

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 606 015**

51 Int. Cl.:

G01N 33/574 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **21.12.2006** **E 12156879 (4)**

97 Fecha y número de publicación de la concesión europea: **26.10.2016** **EP 2485047**

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

30 Prioridad:

22.12.2005 US 753331 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
17.03.2017

73 Titular/es:

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

72 Inventor/es:

RUSSELL, JOHN;
COLPITTS, TRACEY;
RUSSELL, ERIC;
FROST, STEPHEN;
RAMIREZ, JAVIER y
SINGH, BHAWANI

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 606 015 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 la detección de 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 en mujeres en los Estados Unidos, con una proyección de 172.500 nuevos casos estimados a diagnosticar durante 2005 (estadísticas de la Sociedad Americana del Cáncer). Es la causa más común de muerte por cáncer en ambos sexos, con más de 163.000 muertes relacionadas con el cáncer de pulmón esperadas en 2005. El cáncer de pulmón es también un problema de salud importante en otras zonas del mundo. En la Unión Europea se producen aproximadamente 135.000 nuevos casos cada año. Genesis Report, Febrero de 1995. Además, la incidencia está aumentando rápidamente en Europa Central y Oriental, donde los hombres tienen las mayores tasas de consumo de cigarrillos del mundo. T. Reynolds, J. Natl. Cancer Inst. 87:1348-1349 (1995). El tabaco es el único responsable de más de 90% de todos los casos de cáncer de pulmón, tráquea, y bronquios. CPMCnet, Guide to Clinical Preventive Services. La Agencia Internacional para la Investigación del Cáncer de la Organización Mundial de la Salud estima que en 2002, en todo el mundo, había 1.352.000 casos de cáncer de pulmón con 1.179.000 muertes debidas a la enfermedad.

El cáncer de pulmón en fase temprana puede ser detectado por medio de radiografía de tórax y examen citológico de esputo, sin embargo, estos procedimientos no tienen la suficiente precisión como para ser utilizados rutinariamente como ensayos de escrutinio para los individuos asintomáticos. Los problemas técnicos potenciales que pueden limitar la sensibilidad de la radiografía de tórax incluyen técnicas subóptimas, exposición insuficiente, y posicionamiento y cooperación del paciente. T. G. Tape, et al., Ann. Intern. Med. 104:663-670 (1986). Los radiólogos a menudo están en desacuerdo sobre la interpretación de las radiografías de tórax y más de 40% de éstas son significativas o potencialmente significativas. P. G. Herman, et al, Chest 68: 278-282 (1975). Las interpretaciones de los falsos negativos son la causa de la mayoría de los errores y los resultados no concluyentes requieren ensayos de seguimiento para su aclaración. T. G. Tape et al., supra.

La citología de esputo es menos sensible que la radiografía de tórax en la detección temprana del cáncer de pulmón. 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 principales vías respiratorias, el tipo histológico del tumor, y la experiencia y la formación del citopatólogo. R. J. Ginsberg et al. En: Cancer: Principles and Practice of Oncology, cuarta edición, págs. 673-723, Filadelfia, PA: J/B. Lippincott Co. (1993).

La mayoría de los nuevos casos de cáncer de pulmón se detectarán cuando la enfermedad se haya extendido más allá del pulmón. En los Estados Unidos sólo 16% de los nuevos cánceres de pulmón de células no pequeñas se detectan en una fase localizada cuando la supervivencia a 5 años es más alta en 49,7%. En contraste, 68% de los nuevos casos se detectan cuando la enfermedad ya se ha extendido a nivel local o metastasizado a lugares distantes que tienen tasas de supervivencia a 5 años de 18,5% y 1,8%, respectivamente. Del mismo modo, 80% de los cánceres de pulmón de células pequeñas detectados recientemente se descubren con invasión local o metástasis a distancia, que tienen tasas de supervivencia a 5 años de 9,5% y 1,7%, respectivamente. Stat Bite, J. Natl. Cancer Inst. 87: 1662 (1995). Estas estadísticas demuestran que los procedimientos actuales no están logrando detectar el cáncer de pulmón en una etapa temprana y tratable de la enfermedad y que se necesitan mejores métodos de detección y tratamiento para reducir la mortalidad.

Los métodos más utilizados para el seguimiento de pacientes con cáncer de pulmón después de la terapia primaria son la visita a la clínica, la radiografía de tórax, el hemograma completo, las pruebas de función hepática y la tomografía computarizada torácica (CT). La detección de la recurrencia en un seguimiento regular, sin embargo, no afecta en gran medida al modo de tratamiento y al tiempo de supervivencia global conduciendo a la conclusión de que los métodos de seguimiento actuales no son rentables. K.S. Naunheim 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 producido un nuevo examen de la utilización de la TC para escrutar personas asintomáticas que presentan alto riesgo de cáncer de pulmón. C. I. Henschke et al., Clin. Imaging 28: 317-321 (2004). Informaron de dos estudios que indicaban que la TC puede detectar el cáncer de pulmón asintomático sin generar demasiados falsos positivos. J. Gohagan et al, Chest 126: 114-121 (2004) evaluaron un protocolo de prueba para un estudio aleatorizado que comparaba la radiografía de tórax con la tomografía computarizada espiral de baja dosis (TC) y llegaron a la conclusión de que era factible una prueba clínica aleatorizada extensa como para detectar el cáncer de pulmón. Sin embargo, incluso si se aplica en la práctica clínica, el coste de los escrutinios por TC será alto y el número de falsos positivos que conducen a ensayos adicionales será alto. Un análisis de sangre de bajo coste con buena especificidad complementará la TC para la detección temprana del cáncer. Otra estrategia para mejorar la utilidad de la TC implica el uso de un análisis de sangre de alta sensibilidad para el cáncer de

pulmón en estadio temprano. Dicho ensayo se podría ofrecer a los pacientes como una alternativa a la TC o los rayos X; si el ensayo es positivo, se obtendrían imágenes del paciente; si el ensayo es negativo, el paciente no se sometería a barrido, pero podría ser sometido a ensayo de nuevo en el futuro. Si un análisis de sangre ofrece una alta sensibilidad o una alta especificidad o, idealmente, ambas, este ensayo encontrará utilidad en los protocolos actuales que se utilizan para detectar el cáncer de pulmón en estadio temprano.

Adicionalmente, ha habido recientemente un nuevo examen de los marcadores tumorales y su utilidad cuando se combinan en paneles especiales para identificar a los individuos que están en riesgo de padecer cáncer de pulmón. Sin embargo, la falta de sensibilidad que era característica de los marcadores individuales todavía impide que los paneles de marcadores tumorales sean útiles para la detección precoz del cáncer de pulmón. Por el contrario, se sabe que un panel de marcadores de inmunoanálisis conocidos, a saber, CEA, NSE, y proGRP es útil para realizar un diagnóstico histológico del cáncer de pulmón cuando la obtención de una muestra de biopsia es difícil. (C. Gruber et al., *Tumor Biology* 27 (Suplemento 1): 71 (2006) y P. Stieber et al., *Tumor Biology*, 27 (Suplemento 2): S5-4 (2006)).

Se han hecho intentos para descubrir marcadores tumorales mejorados para el cáncer de pulmón identificando primero los componentes celulares expresados diferencialmente en el tejido tumoral de pulmón en comparación con el tejido pulmonar normal. La electroforesis en gel de poli(acrilamida) bidimensional se ha utilizado para caracterizar las diferencias cuantitativas y cualitativas en la composición de polipéptidos. 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 esta técnica, sin embargo, está limitada por el grado de resolución de proteínas de las dos etapas electroforéticas y por la etapa de detección que depende de la tinción de proteínas en geles. Asimismo, la inestabilidad del polipéptido va a generar artefactos en el patrón bidimensional.

También se han realizado intentos para identificar biomarcadores y su uso para ayudar al diagnóstico del cáncer de pulmón, tales como los descritos en J. Bréchot et al., *Eur. J. Cancer* 33(3): 385-391 (1997) Publicación Internacional Núm. WO 2005/098445 A2 del Eastern Virginia Medical School. Los biomarcadores comentados en el documento WO 2005/098445 fueron identificados mediante espectrometría de desorción/ionización por láser de superficie (SELDI). Se describen diversos marcadores, kits, métodos y un método de análisis de árboles de decisión. Sin embargo, estos marcadores, kits y métodos no han sido adoptados para su utilización en la práctica habitual, ya que estos marcadores y métodos no han sido duplicados en ningún laboratorio.

También se han hecho intentos para descubrir una respuesta inmunitaria específica para el cáncer de pulmón mediante inspección de bibliotecas de péptidos expresadas en levaduras o bacterias con sueros de individuos enfermos y no enfermos. Las publicaciones del laboratorio de Hirschowitz (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 son autoantígenos para los pacientes con cáncer de pulmón. Los autores han informado sobre la identificación satisfactoria de pacientes con cáncer de pulmón tanto sintomáticos como asintomáticos en estudios controlados. Sin embargo, el número de casos y controles es limitado (<200 sujetos en total) y el método debe ser validado en una población mucho mayor.

En la actualidad, la identificación de individuos con riesgo de cáncer de pulmón se basa en gran parte en la historia de tabaquismo de la persona. Otras exposiciones ambientales tales como el amianto, partículas, etc. pueden aumentar el riesgo de desarrollar cáncer de pulmón. Estos factores de riesgo conocidos se han combinado en uno o más algoritmos y son accesibles a los médicos y al público para evaluar el riesgo de que los individuos desarrollen el cáncer de pulmón (P. B. Bach et al., *J. Natl. Cancer Inst.* 95: 470-478 (2003)). Desafortunadamente, este algoritmo no es sensible ni es lo suficientemente específico como para ser útil para la detección del cáncer de pulmón en estadio temprano. De hecho, basándose en el algoritmo citado, un individuo con una historia significativa de tabaquismo tiene un riesgo relativo de 1/500 a 1/100 de desarrollar cáncer de pulmón. Esto significa que incluso utilizando el método de Bach et al. tantos como 499 de 500 barridos de tomografía computarizada no conducirán al descubrimiento de un caso de cáncer de pulmón. La Publicación Internacional Núm. WO 2005/017523 A2 de la Université de Genève y B. Lucas describen el uso de la apolipoproteína C-III como un biomarcador para el diagnóstico del accidente cerebrovascular.

En consecuencia, sigue habiendo una necesidad en la técnica de métodos y marcadores útiles para la detección del cáncer de pulmón que sean rápidos, prácticos y rentables de realizar. También sería ventajoso proporcionar métodos y marcadores específicos que pudieran ser utilizados para indicar la predisposición probable de un paciente o el riesgo de desarrollar cáncer de pulmón. Tales métodos podrían incluir un método para someter a ensayo una muestra de biomarcadores indicativos de cáncer de pulmón y la detección de tales marcadores. Tales métodos pueden incluir métodos mejorados de análisis de espectros de masa de una muestra biológica para los marcadores o el análisis de una muestra y la posterior detección de biomarcadores como una indicación del cáncer de pulmón o como un riesgo de desarrollar cáncer de pulmón.

Compendio de la invención

La invención se basa en parte en el descubrimiento de que los métodos rápidos y sensibles para ayudar en la detección del cáncer de pulmón en un sujeto que se sospecha que tiene cáncer de pulmón se pueden basar en ciertas combinaciones de biomarcadores y biomarcadores y parámetros biométricos.

La presente invención proporciona de este modo un método para ayudar en el diagnóstico de un sujeto que se sospecha que tiene cáncer de pulmón, tal como se define en la reivindicación 1 adjunta; definiéndose las realizaciones preferidas de dicho método en las reivindicaciones 2-16 adjuntas. La presente invención también proporciona un kit como se define en la reivindicación 17 adjunta; definiéndose las realizaciones preferidas de dicho kit en las reivindicaciones 18-19 adjuntas.

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

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

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

Opcionalmente, el método anterior puede comprender además la etapa de obtener al menos un parámetro biométrico de un sujeto. Un ejemplo de un parámetro biométrico que se puede obtener es el historial de tabaquismo del sujeto. Si el método anterior comprende además la etapa de obtención de al menos un parámetro biométrico del sujeto, a continuación, el método puede comprender además la etapa de comparar el al menos un parámetro biométrico con un punto de corte predeterminado para cada uno de dichos parámetros biométricos 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 a cada biomarcador cuantificado en la etapa c para llegar a una puntuación total de dicho sujeto en la etapa d, comparando la puntuación total, con una puntuación total predeterminada en la etapa e y determinar si dicho sujeto tiene riesgo de desarrollar cáncer de pulmón basándose en la puntuación total en la etapa f.

Los ejemplos de los biomarcadores que pueden ser cuantificados en el método anterior son uno o más biomarcadores seleccionados del grupo de anticuerpos, antígenos y regiones de interés. Más específicamente, los biomarcadores que se pueden cuantificar incluyen, pero no se limitan a, uno o más de: anti-p53, anti-TMP21, dominio anti-proteína 1 Niemann-Pick tipo C1, péptido C terminal) (dominio anti-NPC1L1C), anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSIN1, anti-RCV1, anti-MAPKAPK3, al menos un anticuerpo contra ciclina E2 inmunorreactiva, citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide A sérico, alfa-1-anti-tripsina, apolipoproteína CIII, Acn6399, Acn9459, Pub11597 Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 y HIC3959.

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

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

En el método anterior, la DFI de los biomarcadores con respecto al cáncer de pulmón es preferiblemente menor de

aproximadamente 0,4.

En el método anterior, el panel puede comprender al menos un anticuerpo seleccionado del grupo que consiste en: anti-p53, anti-TMP21, dominio anti-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSIN1, anti-RCV1, anti-MAPKAPK3 y al menos un anticuerpo contra ciclina E2 inmunorreactiva y al menos un antígeno seleccionado del grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide A sérico, alfa-1-anti-tripsina y apolipoproteína CIII.

En el método anterior, el parámetro biométrico obtenido del sujeto se selecciona del grupo que consiste en el historial de tabaquismo, la edad, la exposición a carcinógenos del sujeto y el género. Preferiblemente, el parámetro biométrico es el número de paquetes-año fumados por el sujeto.

Opcionalmente, el método puede comprender adicionalmente la cuantificación de al menos una región de interés en la muestra de ensayo. Si se va a cuantificar una región de interés en la muestra de ensayo, en ese caso, el panel puede comprender adicionalmente al menos una región de interés seleccionada del grupo que consiste en: Acn6399, Acn9459, Pub 11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 y HIC3959.

Opcionalmente, el método anterior también puede emplear un Método de Puntuación Dividida y Ponderada para determinar si un sujeto se encuentra en riesgo de desarrollar cáncer de pulmón. Si el método anterior emplea un Método de Puntuación Dividida y Ponderada, en ese caso, en dicho método, la etapa b comprende la comparación del 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 e comprende la comparación de la cantidad de cada biomarcador en el panel con un número de puntos de corte predeterminados para dicho marcador biológico y la asignación de una puntuación para cada biomarcador basándose en dicha comparación, la etapa f comprende la combinación de la puntuación asignada para cada parámetro biométrico determinado en la etapa b con la puntuación asignada a cada biomarcador cuantificado en la etapa e para llegar a una puntuación total de dicho sujeto, la etapa g comprende la comparación de la puntuación total determinada en la etapa f con un número de la puntuación total predeterminado y la etapa h comprende la determinación de si dicho sujeto tiene cáncer de pulmón basándose en la puntuación total determinada en la etapa g.

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

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

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

En el método anterior, el panel puede comprender al menos un anticuerpo seleccionado del grupo que consiste en: anti-p53, anti-TMP21, dominio anti-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSIN1, anti-RCV1, anti-MAPKAPK3 y al menos un anticuerpo contra ciclina E2 inmunorreactiva. El panel puede comprender al menos un antígeno seleccionado del grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA 15-3, SCC, CA 19-9, proGRP, amiloide A sérico, alfa-1-anti-tripsina y apolipoproteína CIII.

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

Opcionalmente, el método anterior también puede emplear una Puntuación Dividida y Ponderada para determinar si un sujeto se encuentra en riesgo de desarrollar cáncer de pulmón. Si el método anterior emplea tal Método de Puntuación Dividida y Ponderada, en ese caso, en dicho método, la etapa c comprende la comparación de la cantidad de cada biomarcador en el panel con un número de puntos de corte predeterminados para dicho

biomarcador y la asignación de una puntuación a cada biomarcador basándose en dicha comparación, la etapa d comprende la combinación de la puntuación asignada a cada biomarcador cuantificado en la etapa c para llegar a una puntuación total de dicho sujeto, la etapa e comprende la comparación de la puntuación total determinada en la etapa d con varias puntuaciones totales predeterminadas y la etapa f comprende la determinación de si dicho sujeto

5 tiene cáncer de pulmón basándose en la puntuación total determinada en la etapa e.

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

- a. obtener una muestra de ensayo de un sujeto;
- 10 b. cuantificar en la muestra de ensayo una cantidad de al menos un biomarcador en un panel, comprendiendo el panel al menos un anticuerpo contra Ciclina E2 inmunorreactiva;
- c. comparar la cantidad de cada biomarcador cuantificado en el panel con un punto de corte predeterminado para dicho biomarcador y asignar una puntuación a cada biomarcador basándose en dicha comparación;
- 15 d. combinar la puntuación asignada a cada biomarcador cuantificado en la etapa c para llegar a una puntuación total de dicho sujeto;
- e. comparar la puntuación total determinada en la etapa d con una puntuación total predeterminada; y
- f. determinar si dicho sujeto tiene cáncer de pulmón, basándose en la puntuación total determinada en la etapa e.

20 En el método anterior, la DFI de los biomarcadores con respecto al cáncer de pulmón es preferiblemente menor de aproximadamente 0,4.

Opcionalmente, el método anterior puede comprender además la cuantificación de al menos un antígeno en la muestra de ensayo, la cuantificación de al menos un anticuerpo en la muestra de ensayo, o la cuantificación de una combinación de al menos un antígeno y al menos un anticuerpo en la muestra de ensayo. Acto seguido, si se van a

25 cuantificar al menos un antígeno, al menos un anticuerpo o una combinación de al menos un antígeno y al menos un anticuerpo en la muestra de ensayo, a continuación, el panel puede comprender adicionalmente al menos un antígeno seleccionado del grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide A sérico, alfa-1-anti-tripsina y la apolipoproteína CIII, al menos un

30 anticuerpo seleccionado del grupo que consiste en: anti-p53, anti-TMP21, dominio anti-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACIN1, anti-RCV1, anti-MAPKAPK3 y al menos un anticuerpo contra ciclina E2 inmunorreactiva o cualquier combinación de los mismos.

Opcionalmente, el método puede comprender adicionalmente la cuantificación de al menos una región de interés en la muestra de ensayo. Si se va a cuantificar una región de interés, el panel puede comprender adicionalmente al menos una región de interés seleccionada del grupo que consiste en: Acn6399, Acn9459, Pub 11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub 17338, TFA6453 y HIC3959.

Opcionalmente, el método anterior también se puede emplear una Puntuación Ponderada y Fraccionada para determinar si un sujeto se encuentra en riesgo de desarrollar cáncer de pulmón. Si el método anterior emplea un Método de Puntuación Ponderada y Fraccionada, en dicho método, la etapa c comprende la comparación de la cantidad de cada biomarcador en el panel con un número de puntos de corte predeterminados para dicho biomarcador y la asignación de una puntuación a cada biomarcador basándose en dicha comparación, la etapa d

45 comprende la combinación de la puntuación asignada a cada biomarcador cuantificado en la etapa c para llegar a una puntuación total de dicho sujeto, la etapa e incluye la comparación de la puntuación total determinado en la etapa d con varias puntuaciones totales predeterminadas y la etapa f comprende la determinación de si dicho sujeto tiene cáncer de pulmón, basándose en la puntuación total determinada en la etapa e.

Opcionalmente, el método anterior puede comprender además la etapa de obtención de al menos un parámetro biométrico de un sujeto. Un parámetro biométrico que puede ser obtenido de un sujeto se puede seleccionar del grupo que consiste en: el historial de tabaquismo, la edad, la exposición a carcinógenos y e género de un sujeto. Un parámetro biométrico preferido que se obtiene es los paquetes-año de tabaquismo del sujeto. Si el método anterior comprende además la etapa de obtención de al menos un parámetro biométrico del sujeto, en ese caso, el método

50 puede comprender adicionalmente la etapa de comparación de al menos un parámetro biométrico contra un punto de corte predeterminado para cada uno de dichos parámetros biométricos y la asignación de una puntuación para cada parámetro biométrico basándose en dicha comparación, la combinación de la puntuación asignada a cada parámetro biométrico con la puntuación asignada a cada biomarcador cuantificado en la etapa c para llegar a una puntuación total de dicho sujeto, la comparación de la puntuación total, con una puntuación total predeterminada en

55 la etapa e y la determinación de si dicho sujeto presenta riesgo de cáncer de pulmón basándose en la puntuación total en la etapa f.

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

- a. obtener una muestra de ensayo de un sujeto;
- b. cuantificar en la muestra de ensayo al menos un biomarcador en un panel, comprendiendo el panel al menos un biomarcador seleccionado del grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA 15-3, SCC, CA 19- 9, proGRP, amiloide A sérico, alfa-1-anti-tripsina y apolipoproteína CIII;
- c. comparar la cantidad de cada biomarcador cuantificado en el panel con un punto de corte predeterminado de dicho biomarcador y asignar una puntuación a cada biomarcador basándose en dicha comparación;
- d. combinar la puntuación asignada a cada biomarcador cuantificado en la etapa c para llegar a una puntuación total de dicho sujeto;
- e. comparar la puntuación total cuantificada en la etapa d con una puntuación total predeterminada; y
- f. determinar si dicho sujeto tiene cáncer de pulmón basándose en la puntuación total.

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

Opcionalmente, el método anterior puede comprender adicionalmente la cuantificación de al menos un anticuerpo en la muestra de ensayo. Acto seguido, el panel puede comprender adicionalmente al menos un anticuerpo seleccionado del grupo que consiste en: anti-p53, anti-TMP21, dominio anti-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSLN1, anti-RCV1, anti-MAPKAPK3 y al menos un anticuerpo contra ciclina E2 inmunorreactiva o cualquier combinación de los mismos.

Opcionalmente, el método puede comprender adicionalmente la cuantificación de al menos una región de interés en la muestra de ensayo. Si se va cuantificar una región de interés, en ese caso, el panel puede comprender adicionalmente al menos una región de interés seleccionada del grupo que consiste en: Acn6399, Acn9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 y HIC3959.

Opcionalmente, el método anterior también se puede emplear una Puntuación Ponderada y Fraccionada para determinar si un sujeto se encuentra en riesgo de desarrollar cáncer de pulmón. Si el método anterior emplea semejante Método de Puntuación Ponderada y Fraccionada, en dicho método, la etapa c comprende la comparación de la cantidad de cada biomarcador en el panel con un número de puntos de corte predeterminados para dicho marcador biológico y la asignación de una puntuación a cada biomarcador basándose en dicha comparación, la etapa d comprende la combinación de la puntuación asignada a cada biomarcador cuantificado en la etapa c para llegar a una puntuación total de dicho sujeto, la etapa e comprende la comparación de la puntuación total determinada en la etapa d con varias puntuaciones totales predeterminadas y la etapa f comprende la determinación de si dicho sujeto tiene cáncer de pulmón, basándose en la puntuación total determinada en la etapa e.

Opcionalmente, el método anterior puede comprender adicionalmente la etapa de obtención de al menos un parámetro biométrico de un sujeto. Un parámetro biométrico que puede ser obtenido de un sujeto se puede seleccionar del grupo que consiste en: el historial de tabaquismo, la edad, la exposición a carcinógenos, y el género de un sujeto. Un parámetro biométrico preferido que se obtiene consiste en los paquetes-año de tabaquismo del sujeto. Si el método anterior comprende adicionalmente la etapa de obtención de al menos un parámetro biométrico del sujeto, en ese caso, el método puede comprender adicionalmente la etapa de comparación del al menos un parámetro biométrico frente a un punto de corte predeterminado para cada uno de dichos parámetros biométricos y la asignación de una puntuación a cada parámetro biométrico basándose en dicha comparación, la combinación de la puntuación asignada a cada parámetro biométrico con la puntuación asignada a cada biomarcador cuantificado en la etapa c para llegar a una puntuación total de dicho sujeto, la comparación de la puntuación total, con una puntuación total predeterminada en la etapa e y la determinación de si dicho sujeto presenta riesgo de desarrollar cáncer de pulmón basándose en la puntuación total en la etapa f.

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

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

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

- 5 Opcionalmente, el método anterior puede comprender además la cuantificación de al menos un antígeno en la muestra de ensayo, la cuantificación de al menos un anticuerpo en la muestra de ensayo, o la cuantificación de una combinación de al menos un antígeno y al menos un anticuerpo en la muestra de ensayo. Acto seguido, si se va a cuantificar al menos un antígeno, al menos un anticuerpo o una combinación de al menos un antígeno o anticuerpo en la muestra de ensayo, en ese caso, el panel puede comprender adicionalmente al menos un antígeno seleccionado del grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide A sérico, alfa-1-anti-tripsina y apolipoproteína CIII, al menos un anticuerpo seleccionado del grupo que consiste en: anti-p53, anti-TMP21, dominio anti-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACIN1, anti-RCV1, anti-MAPKAPK3 y al menos un anticuerpo contra ciclina E2 inmunorreactiva o cualquier combinación de los mismos.
- 10
- 15 Opcionalmente, el método anterior también puede emplear una Puntuación Ponderada y Fraccionada para determinar si un sujeto se encuentra en riesgo de desarrollar cáncer de pulmón. Si el método anterior emplea un Método de Puntuación Ponderada y Fraccionada, en ese caso, en dicho método, la etapa c comprende la comparación de la cantidad de cada biomarcador en el panel con un número de puntos de corte predeterminados para dicho marcador biológico y la asignación de una puntuación a cada biomarcador basándose en dicha comparación, la etapa d comprende la combinación de la puntuación asignada a cada biomarcador cuantificada en la etapa c para llegar a una puntuación total de dicho sujeto, la etapa e incluye la comparación de la puntuación total determinada en la etapa d con varias puntuaciones totales predeterminadas y la etapa f comprende la determinación de si dicho sujeto tiene cáncer de pulmón, basándose en la puntuación total determinada en la etapa e.
- 20
- 25 Opcionalmente, el método anterior puede comprender adicionalmente etapa de obtención de al menos un parámetro biométrico de un sujeto. Un parámetro biométrico que puede ser obtenido de un sujeto se puede seleccionar del grupo que consiste en: el historia de tabaquismo, la edad, la exposición a carcinógenos y el género de un sujeto. Un parámetro biométrico preferido que se obtiene consiste en los paquetes-año de tabaquismo del sujeto. Si el método anterior comprende adicionalmente la etapa de obtención de al menos un parámetro biométrico del sujeto, en ese caso, el método puede comprender adicionalmente la etapa de comparación del al menos un parámetro biométrico frente a un punto de corte predeterminado para cada uno de dichos parámetros biométricos y la asignación de una puntuación a cada biométrico parámetro basándose en dicha comparación, la combinación de la puntuación asignada para cada parámetro biométrico con la puntuación asignada a cada biomarcador cuantificado en la etapa c para llegar a una puntuación total de dicho sujeto, la comparación de la puntuación total, con una puntuación total predeterminada en la etapa e y la determinación de si dicho sujeto se encuentra en riesgo de cáncer de pulmón basándose en la puntuación total en la etapa f.
- 30
- 35

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

- 40 a. obtener una muestra de ensayo de un sujeto;
- b. cuantificar en la muestra de ensayo la cantidad de dos o más biomarcadores en un panel, comprendiendo el panel dos o más de: citoqueratina 19, citoqueratina 18, CA 19-9, CEA, CA 15-3, CA 125, SCC, proGRP, ACN9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 y Hic3959;
- 45 c. comparar la cantidad de cada biomarcador en el panel con un punto de corte predeterminado de dicho biomarcador y asignar una puntuación a cada biomarcador basándose en dicha comparación;
- d. combinar la puntuación asignada a cada biomarcador determinada en la etapa c para llegar a una puntuación total de dicho sujeto;
- e. comparar la puntuación total determinada en la etapa d con una puntuación total predeterminada; y
- 50 f. determinar si dicho sujeto tiene cáncer de pulmón, basándose en la puntuación total determinada en la etapa e.

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

- 55 Opcionalmente, el panel en el método anterior puede comprender: citoqueratina 19, CEA, ACN9459, Pub11597, Pub4789 y TFA2759, citoqueratina 19, CEA, ACN9459, Pub11597, Pub4789, TFA2759 y TFA9133, citoqueratina 19, CA19-9, CEA, CA15- 3, CA125, SCC, citoqueratina 18, ProGRP, Pub11597, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 y Hic3959 o citoqueratina 19, CEA, CA125, SCC, citoqueratina 18, ProGRP, ACN9459, Pub 11597, Pub4789, TFA2759, TFA9133.
- 60

Opcionalmente, el método anterior también puede emplear una Puntuación Dividida y Ponderada para determinar si un sujeto se encuentra en riesgo de desarrollar cáncer de pulmón. Si el método anterior emplea tal Método de Puntuación Dividida y Ponderada, en ese caso, en dicho método, la etapa c comprende la comparación de la

cantidad de cada biomarcador en el panel con un número de puntos de corte predeterminados para dicho marcador biológico y la asignación de una puntuación a cada biomarcador basándose en dicha comparación, la etapa d comprende la combinación de la puntuación asignada a cada biomarcador cuantificada en la etapa c para llegar a una puntuación total de dicho sujeto, la etapa e comprende la comparación de la puntuación total determinada en la etapa d con varias puntuaciones totales predeterminadas y la etapa f comprende la determinación de si dicho sujeto tiene cáncer de pulmón, basándose en la puntuación total determinada en la etapa e.

