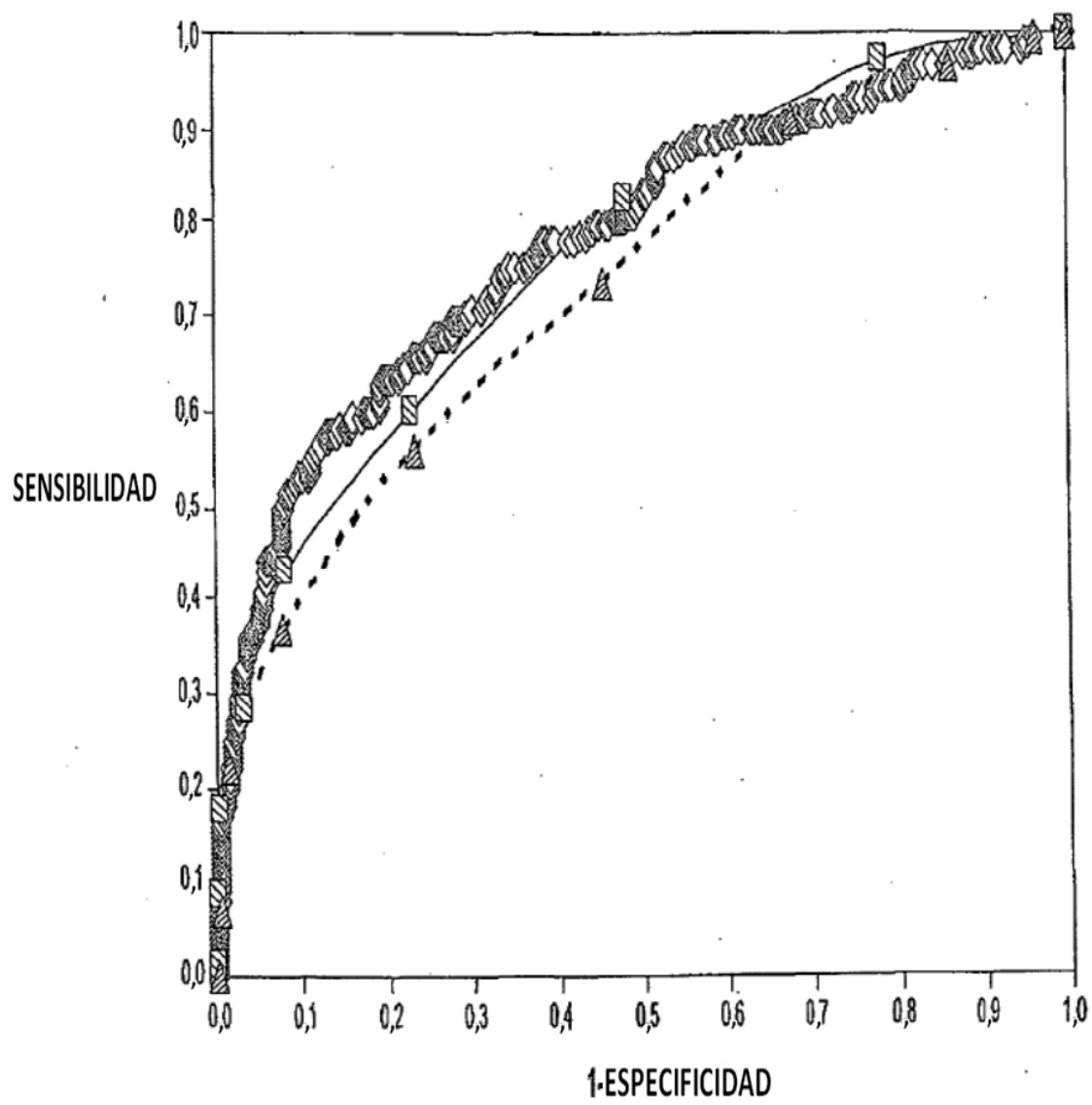


Figura 6