La presente descripción también se refiere a una variedad de kits diferentes de acuerdo con los siguientes aspectos que se pueden utilizar en los métodos descritos anteriormente. En un aspecto, un kit puede comprender un péptido seleccionado del grupo que consiste en: el SEQ ID NO: 1, el SEQ ID NO: 3, el SEQ ID NO: 4, el SEQ ID NO: 5 o cualquiera de sus combinaciones. En otro aspecto, un kit puede comprender al menos un anticuerpo contra 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 la cuantificación de uno o más antígenos en una muestra de ensayo, en donde dichos antígenos son: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, CA19-9, proGRP, amiloide A sérico, alfa-1-anti-tripsina y apolipoproteína CIII; (b) los reactivos que contienen uno o más antígenos para la cuantificación de al menos un anticuerpo en una muestra de ensayo; en donde dichos anticuerpos son: anti-p53, anti-TMP21, dominio anti-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSIN1, anti-RCV1, anti-MAPKAPK3 y al menos un anticuerpo contra ciclina E2 inmunorreactiva; (c) reactivos para la cuantificación de una o más regiones de interés seleccionadas del grupo que consiste en: ACN9459, Pub 11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 y Hic3959; y (d) uno o más algoritmos para la combinación y comparación de la cantidad de cada antígeno, anticuerpo y región de interés cuantificados en la muestra de ensayo frente a un punto de corte predeterminado, y asignación de una puntuación a cada antígeno, anticuerpo y región de interés basándose en dicha comparación, la combinación de la puntuación asignada a cada antígeno, anticuerpo y región de interés cuantificados para obtener una puntuación total, la comparación de la puntuación total con una puntuación total predeterminada y la utilización de dicha comparación como ayuda para determinar si un sujeto tiene cáncer de pulmón. En otro aspecto más, un kit puede comprender: (a) reactivos que contienen al menos un anticuerpo para la cuantificación de uno o más antígenos en una muestra de ensayo, en donde dichos antígenos son citoqueratina 19, citoqueratina 18, CA19-9, CEA, CA-15 -3, CA125, SCC y ProGRP; (b) reactivos para la cuantificación de una o más regiones de interés seleccionadas del grupo que consiste en: ACN9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 y Hic3959; y (c) uno o más algoritmos para la combinación y comparación de la cantidad de cada antígeno y región de interés cuantificados en la muestra de ensayo frente a un punto de corte límite predeterminado, la asignación de una puntuación para cada antígeno y biomarcador cuantificados basándose en dicha comparación, la combinación de la puntuación asignada a cada antígeno y región de interés cuantificados para obtener una puntuación total, la comparación de la puntuación total, con una puntuación total predeterminada y la utilización de dicha comparación como ayuda para determinar si un sujeto tiene cáncer de pulmón. Los ejemplos de los antígenos y las regiones de interés que pueden ser cuantificados 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, Pub 11597, Pub4789 y Tfa2759. En otro aspecto, un kit puede comprender (a) reactivos que contienen al menos un anticuerpo para la cuantificación de uno o más antígenos en una muestra de ensayo, en 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 frente a un punto de corte predeterminado, y la asignación de una puntuación a cada antígeno cuantificado basándose en dicha comparación, la combinación de la puntuación asignada a cada antígeno cuantificado para obtener una puntuación total, la comparación de la puntuación total con una puntuación total predeterminada y el uso de dicha comparación como una ayuda en la determinación de si un sujeto tiene cáncer de pulmón. Los ejemplos de antígenos que pueden ser cuantificados 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 la cuantificación de uno o más biomarcadores, en donde dichos marcadores biológicos son regiones de interés seleccionadas del grupo que consiste en: ACN9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 y Hic3959; y (b) uno o más algoritmos para la combinación y comparación de la cantidad de cada biomarcador cuantificado en la muestra de ensayo frente a un punto de corte predeterminado y la asignación de una puntuación a cada biomarcador cuantificado basándose en dicha comparación, la combinación de la puntuación asignada a cada biomarcador cuantificado para obtener una puntuación total, la comparación de la puntuación total con una puntuación total predeterminada y el uso de dicha comparación como una ayuda en la determinación de si un sujeto tiene cáncer de pulmón. Los ejemplos de las regiones de interés que se pueden cuantificar utilizando el kit se pueden seleccionar del grupo que consiste en: Pub11597, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 y Hic3959.

La presente descripción también se refiere a polipéptidos aislados o purificados. Los polipéptidos aislados o purificados contemplados por la presente invención son: (a) un polipéptido aislado o purificado que tiene (que comprende) una secuencia de aminoácidos seleccionada del grupo que consiste en: el SEQ ID NO: 3 y un polipéptido que tiene una homología de 60% con la secuencia de aminoácidos del SEQ ID NO: 3; (b) un polipéptido aislado o purificado que consiste esencialmente en una secuencia de aminoácidos seleccionada del grupo que

consiste en: el SEQ ID NO: 3 y un polipéptido que tiene una homología de 60% con la secuencia de aminoácidos del SEQ ID NO: 3; (c) un polipéptido aislado o purificado que consiste en una secuencia de aminoácidos del SEQ ID NO: 3; (d) un polipéptido aislado o purificado que tiene una secuencia de aminoácidos seleccionada del grupo que consiste en: el SEQ ID NO: 4 y un polipéptido que tiene una homología de 60% con la secuencia de aminoácidos del SEQ ID NO: 4; (e) un polipéptido aislado o purificado que consiste esencialmente en una secuencia de aminoácidos seleccionada del grupo que consiste en: el SEQ ID NO: 4 y un polipéptido que tiene una homología de 60% con la secuencia de aminoácidos del SEQ ID NO: 4; (f) un polipéptido aislado o purificado que consiste en una secuencia de aminoácidos del SEQ ID NO: 4; (g) un polipéptido aislado o purificado que tiene una secuencia de aminoácidos seleccionada del grupo que consiste en: el SEQ ID NO: 5 y un polipéptido que tiene una homología de 60% con la secuencia de aminoácidos del SEQ ID NO: 5; (h) un polipéptido aislado o purificado que consiste esencialmente en una secuencia de aminoácidos seleccionada del grupo que consiste en: el SEQ ID NO: 5 y un polipéptido que tiene una homología de 60% con la secuencia de aminoácidos del SEQ ID NO: 5; e (i) un polipéptido aislado o purificado que consiste en una secuencia de aminoácidos del SEQ ID NO: 5.

La presente descripción también se refiere a un Método de Puntuación Ponderado y Fraccionado único. Este método se puede utilizar puntura uno o más marcadores obtenidos de un sujeto. Este método puede comprender las etapas de:

- a. obtener al menos un marcador de un sujeto;
- b. cuantificar la cantidad del marcador de dicho sujeto;
- c. comparar la cantidad de cada marcador cuantificado con una serie de puntos de corte predeterminados para dicho marcador y asignar una puntuación a cada marcador basándose en dicha comparación; y
- d. combinar la puntuación asignada para cada marcador cuantificado en la etapa c para llegar a una puntuación total de dicho sujeto.

En el método anterior, los puntos de corte predeterminados se basan en las curvas ROC 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 descripción proporciona un método para determinar el riesgo de un sujeto de desarrollar una afección médica utilizando el Método de Puntuación Ponderada y Fraccionada. Este método puede comprender las etapas de:

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

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

Breve descripción de las figuras

La **Figura 1** es un diagrama de un flujo de trabajo bioinformático. Específicamente, los datos de MS y los datos de IA se sometieron a diversos métodos estadísticos. Se utilizó la regresión logística para generar curvas de Característica Operativa del Receptor (ROC) y obtener el Área Bajo la Curva (AUC) para cada marcador. Los mejores marcadores con la mayor AUC fueron seleccionados como marcadores candidato. El análisis de variables múltiples (MVA) tal como el Análisis Discriminante (DA), el Análisis de Componentes Principales (ACP) y los Árboles de decisión (DT) identificaron marcadores adicionales para la introducción en el modelo. También se pueden incluir parámetros biométricos. Los marcadores robustos que se producen en al menos 50% de las series de entrenamiento se identifican por el método/algoritmo de Puntuación Ponderada y Fraccionada (SSM) y se seleccionan como supuestos biomarcadores. El procedimiento se repite n veces hasta que se obtiene un número adecuado de marcadores para el modelo predictivo final.

La **Figura 2** es un Perfil de MS MALDI-TOF que muestra el biomarcador candidato Pub11597 a) después de concentrar las fracciones de HPLC agrupadas y b) antes del procedimiento de concentración. La muestra es una mezcla compleja incluso después del fraccionamiento mediante HPLC.

La **Figura 3** es un gel teñido que muestra los componentes de las diversas muestras cargadas en el gel. Las

calles A, F y G muestran una mezcla de proteínas patrón de masas moleculares conocidas para fines de calibración. Adicionalmente, las calles b y e muestran una forma altamente purificada de la proteína sospechosa conocida como amiloide A de suero 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 la misma distancia que el patrón de HSAA. Las bandas que tienen la misma distancia de migración que el HSAA se escindieron del gel y se sometieron a digestión en gel y análisis de MS/MS para confirmar su identidad.

La **Figura 4** es un LC-MS/MS del producto digerido con tripsina de Pub11597. Los paneles a-d muestran el MS/MS de las 4 iones precursores principales. Se han anotado los iones producto b e y se proporciona la secuencia de aminoácidos derivada para cada uno de los cuatro iones precursores. La búsqueda de base de datos utilizando las masas moleculares de los iones b e y generados identificó la proteína de origen como HSAA. La secuencia completa del fragmento observado (PM = 11526,51) se proporciona en el SEQ ID NO: 6.

La **Figura 5** proporciona curvas ROC generadas a partir de un panel de 8 biomarcadores de inmunoanálisis realizado en 751 muestras de pacientes descritas en el Ejemplo 1. Los rombos de color negro representan la curva ROC generada a partir de la puntuación total utilizando el Método de Puntuación Ponderada y Fraccionada. Los cuadrados representan la curva ROC generada a partir de la puntuación total utilizando el método de puntuación binaria utilizando puntos de fraccionamiento de una gran cohorte. Los triángulos representan la curva ROC generada a partir de la puntuación total utilizando el método de puntuación binario utilizando puntos de fraccionamiento de una pequeña cohorte.

Descripción detallada de la invención

Definiciones

Según se utiliza en esta solicitud, los siguientes términos tienen los siguientes significados. Todos los demás términos técnicos y científicos tienen el significado comúnmente entendido por los expertos en esta técnica.

El término "adsorbente" se refiere a cualquier material que es capaz de acumular (unirse a) una biomolécula. El adsorbente normalmente recubre una superficie biológicamente activa y está compuesto por un único material o una pluralidad de diferentes materiales que son 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 aniónico, materiales de intercambio catiónico, quelantes de metales, polinucleótidos, oligonucleótidos, péptidos, anticuerpos, polímeros (sintéticos o naturales), papel, etc.

Según se utiliza en la presente memoria, el término "anticuerpo" se refiere a una molécula de inmunoglobulina o una porción inmunológicamente activa de la misma, a saber, una porción de unión a antígeno. Los ejemplos de porciones inmunológicamente activas de moléculas de inmunoglobulina incluyen fragmentos F(ab) y fragmentos F(ab')₂ que pueden ser generados tratando un anticuerpo con una enzima, tal como pepsina. Los ejemplos de los 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 madurado por afinidad, anticuerpos de cadena sencilla, anticuerpos de un solo dominio, fragmentos F(ab), fragmentos F(ab'), Fv unidos por disulfuro ("sdFv"), y anticuerpos antiidiotípicos ("anti-Id") y fragmentos de unión a epítopos funcionalmente activos de cualquiera de los anteriores. Según se utiliza en la presente memoria, el término "anticuerpo" también incluye autoanticuerpos (los Autoanticuerpos son anticuerpos que sintetiza un sujeto o paciente que se dirigen hacia las propias proteínas normales (o los propios antígenos), tales como, pero no limitados a, p53, calreticulina, alfa-enolasa, y HOXB7. Los autoanticuerpos contra una amplia gama 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, el cáncer de mama, el cáncer de próstata, y el cáncer de páncreas entre otros). Un anticuerpo es un tipo de biomarcador.

Según se utiliza en la presente memoria, el término "antígeno" se refiere una molécula susceptible de ser unida por un anticuerpo y que es adicionalmente capaz de inducir la producción de anticuerpos en un animal susceptibles de unirse a al menos un epítipo de ese antígeno. Adicionalmente, una región de interés también puede ser un antígeno (en otras palabras, en última instancia puede ser determinada como 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 mérito para un ensayo en una muestra de población dada y proporciona valores que oscilan de 1 para un ensayo perfecto a 0,5 en el que el ensayo proporciona una respuesta completamente al azar en la clasificación de los sujetos de ensayo. Puesto que el intervalo de las AUC es de sólo 0,5 a 1,0, un pequeño cambio en el AUC tiene mayor trascendencia que un cambio similar en una métrica que oscila de 0 a 1 o de 0 a 100%. Cuando se proporciona un % de cambio en las AUC, éste se calculará basándose en el hecho de que toda el intervalo de la métrica es de 0,5 a 1,0 El paquete estadístico JMP[™] informa del AUC para cada curva ROC generada. Las medidas del AUC son un medio valioso para comparar la precisión del algoritmo de clasificación a través del intervalo completo de datos. Esos algoritmos de clasificación con mayor AUC tienen, por definición, una mayor capacidad para clasificar correctamente incógnitas

entre los dos grupos de interés (enfermo y no-enfermo). El algoritmo de clasificación puede ser tan simple como la medida de una sola molécula o tan complejo como la medida y la integración de múltiples moléculas.

5 El término "enfermedad pulmonar benigna" o "benigna" se refiere a un estado de enfermedad asociado 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, enfermedad pulmonar obstructiva crónica (EPOC), inflamación aguda o crónica, nódulo benigno, neoplasia benigna, displasia, hiperplasia, atipia, bronquiectasias, histoplasmosis, sarcoidosis, fibrosis, granuloma, hematoma, enfisema, atelectasia, histiocitosis y otras enfermedades no cancerosas.

10 El término "superficie biológicamente activa" se refiere a cualquier extensión de bi- o tri-dimensional de un material al que se pueden unir, o con el que pueden interaccionar 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), el carácter hidrófobo, o el carácter hidrófilo.

15 Los términos "muestra biológica" y "muestra de ensayo" se refieren 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, sangre, suero sanguíneo, plasma sanguíneo, aspirado de pezón, orina, semen, líquido seminal, plasma seminal, fluido prostático, excrementos, lágrimas, saliva, sudor, biopsia, ascitis, líquido cefalorraquídeo, leche, linfa, muestras bronquiales y otras muestras de lavado, o muestras de extracto de tejido. Típicamente, sangre, suero, plasma y lavado bronquial son las muestras de ensayo preferidas para su uso en el contexto de la presente invención.

25 El término "biomarcador" se refiere a una molécula biológica (o fragmento de una molécula biológica) que se correlaciona con una afección física. De acuerdo con la presente invención, los biomarcadores están correlacionados con el cáncer, preferiblemente, cáncer de pulmón y se pueden utilizar como ayuda en la detección de la presencia o ausencia del cáncer de pulmón. Tales biomarcadores incluyen, pero no se limitan a, biomoléculas que comprenden nucleótidos, aminoácidos, azúcares, ácidos grasos, esteroides, metabolitos, polipéptidos, proteínas (tales como, pero no limitadas a, antígenos y anticuerpos), carbohidratos, lípidos, hormonas, anticuerpos, regiones de interés que sirven como sustitutos de las moléculas biológicas, combinaciones de las mismas (p. ej., glicoproteínas, ribonucleoproteínas, lipoproteínas) y cualquier complejo que incluya cualquiera de tales biomoléculas, tales como, pero no limitadas a, un complejo formado entre un antígeno y un autoanticuerpo que se une a un epítipo disponible en dicho antígeno. El término "biomarcador" también se puede referir a una porción de una secuencia de polipéptido (parental) que comprende al menos 5 residuos de aminoácido consecutivos, preferiblemente al menos 10 residuos de aminoácido consecutivos, más preferiblemente al menos 15 residuos de aminoácido consecutivos, y conserva una actividad biológica y/o algunas de las características funcionales del polipéptido parental, p. ej., antigenicidad o características de dominios estructurales.

40 El término "parámetro biométrico" se refiere a uno o más rasgos físicos o de comportamiento intrínsecos utilizados para identificar de manera única a los pacientes como pertenecientes a un grupo o población bien definidos. En el contexto de esta invención, "parámetro biométrico" incluye pero no se limita a, descriptores físicos de un paciente. Los ejemplos de un parámetro biométrico incluyen, pero no se limitan a, la altura de un paciente, el peso de un paciente, el sexo de un paciente, el historial de tabaquismo, el historial ocupacional, la exposición a carcinógenos, la exposición a humo ajeno, el historial familiar de cáncer de pulmón, y similares. El historial de tabaquismo por lo general se cuantifica en términos de paquetes-año (Pkyrs). Según se utiliza en la presente memoria, el término 45 "Paquetes Año" se refiere al número de años que una persona ha fumado, multiplicado por el número medio de paquetes fumados por día. Se dice que una persona que ha fumado, de media, 1 paquete de cigarrillos al día durante 35 años tiene un historial de tabaquismo de 35 paquetes-año. Se puede obtener información de los parámetros biométricos de un sujeto utilizando mecanismos de rutina conocidos en la técnica, por ejemplo del propio sujeto mediante el uso de un cuestionario para los pacientes de rutina o un cuestionario del historial de salud, etc. 50 Como alternativa, el parámetro biométrico se puede obtener de una enfermera, una enfermera practicante, asistente médico o un médico del sujeto.

Una "sustitución conservativa de aminoácidos" es una en la que el residuo de aminoácido es reemplazado por un residuo de aminoácido que tiene una cadena lateral similar. Las familias de residuos de aminoácidos que tienen 55 cadenas laterales similares han sido definidas en la técnica. Estas familias incluyen aminoácidos con cadenas laterales alcalinas (p. ej., lisina, arginina, histidina), cadenas laterales ácidas (p. ej., ácido aspártico, ácido glutámico), cadenas laterales polares no cargadas (p. ej., glicina, asparragina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales no polares (p. ej., alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales ramificadas en beta (p. ej., treonina, valina, isoleucina) y cadenas laterales aromáticas (p. ej., 60 tirosina, fenilalanina, triptófano, histidina). Así, un residuo de aminoácido no esencial pronosticado en una proteína se reemplaza preferiblemente por otro residuo de aminoácido de la misma familia de cadena lateral.

La expresión "Análisis de Árbol de Decisiones" se refiere al enfoque clásico en el que una serie de reglas dicotómicas simples (o síntomas) proporcionan una guía a través de un árbol de decisiones para un resultado de

clasificación final o nodo terminal del árbol. Su naturaleza inherentemente sencilla e intuitiva hace que el reparto recursivo sea muy susceptible de un procedimiento de diagnóstico.

El método requiere dos tipos de variables: variables de factor (X) y variables de respuesta (Y 's). Tal como se aplican, 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 del punto de corte, lo que maximiza la pureza de los nodos. Las muestras se dividen en ramas o nodos basándose en valores que están por encima y por debajo de este valor del punto de corte.

Para la variable de respuesta categórica, como en este caso, el valor ajustado se convierte en la probabilidad estimada para cada nivel de respuesta. En este caso, la división se determina por el estadístico chi-cuadrado de razón de verosimilitud (G^2). Esto tiene el efecto de maximizar la diferencia en las respuestas entre las dos ramas de la división. Una discusión más detallada del método y de su aplicación se puede encontrar en la guía JMP Statistics and Graphics.

La construcción de un árbol, sin embargo, tiene sus propios problemas asociados. Una problema es decidir el tamaño óptimo del árbol que proporcionará el mejor modelo predictivo sin un sobreajuste de los datos. Con esto en mente, se desarrolló un método que hiciera uso de la información que se puede extraer en los diferentes nodos del árbol para construir una curva ROC. Tal como se aplica, el método implica la construcción de un árbol de referencia con suficientes nodos que seguramente va a sobreajustar los datos que se están modelando. Posteriormente, el árbol se poda de nuevo, eliminando sucesivamente el peor nodo 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).

El programa de reparto recursivo intenta crear nodos terminales puros, es decir, sólo se incluyen especímenes de un tipo de clasificación. Sin embargo, esto no siempre es posible. A veces los nodos terminales tienen poblaciones mixtas. Por lo tanto, cada nodo terminal tendrá una probabilidad diferente para el cáncer. En un nodo terminal puro para cáncer, la probabilidad de que sea un espécimen de cáncer será de 100% y a la inversa, para un nodo terminal puro para no cáncer, la probabilidad de que sea un espécimen de cáncer será de 0%. La probabilidad de cáncer en cada nodo terminal se representa frente a (1-probabilidad de no cáncer) en cada nodo.

Estos valores se trazan para generar una curva ROC que sea representativa de ese árbol en particular. El AUC calculada para cada árbol representa la "bondad" del árbol o modelo. Al igual que en cualquier aplicación de diagnóstico, cuanto mayor es el AUC, mejor es el análisis, o en este caso el modelo. Un gráfico de AUC frente al tamaño del árbol (número de nodos) tendrá como máximo el mejor modelo para la serie de entrenamiento. Un procedimiento similar se lleva a cabo con una segunda, pero más pequeña subgrupo de los datos para validar los resultados. Los modelos que tienen un rendimiento similar en ambos grupos de validación y entrenamiento se consideran óptimos y, por tanto, se eligen para su posterior análisis y/o validación.

Los términos "conjunto de datos de desarrollo" o "conjunto de datos" se refieren a las características que incluyen los datos del biomarcador completo o el biomarcador y los parámetros biométricos recogidos para un conjunto de muestras biológicas. Estas mismas muestras se extraen de los pacientes con resultados de diagnóstico conocidos. Una característica o conjunto de características se someten a un análisis estadístico con miras a una clasificación de las muestras en dos o más grupos de muestras diferentes (p. ej., cáncer y no cáncer) que se correlacionan con los resultados de los pacientes conocidos. Cuando se utilizan los espectros de masas, los espectros de masas en el conjunto pueden diferir en sus intensidades, pero no en sus masas moleculares aparentes dentro de la precisión del aparato.

El término "clasificador" se refiere a cualquier algoritmo que utiliza las características derivadas de un conjunto de muestras para determinar la enfermedad asociada con la muestra. Un tipo de clasificador se crea por "validación" del algoritmo con datos del conjunto de validación y cuyo rendimiento se evalúa con los datos del conjunto de ensayo. Los ejemplos de clasificadores utilizados junto con la invención son el análisis discriminante, el análisis de árbol de decisión, las curvas de operador receptor o el análisis de división y puntuación.

El término "árbol de decisión" se refiere a un clasificador con una estructura de árbol de tipo diagrama de flujo empleado para la clasificación. Los árboles de decisión consisten en repetidas divisiones de un conjunto de datos en subconjuntos. Cada división consiste en una simple regla aplicada a una variable, p. ej., "si el valor de la 'variable 1' mayor que el 'umbral 1'; en ese caso ir a la izquierda, de lo contrario ir a la derecha". En consecuencia, el espacio de características dadas se reparte en un conjunto de rectángulos asignándose cada rectángulo a una clase.

Los términos "análisis de diagnóstico" y "método de diagnóstico" se refieren a la detección de la presencia o naturaleza de una afección patológica. Los análisis de diagnóstico difieren en su sensibilidad y especificidad. Los sujetos con un ensayo positivo para el cáncer de pulmón y que realmente están enfermos son considerados "verdaderos positivos". En el contexto de la invención, la sensibilidad de un análisis de diagnóstico se define como el

porcentaje de verdaderos positivos en la población enferma. Los sujetos que tienen cáncer de pulmón, pero que no son detectados por el análisis de diagnóstico son considerados "falsos negativos". Los sujetos que no están enfermos y que obtienen un resultado negativo en el análisis de diagnóstico son considerados "verdaderos negativos". El término especificidad de un análisis de diagnóstico, según se utiliza en la presente memoria, se define como el porcentaje de los verdaderos negativos en la población no enferma.

El término "análisis discriminante" se refiere a un conjunto de métodos estadísticos utilizados para seleccionar las características que discriminan de manera óptima entre dos o más grupos que existen naturalmente. La aplicación del análisis discriminante a un conjunto de datos permite al usuario centrarse en las características más discriminantes para su posterior análisis.

La expresión "Distancia del Ideal" o "DFI" se refiere a un parámetro tomado de una curva ROC que es la distancia desde ideal, que incorpora tanto sensibilidad como especificidad y se define como $[(1-\text{sensibilidad})^2 + (1-\text{especificidad})^2]^{1/2}$. La DFI es 0 para un análisis con un rendimiento de sensibilidad de 100% y de especificidad de 100% y aumenta a 1,414 para un análisis con una sensibilidad de 0% y una especificidad de 0%. A diferencia de AUC que utiliza el intervalo completo de datos para su determinación, la DFI mide el rendimiento de un ensayo en particular sobre la curva ROC. Los ensayos con valores más bajos de DFI funcionan mejor que aquellos con valores más altos de DFI. La DFI se discute en detalle en la Publicación de la Solicitud de Patente de los Estados Unidos Núm. 2006/0211019 A1.

Los términos "ensamble", "ensamble de árboles" o "clasificador de ensamble" se pueden utilizar indistintamente y se refieren a un clasificador que se compone de muchos clasificadores elementales más simples, p. ej., un conjunto de árboles de decisión es un clasificador que consiste en árboles de decisión. El resultado del clasificador de ensamble se obtiene mediante la combinación de todos los resultados de sus clasificadores constituyentes, p. ej., mediante votación por mayoría que pondera por igual todos los clasificadores constituyentes. La votación por mayoría es especialmente razonable cuando los clasificadores constituyentes se ponderan a continuación naturalmente por la frecuencia con la que se generan.

Se pretende que el término "epítipo" se refiera a aquella porción de un antígeno susceptible de ser unida por un anticuerpo que también puede ser reconocida por ese anticuerpo. Los determinantes epitópicos consisten habitualmente en agrupaciones superficiales químicamente activas de moléculas tales como cadenas laterales de aminoácidos o azúcares y tienen características estructurales tridimensionales específicas, así como características de carga específicas.

Los términos "característica" y "variable" se pueden utilizar indistintamente y se refieren al valor de una medida de una característica de una muestra. Estas medidas pueden derivar de las características físicas, químicas o biológicas de la muestra. Los ejemplos de las medidas incluyen, pero no se limitan a, un pico del espectro de masas, una señal del espectro de masa, en función de la intensidad de un ROI.

Los cálculos de homología o identidad de secuencia entre secuencias (los términos se utilizan indistintamente en la presente memoria) se realizan como sigue.

Para determinar el porcentaje de identidad de dos secuencias de aminoácidos o de dos secuencias de ácidos nucleicos, las secuencias se alinean con el fin de una comparación óptima (por ejemplo, pueden introducirse huecos en una o ambas de una primera y una segunda secuencias de aminoácidos o de ácido nucleico para un óptimo alineamiento y se pueden desechar secuencias no homólogas a efectos de comparación). En una realización preferida, la longitud de una secuencia de referencia alineada para propósitos de comparación es al menos 30%, preferiblemente al menos 40%, más preferiblemente al menos 50%, incluso más preferiblemente al menos 60%, e incluso más preferiblemente al menos 70%, 80%, 90%, 95%, 99% o 100% de la longitud de los residuos de aminoácidos de la secuencia de referencia alineada. A continuación se comparan los residuos de aminoácidos o los nucleótidos en las posiciones de aminoácidos o correspondientes posiciones de nucleótidos. Cuando una posición en la primera secuencia está ocupada por el mismo residuo de aminoácido o nucleótido que la posición correspondiente en la segunda secuencia, en ese caso las moléculas son idénticas en esa posición (según se utiliza en la presente memoria "identidad" de aminoácido o de ácido nucleico amino "identidad" es equivalente a "homología" de aminoácido o de ácido nucleico). El porcentaje de identidad entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias, teniendo en cuenta el número de huecos, y la longitud de cada hueco, que han de ser introducidas para el alineamiento óptimo de las dos secuencias.

La comparación de secuencias y la determinación del porcentaje de identidad entre dos secuencias se pueden lograr 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 soporte lógico GCG, utilizando ya sea una matriz Blossum 62 ya sea una matriz PAM250, y un peso de hueco de 16, 14, 12, 10, 8, 6, o 4 y un peso de longitud de 1, 2, 3, 4, 5, o 6. En otra realización preferida más, el porcentaje de identidad entre dos secuencias de

nucleótidos se determina utilizando el programa GAP en el paquete de soporte lógico GCG, utilizando una matriz NWSgapdna. CMP y un peso de hueco de 40, 50, 60, 70, o 80 y un peso de longitud de 1, 2, 3, 4, 5, o 6. Un conjunto particularmente preferido de parámetros (y el se debe utilizar si el profesional no está seguro acerca de qué parámetros deben aplicarse para determinar si una molécula está dentro de una limitación de identidad u homología de secuencia de la presente descripción) utiliza una matriz de puntuación Blossum 62 con una penalización por apertura de hueco de 12, una penalización de extensión de hueco de 4, y una penalización por hueco de desplazamiento del marco 5.

El porcentaje de identidad entre dos secuencias de aminoácidos o de nucleótidos se puede 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 residuos de peso PAM 120, una penalización de longitud de hueco de 12 y una penalización por hueco de 4.

Las secuencias de ácido nucleico y de proteínas descritas en la presente memoria se pueden utilizar como una "secuencia problema" para realizar una búsqueda frente a bases de datos públicas para, por ejemplo, identificar otros miembros de la familia o secuencias relacionadas. Dichas búsquedas pueden realizarse 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 realizar 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 descripción. Para obtener alineamientos con huecos con fines de comparación, se puede utilizar Gapped BLAST como describen Altschul et al., Nucleic Acids Res. 25(17): 3389 a 3402 (1997). Cuando se utilizan los programas BLAST y Gapped BLAST, se pueden utilizar los parámetros por defecto de los programas respectivos (por ejemplo, XBLAST y NBLAST).

Según se utiliza en la presente memoria, el término "ciclina E2 inmunorreactiva" se refiere a (1) un polipéptido que tiene una secuencia de aminoácidos de cualquiera de los SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, o SEQ ID NO: 5; (2) cualquier combinación de cualquiera de los 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 homóloga en es al menos 60%, preferiblemente al menos 70%, más preferiblemente al menos 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% al SEQ ID NO: 1, un polipéptido que tiene una secuencia de aminoácidos que es homóloga en al menos 60%, preferiblemente al menos 70%, más preferiblemente al menos 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% al SEQ ID NO: 3, un polipéptido que tiene una secuencia de aminoácidos que es homóloga en al menos 60%, preferiblemente al menos 70%, más preferiblemente al menos 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% homóloga a SEQ ID NO: 4, un polipéptido que tiene una secuencia de aminoácidos que es al menos 60%, preferiblemente al menos 70%, más preferiblemente al menos 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% al SEQ ID NO: 5 y cualquier combinación de los mismos; (4) un polipéptido de ciclina E2 que presenta una inmunorreactividad similar a la de los 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 la del SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 o SEQ ID NO: 5.

Un polipéptido o proteína "aislado" o "purificado" está sustancialmente libre de material celular u otras proteínas contaminantes de la fuente de la célula o tejido de la que deriva la proteína, o sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetizan químicamente. Cuando una proteína o una porción biológicamente activa de la misma se produce recombinantemente, también está preferiblemente sustancialmente libre de medio de cultivo, es decir, el medio de cultivo representa menos de aproximadamente 20%, más preferiblemente menos de aproximadamente 10%, y lo más preferiblemente menos de aproximadamente 5% del volumen de la preparación de proteína.

Según se utiliza en la presente memoria, la expresión "Análisis Discriminante Lineal" se refiere a un tipo de análisis que proporciona una herramienta para identificar aquellas variables o características que son mejores en la categorización correcta de una muestra y que se puede implementar, por ejemplo, mediante el paquete estadístico JMP™. Utilizando las característica por etapas del soporte lógico, se pueden añadir variables a un modelo hasta que este clasifica correctamente todas las muestras. En general, el conjunto de variables seleccionadas de esta manera es sustancialmente menor que el número original de variables en el conjunto de datos. Esta reducción en el número de características simplifica cualquier análisis siguiente, por ejemplo, el desarrollo de un motor de clasificación más general utilizando árboles de decisión, redes neuronales artificiales, o similares.

El término "cáncer de pulmón" se refiere a un estado de cáncer asociado con el sistema pulmonar de un sujeto determinado. En el contexto de la presente invención, los cánceres de pulmón incluyen, pero no se limitan a, adenocarcinoma, carcinoma epidermoide, carcinoma de células escamosas, carcinoma de células grandes, carcinoma de células pequeñas, carcinoma de células no pequeñas, y carcinoma broncoalveolar. En el contexto de la presente invención, los cánceres de pulmón pueden estar en diferentes estadios, así como en diversos grados de clasificación. Los métodos para determinar el estadio de un cáncer de pulmón o de su grado de clasificación son bien conocidos para los expertos en la técnica.

El término "espectrometría de masas" se refiere a la utilización de una fuente de ionización para generar iones en fase gaseosa de una muestra sobre una superficie y la detección de los iones en fase gaseosa con un espectrómetro de masas. El término "espectrometría de masas por desorción láser" se refiere al uso de un láser como una fuente de ionización para generar iones en fase gaseosa de una muestra sobre una superficie y detectar los iones en fase gaseosa con un espectrómetro de masas. Un método preferido de la espectrometría de masas para biomoléculas es la espectrometría de masas con ionización por desorción láser asistida por matriz o MALDI. En MALDI, el analito se mezcla típicamente con un material de matriz que, tras el secado, co-cristaliza con el analito. El material de la matriz absorbe la energía de la fuente de energía que de otro modo fragmentaría las biomoléculas lábiles o analitos. Otro método preferido es espectrometría de masas por desorción/ionización láser inducida en superficie o SELDI. En la SELDI, la superficie sobre la que se aplica el analito desempeña un papel activo en la captura y/o de desorción del analito. En el contexto de la invención, la muestra comprende una muestra biológica que puede haberse sometido a procesamiento cromatográfico u otro procesamiento químico y un sustrato de matriz adecuado.

En la espectrometría de masas la "masa molecular aparente" se refiere a la masa molecular (en Daltons) con respecto a Valor de carga, m/z , de los iones detectados. Cómo se obtiene la masa molecular aparente depende del tipo de espectrómetro de masas utilizado. Con un espectrómetro de masas de tiempo de vuelo, la masa molecular aparente es una función del tiempo desde la ionización a la detección.

El término "matriz" se refiere a una molécula que absorbe energía en forma de fotones a partir de una fuente de luz apropiada, por ejemplo un UV/Vis o láser IR, en un espectrómetro de masas permitiendo de ese modo la desorción de una biomolécula de una superficie. Los derivados del ácido cinámico incluyendo el ácido α -ciano cinámico, el ácido sinapínico y el ácido dihidroxibenzoico se utilizan con frecuencia en forma de moléculas que absorben energía en la desorción por láser de biomoléculas. Las moléculas que absorben energía se describen en la Patente de los Estados Unidos Núm. 5.719.060, que se incorpora a la presente memoria como referencia.

El término "normalización" y sus derivados, cuando se utilizan junto con los espectros de masas, se refieren a los métodos matemáticos que se aplican a un conjunto de espectros de masas para eliminar o reducir al mínimo las diferencias, principalmente debido a parámetros instrumentales, en las intensidades totales de los espectros.

El término "región de interés" o "ROI" se refiere a una adaptación estadística de un subconjunto de un espectro de masas. Una ROI tiene una longitud mínima fija de señales consecutivas. Las señales consecutivas pueden contener huecos de longitud máxima fija en función de cómo se elija la ROI. Las regiones de interés están relacionadas con biomarcadores y pueden servir como sustitutos de los biomarcadores. Más tarde se pueden determinar las regiones de interés para una proteína, polipéptido, antígeno, anticuerpo, lípido, hormona, carbohidrato, etc.

La expresión "Curva de Características Operativas del Receptor" o "curva ROC" se refiere, en su aplicación más simple, a un gráfico de la actuación de una característica particular (por ejemplo, un biomarcador o parámetro biométrico) al distinguir entre dos poblaciones (por ejemplo, los casos (es decir, aquellos sujetos que padecen cáncer de pulmón) y los controles (es decir, aquellos sujetos que son normales o tienen diagnóstico benigno de cáncer de pulmón)). Los datos de la característica a través de toda la población (es decir, los casos y los controles), se clasifican en orden ascendente basándose en el valor de una sola característica. A continuación, para cada valor de dicha característica, se calculan las tasas de verdaderos positivos y falsos positivos para los dato. La tasa de verdaderos positivos se determina contando el número de casos por encima del valor para esa característica en cuestión y luego dividiendo por el número total de casos. La tasa de falsos positivos se determina contando el número de controles por encima del valor para esa característica en cuestión y luego dividiendo por el número total de los controles. Aunque esta definición ha descrito un escenario en el que una característica es elevada en los casos en comparación con los controles, esta definición también abarca un escenario en el que una característica está suprimida en los casos en comparación con los controles. En este escenario, se contarían las muestras por debajo del valor de esa característica en cuestión.

Se pueden generar las curvas ROC para una sola función, así como para otras salidas individuales, por ejemplo, una combinación de dos o más características se combinan matemáticamente (por ejemplo, se suman, restan, multiplican, etc.) juntas para proporcionar un solo valor de suma, este valor de la suma única se puede trazar en una curva ROC. Además, se puede representar en una curva ROC cualquier combinación de características múltiples, por lo que la combinación deriva a un valor de salida único. Estas combinaciones de características pueden comprender un ensayo. La curva ROC es la gráfica de la verdadera tasa positiva (sensibilidad) de un ensayo contra la tasa de falsos positivos (1-especificidad) del ensayo. El área bajo la curva ROC es una figura de mérito para la característica de una población de muestras determinada y proporciona valores que oscilan de 1 para una ensayo perfecto a 0,5 en el que el ensayo proporciona una respuesta completamente al azar en la clasificación de los sujetos del ensayo. Las curvas ROC proporcionan otro medio para escrutar rápidamente un conjunto de datos. Las características que parecen ser de diagnóstico se pueden utilizar preferentemente para reducir el tamaño de los grandes espacios de características.

El término "escrutinio" se refiere a una decisión de diagnóstico en relación con la disposición del paciente hacia el cáncer de pulmón. Se determina que un paciente presenta alto riesgo de cáncer de pulmón con un "ensayo de escrutinio" positivo. Como resultado, se pueden realizar al paciente ensayos adicionales, p. ej., formación de imágenes, ensayos de esputo, ensayos de función pulmonar, broncoscopia y/o procedimientos de biopsia y efectuar un diagnóstico definitivo.

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

La expresión "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 en todas las muestras. Se determina un valor del punto de corte para cada medición. Este valor del punto de corte se puede ajustar para maximizar la exactitud de las clasificaciones correctas entre los grupos de interés (p. ej., enfermo y no enfermo) o se puede configurar para maximizar la sensibilidad o especificidad de un grupo. Para cada medición, se determina si el grupo de interés, p. ej., enfermo, se encuentra por encima del valor corte o por debajo del valor del punto de corte. Para cada medición, se asigna una puntuación a esa muestra cada vez que se encuentra que el valor de la medición está en el lado enfermo del valor del punto de corte. Después de tomar todas las mediciones en una sola muestra, se suman las puntuaciones para producir una puntuación total para el panel de mediciones. Es común ponderar por igual todas las mediciones de tal manera que un panel de 10 mediciones pueda tener una puntuación máxima de 10 (teniendo cada medición una puntuación de 1 o 0) y una puntuación mínima de 0. Sin embargo, puede ser valioso ponderar las mediciones de manera desigual con una puntuación individual mayor para las mediciones más importantes.

Después de determinar las puntuaciones totales, una vez más, se determina un punto de corte para la clasificación de las muestras enfermas de las no enfermas basándose en el panel de mediciones. Una vez más, para un panel de mediciones con una puntuación máxima de 10 y una puntuación mínima de 0, se puede seleccionar 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 entre 0 - 10 como punto de corte).

Según se utiliza en la presente memoria, la expresión "Método de Puntuación Dividida y Ponderada" se refiere a un método que implica la conversión de la medición de un biomarcador o un parámetro biométrico (denominados colectivamente en la presente memoria uno o varios "marcadores") que se identifican y cuantifican en una muestra de ensayo en una de las muchas posibles puntuaciones. Las puntuaciones se obtienen utilizando la siguiente ecuación:

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

en 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 elegido que es menor que o igual a 1. La magnitud del "factor" aumenta para los marcadores que tienen un rendimiento clínico mejorado, tales como, pero no limitados a, los valores de AUC superiores, las desviaciones típicas relativamente pequeñas, la alta especificidad o sensibilidad o la baja DFI. Acto seguido, se puede convertir la medición de un marcador en tantas o tan pocas puntuaciones como se desee. Este método se basa en la curva de Características Operativas del Receptor que refleja el rendimiento del marcador/ensayo en la población de interés. La curva ROC es el gráfico de tasa de verdaderos positivos (sensibilidad) de un ensayo frente a la tasa de falsos positivos (1-especificidad) del ensayo. Cada punto de la curva representa un único valor de la característica/ensayo (marcador) que se está midiendo. Por lo tanto, algunos valores tendrán una tasa de falsos positivos en la población de interés (es decir, los sujetos con riesgo de desarrollar cáncer de pulmón), mientras que otros valores de la característica tendrán altas tasas de falsos positivos en esa población. Este método proporciona puntuaciones mayores para los valores de las características (es decir, biomarcadores o parámetros biométricos) que tienen bajas tasas de falsos positivos (teniendo por ello una alta especificidad) para la población de sujetos de interés. El método consiste en la elección de los niveles deseados de falsa positividad (1-especificidad) por debajo de los cuales el ensayo dará como resultado una puntuación mayor. En otras palabras, a los marcadores que son altamente específicos se les asigna una puntuación mayor o un mayor intervalo de las puntuaciones que a los marcadores que son menos específicos.

Según se utiliza en la presente memoria, el término "sujeto" se refiere a un animal, preferiblemente un mamífero, incluyendo un ser humano o no humano. Los términos paciente y sujeto pueden utilizarse indistintamente en la presente memoria.

La expresión "Validación de modelos de DT de diez grupos" se refiere al hecho de que las buenas prácticas de análisis requieren que los modelos sean validados frente a una nueva población para evaluar su valor predictivo. En lugar de una nueva población, los datos se pueden dividir en conjuntos de entrenamiento y conjuntos de validación independientes. Se generan diez subconjuntos aleatorios para su utilización como conjuntos de validación. Para cada conjunto de validación, hay un correspondiente conjunto de entrenamiento independiente que no tiene muestras en común. Se generan diez modelos DT a partir de los diez conjuntos de entrenamiento, como se ha descrito anteriormente y se interrogan con los conjuntos de validación.

Los términos "conjunto de ensayo" o "desconocido" o "conjunto de validación" se refieren a un subconjunto de todo el conjunto de datos disponibles que consiste en aquellas entradas que no se incluyen en el conjunto de entrenamiento. Datos de ensayo se aplica de datos para evaluar el rendimiento del clasificador.

Los términos "grupo de entrenamiento" o "conjunto conocido" o "conjunto de referencia" se refieren a un subconjunto de la totalidad del respectivo conjunto de datos disponibles. Este subconjunto se selecciona típicamente al azar, y se utiliza únicamente para el propósito de la construcción del clasificador.

El término "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 una serie de características y que permite un análisis de la curva ROC. Este enfoque se aplica mejor a un conjunto reducido de características, ya que supone un modelo simplista de la relación de las características entre sí. Un resultado positivo sugiere que podrían ser satisfactorios métodos de clasificación más sofisticados. Un resultado negativo, aunque sea decepcionante no implica necesariamente el fracaso de otros métodos.

Polipéptidos de ciclina E2

La presente descripción se refiere a polipéptidos de ciclina E2 inmunorreactiva aislados o purificados o a fragmentos biológicamente activos de los mismos que pueden ser utilizados como inmunógenos o antígenos para elevar o someter a ensayo (o más generalmente, para unirse a) anticuerpos que se pueden utilizar en los métodos descritos en la presente memoria. Los polipéptidos de ciclina E2 inmunorreactiva de la presente descripción se pueden aislar de fuentes de células o tejido utilizando técnicas de purificación de proteínas convencionales. Alternativamente, los polipéptidos de ciclina E2 inmunorreactiva aislados o purificados y los fragmentos biológicamente activos de los mismos se pueden producir por técnicas de ADN recombinante o se pueden sintetizar químicamente. Los polipéptidos de ciclina E2 inmunorreactiva aislados o purificados de la presente descripción tienen las secuencias de aminoácidos mostradas en los SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 y SEQ ID NO: 5. El SEQ ID NO: 1 es la secuencia de aminoácidos de una forma expresada de ADNc de ciclina E2 humana (Acceso Genbank BC007015.1). El SEQ ID NO: 3 es una secuencia de 38 aminoácidos que comprende el extremo C de BC007015.1 más un aminoácido (cisteína) y también se denomina en la presente memoria "E2-1". El SEQ ID NO: 4 tiene 37 aminoácidos de longitud y es idéntico al SEQ ID NO: 3, excepto que el SEC ID NO: 4 no contiene, en su extremo amino, la primera cisteína del SEQ ID NO: 3. El SEQ ID NO: 5 es una secuencia de 19 aminoácidos que comprende el extremo C de BC007015.1 y se denomina en la presente memoria "E2-2". Como se describe con más detalle en los Ejemplos, la inmunorreactividad del SEQ ID NO: 1 se comparó con la inmunorreactividad del SEQ ID NO: 2. El SEQ ID NO: 2 es otra forma expresada de ADNc de ciclina E2 humana (Acceso Genbank BC020729.1). Se ha encontrado que el SEQ ID NO: 1 muestra una fuerte inmunorreactividad con varias agrupaciones de muestras de cáncer y muestra una reactividad mucho más baja con agrupaciones (no cancerosas) benignas y normales. En cambio, el SEQ ID NO: 2 mostró poca reactividad con cualquier muestra agrupada de cáncer o sin cáncer. Se determinó que la inmunorreactividad del SEQ ID NO: 1 era el resultado de los primeros 37 aminoácidos presentes en el extremo C del SEQ ID NO: 1 que no están presentes en el SEQ ID NO: 2. Se ha demostrado que los SEQ ID NO: 3 y 5, que son ambos derivados del extremo C del SEQ ID NO: 1, muestran una fuerte inmunorreactividad entre las agrupaciones cancerosas y no cancerosas. Por lo tanto, los anticuerpos generados contra cualquiera de los SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 y SEQ ID NO: 5 o cualquier combinación de estas secuencias (tales como, los anticuerpos generados contra los SEQ ID NO: 1 y SEQ ID NO: 3, los anticuerpos generados contra los SEQ ID NO: 1 y SEQ ID NO: 4, los anticuerpos generados contra los SEQ ID NO: 1 y SEQ ID NO: 5, los anticuerpos generados contra los SEQ ID NO: 1, SEQ ID NO: 3 y SEQ ID NO: 4, los anticuerpos generados contra los SEQ ID NO: 1, SEQ ID NO: 3 y SEQ ID NO: 5, los anticuerpos generados contra los SEQ ID NO: 1, SEQ ID NO: 4 y SEQ ID NO: 5, los anticuerpos generados contra los SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 y SEQ ID NO: 5, los anticuerpos generados contra los SEQ ID NO: 3 y SEQ ID NO: 4, los anticuerpos generados contra los SEQ ID NO: 3 y SEQ ID NO: 5, los anticuerpos generados contra los SEQ ID NO: 3, SEQ ID NO: 4 y SEQ ID NO: 5, los anticuerpos generados contra los SEQ ID NO: 4 y SEQ ID NO: 5) se puede utilizar en los métodos descritos en la presente memoria. Por ejemplo, tales anticuerpos pueden ser anticuerpos sujetos generados contra cualquiera de los 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 pueden ser incluidos en uno o más kits para su uso en los métodos de la presente invención descrita en la presente memoria.

La presente descripción también abarca polipéptidos que difieren de los polipéptidos descritos en la presente

memoria (es decir, los SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 y SEQ ID NO: 5) por una o más sustituciones conservativas de aminoácidos. Además, la presente descripción también abarca polipéptidos que tienen una similitud de secuencia global (identidad) o una homología de al menos 60%, preferiblemente al menos 70%, más preferiblemente al menos 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 de que tiene la secuencia de aminoácidos del SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4 y SEQ ID NO: 5.

Uso de biomarcadores y parámetros biométricos en la detección de la presencia de cáncer de pulmón

La presente invención se refiere a métodos que ayudan eficazmente en la diferenciación entre los sujetos normales y aquellos con cáncer de pulmón o que están en riesgo de desarrollar cáncer de pulmón. Se considera que los sujetos normales son aquellos que no han sido diagnosticados de ninguna afección médica, tal como cáncer, más preferiblemente los que no han sido diagnosticados de cáncer de pulmón.

La presente invención proporciona ventajosamente métodos rápidos, sensibles y fáciles de utilizar para ayudar en el diagnóstico del cáncer de pulmón. Por otra parte, la presente invención se puede utilizar para identificar a los individuos en riesgo de desarrollar una afección médica, para escrutar a los sujetos en riesgo de una afección médica, para controlar a los pacientes diagnosticados de o que reciben tratamiento para una afección médica, y para controlar la eficacia del tratamiento de un paciente que está siendo tratado por una afección médica, en donde la afección médica es cáncer de pulmón.

En general, los métodos de la presente invención implican el uso de una muestra de ensayo de un sujeto. Típicamente, se obtiene de un sujeto una muestra de ensayo y se procesa utilizando métodos convencionales conocidos 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 punción de la vena, o, si se requiere un volumen menor, por un pinchazo en el dedo. En ambos casos, los elementos y los coágulos formados se separan por centrifugación. La orina o las heces pueden ser recogidas directamente del paciente con la condición de que sean tratadas rápidamente o estabilizadas con conservantes si el proceso no puede realizarse inmediatamente. Las muestras más especializadas, como los lavados bronquiales o líquido pleural pueden ser recogidas durante la broncoscopia o mediante biopsia transcutánea o abierta y se procesan de manera similar al suero o plasma una vez que los materiales particulados se han eliminado por centrifugación.

Después del procesamiento, la muestra de ensayo obtenida del sujeto se interroga para determinar la presencia y la cantidad de dos o más biomarcadores que se puede correlacionar con un diagnóstico de cáncer de pulmón, en donde los biomarcadores cuantificados son al menos Acn9459 y TFA9133, en donde Acn9459 tiene una secuencia de aminoácidos como se expone en el SEQ ID NO: 8, y en donde TFA9133 tiene una secuencia de aminoácidos como se expone en el SEQ ID NO: 8 que ha perdido un ácido siálico de un radical glicosilado Thr74. Específicamente, los solicitantes han encontrado que la detección y cuantificación de los dos o más biomarcadores anteriores o una combinación de estos biomarcadores y parámetros biométricos (por ejemplo, al menos 2 biomarcadores, al menos 2 biomarcadores y 1 parámetro biométrico, 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. Los dos o más biomarcadores identificados y cuantificados en los métodos descritos en la presente memoria pueden estar contenidos en uno o más paneles. El número de biomarcadores que comprende un panel no es crítico y puede ser, pero no se limita a, 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.

Como se mencionó anteriormente, los métodos de la presente invención implican la identificación de la presencia de y, a continuación la cuantificación de, dos o más biomarcadores en una muestra de ensayo obtenida de un sujeto, en donde los biomarcadores cuantificados son al menos Acn9459 y TFA9133, en donde Acn9459 tiene una secuencia de aminoácidos como se expone en el SEQ ID NO: 8, y en donde TFA9133 tiene una secuencia de aminoácidos como se expone en el SEQ ID NO: 8 que ha perdido un ácido siálico de un radical glicosilado Thr74. Cualquiera de los biomarcadores que son útiles o se cree que son útiles para ayudar en el diagnóstico de un paciente sospechoso de estar en riesgo de cáncer de pulmón puede ser cuantificado en los métodos descritos en la presente memoria y pueden estar contenidos en uno o más paneles. De acuerdo con la invención, el panel incluye o más biomarcadores, en donde los biomarcadores son al menos Acn9459 y TFA9133, en donde Acn9459 tiene una secuencia de aminoácidos como se expone en el SEQ ID NO: 8, y en donde TFA9133 tiene una secuencia de aminoácidos como se expone en el SEQ ID NO: 8 que ha perdido un ácido siálico de un radical glicosilado Thr74. Los ejemplos de otros biomarcadores que pueden ser incluidos en un panel, incluyen, pero no se limitan a, anti-p53, anti-TMP21, anti-proteína 1 Niemann-Pick tipo C1, dominio del péptido C-terminal (anti-dominio NPC1L1C), anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACIN1, anti-RCV1, anti-MAPKAPK3, al menos un anticuerpo contra ciclina E2 inmunorreactiva (tal como un anticuerpo contra los SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5 o cualquiera de sus combinaciones), antígenos, tales como, pero no limitados a, antígeno carcinoembrionario (CEA),

antígeno canceroso 125 (CA125), antígeno canceroso 15-3 (CA15 -3), péptido liberador de progastrina (proGRP), antígeno de células escamosas (SCC), péptidos o proteínas de citoqueratina 8, citoqueratina 19 (también conocidos simplemente como "CK-19, CYFRA 21-1, Cyfra" en la presente memoria), y péptidos o proteínas de citoqueratina 18 (CK-18, TPS), antígenos de carbohidratos, tales como antígeno canceroso 19-9 (CA19-9), que es el grupo

sanguíneo A de Lewis con residuos de ácido siálico añadidos, amiloide A sérico, alfa-1-anti- tripsina y apolipoproteína CIII, y las regiones de interés, tales como, pero no limitadas a, Acn6399, Pub11597, Pub4789, TFA2759, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub17338, TFA6453 y HIC3959.

Adicionalmente, han encontrado ciertas regiones de interés que están altamente correlacionadas (lo que significa que estas regiones de interés tienen altos coeficientes de correlación entre unas y otras) con algunas otras regiones de interés y por lo tanto son susceptibles de ser sustituidas por otras en el contexto de la presente invención. En concreto, estas regiones de interés altamente correlacionadas han sido ensambladas en ciertas familias de correlación o "grupos". Las regiones de interés contenidas dentro de estos "grupos" pueden ser sustituidas unas por otras en los métodos y kits de la presente invención. Estas familias de correlación o "grupos" de las 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, Pub 12193, Pub13412, Acn10679 y Act10877.

Grupo E: Las regiones de interés: Pub6391, Pub6533, Pub6587, Pub6798, Pub9317 y Pub 13571.

Grupo F: Las regiones de interés: Pub7218, Pub7255, Pub7317, Pub7413, Pub7499, Pub7711, Pub14430 y Pub 15599.

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

Grupo H: Las regiones de interés: Pub6249, Pub12501 y Pub 12717.

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 zonas 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 H: 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: La región de interés: Acn9471.

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

Grupo S: Las regiones de interés: Hic2728 y Hic3276.

Grupo T: las regiones de interés: Hic6381, Hic6387, Hic6450, Hic6649, Hic6816 y Hic6823.

Grupo T: Las regiones de interés: Hic8791 y Hic8897.

Grupo V: Las regiones de interés: Tfa6453 y Tfa6652.

Grupo W: Las regiones de interés: Hic6005 y Hic5376.

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

Los paneles preferidos que están de acuerdo con la presente descripción, incluyen, pero no se limitan a:

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

2. Un panel que comprende al menos un biomarcador, en donde dicho biomarcador comprende al menos un anticuerpo contra ciclina E2 inmunorreactiva. Adicionalmente, el panel puede también opcionalmente comprender además biomarcadores adicionales, tales como, al menos un antígeno, al menos un anticuerpo, al menos un antígeno y al menos un anticuerpo, al menos una región de interés, al menos un antígeno y al menos una región de interés y al menos un anticuerpo y al menos un antígeno, al menos un anticuerpo y al menos una región de interés en la muestra de ensayo.

3. Un panel que comprende al menos un biomarcador, en donde el biomarcador se selecciona del grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, SCC, proGRP, amiloide A sérico, 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 en la muestra de ensayo.

4. Un panel que comprende al menos un biomarcador, en donde el biomarcador es al menos una región de interés seleccionada del grupo que consiste en: Acn6399, Pub11597 Pub4789, TFA2759, Pub3743, Pub8606,

Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub 17338, TFA6453 y HIC3959. El panel también puede comprender opcionalmente además biomarcadores adicionales, tales como, al menos un antígeno, al menos un anticuerpo y al menos un antígeno y al menos un anticuerpo en la muestra de ensayo.

5. Un panel que comprende al menos un biomarcador en un panel, en donde el al menos un biomarcador se selecciona del grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, SCC, proGRP, amiloide A sérico, alfa-1-anti-tripsina, apolipoproteína CIII, Acn6399, Pub11597, Pub4789, TFA2759, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Pub6453, Pub2951, Pub2433, Pub 17338, TFA6453 y HIC3959. El panel también puede comprender opcionalmente además biomarcadores adicionales, tales como al menos un anticuerpo. Los paneles preferidos son paneles que comprenden: citoqueratina 19, CEA, Pub11597, Pub4789 y TFA2759; citoqueratina 19, CEA, ACN9459, Pub11597 Pub4789, TFA2759 y citoqueratina 19, CA 19-9, CEA, CA 15-3, CA 125, SCC, citoqueratina 18 y ProGRP; Pub 11597, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 y Hic3959; y citoqueratina 19, CEA, CA125, SCC, citoqueratina 18, proGRP, ACN9459, Pub11597, Pub4789, TFA2759, TFA9133.

La presencia y cantidad de los biomarcadores en la muestra de ensayo se pueden obtener y cuantificar utilizando mecanismos rutinarios conocidos por los expertos en la técnica. Por ejemplo, los métodos para la cuantificación de antígenos o anticuerpos en muestras de ensayo son bien conocidos 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 una o más inmunoanálisis que son conocidos en la técnica. Los inmunoanálisis comprenden típicamente: (a) proporcionar un anticuerpo (o antígeno) que se una específicamente al biomarcador (es decir, 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 un< específicamente a un antígeno, se pueden utilizar antígenos purificados o sus secuencias de ácido nucleico. Las secuencias de ácido nucleico y de aminoácidos para los antígenos se pueden obtener por medio de la caracterización adicional de estos antígenos. Por ejemplo, los antígenos pueden ser péptidos mapeados con numerosas enzimas (por ejemplo, tripsina, proteasa V8, etc.). Los pesos moleculares de los fragmentos de digestión de cada antígeno se pueden utilizar para buscar en las bases de datos, tales como la base de datos SwissProt, secuencias que coincidan con los pesos moleculares de los fragmentos de digestión generados por diversas enzimas. Utilizando este método, se pueden identificar las secuencias de ácido nucleico y de aminoácidos de otros antígenos si estos antígenos son proteínas conocidas en las bases de datos.

Alternativamente, las proteínas pueden ser secuenciadas utilizando la secuenciación con escalas de proteínas. Las escalas de proteínas se pueden generar, por ejemplo, fragmentando las moléculas y sometiendo los fragmentos a digestión enzimática u otros métodos que eliminan secuencialmente un único aminoácido del extremo del fragmento. Se describen métodos de preparación de escalas de proteínas, por ejemplo, en la Publicación Internacional WO 93/24834 y en la Patente de los Estados Unidos Núm. 5.792.664. La escala se analiza a continuación mediante espectrometría de masas. La diferencia en las masas de los fragmentos de la escala identifica el aminoácido eliminado del extremo de la molécula.

Si los antígenos no son proteínas conocidas en las bases de datos, se pueden determinar las secuencias de ácidos nucleicos y aminoácidos incluso con el conocimiento de una porción de la secuencia de aminoácidos del antígeno. Por ejemplo, las sondas degeneradas se pueden elaborar basándose en la secuencia de aminoácidos N-terminal del antígeno. Estas sondas se pueden utilizar después para escrutar una biblioteca genómica o de ADNc creada a partir de una muestra de la cual se detectó inicialmente un antígeno. Los clones positivos se pueden identificar, amplificar, y se pueden subclonar sus secuencias de ADN recombinantes utilizando mecanismos que son bien conocidos. Véase, por ejemplo, Current Protocols for Molecular Biology (Ausubel et al., Green Publishing Assoc. y Wiley-Interscience 1989) y Molecular Cloning: A Laboratory Manual, segunda Ed. (Sambrook et al., Cold Spring Harbor Laboratory, NY 1989).

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

Después de proporcionar el anticuerpo, se puede detectar y/o cuantificar un antígeno utilizando cualquiera de los numerosos análisis de unión inmunológicos bien reconocidos (Véanse, por ejemplo, las Patentes de los Estados Unidos Números 4.366.241, 4.376.110, 4.517.288 y 4.837.168). Los análisis que se pueden utilizar en la presente invención incluyen, por ejemplo, una enzima de inmovinoabsorción con enzima ligada (ELISA), que también se

conoce como un "análisis sándwich", un inmunoanálisis enzimático (EIA), un radioinmunoanálisis (RIA), un fluoroinmunoanálisis (FIA), un inmunoanálisis quimioluminiscente (CLIA), un inmunoanálisis de recuento (CIA), un inmunoanálisis enzimático con medios de filtro (MEIA), un análisis de inmunoabsorción de fluorescencia ligada (FLISA), inmunoanálisis de aglutinación e inmunoanálisis fluorescentes múltiplex (tales como Luminex™ LabMAP), etc. Para una revisión de los inmunoanálisis generales, véanse también, *Methods in Cell Biology: Antibodies in Cell Biology*, volumen 37 (Asai, ed. 1993); *Basic and Clinical Immunology* (Stites y Terr, eds., 7ª ed. 1991).