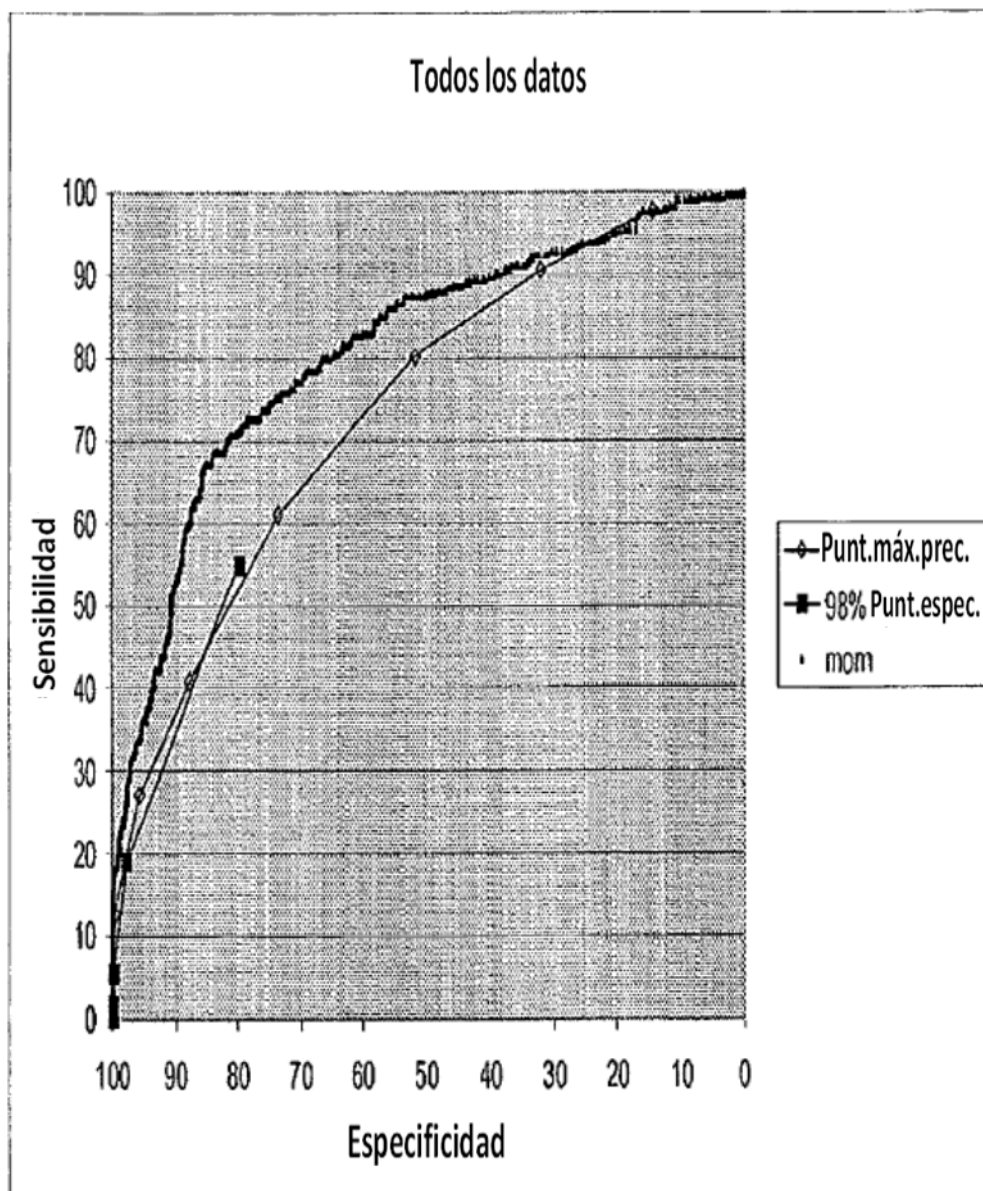


Figura 7

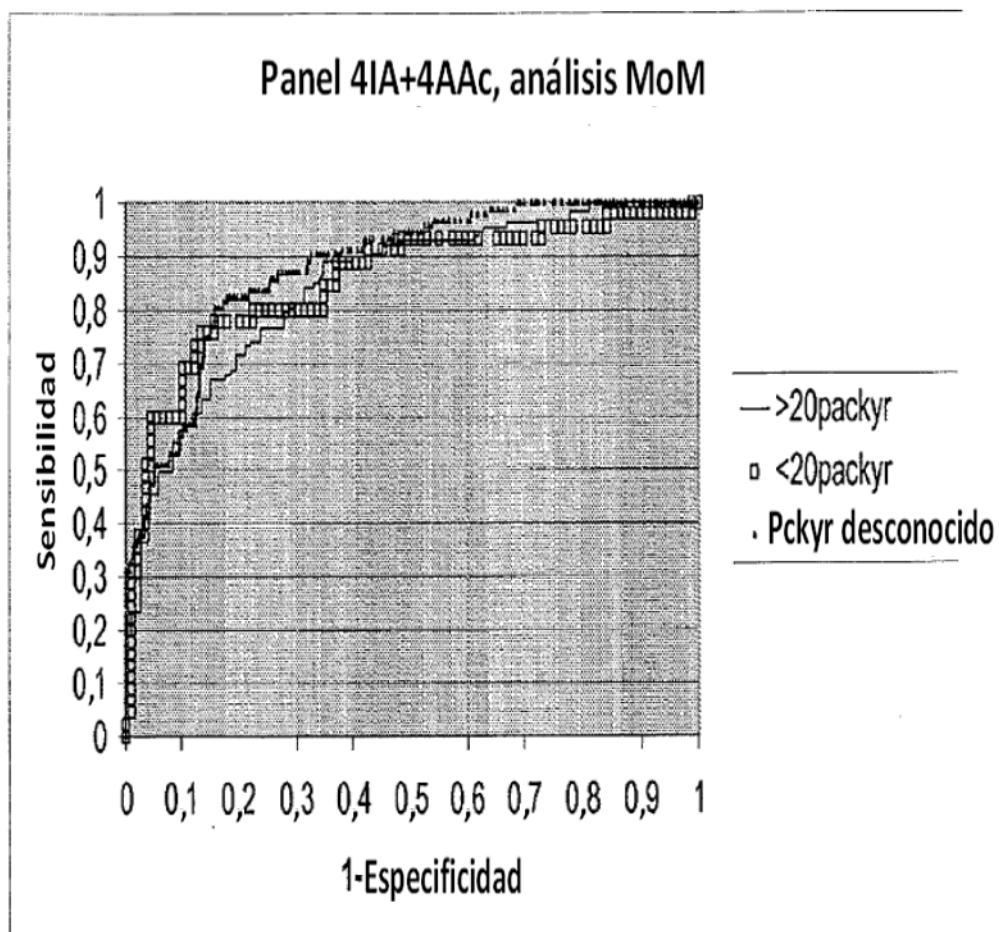