Generalmente, una muestra de ensayo obtenida de un sujeto se puede poner en contacto con el anticuerpo que se une específicamente a un antígeno. Opcionalmente, el anticuerpo puede ser fijado a 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 los soportes sólidos incluyen vidrio o de plástico en forma de, por ejemplo, una placa de microtitulación, un portaobjetos o un cubreobjetos de microscopio de vidrio, una varilla o, una esfera, o una microesfera. Los anticuerpos también se pueden anclar a un sustrato de sonda o una matriz ProteinChip™ como se ha descrito anteriormente (Véase, por ejemplo, Xiao et al, *Cancer Research* 62: 6029-6033 (2001)).

Después de incubar la muestra con los anticuerpos, la mezcla se lava y el complejo antígeno-anticuerpo formado se puede detectar. Esto se puede lograr mediante la incubación de la mezcla lavada con un reactivo de detección. Este reactivo de detección puede ser, por ejemplo, un segundo anticuerpo que está marcado con una marca detectable. En términos de la marca detectable, se puede utilizar cualquier marca detectable conocida en la técnica. Por ejemplo, la marca detectable puede ser una marca radiactiva (tal como, por ejemplo, ^3H , ^{125}I , ^{35}S , ^{14}C , ^{32}P , y ^{33}P), una marca enzimática (tal como, por ejemplo, peroxidasa de rábano picante, fosfatasa alcalina, glucosa 6-fosfato deshidrogenasa, y similares), una marca quimioluminiscente (tal como, por ejemplo, ésteres de acridinio, tioésteres de acridinio, sulfonamidas de acridinio, ésteres de fenantridinio, luminal, isoluminol y similares), una marca de fluorescencia (tal como, por ejemplo, fluoresceína (por ejemplo, 5-fluoresceína, 6-carboxifluoresceína, 3'-carboxifluoresceína, 5(6)-carboxifluoresceína, 6-hexacloro-fluoresceína, 6-tetraclorofluoresceína, isotiocianato de fluoresceína, y similares)), rodamina, ficobiliproteínas, R-ficoeritrina, puntos cuánticos (por ejemplo, seleniuro de cadmio protegido terminalmente con sulfuro de cinc), una marca termométrica, o una marca inmunológica de reacción en cadena de la polimerasa. Una introducción a las marcas, los procedimientos de marcado y la detección de las marcas 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 y catálogo combinado publicado por Molecular Probes, Inc., Eugene, Oregón. Alternativamente, el marcador de la muestra se puede detectar utilizando un análisis indirecto, en donde, por ejemplo, se utiliza un segundo anticuerpo marcado para detectar el anticuerpo específico del marcador unido, y/o en un análisis de competición o inhibición en donde, por ejemplo, un anticuerpo monoclonal que se une a un epítipo distinto del antígeno se incuba simultáneamente con la mezcla.

A lo largo de los análisis, pueden ser necesarias etapas de incubación y/o lavado después de cada combinación de reactivos. Las etapas de incubación pueden variar de aproximadamente 5 segundos a varias horas, preferiblemente de aproximadamente 5 minutos a aproximadamente 24 horas. Sin embargo, el tiempo de incubación dependerá del formato de análisis, el biomarcador (antígeno), el volumen de solución, las concentraciones y similares. Por lo general, los análisis se llevarán a cabo a temperatura ambiente, aunque pueden llevarse a cabo en un intervalo de temperaturas, tal como 10°C a 40°C.

Los mecanismos de inmunoanálisis son bien conocidos en la técnica, y una visión general de la tecnología aplicable se puede encontrar en Harlow y Lane, *supra*.

El inmunoanálisis se puede utilizar para determinar una cantidad de ensayo de un antígeno en una muestra de un sujeto. En primer lugar, se puede detectar una cantidad de ensayo de un antígeno en una muestra utilizando los métodos de inmunoanálisis descritos anteriormente. Si un antígeno está presente en la muestra, se formará un complejo de antígeno-anticuerpo con un anticuerpo que se une específicamente al antígeno en condiciones de incubación adecuadas descritas anteriormente. La cantidad de complejo antígeno-anticuerpo se puede determinar mediante la comparación con un patrón. El AUC para el antígeno se puede calcular a continuación utilizando técnicas conocidas, tales como, pero no limitadas a, un análisis ROC. Alternativamente, se puede calcular la DFI. Si el AUC es mayor que aproximadamente 0,5 o la DFI es menor que aproximadamente 0,5, el inmunoanálisis se puede utilizar para discriminar sujetos con la enfermedad (tales como el cáncer, preferiblemente, cáncer de pulmón) de los sujetos normales (o benignos).

Los kits de inmunoanálisis para numerosos antígenos se encuentran disponibles comercialmente. Por ejemplo, los kits para la cuantificación de citoqueratina 19 están disponibles de F. Hoffmann-La Roche Ltd (Basilea, Suiza) y Brahms Aktiengesellschaft (Hennigsdorf, Alemania), los kits para la cuantificación de citoqueratina 18 están disponibles en IDL Biotech AD (Bromma, Suecia) y en Diagnostic Products Corporation (Los Ángeles, CA), los kits para la cuantificación de CA125, CEA SCC y CA19-9 se encuentran disponibles cada uno en Abbott Diagnostics (Abbott Park, IL) y F. Hoffman-La Roche Ltd., los kits para la cuantificación de amiloide A sérico y la apolipoproteína CIII están disponibles en Linco Research, Inc. (St. Charles, MO), los kits para la cuantificación de ProGRP están

disponibles en Advanced Life Science Institute, Inc. (Wako, Japón) y en IBL Inmuno-Biological Laboratories (Hamburgo, Alemania) y los kits para la cuantificación de alfa 1 antitripsina están disponibles en Autoimmune Diagnostica GMBH (Strassberg, Alemania) y Genway Biotech, Inc. (San Diego, CA).

La presencia y cuantificación de uno o más anticuerpos en una muestra de ensayo se pueden determinar utilizando inmunoanálisis similares a los descritos anteriormente. Tales inmunoanálisis se llevan a cabo de una manera similar a los inmunoanálisis descritos anteriormente, excepto por el hecho de que las funciones del anticuerpo y los antígenos en los análisis descritos anteriormente se invierten. Por ejemplo, un tipo de inmunoanálisis que se puede llevar a cabo es un análisis de esferas de autoanticuerpos. En este análisis, un antígeno, tal como el antígeno p53 disponible comercialmente (que se puede adquirir de BioMol International LP, Plymouth Landing, Pennsylvania), puede ser fijado a un soporte sólido, por ejemplo, una esfera, una microplaca de plástico, un portaobjetos o un cubreobjetos de microscopio de vidrio o una membrana hecha de un material tal como nitrocelulosa que se une a los antígenos de proteínas, utilizando mecanismos rutinarios conocidos en la técnica o utilizando los mecanismos y métodos descritos en el Ejemplo 3 en la presente memoria. Alternativamente, si no está disponible comercialmente un antígeno, en ese caso, el antígeno se puede purificar a partir de líneas de células cancerosas (preferiblemente, líneas celulares de cáncer de pulmón) o de los propios tejidos de cáncer de un sujeto (preferiblemente, los tejidos de cáncer de pulmón) (Véase, S-H Hong, et al., *Cancer Research* 64: 5504-5510 (2004)) o se puede expresar a partir de un clon de ADNc (Véase, Y-L Lee, et al., *Clin. Chim. Acta* 349: 87-96 (2004)). La esfera que contiene el antígeno se pone en contacto con la muestra de ensayo. Después de incubar la muestra de ensayo con la esfera que contiene el antígeno unido, la esfera se lava y se detecta cualquier complejo de antígeno-anticuerpo formado. Esta detección se puede realizar como se ha descrito anteriormente, a saber, mediante la incubación de la esfera lavada con un reactivo de detección. Este reactivo de detección puede ser por ejemplo, un segundo anticuerpo (tal como, pero no limitado a, anti-inmunoglobulina humana G (IgG), anti-inmunoglobulina humana A (IgA), anti-inmunoglobulina humana M (IgM)) que está marcado con una marca detectable. Después de la detección, se puede determinar la cantidad de complejo de anticuerpo-antígeno mediante la comparación de la señal con la generada por un patrón, como se ha descrito anteriormente. Alternativamente, el complejo de anticuerpo-antígeno se puede detectar mediante el aprovechamiento de la naturaleza multivalente de las inmunoglobulinas. En lugar de hacer reaccionar el complejo de anticuerpo-antígeno con un anti-anticuerpo humano, el complejo de anticuerpo-antígeno se puede exponer a un antígeno soluble que está marcado con una marca detectable que contiene el mismo epítipo que el antígeno unido a la fase sólida. Todos los sitios de unión del anticuerpo no ocupados se unirán al antígeno soluble (que está marcado con la marca detectable). Después del lavado, la marca detectable se detecta utilizando mecanismos de rutina conocidos por los expertos normales en la técnica. Cualquiera de los métodos descritos anteriormente permite la cuantificación sensible y específica de un anticuerpo específico en una muestra de ensayo. El AUC para el anticuerpo (y por tanto, la utilidad del anticuerpo, tal como un autoanticuerpo, para la detección del cáncer de pulmón en un sujeto) se puede detectar a continuación utilizando mecanismos rutinarios conocidas por los expertos en la técnica, tales como, pero no limitados a, un análisis ROC. Alternativamente, se puede calcular la DFI. Si el AUC es mayor de aproximadamente 0,5 o la DFI es menor de aproximadamente 0,5, se puede utilizar el inmunoanálisis para discriminar los sujetos con enfermedad (tal como el cáncer, preferiblemente, cáncer de pulmón) de los sujetos normales (o benignos).

La presencia y la cantidad de regiones de interés se pueden determinar utilizando técnicas de espectrometría de masas. Mediante la utilización de la espectroscopia de masas, los solicitantes han encontrado 212 regiones de interés que son útiles como una ayuda en el diagnóstico y la detección del 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 debe primero ser preparada para el análisis de espectrometría de masas. La preparación de la muestra puede tener lugar en una variedad de formas, pero la más utilizada implica poner en contacto la muestra con uno o más adsorbentes anclados a una fase sólida. Los adsorbentes pueden ser grupos aniónicos o catiónicos, grupos hidrófobos, grupos quelantes de metales, con o sin un ligando de metal, anticuerpos, ya sean policlonales o monoclonales, o antígenos adecuadas para la unión a sus anticuerpos cognados. La fase sólida puede ser una superficie plana de metal, vidrio o plástico. La fase sólida puede también ser de naturaleza microparticulada, o bien microesferas, partículas amorfas, o polímeros insolubles para una mayor área de superficie. Además, los materiales microparticulados pueden ser magnética para facilitar la manipulación. Los biomarcadores de interés se adsorben a la fase sólida y las moléculas a granel se eliminan mediante lavado. Para el análisis de masas, los biomarcadores de interés se hacen eluir de la fase sólida con un disolvente que reduce la afinidad del biomarcador hacia el adsorbente. Los biomarcadores se introducen después en el espectrómetro de masas para el análisis. Preferiblemente, se identifican los espectros periféricos y se excluyen en la evaluación de los espectros. Además, los inmunoanálisis, tales como los descritos anteriormente también se pueden utilizar. Al término de un inmunoanálisis, el analito se puede hacer eluir de la superficie inmunológica e introducir en el espectrómetro de masas para el análisis.

Una vez preparada la muestra de ensayo, ésta se introduce en un analizador de masas. La ionización mediante desorción 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 sobre una placa diana con una matriz eficiente en la absorción y transferencia de energía láser a la muestra. Los iones creados se separan, se cuentan y se calibran frente a iones

de masa y carga conocidas. Los datos de masa recogidos para cualquier muestra representan un recuento de iones en una razón de masa/carga específica (m/z). Se prevé que los diferentes métodos de preparación de muestras y las diferentes técnicas de ionización producirán diferentes espectros.

Los ensayos de calificación para los datos del espectro de masas implican típicamente un riguroso proceso de análisis de valores atípicos con un mínimo procesamiento previo de los datos originales. El proceso de identificación de los valores atípicos se inicia con el cálculo de la corriente de iones total (TIC) del espectro bruto. No se aplican algoritmos de suavizado ni corrección de los valores basales a los espectros en bruto antes del cálculo de la TIC. La TIC se calcula mediante la suma de las intensidades en cada valor de m/z a través del intervalo de masa detectado (m/z). Esto evalúa los fallos de los aparatos, problemas de aplicación de la muestra y otros defectos similares. Además de la TIC, se calcula el promedio de CV% (porcentaje de coeficiente de variación) en todo el espectro para cada muestra. Utilizando el número de mediciones repetidas para cada muestra, se calcula un CV% en cada valor de m/z en todo el intervalo de masa detectado. Estos CV% se promedian a continuación en conjunto para obtener un promedio de CV% que sea representativo para esa muestra particular. El CV% promedio puede o ser utilizado o no como una primera etapa de filtrado para la identificación de los valores atípicos. En general, las repeticiones con altos CV% promedio (mayores que 30% o cualquier otro valor aceptable) indican una mala reproducibilidad.

Como se describió anteriormente, el valor calculado de TIC y el CV% promedio de cada espectro se podrían utilizar como predictores para calificar la reproducibilidad y la "bondad" de los espectros. Sin embargo, mientras que estas mediciones proporcionan un buen descriptor para la mayor parte de la propiedad del espectro, no proporcionan ninguna información sobre la reproducibilidad de las características principales de los espectros tales como las intensidades individuales en cada valor de m/z. Este obstáculo fue superado por una adaptación de los cálculos del Ángulo de Contraste Espectral (SCA) referidos Wan et. Al., (J. Am. Soc. Mass Spectrom. 2002,13, 85-88). En los cálculos de SCA, todo el espectro se trata como un vector cuyos componentes son los valores de m/z individuales. Con esta interpretación, el ángulo theta (θ) entre los dos vectores es proporcionado por la fórmula matemática convencional

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

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

En el uso, el número total de cálculos y comparaciones se reduce calculando primero un espectro medio, para cualquiera de las repeticiones de la muestra o para todas las muestras dentro de un grupo en particular (p. ej., cánceres). A continuación, se calcula un SCA entre cada espectro y el espectro promedio calculado. Los espectros que difieren drásticamente del espectro del intervalo se consideran valores atípicos, siempre que cumplan con los criterios que se describen a continuación.

El uso de más de un predictor para seleccionar los valores atípicos es preferible porque un predictor no es suficiente para describir completamente un espectro de masas. Un análisis de valores atípicos multivariante se puede llevar a cabo utilizando múltiples predictores. Estos predictores podrían ser, pero no se limitan a, TIC, el promedio de CV%, y SCA. Utilizando el paquete estadístico JMP™ (SAS Institute Inc., Cary, NC), se calculan las distancias de Mahalanobis para cada repetición de la medición en el grupo (p. ej., Cáncer). Se puede calcular un valor crítico (no un límite de confianza) de manera que aproximadamente 95% de las observaciones caiga por debajo de este valor. El 5% restante que caen por encima del valor crítico se consideran valores atípicos y se excluyen del análisis adicional.

Después de la calificación de los datos espectrales de masa, los espectros son generalmente normalizados, redimensionando las intensidades de modo que TIC sea la misma para todos los espectros en el conjunto de datos o redimensionando las intensidades relativas con respecto a un pico en todos los espectros.

Después de la normalización, los espectros de masas se reducen a un conjunto de características de intensidad. En otras aplicaciones, estos se reducen a una lista de intensidades espectrales en los valores de m/z asociados con la biomoléculas. Preferiblemente, 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 conjuntos de datos de interés. Estos conjuntos de datos representan los grupos de interés (p. ej., enfermos y no enfermos). Se realiza un test de la t sobre los valores de intensidad en todas las muestras en cada m/z. Se identifican esos valores m/z con los valores de p del test de la t menores que un umbral especificado por el operario. De los valores de m/z identificados, los que son contiguos se agrupan y se definen como una región de interés. El número mínimo de valores de m/z contiguos requerido para formar una ROI y cualquiera de los huecos permitidos dentro de ese grupo contiguo puede ser definido por el usuario. Otra clasificación para la ROI es el valor absoluto del logaritmo de la razón de las medias de los dos grupos. Cuando este valor es mayor que algún valor del punto de corte umbral, digamos 0,6 cuando se

utilizan logaritmos de base 10, la ubicación-masa-carga se convierte en un candidato de la inclusión en una ROI. La ventaja de utilizar el método de ROI es que no sólo destaca las diferencias en el patrón de altas intensidades entre los espectros de las dos clases, sino que también encuentra diferencias más sutiles como los hombros y las intensidades muy bajas que se perderían por los métodos de búsqueda de picos.

Una vez que se ha determinado la región de interés, a menudo se utiliza el valor de la media o de la mediana de m/z del intervalo de la ROI como un identificador para la región. Cada región es un marcador potencial que diferencia los conjuntos de datos. Se puede extraer una variedad de parámetros (p. ej., intensidad total, intensidad máxima, intensidad mediana, o intensidad media) de los datos de la muestra y se asocian con la ROI. Por lo tanto, cada espectro de la muestra se ha reducido de muchos miles de pares de m/z , intensidad a 212 ROI y su identificador, pares de función de intensidad. Estos descriptores se utilizan como variables de entrada para las técnicas de análisis de datos.

Opcionalmente, antes o después de la identificación y la cuantificación de los dos o más biomarcadores en una muestra de ensayo obtenida de un sujeto, los métodos de la presente invención pueden incluir la etapa de obtención de al menos un parámetro biométrico de un sujeto. El número de parámetros biométricos obtenidos de un sujeto no es crítico. Por ejemplo, se pueden obtener 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc. parámetros biométricos de un sujeto. Alternativamente, los métodos de la presente invención no tienen que incluir una etapa de obtención de cualquiera de los parámetros biométricos de un sujeto. El parámetro biométrico preferido obtenido de un sujeto es el historial de tabaquismo del sujeto, en concreto, los paquetes-año que fuma el sujeto. Otros parámetros biométricos que se pueden obtener del sujeto incluyen, pero no se limitan a, la edad, la exposición a carcinógenos, el sexo, los antecedentes familiares de tabaquismo, etc.

Como se mencionó anteriormente, en los métodos de la presente invención, se analiza la muestra de ensayo para determinar la presencia de dos o más biomarcadores contenidos en el panel. Si se determina que un biomarcador está presente en la muestra de ensayo, en ese caso, se cuantifica la cantidad de cada uno de dichos biomarcadores detectado (utilizando las técnicas descritas anteriormente en la presente memoria). Una vez que se cuantifica la cantidad de cada biomarcador en la muestra de ensayo, se compara la cantidad de cada biomarcador cuantificado con un punto de corte predeterminado (que es típicamente, un valor o un número, tal como un número entero, y se denomina alternativamente en la presente memoria "punto de división") para ese biomarcador específico. El corte predeterminado empleado en los procedimientos de la presente invención se puede determinar utilizando mecanismos de rutina conocidos en la técnica, tales como, pero no limitados a, análisis de variables múltiples (Véase la Figura 1), Regresión Logística Transformada, un Método de División y Puntuación o cualquier combinación de los mismos. Por ejemplo, cuando se utiliza el Método de División y Puntuación, el valor o número de corte predeterminado dependerá del resultado que se desea alcanzar. Si el resultado que se desea lograr es maximizar la exactitud de las clasificaciones correctas de cada marcador en un grupo de interés (es decir, la correcta identificación de los sujetos en riesgo de desarrollar cáncer de pulmón y los que no están en riesgo de desarrollar cáncer de pulmón), en ese caso, se elegirá un valor o número específico para el corte predeterminado para ese biomarcador basándose en el resultado deseado. Por el contrario, si el resultado deseado es el de maximizar la sensibilidad de cada marcador, en ese caso, se puede elegir un valor o número diferente para el corte predeterminado para ese biomarcador basándose en ese resultado deseado. Asimismo, si el resultado deseado es el de maximizar la especificidad de cada marcador, en ese caso, se puede elegir un valor diferente para el corte predeterminado para ese biomarcador basándose en ese resultado deseado. Una vez que se cuantifica la cantidad de cualquiera de los biomarcadores presentes en la muestra de ensayo, se puede utilizar esta información para generar las curvas ROC, AUC y otra información que puede ser utilizada por un experto en la técnica utilizando mecanismos de rutina para determinar el punto de corte predeterminado apropiado para cada biomarcador dependiendo del resultado deseado. Después de comparar la cantidad de cada biomarcador con el corte predeterminado, se asigna una puntuación (es decir, un número, que puede ser cualquier número entero, tal como de 0 a 100) a cada biomarcador basándose en la comparación. Por otra parte, si además de los dos o más biomarcadores, se obtienen uno o más parámetros biométricos para un sujeto, en ese caso cada parámetro biométrico se compara con un punto de corte predeterminado para dicho parámetro biométrico. El corte predeterminado para cualquier parámetro biométrico se puede determinar utilizando las mismas técnicas que se describen en la presente memoria con respecto a la determinación de los puntos de corte predeterminados para uno o más biomarcadores. Al igual que con la comparación de biomarcadores, a continuación se asigna una puntuación (es decir, un número, que puede ser cualquier número entero, tal como 0 a 100) a ese parámetro biométrico basándose en dicha comparación.

Alternativamente, en lugar de utilizar el método de puntuación descrito anteriormente, se puede utilizar un Método de División y Puntuación Ponderada. Si se utiliza un Método de División y Puntuación Ponderada, a continuación, una vez que se cuantifica la cantidad de cada biomarcador en una muestra de ensayo, se compara la cantidad de cada biomarcador detectado en la muestra de ensayo con numerosos puntos de corte predeterminados para ese biomarcador específico. De todos los diferentes puntos de corte predeterminados disponibles, se asigna a ese biomarcador una única puntuación (es decir, un número, que puede ser cualquier número entero, tal como de 0 a 100). Este Método de División y Puntuación Ponderada también se puede utilizar con uno o más parámetros

biométricos.

Una vez que se asigna una puntuación a cada uno de los biomarcadores cuantificados, y, opcionalmente, a cualquiera de los parámetros biométricos obtenidos del sujeto, se combina la puntuación para cada biomarcador o cada biomarcador y cada parámetro biométrico para llegar a una puntuación total para el sujeto. Esta puntuación total se compara después con una puntuación total predeterminada. Basándose en esta comparación, se puede llevar a cabo una determinación de si el sujeto está o no en riesgo de padecer cáncer de pulmón. La determinación de si un sujeto se encuentra o no en riesgo de desarrollar cáncer de pulmón puede ser en función de si la puntuación total es o no mayor o menor que la puntuación total predeterminada. Por ejemplo, dependiendo del valor asignado a la puntuación total predeterminada, se puede considerar que un sujeto con una puntuación total que es superior a la puntuación total predeterminada se encuentra en un riesgo más alto y de este modo puede ser referido para más pruebas o procedimientos de seguimiento. La puntuación total predeterminada (denominada alternativamente en la presente memoria "umbral") que se va a utilizar en el método puede ser determinada utilizando las mismas técnicas descritas 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 la única salida de marcadores combinados, sin embargo, un solo marcador produciría una curva ROC similar. Las curvas ROC van desde una baja sensibilidad y baja tasa de falsos positivos (1-especificidad) en un extremo a una alta sensibilidad y alta tasa de falsos positivos en el otro extremo. La forma de la curva entre estos dos extremos puede variar de manera significativa. Si se requirió un método para tener una sensibilidad de al menos 90%, en ese caso, basándose en las curvas ROC mostradas en la Figura 5, la tasa de falsos positivos sería de 60 a 70% dependiendo de la curva seleccionada. Si se requirió un método para tener como máximo una tasa de falsos positivos de 10%, en ese caso la sensibilidad sería de 40-55% dependiendo de la curva elegida. Ambos métodos derivan del mismo panel de marcadores, sin embargo, con el fin de proporcionar diferentes características de rendimiento clínico, el umbral (o puntuación total predeterminada) del panel ha sido cambiado. A modo de cálculo, subyacente a 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 inferior del intervalo de datos, todas las muestras son positivas, y esto genera un punto de la curva ROC con una alta sensibilidad y una alta tasa de falsos positivos. Cuando el umbral (o puntuación total predeterminada) está en el extremo superior del intervalo de datos, todas las muestras son negativas, y esto genera un punto de la curva ROC con baja sensibilidad y baja tasa de falsos positivos. A menudo se requiere un método para tener una característica clínica deseada, tal como un nivel mínimo de sensibilidad (es decir, 90%), un nivel mínimo de especificidad (es decir, 90%), o ambos. Cambiando el umbral (o puntuación total predeterminada) de los marcadores se puede ayudar a lograr las características clínicas deseadas.

Las etapas descritas anteriormente de (a) la comparación de la cantidad de cada biomarcador en un panel con corte predeterminado (o una serie de puntos de corte predeterminados si se utiliza el Método de División y Puntuación Ponderada), asignando una puntuación (o una puntuación de una de las numerosas puntuaciones posibles si se utiliza el Método de División y Puntuación Ponderada) para cada biomarcador basándose en la comparación, la combinación de la puntuación asignada a cada parámetro biométrico en un panel para llegar a una puntuación total para el sujeto, la comparación de la puntuación total con una puntuación total predeterminada y la determinación de si un sujeto se encuentra en riesgo de cáncer de pulmón basándose en la puntuación total; o (b) la comparación de al menos un parámetro biométrico frente a un punto de corte predeterminado (o un número de puntos de corte predeterminados si se utiliza el Método de División y Puntuación Ponderada) a cada parámetro biométrico y la asignación de una puntuación (o una puntuación de una de una serie de posibles puntuaciones si se utiliza el Método de División y Puntuación Ponderada) a cada parámetro biométrico basándose en dicha comparación, la comparación de la cantidad de cada biomarcador en un panel con un punto de corte predeterminado, asignando una puntuación a cada biomarcador basándose en la comparación, la combinación de la puntuación asignada a cada parámetro biométrico con la puntuación asignada a cada biomarcador cuantificada para llegar a una puntuación total para el sujeto, la comparación de la puntuación total con una puntuación total predeterminada y la determinación de si un sujeto se encuentra en riesgo de cáncer de pulmón basándose en la puntuación total se puede realizar manualmente, por ejemplo por un ser humano, o se puede realizar, total o parcialmente por medio de un programa de ordenador o algoritmo, junto con el equipo necesario, tal como un dispositivo de entrada, memoria, procesador, visualización y salida.

Sólo con fines ilustrativos, se proporcionará a continuación un ejemplo de cómo se puede realizar el método de la presente descripción. En este ejemplo, un paciente se somete a un ensayo para determinar la probabilidad del paciente de tener 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, CA 15-3, CA 19-9, SCC, proGRP y citoqueratina 18. La puntuación total predeterminada (o umbral) para el panel es 3. Después de obtener una 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 de ensayo del paciente. A los efectos de este ejemplo, se determina que la cantidad de cada uno de los 8 biomarcadores en la muestra de ensayo es: 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 compara después con el corte

predeterminado correspondiente (o punto de división). Los puntos de corte predeterminados para cada uno de los biomarcadores es: 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. A cada biomarcador que tiene una cantidad que es mayor que su correspondiente corte predeterminado (punto de división), se le puede asignar una puntuación de 1. A cada biomarcador que tiene una cantidad que es menor que o igual a su correspondiente corte predeterminado, se le puede asignar una puntuación de 0. A continuación, basándose en dicha comparación, a cada biomarcador se le asigna una puntuación de la siguiente manera: citoqueratina 19: 1, CEA: 0, CA125: 1, CA 153: 0, CA 199: 0, SCC: 1, proGRP: 1, y citoqueratina 18: 1. Las puntuaciones para cada uno de los 8 biomarcadores se combinan después matemáticamente (es decir, mediante la adición de cada una de las puntuaciones de los biomarcadores en conjunto) para llegar a la puntuación total para el paciente. La puntuación total para el paciente es de 5 (La puntuación total se calcula como sigue: $1 + 0 + 1 + 0 + 0 + 1 + 1 + 1 = 5$). La puntuación total para el paciente se compara con la puntuación total predeterminada, que es 3. Una puntuación total superior a la puntuación total predeterminada de 3 indicaría un resultado positivo para el paciente. Una puntuación total de menos de o igual a 3 indicaría un resultado negativo para el paciente. En este ejemplo, debido a que la puntuación total del paciente es mayor que 3, se considerará que el paciente tenía un resultado positivo y por lo tanto se le remitiría a ensayos adicionales por un indicio o sospecha de cáncer de pulmón. Por el contrario, si la puntuación total del paciente era de 2, se consideraría que el paciente tenía un resultado negativo y no se le remitiría a ensayos adicionales.

En un ejemplo adicional, se utiliza de nuevo el panel de 8 biomarcadores descrito anteriormente, sin embargo, en este ejemplo, se emplea el Método de División y Puntuación Ponderada. En este ejemplo, la puntuación total predeterminada (o umbral) para el panel es de 11,2 y las cantidades de los biomarcadores cuantificados en la muestra de ensayo son las mismas que se ha descrito anteriormente. La cantidad de cada uno de los biomarcadores se compara en ese caso con 3 puntos de corte predeterminados diferentes para cada uno de los biomarcadores. Por ejemplo, se proporcionan los puntos de corte predeterminados para cada uno de los biomarcadores a continuación en la Tabla A.

Tabla A

	CEA	Citoqueratina 18	ProGRP	CA15-3	CA125	SCC	Citoqueratina 19	CA19-9
Corte predeterminado @ 50% de especificidad	2,02	47,7	11,3	16,9	15,5	0,93	1,2	10,6
Corte predeterminado @ 75% de especificidad	3,3	92,3	18,9	21,8	27	1,3	1,9	21,9
Corte predeterminado @ 90% de especificidad	4,89	143,3	28,5	30,5	38,1	1,98	3,3	45,8
Puntuación por debajo de 50% de especificidad	0	0	0	<u>0</u>	<u>0</u>	0	0	<u>0</u>
Puntuación por encima de 50% de 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 de 75% de especificidad	5,36	5,2	<u>4,96</u>	2,32	5,36	4,96	<u>8,4</u>	2,2
Puntuación por encima de 90% de especificidad	13,4	13	12,4	5,8	13,4	12,4	21	5,5

Por lo tanto, se pueden proporcionar 4 puntuaciones posibles para cada biomarcador. La cantidad de cada biomarcador cuantificado se compara con los puntos de corte predeterminados (puntos de división) proporcionados en la Tabla A anterior. Por ejemplo, para CEA, ya que la cantidad de CEA cuantificada en la muestra de ensayo era 2,75, caía entre el corte predeterminado de 2,02 para 50% de especificidad y 3,3 para 75% de especificidad en la Tabla A. Por lo tanto, a CEA se le asigna una puntuación de 2.68. Esto se repite para los restantes biomarcadores que son evaluados de manera similar y a cada uno se le asignan las siguientes puntuaciones: citoqueratina 18: 2,6, proGRP: 4,96, CA15-3: 0, CA125: 0, SCC: 2,48, citoqueratina 19: 8,4 y CA 19-9: 0. Las puntuaciones para cada uno de los 8 biomarcadores se combinan después matemáticamente (es decir, mediante la adición de cada una de las puntuaciones de los biomarcadores en conjunto) 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 para el paciente se compara con la puntuación total predeterminada, que es de 11,2. Una puntuación total superior a la puntuación total predeterminada de 11,2 indicaría un resultado positivo para el paciente. Una puntuación total de menos de o igual a 11,2 indicaría un resultado negativo para el paciente. En este ejemplo, debido a que la puntuación total del paciente fue mayor de 11,2, se consideraría que el paciente tenía un resultado positivo y por lo tanto se le remitiría a ensayos adicionales por un indicio o sospecha de cáncer de pulmón.

Además, de acuerdo con la presente descripción el Método de División y Puntuación Ponderada descrito en la presente memoria también se puede utilizar para puntuar uno o más marcadores obtenidos de un sujeto. Preferiblemente, dichos marcadores, ya sean o uno o más biomarcadores, uno o más parámetros biométricos o una combinación de biomarcadores y parámetros biométricos pueden ser utilizados como una ayuda en el diagnóstico o la evaluación de si un sujeto se encuentra en riesgo de desarrollar una afección médica, tal como un cáncer o alguna otra enfermedad. Cualquier afección médica en la que se utilizan marcadores o se pueden utilizar para evaluar el riesgo puede ser utilizada en los métodos descritos en la presente memoria. Dicho método puede comprender las etapas de:

- a. obtener al menos un marcador de un sujeto;
- b. cuantificar la cantidad del marcador de dicho sujeto;
- c. comparar la cantidad de cada marcador cuantificado con una serie de puntos de corte predeterminados para dicho marcador y asignar de una puntuación a cada marcador basándose en dicha comparación; y
- d. combinar la puntuación asignada a cada marcador cuantificado en la etapa c para llegar a una puntuación total de dicho sujeto.

Preferiblemente, el método comprende las etapas de:

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

DFI

Como se discutió anteriormente en la presente memoria, los solicitantes han encontrado que la detección y cuantificación de uno o más biomarcadores o una combinación de biomarcadores y parámetros biométricos es útil como una ayuda en el diagnóstico del cáncer de pulmón en un paciente. Además, los solicitantes también han encontrado que los uno o más biomarcadores y las combinaciones de uno o más biomarcadores y uno o más parámetros biométricos descritos en la presente memoria tienen una DFI con respecto al cáncer de pulmón que es menor de aproximadamente 0,5, preferiblemente menor de aproximadamente 0,4, más preferiblemente, menor de aproximadamente 0,3 y aún más preferiblemente, menor de aproximadamente 0,2. Las Tablas 25-29 proporcionan ejemplos de paneles que contienen varios biomarcadores o combinaciones de biomarcadores y parámetros biométricos que exhiben una DFI que es menor que aproximadamente 0,5, menor de aproximadamente 0,4, menor de aproximadamente 0,3 y menores de aproximadamente 0,2.

Kits

Dos o más biomarcadores, uno o más de los polipéptidos de ciclina E2 inmunorreactiva, los parámetros biométricos y cualquier combinación de los mismos son susceptibles de la formación de kits (tales como paneles) tales como los definidos en la reivindicación adjunta 17. En un aspecto, el kit puede comprender un péptido seleccionado del grupo que consiste en: los SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5 o combinaciones de los mismos.

En otro aspecto, el kit puede comprender al menos un anticuerpo contra ciclina E2 inmunorreactiva o cualquier combinación de los mismos.

En un aspecto adicional, el kit puede comprender (a) reactivos que contienen al menos un anticuerpo para la cuantificación de 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, CA 19-9, proGRP, amiloide A sérico, alfa-1-anti-tripsina y apolipoproteína CIII; (b) reactivos que contienen uno o más antígenos para la cuantificación de al menos un anticuerpo en una muestra de ensayo; en donde dichos anticuerpos son: anti-p53, anti-TMP21, dominio anti-NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACIN1, anti-RCV1, anti-MAPKAPK3 y al menos un anticuerpo contra ciclina E2 inmunorreactiva; (c) reactivos para la cuantificación de una o más regiones de interés seleccionadas del grupo que consiste en: Pub11597, Pub4789, TFA2759, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 y Hic3959; y (d) uno o más algoritmos o programas de ordenador para realizar las etapas de combinación y comparación de la cantidad de cada antígeno, anticuerpo y región de interés cuantificados en la muestra de ensayo frente a un punto de corte predeterminado (o frente una serie de puntos de corte predeterminados) y la asignación de una puntuación a cada antígeno, anticuerpo y región de interés (o una puntuación de una de las numerosas puntuaciones posibles), cuantificada de acuerdo con dicha comparación, la

combinación de la puntuación asignada a cada antígeno, anticuerpo y región de interés cuantificados para obtener una puntuación total, la comparación de la puntuación total con una puntuación total predeterminada y el uso de dicha comparación como una ayuda en la determinación de si un sujeto tiene cáncer de pulmón. Alternativamente, en lugar de uno o más algoritmos o programas de ordenador, se pueden proporcionar una o más instrucciones para la realización manual de las etapas anteriores por parte de un ser humano. Los reactivos incluidos en el kit para la cuantificación de una o más regiones de interés puede incluir un adsorbente que se une y retiene al menos una región de interés contenida en un panel, soportes sólidos (tales como esferas) para su utilización en conexión con dichos adsorbentes, una o más marcas detectables, etc. El adsorbente puede ser cualquiera de los muchos adsorbentes utilizados en la química y la inmuquímica analíticas, incluyendo quelatos de metales, 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 la cuantificación de uno o más antígenos en una muestra de ensayo, en donde dichos antígenos son citoqueratina 19, citoqueratina 18, CA 19-9, CEA, CA15-3, CA125, SCC y ProGRP; (b) reactivos para la cuantificación de una o más regiones de interés seleccionadas del grupo que consiste en: ACN9459, Pub11597, Pub4789, TFA2759, TFA9133, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 y Hic3959; y (c) uno o más algoritmos o programas de ordenador para realizar las etapas de combinación y comparación la cantidad de cada antígeno y región de interés cuantificados en la muestra de ensayo frente a un predeterminado (o frente a varios puntos de corte predeterminados) y la asignación de una puntuación a cada antígeno y región de interés (o una puntuación de una de las numerosas posibles puntuaciones) cuantificado basándose en dicha comparación, la combinación de la puntuación asignada a cada antígeno y región de interés 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 en la determinación de si un sujeto tiene cáncer de pulmón. Alternativamente, en lugar de uno o más algoritmos o programas de ordenador, se pueden proporcionar una o más instrucciones para la realización manual de las etapas anteriores por parte de un ser humano. Los reactivos incluidos en el kit para la cuantificación de 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 esferas) para su utilización en conexión con dichos adsorbentes, una o más marcas detectables, etc. Preferiblemente, 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 la cuantificación de uno o más antígenos en una muestra de ensayo, en donde dichos antígenos son citoqueratina 19, citoqueratina 18, CA 19-9, CEA, CA15-3, CA125, SCC y ProGRP; y (b) uno o más algoritmos o programas de ordenador para realizar las etapas de combinación y comparación de la cantidad de cada antígeno cuantificado en la muestra de ensayo frente a un punto de corte predeterminado (o frente a varios puntos de corte predeterminados) y la asignación de una puntuación a cada antígeno (o una puntuación de una de las numerosas puntuaciones posibles), cuantificado basándose en dicha comparación, la combinación de la puntuación asignada a cada antígeno cuantificado para obtener una puntuación total, la comparación de la puntuación total con una puntuación total predeterminada y el uso de dicha comparación como una ayuda en la determinación de si un sujeto tiene cáncer de pulmón. Alternativamente, en lugar de uno o más algoritmos o programas de ordenador, se pueden proporcionar una o más instrucciones para la realización manual de los pasos anteriores por parte de un ser humano. El kit también puede contener uno o más marcadores detectables. Preferiblemente, el kit contiene los reactivos necesarios para cuantificar los siguientes antígenos citoqueratina 19, citoqueratina 18, CA 19-9, CEA, CA-15-3, CA125, SCC y ProGRP.

En otro aspecto, un kit puede comprender (a) reactivos para la cuantificación de uno o más biomarcadores, en donde dichos biomarcadores son regiones de interés seleccionadas del grupo que consiste en: Pub11597, Pub4789, TFA2759, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 y Hic3959; y (b) uno o más algoritmos o programas de ordenador para realizar las etapas de combinación y comparación de la cantidad de cada biomarcador cuantificado en la muestra de ensayo frente a un punto de corte predeterminado (o frente a varios puntos de corte predeterminados) y la asignación de una puntuación a cada biomarcador (o una puntuación de una de las numerosas puntuaciones posibles), cuantificado basándose en dicha comparación, la combinación de la puntuación asignada a cada biomarcador cuantificado para obtener una puntuación total, la comparación de la puntuación total, con una puntuación total predeterminada y el uso de dicha comparación como una ayuda en la determinación de si un sujeto tiene cáncer de pulmón. Alternativamente, en lugar de uno o más algoritmos o programas de ordenador, se pueden proporcionar una o más instrucciones para la realización manual de las etapas anteriores por parte de un ser humano. Preferiblemente, las regiones de interés que se van a cuantificar en el kit se seleccionan del grupo que consiste en: Pub11597, Pub3743, Pub8606, Pub4487, Pub4861, Pub6798, Tfa6453 y Hic3959. Los reactivos incluidos en el kit para la cuantificación de 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 esferas) que se van a utilizar en conexión con dichos adsorbentes, una o más marcas detectables, etc.

Identificación de biomarcadores

Los biomarcadores pueden ser aislados, purificados e identificados por medio de mecanismos bien conocidos por los expertos en la técnica. Estos incluyen mecanismos cromatográficos, electroforéticos y de centrifugación. Estos mecanismos se comentan en Current Protocols in Protein Science, J. Wiley and Sons, Nueva York, NY, Coligan et al. (Eds) (2002) y Harris, E.L.V., S. Angal en Protein Purification Applications: A Practical Approach, Oxford University Press, Nueva York, Nueva York (1990) y en otros lugares.

A modo de ejemplo, y no de limitación, se proporcionarán a continuación ejemplos de la presente invención:

Ejemplos

Se recogieron muestras clínicas de suero sanguíneo de pacientes (Ejemplo 1) y se analizaron para someter a inmunoanálisis los marcadores antigénicos (Ejemplo 2), para someter a inmunoanálisis los marcadores de anticuerpos utilizando esferas (Ejemplo 3) o portaobjetos (Ejemplo 4), y para determinar los biomarcadores identificados mediante espectrometría de masas (Ejemplo 5). Los marcadores identificados se clasificaron y se jerarquizaron utilizando una variedad de algoritmos (Ejemplo 6). Estos marcadores jerarquizados se combinaron utilizando un método de puntuación (Ejemplo 7) para identificar modelos predictivos (Ejemplo 8) para evaluar su utilidad clínica. Los ejemplos del uso de los métodos de ayuda para la detección de cáncer de pulmón en pacientes que se sospecha que tienen cáncer de pulmón se ilustran en el Ejemplo 9. Los biomarcadores identificados mediante Regiones de Interés de espectrometría de masas se analizaron para determinar su composición e identidad (Ejemplo 10). El Ejemplo 11 es un ejemplo profético que describe cómo y medir se pueden detectar los biomarcadores utilizando técnicas de inmunoanálisis e inmunotécnicas de espectrometría de masas.

Ejemplo 1: Especímenes clínicos

Se recogieron muestras clínicas de suero del paciente bajo un protocolo aprobado por la Junta de Revisión Institucional. Todos los sujetos que aportaron un espécimen dieron su consentimiento informado para que la muestra fuera recogida y utilizada en este proyecto. Las muestras de suero se extrajeron en un tubo separador de suero y se dejaron coagular durante 15 minutos a temperatura ambiente. El coágulo se centrifugó y la muestra se vertió a alícuotas de 2 mL. En el plazo de 24 horas las muestras se congelaron a - 80°C y se mantuvieron a esa temperatura hasta que se llevó a cabo un procesamiento adicional. Tras la recepción, las muestras se descongelaron y se volvieron a repartir en alícuotas en volúmenes más pequeños por conveniencia y se volvieron a congelar. Las muestras se descongelaron a continuación un tiempo final inmediatamente antes de su análisis. Por lo tanto, todas las muestras del conjunto se congelaron y descongelaron dos veces antes de su análisis.

Se recolectaron y se analizaron un total de 751 muestras. El grupo estaba compuesto por 250 biopsias de pacientes con cáncer de pulmón confirmado, 274 biopsias de pacientes con enfermedades pulmonares benignas confirmadas, y 227 sujetos aparentemente normales. Los pacientes con cáncer y con enfermedades benignas fueron confirmados en su diagnóstico mediante una biopsia definitiva. Los sujetos normales no fueron sometidos a tal procedimiento diagnóstico definitivo y fueron considerados "normales" por la falta de una enfermedad maligna manifiesta. Después de este procedimiento de diagnóstico definitivo, sólo se seleccionaron los pacientes ≥ 50 años de edad. Después de esta selección, quedaron 231 cánceres, 182 enfermedades benignas, y 155 normales. Esta gran cohorte de sujetos con cáncer, enfermedad pulmonar benigna, y aparentemente normales, será referida colectivamente en lo sucesivo como "gran cohorte". Un subconjunto de la gran cohorte se utilizó para centrarse en la diferenciación entre enfermedad pulmonar benigna y cáncer de pulmón. Esta cohorte, en lo sucesivo referida como "pequeña cohorte", consistió en 138 sujetos con cáncer, 106 con enfermedades benignas, y 13 aparentemente normales. Después de retirar la "pequeña cohorte" de la "gran cohorte", seguía habiendo 107 sujetos con cánceres, 74 con enfermedades benignas, y 142 aparentemente normales. Esta cohorte, en lo sucesivo referida como "cohorte de validación" es independiente de la pequeña cohorte y se utilizó para validar los modelos de predicción generados. Las muestras clínicas preparadas como se ha descrito se utilizaron en los Ejemplos 2-10.

Ejemplo 2. Detección de biomarcadores mediante inmunoanálisis

Análisis Architect™ de A. Abbott Laboratories (Abbott Park, IL, en lo sucesivo, "Abbott")

Se adquirieron kits Architect™ para los siguientes antígenos: CEA, CA125, SCC, CA19-9 y CA 15-3. Todos los análisis se realizaron de acuerdo con las instrucciones del fabricante. Las concentraciones de los analitos en las muestras fueron proporcionadas por el aparato Architect™. Estas concentraciones se utilizaron para generar los datos de AUC mostrados a continuación en la Tabla 1.

Tabla 1. Rendimiento clínico (AUC) de CA125, CEA, CA15-3, CA19-9, y SCC en las pequeñas y grandes cohortes. El Núm. obs se refiere al número total de individuos o muestras clínicas en cada grupo.

Marcador	gran cohorte		pequeña cohorte	
	Núm. obs.	AUC	Núm. 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. Análisis Elecsys™ de Roche

Se realizaron mediciones de Cyfra 21-1 (Citoqueratina 19, CK-19) en el sistema **Elecsys™** 2010 (Roche Diagnostics GmbH, Mannheim, Alemania) de acuerdo con las instrucciones del fabricante. La concentración de Cyfra 21-1 fue proporcionada por el aparato **Elecsys™**. Se generó una curva ROC con los datos y las AUC para las grandes y pequeñas cohortes se presentan a continuación en la Tabla 2.

Tabla 2. Rendimiento clínico (AUC) de Citoqueratina 19.

Marcador	gran cohorte		pequeña cohorte	
	Núm. obs.	AUC	Núm. obs.	AUC
ICK-19	537	0,68	248	0,718

C. Análisis de placa de microtitulación

Se adquirieron los siguientes kits de ELISA: ProGRP de Advanced Life Science Institute, Inc. (Japón), TPS (Citoqueratina 18, CK-18) de IDL Biotech AB (Bromma, Suecia) y ELISA de IgG de Parainfluenza 1/2/3 de IBL Immuno Biological Laboratories (Minneapolis, MN, EE.UU.). Los análisis se realizaron de acuerdo con las instrucciones del fabricante. Las concentraciones de los analitos se obtuvieron a partir de cálculos indicados y previstos en el protocolo del fabricante. Las AUC obtenidas para los análisis individuales se muestran a continuación en la Tabla 3.

Tabla 3. Rendimiento clínico (AUC) de citoqueratina 18, ProGRP, y parainfluenza 1/2/3.

Marcador	gran cohorte		pequeña cohorte	
	Núm. obs.	AUC	Núm. 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 esferas de autoanticuerpos

A. Se anclaron proteínas humanas disponibles comercialmente (véase la Tabla 4, a continuación) a esferas Luminex™ SeroMap™ (Austin, Texas) y los conjuntos de esferas individuales se combinaron para preparar el reactivo. Porciones del reactivo se expusieron a las muestras de suero humano en condiciones que permitían que cualquiera de los anticuerpos presentes se uniera a las proteínas. El material no unido se retiró mediante lavado y las esferas se expusieron a continuación a un producto conjugado fluorescente de R-ficoeritrina conectado a un anticuerpo que se une específicamente a IgG humana. Después del lavado, las esferas se pasaron a través de un aparato Luminex 100™, que identifica cada esfera de acuerdo con sus colorantes internos, y se midió la fluorescencia unida a la esfera, correspondiente a la cantidad de anticuerpo unido a la esfera. De esta manera, se evaluaron las respuestas inmunitarias de 772 muestras (251 de cáncer de pulmón, 244 normales, 277 benignos) contra 21 proteínas humanas, así como contra varias proteínas no humanas para los controles (albúmina de suero bovino (BSA) y toxina del tétanos).

Los antígenos MUC-1 (Fujirebio Diagnostics Inc., Malvern, PA), Citoqueratina 19 (Biodesign, Saco, ME), y CA-125 (Biodesign, Saco, ME) se obtuvieron como fracciones de intercambio iónico de cultivos celulares (Véase la Tabla 4, a continuación). Estas preparaciones relativamente brutas se sometieron a fraccionamiento adicional por su peso

molecular utilizando HPLC con una columna de exclusión por tamaño (BioRad SEC-250, Hercules, CA) con fase móvil = PBS a 0,4 mL/ minuto. Se recogieron las fracciones a partir de 15 minutos con 1 minuto para cada fracción para un total de 23 fracciones para cada antígeno. Para MUC-1, se inyectaron 250 µL; para la Citoqueratina 19 y CA-125, se inyectaron 150 µL. Las tres muestras mostraron señales que indican diversas concentraciones de proteínas de PM más altos que eluían a partir de 15-24 minutos, con señales demasiado altas para medir a tiempos superiores a 24 minutos, lo que indica un altas concentraciones de materiales de PM más bajos. Para el recubrimiento sobre esferas se combinaron las siguientes fracciones: fracciones 6,7 de MUC-1-A; fracciones 10,11 de MUC-1-B; fracciones 12,13 de MUC-1-C; fracciones 4,5 de citoqueratina 19-A; fracciones 8,9 de citoqueratina 19-B; fracciones 16,17 de citoqueratina 19-C; fracciones 5,6 de CA125-A; fracciones 12,13 de CA125-B.

Tabla 4. Lista de proteínas.

Identificación de la esfera	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
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	Vision Lab, 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	RECAF leche	BioCurex, Vancouver, BC Canadá
28	BSA	Sigma, St. Louis, MO

B. Recubrimiento de esferas Luminex SeroMap™ con antígenos

A los pocillos de una placa de ultrafiltración Omega10K (Pall Corporation, Ann Arbor, Michigan) se les añadieron 50 µL de agua. Después de 10 minutos la placa se colocó a vacío. Cuando los pocillos se vaciaron, se añadieron 10 µL de agua para mantener la hidratación. A cada pocillo se le añadieron 50-100 µL de ácido morfolinoetanofónico 5 mM (MES) pH 5,6, 50 µL de la esfera Luminex™ SeroMAP™ indicada y el volumen apropiado correspondiente a 10-

20 µg de cada antígeno indicado en la Tabla 4. Las esferas se suspendieron con la pipeta. A las esferas se les añadieron 10 µL 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. Después de 14 horas, la placa se filtró con succión por vacío, se lavó con agua, y, finalmente, las esferas se resuspendieron en 50 µL de trietanolamina (TEA) 20 mM de pH 5,6. La placa se agitó en el agitador en oscuridad. Se añadieron unos segundos 10 µL de EDAC (2,0 mg en 1,0 mL de MES 5 mM de pH 5,6) a cada pocillo, y la placa se colocó en un agitador en oscuridad durante una hora. Después del lavado, se añadieron a cada pocillo) 200 µL de tampón PBS que contenía BSA al 1% y azida de sodio al 0,08% (PBN, seguido de sonicación con sonda, y se colocó en oscuridad.

10 D. Ensayo de muestras de suero con esferas recubiertas

Las muestras de suero se prepararon en microplacas a una dilución 1:20 en PBN, con 80 muestras por microplaca. A 50 µL del conjunto de esferas descrito anteriormente se les añadieron 5 µL de suero de conejo (de un conejo inmunizado con un antígeno no relacionado con los sometidos a ensayo aquí). El conjunto de esferas se agitó a vórtice y se colocó a 37°C. Después de 35 minutos, se añadió 1 mL de PBN que contenía suero de conejo al 5% y CHAPS al 1% (BRC). El conjunto de esferas se agitó a vórtice, se centrifugó y se resuspendió en 1,05 mL de BRC. Los pocillos de una placa de filtro Supor de 1.2µ (Pall Corporation) se lavaron con 100 µL de PBN. A cada pocillo se le añadieron 50 µL de BRC, 10 µL de cada muestra de suero 1:20, y 10 µL de esferas resuspendidas. La placa se agitó a la temperatura ambiente en oscuridad durante 1 hora, se filtró y después se lavó 3 veces durante 10 minutos con 100 µL de BRC. Se añadieron 50 µL de producto conjugado de detección (20 µL de anti-IgG humana-RPE en 5,0 mL de BRC) y la placa se agitó en oscuridad durante 30 minutos después de que las esferas se volvieron a suspender mediante una pipeta. A continuación se añadieron 100 µL de BRC, las esferas se agitaron mediante una pipeta y las muestras se analizaron en un aparato Luminex™ 100.

25 Los resultados (intensidad mediana de las esferas para cada muestra y antígeno) se evaluaron mediante análisis ROC con los siguientes resultados para las grandes y pequeñas cohortes mostradas a continuación en la Tabla 5:

Tabla 5. Rendimiento clínico de la matriz de esferas de autoanticuerpos que contiene proteínas de la Tabla 4 en las grandes y pequeñas cohortes.

	gran cohorte		pequeña cohorte	
Biomarcador	Núm. obs	AUC	Núm. 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
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

	gran cohorte		pequeña cohorte	
Biomarcador	Núm. obs	AUC	Núm. obs	AUC
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
RECAF62	579	0,54	253	0,53
RECAF50	579	0,53	253	0,53
RECAF leche	579	0,54	253	0,62
BSA	579	0,57	253	0,59

Ejemplo 4. Matriz en portaobjetos de autoanticuerpos

A. Preparación de antígeno

Se prepararon aproximadamente 5.000 proteínas derivadas de Ultimate ORF Collection™ de Invitrogen (Invitrogen, Grand Island, NY) como fusiones recombinantes de la secuencia de glutatión-S-transferasa (GST) con una proteína humana completa. La etiqueta de GST permitió la evaluación de la cantidad de cada proteína unida a la matriz independiente de otras características de la proteína.

B. Recubrimiento de portaobjetos con antígeno

La ProtoMatriz consta de una superficie de vidrio (portaobjetos) recubierto con nitrocelulosa donde se habían aplicado las aproximadamente 5000 proteínas mencionadas anteriormente, así como numerosos rasgos de control.

C. Ensayo de muestras de suero con portaobjetos recubiertos

La matriz se bloqueó primero con PBS/BSA al 1%/Tween 20 al 0,1% durante 1 hora a 4°C. A continuación, se expuso a la muestra de suero diluido 1:120 en Profiling Buffer (el "Profiling Buffer" examinado en la presente memoria contenía PBS, MgCl₂ 5 mM, ditioneitol 0,5 mM, Triton X-100 al 0,05, glicerol al 5%, BSA al 1%) durante 90 minutos a 4°C. A continuación, la matriz se lavó tres veces con Profiling Buffer durante 8 minutos por lavado. A continuación, la matriz se expuso a anti- IgG humano conjugado con AlexaFluor a 0,5 µg/mL en Profiling Buffer durante 90 minutos a 44°C. A continuación, la matriz se lavó tres veces con Profiling Buffer durante 8 minutos por lavado. Después de secar en una centrifuga se exploró utilizando un escáner de micromatrices fluorescentes Axon GenePix 4000B (Molecular Devices, Sunnyvale, CA).

D. Selección de biomarcadores

Mediante la comparación de la distribución de señales positivas de suero de pacientes con cáncer con la de los pacientes normales se determinó la identidad de las proteínas que provocan autoanticuerpos característicos de pacientes con 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 conjuntos que contenían cada uno suero de 4 o 5 pacientes con cáncer de pulmón, 10 conjuntos que contenían cada uno suero de 4 o 5 pacientes normales y 10 conjuntos que contenían cada uno suero de 4 o 5 pacientes con enfermedades pulmonares benignas. Estos conjuntos fueron enviados a Invitrogen para su procesamiento como se ha descrito anteriormente. Las intensidades de fluorescencia correspondientes a cada proteína para cada grupo se presentan en una hoja de cálculo. Cada proteína se representó dos veces, correspondiente a aplicaciones duplicadas en la matriz.

En un algoritmo para la evaluación de la especificidad contra el cáncer de la respuesta inmunitaria para una proteína particular, se suministró un valor del punto de corte por el fabricante (Invitrogen), que distingue mejor las intensidades de señal de las muestras de cáncer de las de las muestras negativas para cáncer. Se determinó el número de muestras de cada grupo con intensidades por encima de este punto de corte (Recuento para Cáncer y Recuento Negativo para Cáncer respectivamente) y se colocó en la hoja de cálculo como parámetros. Además, se calculó un valor de p, que representa la probabilidad de que no hubiera un aumento de la señal en un grupo en comparación con el otro. Los datos se ordenaron a continuación para llevar a la cabeza aquellas proteínas con menor cantidad de positivos en el grupo Negativo para Cáncer y la mayoría de los positivos en el grupo de cáncer, y ordenados adicionalmente por el valor p de menos a más. La clasificación por esta fórmula proporcionó la siguiente información proporcionada a continuación en la Tabla 7.