Figura 8

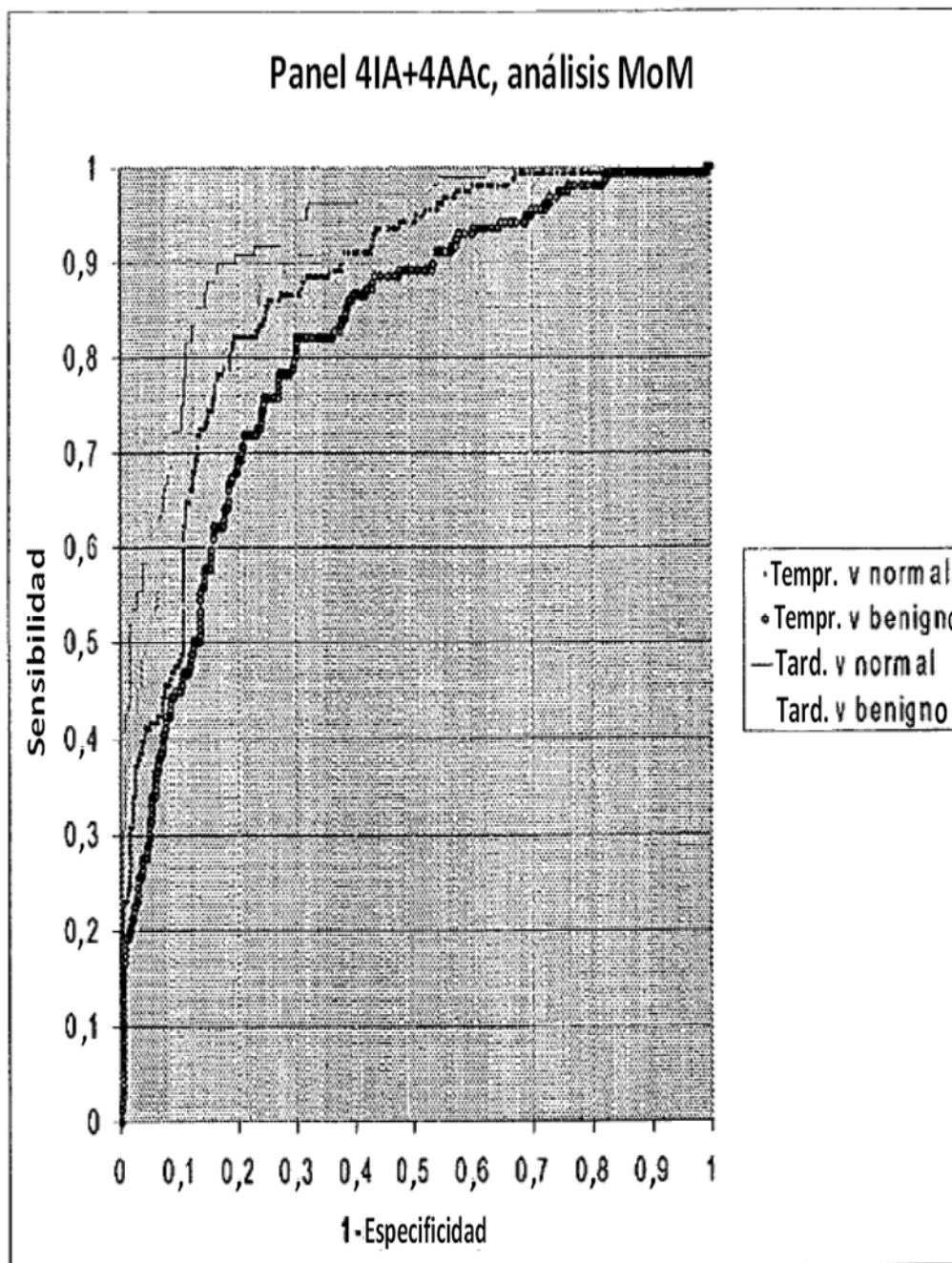