Tabla 7. Lista de ID. de antígeno

Identificación de antígenos	Recuento de Cáncer	Recuento Negativo para cáncer	P-Valor
proteína 1 de la vesícula acrosómica (ACRV1)	6	0	0,0021
caja de cabeza de horquilla A3 (FOXA3)	6	0	0,0072
factor de transcripción IIA general	6	0	0,5539
dominio WW que contiene la proteína ubiquitina E3 ligasa 2	5	0	0,0018
dominio PDZ que contiene 1 (PDZK1)	5	0	0,0018
ciclina E2	5	0	0,0018
ciclina E2	5	0	0,0018
dominio de tipo 2 de ácido fosfatídico fosfatasa que contiene 3 (PPAPDC3)	5	0	0,0088
repetición de anquirina y dominio del motivo alfa estéril que contiene 3	5	0	0,0563
dedo de zinc	5	0	0,0563
cisteinil ARNt sintetasa	4	0	0,0077
cisteinil ARNt sintetasa	4	0	0,0077
factor de transcripción que se une al promotor 3 de IGHM (TFE3)	4	0	0,0077
dominio WW que contiene la proteína ubiquitina E3 ligasa 2	4	0	0,0077
Marco de lectura abierto 7 del cromosoma 21	4	0	0,0077
Marco de lectura abierto 7 del cromosoma 21	4	0	0,0077
motivo IQ que contiene F1 (IQCF1)	4	0	0,0077
proteína 1citósólica de linfocitos (L-plastina) (LCP1)	4	0	0,0077
proteína de la vesícula acrosómica 1 (ACRV1)	4	0	0,0077
homólogo de DnaJ (Hsp40)	4	0	0,0077
homólogo de 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 PDZ que contiene 1 (PDZK1)	4	0	0,0077
proteína quinasa C y sustrato de caseína quinasa en las neuronas 2	4	0	0,0077
dominio quinasa 2 de LIM	4	0	0,0077
polipéptido D de (ARN)polimerasa III (dirigida por ADN)	4	0	0,0077
proteína del motivo de unión a ARN	4	0	0,0077
4 asociada al ciclo de división celular (CDCA4)	4	0	0,0312
factor de intercambio de nucleótidos de guanina Rho (GEF) 1	4	0	0,076
LUC7 tipo 2 (S. cerevisiae)	4	0	0,2302
similar a RIKEN cDNA 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 de receptores de tipo lectina de células asesinas	3	0	0,0296

Identificación de antígenos	Recuento de Cáncer	Recuento Negativo para cáncer	P-Valor
subfamilia C de receptores de tipo lectina de células asesinas	3	0	0,0296
LATS	3	0	0,0296
N-acilesfingosina amidohidrolasa (ceramidasa ácida) 1 (ASAH1)	3	0	0,0296
N-acilesfingosina amidohidrolasa (ceramidasa ácida) 1 (ASAH1)	3	0	0,0296
paralemmína	3	0	0,0296
paralemmína	3	0	0,0296
proteína 1 de interacción con PIN2	3	0	0,0296
proteína ribosomal S6 quinasa	3	0	0,0296
proteína ribosomal S6 quinasa	3	0	0,0296
dominio SH3 y PX que contiene 3 (SH3PX3)	3	0	0,0296
dominio SH3 y PX que contiene 3 (SH3PX3)	3	0	0,0296
compañero de fusión TCF3 (E2A) (en la leucemia infantil) (TFPT)	3	0	0,0296
compañero de fusión TCF3 (E2A) (en la leucemia infantil) (TFPT)	3	0	0,0296
de factor de transcripción que se une al potenciador 3 de IGHM (TFE3)	3	0	0,0296
Marco de lectura abierto 117 del cromosoma 1	3	0	0,0296
Marco de lectura abierto 117 del cromosoma 1	3	0	0,0296
proteína expresada en exceso asociada a la resistencia al cisplatino	3	0	0,0296
Proteína de interacción hsp70	3	0	0,0296
proteína hipotética FLJ22795	3	0	0,0296
proteína hipotética FLJ22795	3	0	0,0296
interferón inducido por proteína transmembrana 1 (9-27)	3	0	0,0296
interferón inducido por proteína transmembrana 1 (9-27)	3	0	0,0296
Motivo IQ que contiene F1 (IOCF1)	3	0	0,0296
repeticiones ricas en leucina y Motivo IQ que contiene 2 (LRRIQ2)	3	0	0,0296
repeticiones ricas en leucina y Motivo IQ que contiene 2 (LRRIQ2)	3	0	0,0296
paralemmína 2	3	0	0,0296
paralemmína 2	3	0	0,0296
Dominio RWD que contiene 1	3	0	0,0296
familia de transportadores de soluto 7	3	0	0,0296
familia de transportadores de soluto 7	3	0	0,0296
tropomiosina 1 (alfa)	3	0	0,0296
tropomiosina 1 (alfa)	3	0	0,0296
candidato 4 subtransferible supresor de tumores	3	0	0,0296
4A de tipo ubiquitina	3	0	0,0296
vestigial de tipo 4 (Drosophila) (VGLL4)	3	0	0,0296
dominio de la repetición WD 16	3	0	0,0296
dominio de la repetición WD 16	3	0	0,0296

Identificación de antígenos	Recuento de Cáncer	Recuento Negativo para cáncer	P-Valor
proteína quinasa activada por proteína quinasa 3 activada por mitógenos	3	0	0,0296
proteína quinasa activada por proteína quinasa 3 activada por mitógenos	3	0	0,0296
proteína quinasa 1 asociada a la muerte (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 de antígeno de melanoma	3	0	0,0296
proteína quinasa activada por proteína quinasa activada por mitógeno 3 (MAPKAPK3)	3	0	0,0296
nei de tipo 2 (E. coli) (NEIL2)	3	0	0,0296
proteína quinasa C y substrato de 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 TIA	3	0	0,0296
factor trébol 2 (proteína espasmolítica 1) (TFF2)	3	0	0,0296
uroporfirinógeno III sintasa (porfiria eritropoyética congénita) (UROS)	3	0	0,0296
proteína 29 kDa inducida por citoquina (CIP29)	3	0	0,0296
proteína 106C transmembrana (TMEM106C)	3	0	0,0296
Marco de lectura abierto 11 del cromosoma 9	3	0	0,0296
O-6-metilguanina-ADN metiltransferasa (MGMT)	3	0	0,0296
Proteína 1 asociada a PDGF (PDAP1)	3	0	0,0296
Proteína 1 asociada a PDGF (PDAP1)	3	0	0,0296
polipéptido D de (ARN) polimerasa III (dirigida por ADN)	3	0	0,0296
Asociado a Rho	3	0	0,0296
Asociado a Rho	3	0	0,0296
proteína del motivo de unión a ARN	3	0	0,0296
tetraspanina 17	3	0	0,0296

Un segundo algoritmo calculó la especificidad del cáncer de la respuesta inmunitaria para una proteína como la diferencia entre la señal media para el cáncer y la señal media de las muestras negativas para cáncer dividido por la desviación típica de las intensidades de señal de las muestras negativas para cáncer. Esto tiene la ventaja de que las respuestas inmunitarias fuertes afectan el resultado más de las débiles. Los datos se ordenan para llevar a la cabeza aquellas proteínas con los valores más altos. Las 100 primeras enumeraciones identificadas por esta clasificación se muestran a continuación en la Tabla 8:

Tabla 8. Lista de ID de antígenos ordenados para llevar a la cabeza las proteínas de la más alta proporción S/N. El S/N se calculó dividiendo la diferencia de la intensidad media de la señal de los dos grupos (Media con Cáncer - Media Sin Cáncer) mediante la desviación típica del grupo Sin Cáncer (DT Sin Cáncer).

Identificación de antígenos	Dif. media/DT (sin cáncer)
Compañero de fusión TCF3 (E2A) (en leucemia infantil) (TFPT)	21,4
proteasa 45 específica de ubiquitina (USP45)	16,1

ES 2 606 015 T3

Identificación de antígenos	Dif. media/DT (sin cáncer)
proteasa 45 específica de ubiquitina (USP45)	15,6
enzima E2O de conjugación con ubiquitina	15,1
Compañero de fusión TCF3 (E2A) (en leucemia infantil) (TFPT)	13,9
enzima E2O de conjugación con ubiquitina	12,3
bobina en espiral 1 rica en prolina (PRRC1)	11,5
bobina en espiral 1 rica en prolina (PRRC1)	10
LLC/linfoma 10 de células B	9,8
familia 7 de transportadores de soluto	8,8
LLC/linfoma 10 de células B	8,7
homólogo de DnaJ (Hsp40)	8,2
homólogo de DnaJ (Hsp40)	8
familia 7 de transportadores de soluto	7,9
vestigial de tipo 4 (Drosophila) (VGLL4)	6,5
dominio SH3 y PX que contiene 3 (SH3PX3)	6,3
ciclina E2	6,1
dominio SH3 y PX que contiene 3 (SH3PX3)	6,1
ciclina E2	6
clon de ADNc IMAGE:3941306	5,9
Paralemmina	5,8
proteína 1 transmembrana inducida por interferón (9-27)	5,6
Paralemmina	5,4
ribulosa-5-fosfato-3-epimerasa	5,4
repeticiones ricas en leucina y Motivo IQ que contiene 2 (LRRIQ2)	5,3
ribulosa-5-fosfato-3-epimerasa	5,3
4 asociada al ciclo de división celular (CDCA4)	5,2
interferón inducido por proteína transmembrana 1 (9-27)	4,8
repeticiones ricas en leucina y Motivo IQ que contiene 2 (LRRIQ2)	4,7
proteína quinasa activada por proteína quinasa 3 activada por mitógenos	4,5
calcio/calmodulina dependiente de la proteína quinasa I (CAMK1)	4,4
proteína de interacción RAB3A (rabin3) -de tipo 1 (RAB3IL1)	4,3
dimetilarginina dimetilaminohidrolasa 2 (DDAH2)	4,2
Proteína de interacción con hsp70	4,1
Marco de lectura abierto 11 del cromosoma 9	4,1
proteína quinasa activada por proteína quinasa 3 activada por mitógenos	4,1
proteína 1 de la vesícula acrosómica (ACRV1)	4,1
triosafosfato isomerasa 1	4
triosafosfato isomerasa 1	3,8
uroporfirinógeno III sintasa (porfiria eritropoyética congénita) (UROS)	3,7
subfamilia C de receptores de tipo lectina de células asesinas	3,7

ES 2 606 015 T3

Identificación de antígenos	Dif. media/DT (sin cáncer)
receptor alfa relacionado con estrógenos (ESRRA)	3,6
proteína de la vesícula acrosómica 1 (ACRV1)	3,6
4 asociada al ciclo de división celular (CDCA4)	3,6
proteínas de interacción con RAB3A (rabin3) - de tipo 1 (RAB3IL1)	3,5
proteína quinasa 1 asociada a muerte (DAPK1)	3,5
proteína quinasa C y sustrato de caseína quinasa en neuronas 2	3,5
Tropomodulina 1	3,4
Tropomodulina 1	3,4
Marco de lectura abierto 117 del cromosoma 1	3,4
dimetilarginina dimetilaminohidrolasa 2 (DDAH2)	3,4
receptor alfa relacionado con estrógenos (ESRRA)	3,2
que contiene dominio de homología con pleckstrina	3,1
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 proteína quinasa activada por mitógeno 3 (MAPKAPK3)	3,1
homólogo 37 del ciclo de división celular Cdc37 (S. cerevisiae) - de tipo 1	3
candidato 4 subtransferible supresor de tumores	3
Dominio RWD que contiene 1	3
proteína hipotética FLJ22795	3
homólogo 37 del ciclo de división celular Cdc37 (S. cerevisiae) - de tipo 1	2,9
dominio WW que contiene proteína ubiquitina E3 ligasa 2	2,9
dominio PDZ que contiene 1 (PDZK1)	2,9
proteína quinasa activada por proteína quinasa activada por mitógeno 3 (MAPKAPK3)	2,9
factor de transcripción que se une al potenciador 3 de IGHM (TFE3)	2,9
caja de cabeza de horquilla A3 (FOXA3)	2,8
Marco de lectura abierto 117 del cromosoma 1	2,8
repetición de anquirina y dominio del motivo alfa estéril que contiene 3	2,8
dominio OCIA que contiene 1 (OCIAD1)	2,8
polimerasa (dirigida por ADN)	2,8
SMAD	2,8
KIAA0157 (KIAA0157)	2,8
LLC/linfoma 7C de células B (BCL7C)	2,8
proteína ribosomal S6 quinasa	2,8
Marco de lectura abierto 11 del cromosoma 9	2,7
proteína ribosomal S6 quinasa	2,7
proteína inducida por citoquina 29 kDa (CIP29)	2,7
factor 2 de unión al receptor nuclear	2,7

Identificación de antígenos	Dif. media/DT (sin cáncer)
regulador 1 del factor C1 de células anfitrionas (dependiente de XPO1) (HCFC1R1)	2,7
quinasa de tipo STE20 (levadura) (SLK)	2,7
dominio OCIA que contiene 1 (OCIAD1)	2,6
proteína quinasa C y substrato de caseína quinasa en neuronas 2	2,6
homólogo de quaking	2,6
nexina 16 de clasificación (SNX16)	2,6
proteína citosólica de linfocitos 1 (L-plastina) (LCP1)	2,6
marco de lectura abierto 7 del cromosoma 21	2,5
quinasa de tipo STE20 (levadura) (SLK)	2,5
regulador 1 del factor C1 de células anfitrionas (dependiente de XPO1) (HCFC1R1)	2,5
Proteína de interacción con hsp70	2,5
homólogo de quaking	2,5
de factor de transcripción de unión al potenciador 3 de IGHM (TFE3)	2,5
SMAD	2,4
dominio WW que contiene la proteína ubiquitina E3 ligasa 2	2,4
marco de lectura abierto 7 del cromosoma 21	2,4
dominio PDZ que contiene 1 (PDZK1)	2,4
de tipo acetilserotonina O-metiltransferasa	2,4
LLC/linfoma 7C de células B (BCL7C)	2,3
proteína ribosomal S19 (RPS19)	2,3
O-6-metilguanina-ADN metiltransferasa (MGMT)	2,3

Mediante la comparación de los resultados de clasificación de las Tablas 7 y 8 y el examen de las señales generadas por las muestras de cáncer y negativas para cáncer para cada proteína se seleccionaron las siguientes 25 proteínas mostradas a continuación en la Tabla 9 para investigación adicional.

5

Tabla 9. Veinticinco primeras proteínas seleccionadas para investigación adicional.

Clon	identificación de antígenos
BC007015.1	ciclina E2
NM_002614.2	dominio PDZ que contiene 1 (PDZK1)
NM_001612.3	proteína de la vesícula acrosómica 1 (ACRV1)
NM_006145.1	homólogo de DnaJ (Hsp40)
BC011707.1	factor 2 de unión al receptor nuclear
BC008567.1	marco de lectura abierto 7 del cromosoma 21
BC000108.1	dominio WW que contiene la proteína ubiquitina E3 ligasa 2
BC001662.1	proteína quinasa activada por proteína quinasa 3 activada por mitógenos
BC008037.2	proteína quinasa C y substrato de caseína quinasa en neuronas 2
NM_005900.1	SMAD
NM_013974.1	dimetilarginina dimetilaminohidrolasa 2 (DDAH2)

Clon	identificación de antígenos
NM_000375.1	uroporfirinógeno III sintasa (porfiria eritropoyética congénita) (UROS)
NM_145701.1	4 asociada al ciclo de división celular (CDCA4)
BC016848.1	Marco de lectura abierto 117 del cromosoma 1
BC014307.1	marco de lectura abierto 11 del cromosoma 9
BC000897.1	interferón inducido por proteína transmembrana 1 (9-27)
NM_024548.2	repeticiones ricas en leucina y Motivo IQ que contiene 2 (LRR1Q2)
BC013778.1	familia 7 de transportadores de soluto
BC032449.1	Paralemmena
NM_153271.1	dominio SH3 y PX que contiene 3 (SH3PX3)
NM_013342.1	Compañero de fusión TCF3 (E2A) (en leucemia infantil) (TFPT)
NM_006521.3	factor de transcripción que se une al potenciador 3 de IGHM (TFE3)
BC016764.1	ribulosa-5-fosfato-3-epimerasa
BC014133.1	homólogo 37 del ciclo de división celular Cdc37 (S. cerevisiae) - de tipo 1
BC053545.1	tropomiosina 1 (alfa)

E. Ciclina E2

5 Se encontró que estaban presentes dos formas de la ciclina E2 en el ProtoArray™. La forma identificada como BC007015.1 Genbank (SEQ ID NO: 1) mostró una fuerte inmunorreactividad con varios de los conjuntos de muestras de cáncer y reactividad mucho más baja con los conjuntos de enfermedad benigna y normales (negativas para cáncer). En contraste, la forma identificada como BC020729.1 Genbank (SEQ ID NO: 2) mostró poca reactividad con cualquiera de las muestras reunidas de cáncer o negativas para cáncer. Como se muestra más abajo, un alineamiento de secuencia de las dos formas mostró identidad a lo largo de 259 aminoácidos, con diferencias en ambas regiones N-terminales y C-terminales. BC020729.1 tiene 110 aminoácidos en el extremo N-terminal y 7 aminoácidos en el C-terminal que no están presentes en BC007015.1. BC007015.1 tiene 37 aminoácidos en el C-terminal que no están presentes en BC020729.1. Debido a que sólo la forma BC007015.1 muestra inmunoreactividad, esto se atribuye a la porción de 37 aminoácidos en el extremo C-terminal.

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

20 BC007015.1 1 M

BC020729.1 1 MSRRSSRLQAKQQPQPSQTESPQEAQIIQAKKRKTTQDVKKRREEVTKKHQYEIRNCWPP *

25 BC007015.1 BC020729.1 61 VLGGISPCIIETPHKEIGTSDFSRFTNYRFKNLFINPSPLPDLSWG C

BC007015.1 2 SKEVWLNMLKKESRYVHDKHFEVLHSDLEPQMRSILLDWLLEVCVYTLHRETFYLAQDF
BC020729.1 110 SKEVWLNMLKKESRYVHDKHFEVLHSDLEPQMRSILLDWLLEVCVYTLHRETFYLAQDF

30 BC007015.1 62 FDRFMLTQKDINKNMLQLIGITSLFIASKLEEIYAPKLQEFAYVTDGACSEEDILRMELI
BC020729.1 170 FDRFMLTQKDINKNMLQLIGITSLFIASKLEEIYAPKLQEFAYVTDGACSEEDILRMELI

BC007015.1 122 ILKALKWELCPVTIISWNLFLQVDALKDAPKVLLPQYSQETFIQIAQLLDLCILAIDSL
BC020729.1 230 ILKALKWELCPVTIISWNLFLQVDALKDAPKVLLPQYSQETFIQIAQLLDLCILAIDSL

35 BC007015.1 182 EFQYRILTAAALCHFTSIEVVKKASGLEWDSISECVDWMVPFVNVVKSTSPVKLKTFFKI
BC020729.1 290 EFQYRILTAAALCHFTSIEVVKKASGLEWDSISECVDWMVPFVNVVKSTSPVKLKTFFKI

40

BC007015.1 242 PMEDRHNIQTHNTYLAMLEEVNYINTFRKGGQLSPVCNGGIMTPPKSTEKPPGKH
 BC020729.1 350 PMEDRHNIQTHNTYLAMLCMISSHV *****

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

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

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

Péptidos derivados de BC007015.1

Cada uno de los péptidos E2-1 y E2-2 fue ligado a BSA mediante la activación de la BSA con maleimida seguido de acoplamiento del péptido. La BSA activada se prepara de acuerdo con el siguiente protocolo: a 8,0 mg de BSA en 200 µL de PBS se les añadió 1 mg de GMBS (N-(gamma-maleimido-butilil-oxi)succinimida, Pierce, Rockford IL) en 20 µL de DMF y 10 µL de trietanolamina 1 M de pH 8.4. Después de 60 minutos, la mezcla se hizo pasar a través de una columna Sephadex G50 con tampón PBS recogiendo fracciones 400 µL. A la BSA-Mal activada (100 µL) se le añadieron o bien 2,5 mg de péptido E2-1 o bien 3,2 mg de péptido E2-2. En ambos casos, la mezcla se agitó a vórtice y se colocó sobre hielo durante 15 minutos, después de lo cual la mezcla se trasladó a temperatura ambiente durante 25 minutos. Los productos acoplados, BSA-Mal-E2-1 (BM-E2-1) y BSA-Mal-E2-2 (BM-E2-2), se hicieron pasar a través de una columna Sephadex G50 para su limpieza.

Las proteínas y los péptidos se acoplaron a microesferas Luminex™ utilizando dos métodos. El primer método se describe en el Ejemplo 10C y se conoce como el "método directo". El segundo método se conoce como el "método de pre-activación" y utiliza el siguiente protocolo: A los pocillos de una placa de trafiltración Omega 10k se les añadieron 100 µL de agua; después de 10 minutos se colocaron a vacío. Cuando pocillos estuvieron vacíos, se añadieron 20 µL de MES (100 mM) de pH 5,6 y 50 µL a cada uno de conjuntos de esferas 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 a los pocillos en la columna 2 filas A, B, C, D y E se les añadieron 10 µL de NHS (20 mg/mL) en MES y 10 µL de EDAC (10 mg/mL) en MES. Después de 45 minutos de agitación en oscuridad, la placa se colocó a vacío para aspirar el tampón y los reactivos que no habían reaccionado. Cuando los pocillos estuvieron vacíos se añadieron 100 µL de MES y se permitió que pasaran a través de las membranas. La placa se retiró del vacío y se añadieron 20 µL de MES y 50 µL de agua. A los pocillos indicados en la Tabla 10 se les añadieron 4 µL de cada proteína o péptido (excepto DNAJB1, añadidos 2 µL) y se agitó con pipetas para dispersar las esferas. La placa se agitó durante 30 minutos en un agitador, a continuación se añadieron 10 mg de EDAC de 5 µL/mL en MES a la columna 1, filas EFGH (para el acoplamiento directo), y la placa se agitó en un agitador durante 30 minutos, después se pusieron a vacío para eliminar el tampón y los reactivos que no habían reaccionado. Cuando los pocillos estuvieron vacíos se les añadieron 50 µL de PBS y las mezclas se agitaron y la placa se colocó a vacío. Cuando los pocillos estuvieron vacíos se añadieron 50 µL de PBS y las mezclas se agitaron con pipetas para dispersar las esferas, y se incubaron durante 60 minutos en el agitador. Para detener la reacción se añadieron de 200 µL de PBN y las mezclas se sometieron a sonicación.

La Tabla 10 siguiente resume las diferentes presentaciones péptidos y proteínas de ciclina E2 en los diferentes conjuntos de esferas. Los péptidos, E2-1 y E2-2, se acoplaron a BSA que a continuación se acoplaron a las esferas utilizando el método de preactivación (ID de esferas 25 y 26) o el método directo (ID de esferas 30 y 31). Los péptidos, E2-1 y E2-2, también se acoplaron a las esferas sin BSA utilizando el método de preactivación (ID de esferas 28 y 29) o el método directo (ID de esferas 33 y 34). Las esferas 35, 37, 38, 39, y 40 se recubrieron con la proteína utilizando el método de preactivación.

Tabla 10. Resumen de las diferentes presentaciones de péptidos y proteínas de ciclina E2 sobre diferentes esferas.

Columna	Fila	ID. de la esfera	antígeno	Fuente	Método de acoplamiento
1	A	25	BM-E2-1	3,9 mg/mL	Preactivación
1	B	26	BM-E2-2	2,4 mg/mL	Preactivación
1	C	28	E2-1	21 mg/mL	Preactivación
1	D	29	E2-2	40 mg/mL	Preactivación
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	Preactivación

Columna	Fila	ID. de la esfera	antígeno	Fuente	Método de acoplamiento
2	B	37	MAPKAPK3	(Genway, San Diego, CA) 0,5 mg/mL	Preactivación
2	C	38	p53	(Biomol, Plymouth Meeting, PA) 0,25 mg/mL	Preactivación
2	D	39	TMOD1	(Genway, San Diego, CA) 0,8 mg/mL	Preactivación
2	E	40	DNAJB1	(Axxora, San Diego, CA) 1 mg/mL	Preactivación

Las esferas se sometieron a ensayo con sueros de pacientes de la siguiente manera: a 1 mL de PBN se le añadieron 5 µL de cada preparación de esferas. La mezcla se sometió a sonicación y se centrifugó, y las esferas sedimentadas se lavaron con 1 mL de BSA al 1% en PBS y se resuspendieron en 1 mL del mismo tampón. A una placa de filtro Supor de 1.2u (Pall Corporation, East Hills, NY) se le añadieron 100 µL de PBN/Tween (BSA al 1% en PBS que contenía Tween 20 al 0,2%). Después de 10 minutos se filtró la placa, y se añadieron 50 µL de PBN/Tween al 0,2% (BSA al 1% en PBS que contenía Tween 20 al 0,2%). A cada pocillo se añadieron 20 µL de mezcla de microesferas y 20 µL de suero (1:50) como se muestra en la Tabla 11. El suero era o bien suero de paciente humano o suero anti-GST de conejo. La placa se colocó en un agitador en oscuridad. Después de 1 hora, la placa se filtró y se lavó con 100 µL de PBN/Tween tres veces. Se añadieron 50 µL de anti-IgG humana-RPE (1:400) (Sigma, St. Louis, MO) para detectar anticuerpos humanos, mientras que se añadieron 50 µL de anti-IgG de conejo-RPE (1:200) para detectar los anticuerpos anti-GST de conejo. La placa se colocó en un agitador en oscuridad durante 30 minutos después de lo cual las esferas se filtraron, se lavaron y se analizaron en Luminex™.

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

Tabla 11: Resultados de Luminex para esferas recubiertas con péptidos y proteínas de ciclina E2, expuestos a sueros de pacientes.

ID. de la esfera									
	25	26	28	29	35	30	31	33	34
	Preactivación					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
anti-GST de conejo	13	14	40	21	1358	10	13	56	22

Resulta evidente a partir de la Tabla 11 anterior que las esferas 25 y 30, que contenían el péptido E2-1 ligado a BSA y acoplado directamente (utilizando el método directo) o por medio de la activación previa (o método de Preactivación) de las esferas, respectivamente, proporcionaron las señales más fuertes. El péptido E2-1 acoplado sin el portador de BSA también proporcionó señales fuertes, aunque sólo aproximadamente la mitad que las proporcionadas con el portador de BSA. El péptido E2-2 produjo señales mucho más bajas cuando se acopló a través del portador de BSA, y señales casi indetectables sin el portador de BSA. La proteína CCNE2 completa (que contenía un marcador de fusión GST N-terminal) mostró señales muy por encima de las de cualquier forma de péptido E2-2, pero aún muy por debajo de las del péptido E2-1, lo que sugiere que contiene la porción inmunorreactiva de la secuencia, pero a menor densidad en la esfera. Su señal con anti-GST de conejo muestra que esta proteína de fusión de GST se acoplaba con éxito a la microesfera.

Las proteínas mostradas en la Tabla 12, a continuación, se aplicaron como recubrimiento sobre esferas Luminex SeroMap™ mediante preactivación y métodos directos como se ha descrito anteriormente, y mediante recubrimiento pasivo. Para el recubrimiento pasivo, se añadieron 5 µg de la proteína, en solución tal como se recibió del proveedor, a 200 µL de esferas SeroMap™, la mezcla se sometió a vórtice, y se incubó durante 5 horas a temperatura ambiente, después durante 18 horas a 4°C, a continuación se centrifugaron hasta sedimentación y el sedimento se lavó y se resuspendió en PBN.

Tabla 12. Proteínas aplicadas como recubrimiento sobre esferas Luminex SeroMap™ mediante preactivación y métodos directos.

Revestimiento	Proteína	Cuenta	Fuente
Preactivación	TMP21-ECD	1	Abbott, North Chicago, IL
Preactivación	dominio NPC1L1C	5	Abbott, North Chicago, IL
Preactivación	PSEN2(1-86aa)	14	Abbott, North Chicago, IL
Preactivación	IgG humana	22	Abbott, North Chicago, IL
Preactivación	BM-E2-2	26	Abbott, North Chicago, IL
Directo	BM-E2-1	30	Abbott, North Chicago, IL
Preactivación	TMOD1	39	Genway, San Diego, CA
Preactivación	DNAJB1	40	Axxora, San Diego, CA
Preactivación	PSMA4	41	Abnova, Taipei City, Taiwán
Preactivación	RPE	42	Abnova, Taipei City, Taiwán
Preactivación	CCNE2	43	Abnova, Taipei City, Taiwán
Preactivación	PDZK1	46	Abnova, Taipei City, Taiwán
Directo	CCNE2	49	Genway, San Diego, CA
Preactivación	Paxilina	53	BioLegend, San Diego, CA
Directo	anfifisina	54	LabVision, Fremont, CA
Preactivación	CAMK1	55	Upstate, Charlottesville, VA
Pasivo	DNAJB11	67	Abnova, Taipei City, Taiwán
Pasivo	RGS1	68	Abnova, Taipei City, Taiwán
Pasivo	PACSIN1	70	Abnova, Taipei City, Taiwán
Pasivo	SMAD1	71	Abnova, Taipei City, Taiwán
Pasivo	p53	72	Biomol, Plymouth Meeting, PA
Pasivo	RCV1	75	Genway, San Diego, CA
Pasivo	MAPKAPK3	79	Genway, San Diego, CA

- 5 Se sometieron a ensayo muestras de suero de 234 pacientes (87 cánceres, 70 de enfermedad benigna, y 77 normales). Los resultados de este ensayo se analizaron mediante curvas ROC. Las AUC calculadas para cada antígeno se muestra en la Tabla 13 a continuación.

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

Proteína	AUC
péptido 1 de ciclina E2	0,81
proteína ciclina E2 (Genway)	0,74
peptide2 ciclina E2	0,71
TMP21-ECD	0,66
dominio NPC1L1C	0,65
PACSIN1	0,65
p53	0,63
proteína quinasa activada por proteína quinasa activada por mitógeno (MAPKAPK3)	0,62
Tropomodulina 1 (TMOD1)	0,61
PSEN2(1-86aa)	0,60

Proteína	AUC
proteína de unión a ADN J 1 (DNAJB1)	0,60
proteína de unión a ADN J 11 (DNAJB11)	0,58
RCV1	0,58
(proteína quinasa dependiente de calcio/calmodulina - CAMK1 1)	0,57
SMAD1	0,57
ANFIFISINA 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 1 que contiene el dominio PDZ (PDZK1)	0,47

Ejemplo 5. Espectrometría de masas

A. Preparación de la muestra mediante elución secuencial de una esfera magnética mixta (MMB)

Las muestras de suero se descongelaron y se mezclaron con un volumen igual de tampón Sol B de Invitrogen. La mezcla se agitó a vórtice y se filtró a través de un filtro de 0,8 µm (Sartorius, Goettingen, Alemania) para aclarar y eliminar los residuos antes de su tratamiento posterior. Se realizó una preparación de Muestra Automatizada en una placa de 96 pocillos KingFisher® (Thermo Fisher, Scientific, Inc., Waltham, MA) utilizando una mezcla de esferas magnéticas de intercambio aniónico fuerte Dynal® (Invitrogen) y de intercambio catiónico débil Abbott Laboratories (Abbott, Abbott Park, IL). Típicamente las esferas de intercambio aniónico tienen aminas cuaternarias o grupos dietilamina basados en hidrocarburos como grupos funcionales terminales y las esferas de intercambio catiónico débil tienen típicamente grupos funcionales basados en ácido sulfónico (ácido carboxílico). Las esferas de intercambio catiónico de Abbott (esferas CX) estaban a una concentración de 2,5% (masa/volumen) y las esferas de intercambio aniónico fuerte DYNAL® (cuentas AX) estaban a una concentración de 10 mg/mL. Inmediatamente antes de la preparación de las muestras, las esferas de intercambio catiónico se lavaron una vez con Tris.HCl 20 mM, pH 7,5, Triton X100 al 0,1% reducido (tampón Tris-Triton). Se utilizaron otros reactivos, Tris.HCl 20 mM, pH 7,5 (tampón Tris), ácido trifluoroacético al 0,5% (en lo sucesivo "solución de TFA") y acetonitrilo al 50% (en lo sucesivo "solución de acetonitrilo"), en esta preparación de la muestra, y fueron de preparación propia. Los reactivos y las muestras se configuraron en la placa de 96 pocillos de la siguiente manera:

La Fila A contenía una mezcla de 30 µL de esferas AX, 20 µL de esferas CX y 50 µL de tampón Tris.
 La Fila B contenía 100 µL de tampón Tris.
 La Fila C contenía 120 µL de tampón Tris y 30 µL de la 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 de TFA.
 La Fila G contenía 50 µL de solución de Acetonitrilo.
 La Fila H estaba vacía.

Las esferas y el tampón de la fila A se premezclan y las cuentas se recolectan con el recuento de Recolección de 3 (parámetro de aparato que indica cuántas veces la sonda magnética entra en solución para recolectar las esferas magnéticas) y se transfieren a la fila B para el lavado en tampón Tris - con ajuste de liberación "rápido", ajuste de lavado - medio y tiempo de lavado de 20 segundos. Al final de la etapa de lavado de esferas, las esferas se recogen con el recuento de Recolección de 3 y se transfieren a la fila C para unir la muestra. El ajuste de liberación de la esfera es rápido. La unión de la muestra se realiza con ajuste "lento", con tiempo de unión de 5 minutos. Al final de la etapa de unión, las esferas se recogen con el recuento de Recolección de 3. Las esferas recogidas se transfieren a la fila D para la primera etapa de lavado -ajuste de liberación "rápido", ajuste de lavado - medio, con tiempo de lavado de 20 segundos. Al final de la primera etapa de lavado, las esferas se recogen con el recuento de Recolección 3. Las esferas recogidas se transfieren a la fila E, para la segunda etapa de lavado -ajuste de liberación "rápido", el ajuste de lavado - medio, con tiempo de lavado de 20 segundos. Al final de la segunda etapa de lavado, las esferas se recogen con el recuento de Recolección 3. Las esferas recogidas se transfieren a la fila F para la elución en la liberación en solución de TFA -con ajuste de liberación "rápido", ajuste de elución - rápido y tiempo de

elución de 2 minutos. Al final de la etapa de elución de TFA, las esferas se recogen con el recuento de Recolección 3. Este eluyente de TFA se recogió y se analizó mediante espectrometría de masas. Las esferas recogidas se transfieren a la fila G para la elución en la de solución de acetonitrilo -con ajuste de liberación "rápido", ajuste de elución - rápido y tiempo de elución de 2 minutos. Después de la elución, las esferas se retiran con el recuento de Recolección 3 y se vierten a la fila A. El eluyente acetonitrilo (ACN) se recogió y se analizó mediante espectrometría de masas.

Todas las muestras se analizan 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 preparación de muestras diana MALDI automatizada. Por lo tanto, cada muestra proporcionó dos eluyentes para el análisis de espectrometría de masas.

Se utilizó un robot CLINPROT (Bruker Daltonics Inc., Billerica, MA) para la preparación de las dianas MALDI MS antes de la consulta. En resumen, el proceso implicó la carga de la placa de muestra que contenía las muestras de suero eluidas y los viales que contenían la solución matriz de MALDI (10 mg/mL de ácido Sinapínico en Acetonitrilo al 70%) en las posiciones designadas en el robot. Se cargó un archivo que contenía el procedimiento de aplicación y se inició desde el ordenador que controlaba el robot. En este caso, el procedimiento de aplicación implicó la aspiración de 5 µL de solución de la matriz y su dispensación en la placa de matriz, seguido de 5 µL de muestra. La mezcla previa de la muestra y de la matriz se lleva a cabo mediante la aspiración de 5 µL de la mezcla y su dispensación varias veces en la placa de la matriz. Después de la premezcla, se aspiraron 5 µL de la mezcla y 0,5 µL se depositaron en cuatro aplicaciones contiguas en la diana de chips de anclaje (Bruker Daltonics Inc., Billerica, MA). Los 3 µL restantes de la solución se vertieron en el contenedor de residuos. La aspiración de más muestra de la necesaria minimizó la formación de burbujas de aire en las puntas desechables que pueden dar lugar a aplicaciones perdidas durante el vertido de la muestra sobre la diana de chips de anclaje.

B. Preparación de la muestra mediante cromatografía de interacción hidrófoba de esferas magnéticas C8 (HIC-C8 MB)

Las muestras de suero se mezclaron con tampón SOLB y aclararon con filtros como se describe en el Ejemplo 5A. La preparación de la muestra automatizada se realizó en una placa de 96 pocillos KingFisher[®] utilizando kits de purificación CLINPROT conocidos como 100 MB-HIC 8 (Bruker Daltonics Inc., Billerica, MA). El kit incluye esferas magnéticas C8, solución de unión, y solución de lavado. Todos los demás reactivos fueron adquiridos de Sigma Chem. Co., si no se indica lo contrario. Los reactivos y las muestras se dispusieron en la placa de 96 pocillos de la siguiente manera:

La Fila A contenía una mezcla de 20 µl de esferas 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 inmediatamente antes de la etapa de elución para minimizar la evaporación del disolvente orgánico).

La Fila G contenía 100 µl de agua DI.

La Fila H estaba vacía.

Las esferas de la fila A se mezclaron previamente y se recogen con un "recuento de Recolección" de 3 y se transfirieron a la fila B para unir la muestra. El ajuste de liberación de la esfera se establece en "rápido" con un tiempo de liberación de 10 segundos. La unión de la muestra se realizó con el ajuste "lento" durante 5 minutos. Al final de la etapa de unión, las esferas se recogieron 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 tiempo = 10 segundos, ajuste de lavado = medio con tiempo = 20 segundos). Al final de la primera etapa de lavado, las esferas se recogieron con un "recuento de Recolección" de 3 y se transfirieron a la fila D para una segunda etapa de lavado con los mismos parámetros que en la primera etapa de lavado. Al final de la segunda etapa de lavado, las esferas se recogieron una vez más y se transfirieron a la fila E para una tercera y última etapa de lavado como se ha descrito anteriormente. Al final de la tercera etapa de lavado, el KingFisher[™] se detuvo 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. Después de la adición del acetonitrilo, se reanudó el procedimiento. Las esferas recogidas de Fila E se transfirieron a la Fila F para la etapa de elución (ajuste de liberación = rápido con tiempo = 10 segundos, ajuste de elución = rápido con tiempo = 2 minutos). Después de la etapa de elución, las esferas se retiraron y se vertieron a la fila G. Todas las muestras fueron por duplicado, como se ha descrito anteriormente en el Ejemplo 5a.

Se utilizó un robot CLINPROT (Bruker Daltonics Inc., Billerica, MA) para la preparación de las dianas MALDI antes de la consulta mediante MS como se ha descrito en la sección anterior con sólo modificaciones mínimas en la matriz MALDI utilizada. En este caso, en lugar de SA, se utilizó HCCA (HCCA 1 mg/mL en ACN al 40%/MeOH al 50%/agua al 10%, v/v/v). Todos los demás parámetros siguieron siendo los mismos.

C. Preparación de la muestra utilizando el chip SELDI

Se utilizaron los siguientes reactivos:

- 5 1. tampón de fosfato 100 mM, 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. solución de ácido sinapínico de 10 mg/mL, preparada disolviendo una cantidad pesada de ácido sinapínico en una cantidad suficiente de una solución preparada mezclando volúmenes iguales de acetonitrilo y ácido trifluoroacético acuoso al 0,4% (v/v) para proporcionar una concentración final de 10 mg de ácido sinapínico por mL de solución.
- 10 3. el agua desionizada, el ácido sinapínico y el ácido trifluoroacético eran de Fluka Chemicals. El acetonitrilo era de Burdick and Jackson.

15 Se utilizan matrices Q10 ProteinChip en la configuración de ocho aplicaciones y se obtuvieron Bioprocesadores para mantener las matrices en una matriz 12 X 8 con una huella idéntica a una microplaca convencional de Ciphergen. La superficie activa de Q10 es un intercambiador aniónico fuerte de amina cuaternaria. Se utilizó un espectrómetro de masas de tiempo de vuelo de ionización/desorción láser asistida por matriz (MALDI) Ciphergen ProteinChip System, Serie 4000 para analizar los péptidos unidos a la superficie del chip. Todos los productos Ciphergen se obtuvieron de Ciphergen Biosystems, Dumbarton, California.

20 Todos las transferencias de líquido, diluciones, y lavados se realizaron mediante una pipeta robótica Hamilton Microlab STAR de Hamilton Company, Reno, Nevada.

25 Las muestras de suero se descongelaron a temperatura ambiente y se mezclaron mediante agitación suave. Los viales que contenían la muestra se cargaron en soportes de muestra de 24 posiciones sobre la pipeta Hamilton; Se cargaron cuatro soportes de muestra con un total de 96 muestras. Dos Bioprocesadores que albergaban los chips Q10 (192 aplicaciones en total) se colocaron en la plataforma de la pipeta Hamilton. Los recipientes con tampón fosfato 100 mM y agua desionizada se cargaron en la pipeta Hamilton. También se colocaron puntas de pipeta desechables en la plataforma del aparato.

30 Todo el procesamiento de las muestras fue totalmente automatizado. Cada muestra se diluyó 1 a 10 en dos alícuotas separadas mezclando 5 microlitros de suero con 45 microlitros de tampón de fosfato en dos pocillos separados de una microplaca en la plataforma de la pipeta Hamilton. Los chips Q10 se activaron mediante la exposición de cada aplicación a dos alícuotas de 150 microlitros de tampón de fosfato. Se permitió que el tampón activara la superficie durante 5 minutos después de cada adición. Después de aspirar la segunda alícuota de cada aplicación, se añadieron 25 microlitros de suero diluido a cada aplicación y se incubaron durante 30 minutos a temperatura ambiente. Cada muestra se diluyó dos veces con una sola alícuota de cada dilución colocada en una aplicación de un chip Q10. Después de la aspiración del suero diluido, cada aplicación se lavó cuatro veces con 150 microlitros de tampón de fosfato y, finalmente, con 150 microlitros de agua desionizada. Los chips procesados se secaron al aire y se trataron con ácido sinapínico, la matriz se utilizó para permitir el proceso MALDI en el Ciphergen 4000. La solución matriz de ácido sinapínico se cargó en la pipeta Hamilton mediante la colocación de una microplaca de 96 pocillos, cada pocillo se cargó con solución de ácido sinapínico, sobre la plataforma del aparato. Se utilizó una pipeta con cabezal de 96 para agregar 1 microlitro de la matriz de ácido sinapínico a cada aplicación en un Bioprocesador simultáneamente. Después de un período de secado 15 minutos, se añadió una segunda alícuota de 1 microlitro a cada aplicación y se dejó secar.

D. Adquisición de datos MALDI-TOF AutoFlex de preparación de muestras de esferas mixtas

50 El intervalo de adquisición por parte del aparato se ajustó de m/z 400 a 100.000. El aparato se calibró externamente en el modo lineal utilizando patrones de calibración de Bruker que cubren un intervalo de masa de 2-17 kDa. Con el fin de recoger espectros de alta calidad, las adquisiciones fueron totalmente automatizadas con el control difuso conectado, excepto para el láser. El control difuso del láser se apagó de forma que la potencia del láser se mantuvo constante durante la duración del experimento. Dado que el aparato se calibra generalmente a una potencia de láser fija, la precisión se beneficia de mantener una potencia de láser constante. Los otros ajustes de control difuso controlan la resolución y la S/N de los picos en el intervalo de masa de 2-10 kDa. Estos valores se optimizaron antes de cada adquisición y eligieron para maximizar la calidad de los espectros y reducir al mínimo el número de adquisiciones fallidas de muestra a muestra o de aplicación a aplicación. El deflector también se conectó para desviar iones de baja masa molecular (<400 m/z) para evitar la saturación del detector con los iones de la matriz y maximizar la señal procedente de la muestra. Además, antes de cada adquisición, se realizaron 5 disparos de calentamiento (LP aprox. 5-10% por encima del umbral) para eliminar cualquier exceso de matriz a medida que el haz de láser formaba la trama a través de la superficie de la muestra. Para cada espectro de masas, se añadieron simultáneamente 600 disparos de láser sólo si cumplían los criterios de resolución y S/N, establecidos anteriormente. Todos los demás espectros de calidad inferior se ignoraron y descartaron y no se utilizaron correcciones de los valores basales o algoritmos de ajuste durante la adquisición de los espectros sin tratamiento.

Los datos se archivaron, se transformaron en un eje de m/z común para facilitar la comparación y se exportaron en un formato ASCII portátil que podría ser analizado por medio de varios paquetes de soporte lógico estadístico. La transformación en un eje m/z común se llevó a cabo mediante el uso de un algoritmo de interpolación de desarrollo propio.

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

El intervalo de adquisición por parte del aparato se ajustó de m/z 1000 a 20.000 y se optimizó la sensibilidad y resolución. Todos los demás parámetros de adquisición y métodos de calibración se ajustaron como se ha descrito anteriormente en el Ejemplo 5d, con la excepción de que se añadieron simultáneamente 400 disparos de láser para cada espectro de masas.

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

Los Bioprocesadores se cargaron en un espectrómetro de masas de tiempo de vuelo MALDI Ciphergen 4000 utilizando los parámetros optimizados para el intervalo de masa entre 0-50.000 Da. Los datos se digitalizaron y se promediaron sobre las 530 adquisiciones aplicación para obtener un único espectro de corriente de iones vs. masa/carga (m/z). Cada espectro se exportó a un servidor y posteriormente se recuperó en forma de un archivo ASCII para el análisis post-adquisición.

G. Análisis de datos los de espectrometría de masas de la región de Interés

Los datos de espectrometría de masas consisten en los valores de masa/carga de 0-50.000 y sus correspondientes valores de intensidad. Se construyeron los conjuntos de datos de Cáncer y Negativos para Cáncer. El conjunto de datos de Cáncer consiste en los espectros de masas de todas las muestras de cáncer, mientras que conjunto de datos Negativos para Cáncer consiste en los espectros de masas de todas las muestras negativas para cáncer, incluyendo sujetos normales y pacientes con enfermedad pulmonar benigna. Los conjuntos de datos de Cáncer y Negativos para Cáncer se cargaron por separado en un programa de soporte lógico que realiza lo siguiente:

- la prueba t de Student se determina en cada valor de masa/carga registrado para proporcionar un valor de p.
- Los espectros de cáncer y negativos para cáncer se promedian para uno representativo de cada grupo.
- Se determina la razón logarítmica (Razón Log) de la intensidad de los espectros medios de cáncer y los espectros medios negativos para cáncer.

Se especificó que las ROI tenían diez o más valores de masa consecutivos con un valor p de menos de 0,01 y una razón Log absoluta de más de 0,1. Se encontraron 18, 36, y 26 regiones de interés en los conjuntos de datos MMB-TFA, MMB-ACN, y MB-HIC, respectivamente (Tablas 14a a 14c). Adicionalmente, se encontraron 124 ROI (<20 kDa) en los datos de SELDI como se muestra en la Tabla 14d. Las Tablas 14a a 14d enumeran las ROI de la presente invención, ordenadas por el aumento del valor medio de masa. La ROI proporcionada en la tabla es el valor medio de masa para el intervalo calculado (media del valor de la masa al inicio y la finalización para el intervalo proporcionado). La masa media de la ROI es referida simplemente como ROI en adelante. Las intensidades de cada ROI para cada muestra se sometieron a análisis ROC. El AUC para cada marcador fue referida también en las Tablas 14a-14d siguientes. En las Tablas 14a a 14c siguientes, la ROI calculada se obtuvo a partir del análisis de los perfiles de los grupos enfermos y no enfermos. Las muestras individuales se procesaron utilizando tres métodos diferentes: cromatografía de intercambio aniónico/catiónico de esferas magnéticas mixtas eluyendo con a) TFA (TFA) y eluyendo sucesivamente con b) acetonitrilo (ACN), c), utilizando la cromatografía de interacción hidrófoba (HIC). Cada método de preparación de la muestra se analizó de forma independiente con el propósito de obtener la ROI. Todos los espectros fueron recogidos con un espectrómetro de masas MALDI-TOF Bruker Autoflex. En la Tabla 14d siguiente, la ROI calculada se obtuvo a partir del análisis de los perfiles de los grupos enfermos y no enfermos. Todas las muestras se procesaron utilizando un chip Q-10. Todos los espectros se obtuvieron mediante un espectrómetro de masas SELDI-TOF Ciphergen 4000.

Tabla 14a.

ROI	ROI	Promedio	ROI	gran cohorte		pequeña cohorte	
inicio m/z	final m/z	ROI	nombre	Núm. obs	AUC	Núm. obs	AUC
2322,911	2339,104	2331	tfa2331	538	0,66	236	0,52
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

ROI	ROI	Promedio	ROI	gran cohorte		pequeña cohorte	
inicio m/z	final m/z	ROI	nombre	Núm. obs	AUC	Núm. obs	AUC
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	
iniciar m/z	extremo m/z	ROI	nombre	# orbes	AUC	# 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

ES 2 606 015 T3

ROI	ROI	Promedio	ROI	gran cohorte		pequeña cohorte	
iniciar m/z	extremo m/z	ROI	nombre	# orbes	AUC	# obs	AUC
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
22764,3	22771,69	22768	acn22768	519	0,68	244	0,62
22778,44	22788	22783	acn22783	519	0,68	244	0,63
22791,38	23147,21	22969	acn22969	519	0,70	244	0,63

Tabla 14c.

ROI	ROI	Promedio	ROI	gran cohorte		pequeña cohorte	
iniciar m/z	extremo m/z	ROI	nombre	# obs	AUC	# 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

ROI	ROI	Promedio	ROI	gran cohorte		pequeña cohorte	
iniciar m/z	extremo m/z	ROI	nombre	# obs	AUC	# obs	AUC
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	
iniciar m/z	extremo m/z	ROI	nombre	# obs	AUC	# 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
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

ES 2 606 015 T3

ROI	ROI	Promedio	ROI	gran cohorte		pequeña cohorte	
iniciar m/z	extremo m/z	ROI	nombre	# obs	AUC	# obs	AUC
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	S194	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

ES 2 606 015 T3

ROI	ROI	Promedio	ROI	gran cohorte		pequeña cohorte	
iniciar m/z	extremo m/z	ROI	nombre	# obs	AUC	# obs	AUC
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
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	pub1066	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

ROI	ROI	Promedio	ROI	gran cohorte		pequeña cohorte	
iniciar m/z	extremo m/z	ROI	nombre	# obs	AUC	# obs	AUC
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 multivariante de programa del paquete estadístico JMP™ (SAS Institute Inc., Cary, NC) para identificar las ROI que están altamente correlacionadas. Se extrajo una matriz del coeficiente de correlación de dos dimensiones del programa JMP y se analizó adicionalmente mediante Microsoft Excel. Por cada ROI, se identificó un conjunto de ROI para las que el coeficiente de correlación superaba 0,8. Estas regiones de interés juntas se convierten en una familia de ROI correlacionadas. La Tabla 15 muestra las familias que se correlacionan, sus correspondientes ROI miembros, el valor de AUC para las ROI miembros en la gran cohorte, y el promedio de los coeficientes de correlación para los demás miembros de la familia. Por lo tanto, se puede ver que las regiones de interés con masas de 3449 y 3494 están altamente correlacionados y pueden ser sustituidos unos por otros en el contexto de la presente invención.

Tabla 15. Familias de regiones de interés correlacionadas.

Grupo A (n = 2)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.
Pub3448	3449	0,72	0,81
Pub3493	3494	0,72	0,81
Grupo B (n = 2)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.
Pub4487	4488	0,75	0,8
Pub4682	4682	0,72	0,8
Grupo C (n = 9)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.
Pub8776	8776	0,56	0,8

Grupo C (n = 9)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.
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. de 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
Pub11066	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. de 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. de Corr.
Pub7218	7219	0,73	0,82
Pub7255	7255	0,72	0,73

Grupo E (n = 6)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.
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. de 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. de Corr.
Pub6249	6249	0,66	0,82
Pub12501	12502	0,66	0,87
Pub12717	12717	0,73	0,87
Grupo I (n = 5)			
Nombre de ROI	Miembros	AUC	Coef. de 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. de 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. de Corr.
Pub17858	17858	0,81	0,84
Pub18422	18422	0,73	0,92

Grupo K (n = 4)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.
Pub18766	18766	0,69	0,89
Pub18986	18986	0,65	0,91
Grupo L (n = 12)			
Nombre de ROI	Miembros	AUC	Coef. de 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. de Corr.
Pub2331	2331	0,65	0,9
tfa2331	tfa2331	0,66	0,9
Grupo N (n = 2)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.
Pub4557	4557	0,73	0,81
Pub4592	4592	0,71	0,81
Grupo O (n = 6)			
Nombre de ROI	Miembros	AUC	Coef. de 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. de 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

Grupo O (n = 6)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.
acn9371	acn9371	0,65	0,83
acn9662	acn9662	0,66	0,79
Grupo Q (n = 2)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.
acn9459	acn9459	0,66	0,91
acn9471	acn9471	0,7	0,91
Grupo R (n = 4)			
Nombre de ROI	Miembros	AUC	Coef. de 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. de Corr.
hic2728	hic2728	0,61	0,81
hic3276	hic3276	0,64	0,81
Grupo T (n = 6)			
Nombre de ROI	Miembros	AUC	Coef. de 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. de Corr.
hic8791	hic8791	0,58	0,8
hic8897	hic8897	0,61	0,8
Grupo V (n = 2)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.
a6453	tfa6453	0,74	0,84
a6652	a6652	0,72	0,84
Grupo W (n = 2)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.
hic6005	hic6005	0,69	0,74
hic5376	hic5376	0,68	0,74
Grupo X (n = 3)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.

Grupo W (n = 2)			
Nombre de ROI	Miembros	AUC	Coef. de Corr.
Pub4713	4714	0,73	0,83
Pub4750	4751	0,76	0,66
Pub4861	4861	0,72	0,65

Ejemplo 6. Análisis multivariante de biomarcadores utilizando análisis discriminante, análisis de árbol de decisiones y el análisis de componentes principales

- 5 Se llevaron a cabo análisis multivariantes, sobre los biomarcadores de inmunoanálisis y las Regiones de Interés. Todos los diferentes análisis se realizaron con el paquete estadístico JMP. Por simplicidad, el análisis discriminante (DA), el análisis de componentes principales (PCA) y el árbol de decisiones (DT) son referidos generalmente en la presente como memoria métodos multivariantes (MVM). Es notable mencionar que en el PCA, solamente se extrajeron los primeros 15 componentes principales, que representan más de 90% de la variabilidad total de los datos. Los pesos y/o comunales de factores se utilizan para extraer sólo uno de los factores (biomarcador) que más contribuyó a cada componente principal. Puesto que el cuadrado de los pesos de los factores refleja la contribución relativa de cada factor en cada componente principal, estos valores se utilizaron como base para seleccionar el marcador que más contribuyó a cada componente principal. De este modo, se extrajeron 15 factores (biomarcadores) que contribuyeron a la mayor parte de los primeros 15 componentes principales. En el DA, el proceso de selección de los marcadores se lleva a cabo hasta que la adición de más marcadores no tuvo ningún efecto en el resultado de clasificación. En general, el DA utiliza entre 5 y 8 biomarcadores. En el caso de los DT, se construyeron y evaluaron árboles de 6 nodos con aproximadamente 5 biomarcadores.

20 Los biomarcadores se evaluaron utilizando los métodos de muestreo autodocimante y de validación dejando uno fuera bien establecidos (Richard O. Duda et al. En Pattern Classification, 2ª edición, págs. 485, Wiley-Interscience (2000)). Se utilizó un procedimiento de entrenamiento de diez grupos para identificar los biomarcadores robustos que aparecen regularmente. Los biomarcadores robustos se definieron como aquellos marcadores que surgieron en al menos 50% de los conjuntos de entrenamiento. Por lo tanto, se seleccionaron biomarcadores con una frecuencia mayor que o igual a 5 en el procedimiento de entrenamiento de diez grupos de los autores de la presente invención para una evaluación adicional. En la Tabla 16 siguiente se resumen los biomarcadores que se presentaron con regularidad en cada método en cada cohorte.

30 El enfoque para el descubrimiento de biomarcadores utilizando varios métodos estadísticos ofrece una clara ventaja al proporcionar un repertorio más amplio de biomarcadores candidatos (Figura 1). Si bien algunos métodos como el DA y el PCA funcionan bien con los datos distribuidos normalmente, otros métodos no paramétricos, tales como árboles de regresión logística y de decisión funcionan mejor con datos que son discretos, no distribuidos de manera uniforme o que tienen variaciones extremas. Tal enfoque es ideal cuando los marcadores (tales como biomarcadores y parámetros biométricos) de diversas fuentes (espectrometría de masas, inmunoanálisis, historia clínica, etc.) han de ser combinados en un solo panel ya que los marcadores pueden ser distribuidos normalmente o no en la población.

Tabla 16. Marcadores identificados mediante análisis multivariante (MVM). Sólo se seleccionaron los marcadores que aparecen al menos 50% del tiempo para un examen ulterior. En la tabla anterior, no existen diferencias entre "x" y "X".

			pequeña cohorte		
	AUC	Primeros Marcadores	DA	PCA	DT
1	0,76	acn9459	x		
2	0,75	pub4861	x		x
3	0,66	CEA			x
4	0,65	pub9433		x	
5	0,64	pub9648			x
6	0,64	pub2951	x		
7	0,63	pub6052	x		
8	0,6	tfa2759	x		

			pequeña cohorte		
	AUC	Primeros Marcadores	DA	PCA	DT
9	0,6	tfa9133	x		
10	0,59	acn4132		x	
11	0,58	acn6592		x	
12	0,57	pub7775		x	
13	0,56	pub4213		x	
14	0,55	acn9371		x	
		Total	6	6	3
			gran cohorte		
	AUC	Primeros Marcadores	DA	PCA	DT
1	0,81	pub17858	X		x
2	0,81	pub17338			x
3	0,8	pub8606	X		
4	0,72	pub4861	X		x
5	0,69	pub3743	X		x
6	0,67	acn6399		x	
7	0,66	tfa2331		x	
8	0,65	pub9433		x	
9	0,58	acn6592		x	
10	0,56	pub4213		x	
11	0,55	acn9371		x	
		Total	4	6	4

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

A. Mejora del Método de división y puntuación (SSM)

5

El soporte lógico interactivo que implementa el método de puntuación del punto de división descrito por Mor et al. (Véase, PNAS, 102 (21):7677 (2005)) se ha escrito para ejecutarse en Microsoft © Windows. Este soporte lógico lee hojas de cálculo de Microsoft® Excel que son vehículos naturales para el almacenamiento de los resultados del análisis de marcadores (biomarcadores y parámetros biométricos) de un conjunto de muestras. Los datos se pueden almacenar en una sola hoja de cálculo con un campo para designar la enfermedad de la muestra, almacenar en dos hojas, una para muestras enfermas y otra para muestras no enfermas, o en cuatro hojas de trabajo, 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 dos primeros casos, el usuario puede utilizar el soporte lógico para generar automáticamente el entrenamiento seleccionado al azar y los pares de ensayo de la entrada. En el último caso, se pueden leer a la vez varios archivos de Excel y analizados en una sola ejecución.

10

15

20

El soporte lógico presenta una lista de todos los marcadores recogidos en los datos. El usuario selecciona un conjunto de marcadores de esta lista para utilizarlos en el análisis. El soporte lógico calcula automáticamente los puntos de división para cada marcador entre los conjuntos de datos de entrenamiento de enfermos y no enfermos, así como la determinación de si el grupo de enfermos se eleva o se reduce con respecto a los no enfermos. Se elige el punto de división para maximizar la precisión de cada marcador individual. Los puntos de división también se pueden definir y ajustar manualmente.

25

En todos los análisis, se calculan la precisión, especificidad y sensibilidad en cada posible valor umbral utilizando el conjunto seleccionado de marcadores para los conjuntos tanto de formación como de ensayo. En los análisis que producen múltiples resultados, estos resultados se ordenan por la precisión del conjunto de entrenamiento.

Están disponibles tres modos de análisis. El modo más simple calcula los resultados normalizados utilizando únicamente los marcadores seleccionados. Un segundo modo determina el marcador menos valioso en la lista seleccionada. Se llevan a cabo múltiples cálculos, uno para cada posible subconjunto de marcadores formados mediante la eliminación de un único marcador. El subconjunto con la mayor precisión sugiere que el marcador eliminado para crear el subconjunto hace la menor contribución de todo el conjunto. Los resultados de estos dos primeros modos son esencialmente inmediatos. El cálculo más implicado explora todas las combinaciones posibles de los marcadores seleccionados. Se informa sobre los veinte mejores resultados. Esta última opción puede implicar un gran número de candidatos. Por lo tanto, es bastante intensivo desde el punto de vista computacional y puede tomar algún tiempo para completarse. Cada marcador adicional utilizado duplica el tiempo de ejecución.

Para aproximadamente 20 marcadores, a menudo se ha encontrado que hay por lo general de 6 a 10 marcadores que aparecen en los 20 mejores resultados. Estos luego se emparejan con 2 a 4 marcadores diferentes del conjunto. Esto sugiere que puede haber cierta flexibilidad en la selección de marcadores para un panel de diagnóstico. Los veinte primeros mejores resultados son generalmente similares en precisión, pero pueden diferir de manera significativa en sensibilidad y especificidad. Si se observan todas las posibles combinaciones de marcadores de esta manera se proporciona una idea de las combinaciones que podrían ser las más útiles clínicamente.

B. Método de división y puntuación ponderada (en adelante "SWSM")

Como se comentó anteriormente en la presente memoria, este método es un método de puntuación ponderada que implica la conversión de la medición de un marcador en uno de muchos posibles resultados. Esas puntuaciones se obtienen utilizando la ecuación:

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

El marcador Citoqueratina 19 se puede utilizar como ejemplo ilustrativo. Los niveles de Citoqueratina 19 oscilan de 0,4 a 89,2 ng/ml en la pequeña cohorte. Utilizando el soporte lógico Analyze-it, se generó una curva ROC con los datos de Citoqueratina 19 de manera que los cánceres fueran positivos. La tasa de falsos positivos (1-especificidad) se representa en el eje X y la tasa de verdaderos positivos (sensibilidad) se representa en el eje y, y se generó una hoja de cálculo con el valor de Citoqueratina 19 correspondiente a cada punto de la curva. En un punto de corte de 3,3 ng/ml, la especificidad fue de 90% y la tasa de falsos positivos fue de 10%. Se asignó arbitrariamente un factor de tres para este marcador puesto que su AUC fue mayor de 0,7 y menor de 0,8 (véase la Tabla 2). Sin embargo, se puede utilizar cualquier número entero como factor. En este caso, se utilizan números cada vez mayores, indicando los biomarcadores que tienen el valor de AUC un mejor rendimiento clínico. La puntuación para un individuo con un valor de Citoqueratina 19 mayor que o igual a 3,3 ng/ml se calculó de este modo.

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

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

$$\text{Puntuación} = 21$$

A cualquier valor de Citoqueratina 19 mayor de 3,3 ng/ml, se le asignó por lo tanto una puntuación de 21. A cualquier valor de Citoqueratina 19 mayor de 1,9 pero menor de 3,3, se le asignó una puntuación de 8,4 y así sucesivamente (Véase la Tabla 17a, a continuación).

Tabla 17a. Las 4 posibles puntuaciones otorgadas para Citoqueratina 19.

Citoqueratina		
	19	
	AUC	0,70
Corte	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 su valor a medida que aumenta el nivel de especificidad. Los valores elegidos de especificidad se pueden adaptar a cualquier marcador. El número de niveles de especificidad elegidos para cualquier marcador se puede adaptar. Este método permite que la especificidad mejore la contribución de un

biomarcador a un panel.