Figura 9

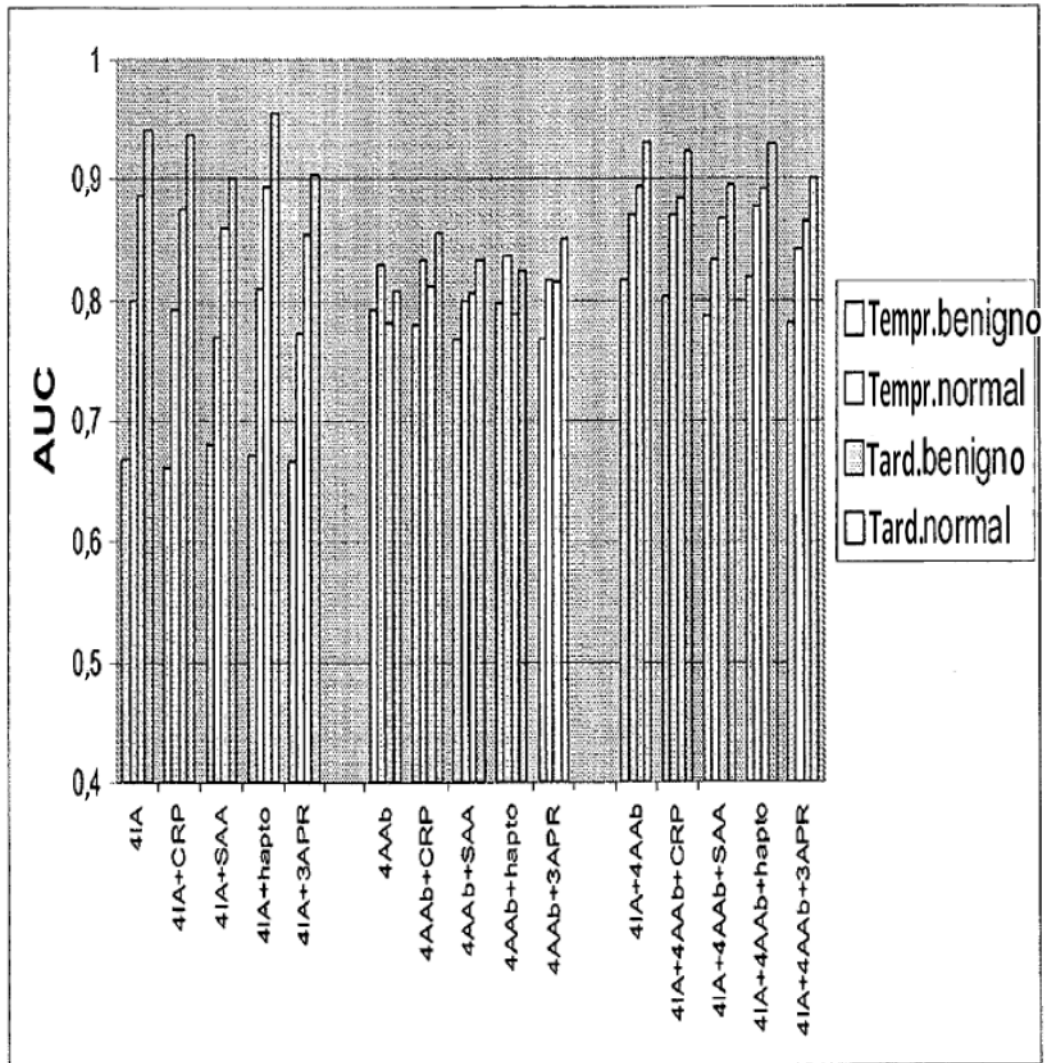


Figura 10