Se realizó una comparación de la 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 biomarcadores de inmunoanálisis: CEA, Citoqueratina 19, Citoqueratina 18, CA125, CA15-3, CA19-9, proGRP, y SCC. Las AUC, factores, niveles de especificidad escogidos, y resultados de cada uno de estos niveles de especificidad se tabulan para cada uno de los marcadores más adelante en la Tabla 17b. Utilizando estos puntos de corte y puntuaciones individuales, cada muestra se tabuló para los ocho biomarcadores. La puntuación total de cada muestra se resumió y se representó en una curva ROC. Esta curva ROC se comparó con las curvas ROC generadas utilizando el método de puntuación binario, con los puntos de división de la pequeña cohorte o los puntos de división de la gran cohorte proporcionados en la Tabla 18 (Véase, Ejemplo 8A). Los valores de AUC para el método de puntuación ponderada, los puntos de división de la gran cohorte del método de puntuación binario, y los puntos de división de la pequeña cohorte del método de puntuación binario 0,78, 0,76, y 0,73, respectivamente. Aparte de la mejora del rendimiento global del panel, como se indica por el valor AUC, el método de puntuación ponderada ofrece un mayor número de posibles valores de puntuación para el panel. Una de las ventajas del mayor número de posibles puntuaciones del panel es que hay más opciones para definir el punto de corte para un ensayo positivo (Véase, la **Figura 5**). El método de puntuación binario aplicado a un panel de 8 biomarcadores puede tener como salida del panel valores que oscilan de 0 a 8 con incrementos de 1 (Véase, la **Figura 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 de 50% de especificidad	0	0	0	0	0	0	0	0
puntuación por encima de 50% de especificidad	2,68	2,6	2,48	1,16	2,68	2,48	4,2	1,1
puntuación por encima de 75% de especificidad	5,36	5,2	4,96	2,32	5,36	4,96	8,4	2,2
puntuación por encima de 90% de especificidad	13,4	13	12,4	5,8	13,4	12,4	21	5,5
* Cada uno de estos valores representa un punto de división.								

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

A. SSM de biomarcadores de inmunoanálisis

Como se comenta en el Ejemplo 2, algunos biomarcadores se detectaron mediante análisis inmunológicos. Estos incluyen Citoqueratina 19, CEA, CA125, SCC, proGRP, citoqueratina 18, CA19-9 y CA15-3. Estos datos fueron evaluados utilizando el SSM. Estos biomarcadores en conjunto exhiben una utilidad clínica limitada. En la pequeña cohorte, que representa la enfermedad pulmonar benigna y cáncer de pulmón, la precisión del panel de 8 biomarcadores con un umbral de 4 o superior como resultado positivo, alcanzó una media de 64,8% de precisión (AUC 0,69) en los 10 conjuntos de ensayo de la pequeña cohorte. En la gran cohorte, que representaba normales, así como enfermedad pulmonar benigna y cáncer de pulmón, la precisión del panel de 8 biomarcador con un umbral de 4 o superior como resultado positivo, alcanzó un promedio de 77,4% (AUC 0,79) en los 10 conjuntos de ensayo de la gran cohorte.

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

Tabla 18. Puntos de división calculados para cada marcador de Inmunoanálisis individual que utiliza el algoritmo de SSM.

	pequeña cohorte		gran cohorte		
	punto de división promedio (límite predeterminado)	Desv. Típ.	punto de división promedio (límite predeterminado)		grupo de control Desv. Típ.
CEA	4,82	0	9,2	0	norma <= punto de división
CK 19	1,89	0,45	2,9	0,3	norma <= punto de división
CA125	13,65	8,96	26	2,6	norma <= punto de división
CA15-3	13,07	3,39	20,1	2,6	norma <= punto de división
CA19-9	10,81	11,25	41,1	18,5	norma <= punto de división
SCC	0,92	0,11	1,1	0,1	norma <= punto de división
ProGRP	14,62	8,53	17,6	0	norma <= punto de división
CK-18	57,37	2,24	67,2	9,5	norma <= punto de división
parainfluenza	103,53	32,64	79,2	9,8	norma > = punto de división
Paquetes-año	30		30		Norma <= punto de división

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

En contraste con el Ejemplo 6, donde se identificaron los supuestos biomarcadores utilizando métodos estadísticos multivariantes, se utilizó en este caso un método simple, no paramétrico que implicó el análisis ROC/AUC para identificar supuestos biomarcadores. Mediante la aplicación de este método, se eligieron marcadores individuales con rendimiento clínico aceptable ($AUC > 0,6$) para su posterior análisis. Sólo se seleccionaron los 15 primeros biomarcadores y el parámetro biométrico (paquetes-año) y los grupos serán referidos como los grupos 16AUC (pequeños y grandes) en adelante. Estos marcadores se enumeran en la Tabla 19 a continuación.

Tabla 19. Primeros 15 biomarcadores y un parámetro biométrico (paquetes-año)

gran cohorte			pequeña cohorte		
Marcador	Núm. obs.	AUC	Marcador	Núm. 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	paquete-año	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

gran cohorte			pequeña cohorte		
Marcador	Núm. obs.	AUC	Marcador	Núm. obs.	AUC
tfa6453	538	0,744	acn9471	244	0,712
pack 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 los marcadores de la pequeña cohorte 16AUC se determinaron utilizando el SSM en cada uno de los 10 subconjuntos de entrenamiento. Este procedimiento se llevó a cabo tanto en ausencia (Tabla 20a) como en presencia (Tabla 20b) de historial de tabaquismo como parámetro biométrico (paquetes-año) utilizando el SSM. De este modo, los 15 biomarcadores (excluyendo el parámetro biométrico paquetes-año) o 15 biomarcadores y el parámetro biométrico (paquetes-año) (16AUC) fueron variables de entrada para el método de división y puntuación. El panel óptimo para cada uno de los 10 conjuntos de entrenamiento se determina basándose en la precisión global. Cada panel se sometió a ensayo frente a las restantes muestras no sometidas a ensayo, y se registraron las estadísticas de rendimiento. Los 10 paneles se compararon y se observó la frecuencia de cada biomarcador. El procedimiento se realizó dos veces, incluyendo y excluyendo el parámetro biométrico de paquetes-año. Los resultados de estos dos procedimientos se presentan en las Tablas 20a y 20b, a continuación. Una vez más, se seleccionaron marcadores robustos con una frecuencia mayor que o igual a 5 para su consideración ulterior. El procedimiento se repitió para la gran cohorte y los resultados se presentan en la Tabla 20c. Las Tablas 20a y 20b contienen una lista parcial de los resultados del MSE de la pequeña cohorte que muestra la frecuencia de los marcadores para a) los biomarcadores 15AUC solamente y b) los biomarcadores 15AUC y el parámetro biométrico de paquetes-año. Obsérvese que en la primera tabla (20a) solamente 5 marcadores tienen frecuencias mayores o iguales a 5. En la Tabla 20b, 7 marcadores se ajustan a ese criterio. La Tabla 20c contiene una lista parcial de los resultados de SSM de la gran cohorte que muestra la frecuencia de los marcadores para los marcadores 15AUC. Obsérvese que 11 marcadores tienen frecuencias mayores o iguales a 5.

Tabla 20a.

Conjunto de entrenamiento Núm.	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.

Conjunto de entrenamiento Núm.	acn 9459	CK 19	pkys	Pub 11597	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

Conjunto de entrenamiento Núm.	acn 9459	CK 19	pkysr	Pub 11597	pub 4789	pub 2433	pub 4861	tfa 6652	acn 9471
5	x	x	x	x			x	x	
6	x		x	x	x				x
7	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

Tabla 20c.

Conjunto de entrenamiento Núm.	pub 11597	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 mediante MVM

Un ejemplo de un método multivariante es el análisis del árbol de decisión. Los biomarcadores identificados utilizando solo el análisis del árbol de decisión se toman en conjunto y utilizan en el SSM. Este grupo de biomarcadores demostró utilidad clínica similar a la del grupo de biomarcadores designado como 16AUC. A modo de ejemplo, el conjunto de ensayo 1 (de 10) tiene AUC de 0,90 (ensayo) sin el parámetro biométrico de paquetes-año, y 0,91 (ensayo) con el parámetro biométrico de paquetes-año.

Los biomarcadores de DT se combinaron con los biomarcadores identificados utilizando PCA y DA para generar el grupo MVM. El grupo 14MVM se evaluó con y sin el historial de tabaquismo como parámetro biométrico (paquetes-año) utilizando el SSM. Una vez más, se seleccionaron marcadores robustos con una frecuencia mayor de o igual a 5 para su ulterior examen (resultados no mostrados). Como se puede observar en las tablas anteriores, los paquetes-año (historial de tabaquismo) tienen un efecto sobre el número y tipo de biomarcadores que emergen como marcadores robustos. Esto no es totalmente inesperado ya que algunos biomarcadores pueden tener efectos sinérgicos o deletéreos sobre otros biomarcadores. Un aspecto de esta descripción implica la búsqueda de los marcadores que trabajan juntos como un panel en la mejora de la capacidad predictiva del modelo. En la misma línea, los biomarcadores que se identificó que trabajaban trabajar sinérgicamente con el parámetro biométricos de paquetes-año en ambos métodos (AUC y MVM) se combinaron en un esfuerzo por identificar un panel superior de marcadores (Véase, Ejemplo 8D).

Los marcadores multivariantes identificados para la gran cohorte fueron evaluados con el SSM. Una vez más, se seleccionaron únicamente los marcadores con frecuencias mayores o iguales a 5 para su consideración ulterior. En la siguiente Tabla 21 se resumen los resultados del SSM para la gran cohorte.

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

Conjunto de entrenamiento Núm.	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
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-año)

- 5 En una etapa posterior, se combinaron todos los marcadores (biomarcadores y parámetros biométricos) con frecuencias mayores o iguales a 5 (en los 10 conjuntos de entrenamiento) para producir una segunda lista de marcadores que contenía marcadores de ambos grupos de AUC y MVM para ambas cohortes. A partir de los resultados de SSM, se seleccionaron 16 marcadores únicos de la pequeña cohorte y 15 marcadores únicos de la gran cohorte con frecuencias iguales o superiores a cinco. En la siguiente Tabla 22 se resumen los marcadores que se habían seleccionado.

Tabla 22. Marcadores combinados de ambos grupos de AUC y MVM.

pequeña cohorte				
	AUC	marcadores	16AUC	14MVM
1	0,77	Publ1597	x	
2	0,76	Acn9459	x	x
3	0,75	Pub4861	x	x
4	0,74	pkyrs	x	x
5	0,72	Pub2433	x	
6	0,72	CK 19	x	
7	0,72	Pub4789	x	
8	0,71	Tfa6652	x	
9	0,66	cea		x
10	0,64	Pub2951		x
11	0,63	Pub6052		x
12	0,6	Tfa2759		x
13	0,6	Tfa9133		x
14	0,59	Acn4132		x
15	0,58	Acn6592		x
16	0,57	Pub7775		x
		Total	8	11

gran cohorte				
	AUC	marcadores	15AUC	11MVM
1	0,813	Pub17338	x	x
2	0,812	pub17858	x	x
3	0,798	pub8606	x	x
4	0,796	pub8662	x	
5	0,773	pub4628	x	
6	0,765	pub6798	x	
7	0,76	pub4750	x	
8	0,751	pub11597	x	
9	0,747	pub4487	x	
10	0,744	tfa6453	x	
11	0,741	hic3959	x	
12	0,72	pub4861		x
13	0,69	pub3743		x
14	0,67	acn6399		x
15	0,66	tfa2331		x
		Total	11	7

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

pequeña cohorte				gran cohorte			
	AUC	Marcadores	Frecuencia		AUC	Marcadores	Frecuencia
1	0,718	CK19	9	1	0,67	acn6399	10
2	0,761	acn9459	8	2	0,69	pub3743	8
3	0,74	pkys	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	pub4487	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				

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

Para cada marcador, se determinó un punto de división mediante la evaluación de cada conjunto de datos de entrenamiento para la más alta precisión en la clasificación a medida que se optimizó el nivel de marcador. Los puntos de división de los ocho marcadores más frecuentes utilizados en la pequeña cohorte se enumeran a continuación.

Tabla 23b.

	Marcadores	Grupo de control	Prom.	Desv. Típ.
--	------------	------------------	-------	------------

1	CK 19	Norma <= SP	1,89	0,45
2	acn9459	Norma > = SP	287,3	23,67
3	pkyrs	Norma <= SP	30,64	4,21
4	cea	Norma <= SP	4,82	0
5	tfa2759	Norma > = SP	575,6	109,7
6	pub11597	Norma <= SP	34,4	2,52
7	pub4789	Norma <= SP	193,5	18,43
8	tfa9133	Norma > = SP	203,6	46,38

La Tabla 23b muestra la lista de los 8 marcadores más frecuentes con sus medios puntos de división (Prom.) (cada uno o a un punto de corte predeterminado). También se incluyen las desviaciones típicas para cada punto de división (Desv. Típ.). La posición del grupo de control con respecto al punto de división se proporciona en la segunda columna desde la izquierda. Como ejemplo, en la Citoqueratina 19, el grupo normal o grupo de control (negativo para cáncer) es inferior o igual al valor del punto de división de 1,89.

Ejemplo 9. Validación de los modelos predictivos.

Los subconjuntos de la lista de 13 biomarcadores y parámetros biométricos para la pequeña cohorte (Véase la Tabla 23a anterior) proporcionan una buena utilidad clínica. Por ejemplo, los 8 biomarcadores y parámetros biométricos más frecuentes utilizados en conjunto como un panel en el método de división y puntuación tienen una AUC de 0,90 para el subconjunto de ensayo 1 (véase la Tabla 23b anterior).

Los modelos predictivos que comprenden un panel de 7 marcadores (marcadores de 1-7, Tabla 23b) y un panel de 8 marcadores (marcadores de 1-8, Tabla 23b) fueron validados utilizando 10 conjuntos de ensayo al azar. Las Tablas 24a y 24b siguientes resumen los resultados de los dos modelos. Todas las condiciones y parámetros de cálculo fueron idénticos en ambos casos con la excepción del número de marcadores en cada modelo.

Tabla 24a.

Núm. de conjunto de ensayo	AUC	Precisión (%)	Sensibilidad (%)	Especificidad (%)	Núm. de marcadores	Umbral
1	0,91	85	80,7	90,7	7	3
2	0,92	85	78,2	93,3	7	3
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
Cra	0,90	83,9	79,1	89,4		
desv. típ.	0,01	1,9	1,4	3,5		

La Tabla 24a muestra el rendimiento clínico del panel de 7 marcadores con diez conjuntos de ensayo al azar. Los 7 marcadores y los puntos de división promedio utilizados en los cálculos se proporcionan en la Tabla 16b. Se utilizó un valor de umbral de 3 para separar el grupo con enfermedad del grupo sin enfermedad. La AUC media para el modelo es 0,90, que corresponde a una precisión media de 83,9% y sensibilidad y especificidad de 79,1% y 89,4%, respectivamente.

Tabla 24b.

Núm. de conjunto de ensayo	AUC	Precisión (%)	Sensibilidad (%)	Especificidad (%)	Núm. 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
Cra	0,91	84,1	91,5	75,1		
desv. típ.	0,01	1,8	1,4	4,7		

La Tabla 24b muestra el rendimiento clínico del panel de 8 marcadores con diez conjuntos de ensayo al azar. Los marcadores 8 y los puntos de división promedio utilizados en los cálculos se proporcionan en la Tabla 16b. Se utilizó un valor de umbral de 3 (una puntuación total predeterminada) para separar el grupo con enfermedad del grupo sin enfermedad. La AUC media para el modelo es de 0,91, que corresponde a una precisión media de 84,1% y sensibilidad y especificidad de 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 sólo se diferencian en la sensibilidad y especificidad. Como se puede observar en la Tabla 24a, el panel de 7 marcadores muestra una mayor especificidad (89,4% frente a 75,0%). En contraste, el panel de 8 marcadores muestra mejor sensibilidad (91,5% frente a 79,0%) a juzgar por sus valores medios (Prom.). Cabe señalar que se seleccionó el umbral (o puntuación total predeterminada) que maximiza la precisión de la clasificación, que es similar a la maximización de la AUC de una curva ROC. Por lo tanto, el umbral elegido de 3 (una puntuación total predeterminada) no sólo maximiza la precisión, sino también ofrece el mejor compromiso entre la sensibilidad y especificidad del modelo. En la práctica, esto significa que se considera que un individuo normal tiene bajo "riesgo" de desarrollar cáncer de pulmón, si dichos ensayos individuales son positivos para menos que o igual a 3 de los 7 marcadores posibles en este modelo (o inferior o igual a 3 de 8 para el segundo modelo). Se considera que los individuos con una puntuación más alta (puntuación total) que el umbral establecido (o puntuación total predeterminada) tienen mayor riesgo y se convierten en candidatos para más pruebas o procedimientos de seguimiento. Cabe señalar que el umbral del modelo (es decir, la puntuación total predeterminada) se puede aumentar o disminuir 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 que el modelo se pueda modificar para hacer frente a diferentes cuestiones de diagnóstico y/o poblaciones en riesgo, p. ej., diferenciando a los individuos normales de individuos sintomáticos y/o asintomáticos.

Varios modelos predictivos se resumen en las Tablas 25a y 25b a continuación. Para cada modelo predictivo, se indican los biomarcadores y parámetros biométricos que constituyen el modelo, como lo es el umbral (es decir, la puntuación total predeterminada), la AUC media, la precisión, la sensibilidad y la especificidad con sus desviaciones típicas correspondientes (entre paréntesis) en los 10 conjuntos de ensayo. El panel de 8 marcadores descrito anteriormente, es el Modelo Mixto 2 y el panel de 7 marcadores descrito anteriormente es el Modelo Mixto 3. El Modelo Mixto 1A y el Modelo Mixto 1B contienen los mismos marcadores. La única diferencia entre el Modelo Mixto 1A y el Modelo Mixto 1B está en el umbral (es decir, la puntuación total predeterminada). Del mismo modo, El Modelo Mixto 10A y el Modelo Mixto 10B contienen los mismos marcadores. La única diferencia entre el modelo Mixto 10A y el Modelo Mixto 10B está en el umbral (es decir, la puntuación total predeterminada).

Tabla 25a.

	pequeña cohorte										
Marcadores	Modelo IA 8	Modelo IA 9	Modelo IA pkys	Modelo MS	Modelo pkys MS	Modelo mixto 1A	Modelo Mixto 1B	Modelo Mixto 2	Modelo Mixto 3	Modelo Mixto 4	Modelo Mixto 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
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 Mixto 6	Modelo Mixto 7	Modelo Mixto 8	Modelo Mixto 9	Modelo Mixto 10A	Modelo Mixto 10B
CK 19		x		x	x	x

pequeña cohorte						
Marcadores	Modelo Mixto 6	Modelo Mixto 7	Modelo Mixto 8	Modelo Mixto 9	Modelo Mixto 10A	Modelo Mixto 10B
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						
Pkys	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.						

Tabla 25a y b. Resumen de los diversos modelos predictivos.

Del mismo modo, para la gran cohorte, se pueden varios modelos predictivos optimizar en cuanto a la precisión, la sensibilidad, o la especificidad globales. En la siguiente Tabla 26 se resumen cuatro modelos posibles.

5

Tabla 26. Cuatro modelos posibles.

gran cohorte				
marcadores	Modelo 1 de MS	Modelo 2 de MS	Modelo 3 de MS	Modelo 4 de MS
acn6399	x	X	x	x
pub3743	x	X	x	x
pub8606	x	X	x	x
pub11597	x	X	x	x

	gran cohorte			
marcadores	Modelo 1 de MS	Modelo 2 de MS	Modelo 3 de MS	Modelo 4 de MS
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)	76,0 (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".

Del mismo modo, los modelos predictivos para la cohorte de ciclina (subconjunto de individuos con anticuerpos anti-proteína ciclina E2 y anticuerpos anti-péptido de ciclina E2 medidos) se resumen en las Tablas 27a y 27b a continuación.

5

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
CK 19									x	x	
CA 19-9											
CEA											
CA15-3											
CA125						x	x	x	x		
SCC								x	x		
CK 18							x	x	x		
ProGRP					X	x	x	x	x		
Parainflu											
Pkyrs				x	X	x	x	x	x		
Acn9459											
Pub11597										x	x
Pub4789											
TFA2759											
TFA9133											
Pub6453											x
Pub2951										x	
Pub4861										x	
Pub2433										x	

ES 2 606 015 T3

	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
Pub3743											
Pub17338											
TFA6652											
Pep de ciclina E2-1	x			x	X	x	x	x	x	x	x
Proteína ciclina E2		x									
Pep de 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
CK 19		x	x	x	X						
CA 19-9											
CEA			x	x	X	x	x				
CA15-3											
CA125					X						
SCC					X						
CK 18					X	x					
ProGRP					X	x	x	x	x	x	
Parainflu											
Pkyrs											
Acn9459											
Pub11597	x			x							
Pub4789											
TFA2759											
TFA9133											
Pub6453	x										
Pub2951											
Pub4861									x	x	
Pub2433									x		
Pub3743									x	x	x
Pub17338									x	x	x

Marcadores	modelo L	modelo M	modelo N	modelo O	modelo P	modelo Q	modelo R	modelo S	modelo T	modelo U	modelo V
TFA6652									x		
Pep de ciclina E2-1		x	x	x	X	x	x	x	x	x	
Proteína Ciclina E2	x										
Pep de Ciclina E2-2											
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.											

Del mismo modo, los modelos predictivos utilizando análisis de autoanticuerpos se resumen en la Tabla 28 a continuación.

5

Tabla 28. Modelos predictivos utilizando análisis de autoAb. * Puntuación total Predeterminada.

Marcadores	modelo AAb1	modelo AAb2
TMP21	X	x
dominio 1NPC1L1C	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

Cinco de estos modelos se utilizaron frente a la cohorte de validación. La siguiente Tabla 29 resume el rendimiento clínico de cada uno de los modelos predictivos para las cohortes independientes, la pequeña cohorte y la cohorte de validación.

10

Tabla 29.

	Modelo Mixto 7	Modelo Mixto 1	Modelo IA 8	Modelo 5 MS	Modelo Mixto 9
CK 19	x	x	x		x
CEA	x	x	x		x
CA19-9			x		

	Modelo Mixto 7	Modelo Mixto 1	Modelo IA 8	Modelo 5 MS	Modelo Mixto 9
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
a2759	x	x			x
a9133		x			
pub3743				X	
pub8606				X	
pub4487				X	
pub4861				X	
pub6798				X	
tfa6453				X	
hic3959				X	
paquetes-año					
límite	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
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

5 A. Fraccionamiento mediante HPLC

Con el fin de obtener la identidad de los candidatos de biomarcadores de MS de la Tabla 22, en primer lugar fue necesario fraccionar las muestras de suero agrupadas y/o individuales mediante HPLC de fase inversa utilizando protocolos convencionales. La obtención de suficiente material para la electroforesis en gel y para el análisis de MS

necesitó varios ciclos de fraccionamiento. Las fracciones individuales fueron perfiladas mediante MS MALDI-TOF y las fracciones que contenían los picos de interés se agruparon y se concentraron en un SpeedVac. Todos los otros candidatos a biomarcadores se procesaron como se ha descrito anteriormente.

- 5 La **Figura 2** muestra un supuesto biomarcador (pub 11597) antes y después de la concentración. Obsérvese que el candidato a biomarcador en 11kDa en la muestra de partida está muy diluido. Después de la concentración la intensidad es mayor, pero la muestra no es lo suficientemente pura para el análisis y necesitó separación adicional mediante SDS-PAGE con el fin de aislar el biomarcador de interés.

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

Después de la concentración, las fracciones que contenían los biomarcadores candidatos se sometieron a SDS-PAGE para aislar la proteína/péptido deseado que tiene la masa molecular correspondiente al biomarcador candidato. La electroforesis en gel (SDS-PAGE) se llevó a cabo utilizando la metodología convencional proporcionada por el fabricante (Invitrogen, Inc.) En resumen, el procedimiento implicaba cargar las muestras que contienen los biomarcadores candidatos y proteínas patrón de masa molecular conocida en diferentes pocillos en el mismo gel como se muestra en **Figura 3**. Mediante la comparación de las distancias de migración de las proteínas patrón con la de la muestra "desconocida", se identificó la banda con la masa molecular deseada y escindió del gel.

20 A continuación, la banda de gel escindida se sometió a digestión trípica en gel automatizada utilizando una estación Waters MassPREP™. Posteriormente, la muestra digerida se extrajo del gel y se sometió a ESI-LC-MS/MS en línea de fase inversa. Los espectros de iones producto fueron utilizados para la búsqueda en la base de datos. Cuando fue posible, se obtuvo comercialmente la proteína identificada y se sometió a SDS-PAGE y a digestión en gel como se describió anteriormente. Una buena concordancia en la electroforesis en gel, los resultados de MS/MS y de búsqueda en la base de datos entre las dos muestras era evidencia adicional de que el biomarcador se identificó correctamente. Como se puede observar en la **Figura 3**, existe una buena concordancia entre el amiloide A sérico humano disponible comercialmente (HSAA) y el supuesto biomarcador en la muestra fraccionada a 11.5kDa. El análisis MS/MS y la búsqueda en la base de datos de confirmó que ambas muestras eran la misma proteína. La **Figura 4** muestra los espectros de MS/MS del biomarcador candidato Pub11597. Las secuencias de aminoácidos derivada de los iones b e y son anotadas en la parte superior de cada panel. El candidato biomarcador fue identificado como un fragmento de la proteína amiloide de suero humano A (HSAA).

Los biomarcadores candidatos pequeños que no eran susceptibles de digestión se sometieron a ESI-Q-TOF y/o fragmentación MALDI-TOF-TOF seguido de secuenciación de novo y búsqueda en la base de datos (BLAST) para obtener información de la secuencia y el ID de proteínas.

35 C. Búsqueda en la Base de datos e ID. de proteínas

Con el fin de caracterizar completamente los biomarcadores candidatos era imperativo identificar las proteínas a partir de los cuales se derivaron. La identificación de proteínas desconocidas implicó la digestión en gel seguida de espectrometría de masas de los fragmentos trípticos. Los iones de productos resultantes del procedimiento de MS/MS se buscaron en la base de datos Swiss-Prot de proteínas para identificar la proteína de origen. Para los candidatos a biomarcadores con masas moleculares bajas, la espectrometría de masas en tándem seguida de secuenciación de novo y búsqueda en la base de datos fue el método de elección para identificar la proteína de origen. Búsquedas consideraron sólo el genoma *Homo sapiens* y precisiones de masa de $\pm 1,2$ Da para los iones precursores y $\pm 0,8$ Da para los iones producto (MS/MS). Sólo se permitió una escisión de menos para la tripsina. Las dos únicas modificaciones variables permitidas para las búsquedas en las bases de datos fueron la carbamidometilación (C) y la oxidación (M). Un ID de proteína final fue atribuido después de la conciliación de los resultados del motor de búsqueda Mascot y la interpretación manual de los espectros de MS y MS/MS relacionados. La precisión de los resultados se verificó por medio de mediciones repetidas.

Tabla 30.

Marcador candidato	Núm. de acceso	Nombre de la proteína	Secuencia peptídica observada	PM medio (Da)
Pub11597	Q6FG67	Amiloide Humana	SFFSFLGEAFDGMWRAYSMDREA NYIGSDKYFHARGNYDAAKRGPGBA	11526,51
		Proteína A	WAAEVIDARENIQRFFGHGAEDSLAD QAANEWGRSGKDPNHFRPAGLPEKY (SEQ ID NO:7)	

Marcador candidato	Núm. de acceso	Nombre de la proteína	Secuencia peptídica observada	PM medio (Da)
ACN9459	P02656	ApoCIII ₁	SEAEDASLLSFMQGYMKHATKTAKDA LSSVQESQVAQQARGWVTDGFSSSLKD YWSTVKDKFSEFWDLDPVEVRP *(T) SAVAA (SEQ ID NO:8)	9421,22
			* (Sitio glicosilada)	
TFA9133	P02656	ApoCIII ₁	ApoCIII ₁ después de la pérdida de ácido siálico	9129,95
Pub4789	P01009	alfa-1 antitripsina	LEAIPMSIPPEVKFN *(E) PFVFLMIDQ NTKSPLFMGKVVNPTQK (SEO ID NO:9)	4776,69
			* (posible sustitución de K por E)	
TFA2759	Q56G89	Péptido de Albúmina humana	DAHKSEVAHRFKDLGEENFKAL VL (SEQ ID NO:10)	2754,10

Tabla 30 anterior muestra la proteína de origen de los diferentes biomarcadores candidatos con su ID de proteínas. Los marcadores se identificaron mediante digestión en gel y LC-MS/MS y/o secuenciación de novo. Obsérvese que sólo se muestran las secuencias de aminoácidos de los fragmentos observados y el PM medio incluye el PTM cuando se indique. Los números de acceso se obtuvieron de la base de datos Swiss-Prot y se proporcionan como única referencia. Es interesante observar que ACN9459 y TFA9133 son los mismos fragmentos de proteína con la excepción de que este último ha perdido un ácido siálico (-291,3 Da) de la fracción glicosilada. Tanto ACN9459 como TFA9133 se identificaron como una variante de la apolipoproteína C III. Los resultados de los autores de la presente invención están de acuerdo con la secuencia conocida publicada y la masa molecular de esta proteína (Bondarenko et. al, J. Lipid Research, 40: 543-555 (1999)). Pub4789 se identificó como la proteína alfa-1-antitripsina. Un examen de los espectros de iones producto sugiere que podría ser una sustitución de K por E en el sitio indicado en la Tabla 30. La incertidumbre en la precisión de la masa impidió la asignación.

Ejemplo 11: Detección del cáncer de pulmón

A. Inmunoanálisis del péptido o proteína: Los biomarcadores descritos en el Ejemplo 9 anterior se puede detectar y medir mediante técnicas de inmunoanálisis. Por ejemplo, se utiliza el sistema de inmunoanálisis Architect™ de Abbott Diagnostics para el análisis automático de un biomarcador desconocido en una muestra que se sospecha que contiene un biomarcador. Como se sabe en la técnica, el sistema utiliza micropartículas magnéticas recubiertas con anticuerpos, que son capaces de unirse a los biomarcadores de interés. Bajo el control del aparato, una alícuota de muestra se mezcla con un volumen igual de micropartículas magnéticas recubiertas con anticuerpo y el doble de volumen de diluyente de muestra, que contiene tampones, sal, agentes tensioactivos, y proteínas solubles. Después de la incubación, las micropartículas se lavan con un tampón de lavado que comprende tampón, sal, agente tensioactivo y conservante. Se añade una alícuota de producto conjugado marcado con acridinio junto con un volumen igual de diluyente de muestra y las partículas se vuelven a dispersar. La mezcla se incuba y a continuación se lava con tampón de lavado. Las partículas lavadas se vuelven a dispersar en predisparador ácido que contiene ácido nítrico y peróxido de hidrógeno para disociar el producto conjugado de acridinio de las micropartículas. A continuación se añade una solución de NaOH para desencadenar la reacción quimioluminiscente. La luz se mide por medio de un fotomultiplicador y el resultado desconocido se cuantifica por comparación con la luz emitida por una serie de muestras que contienen cantidades conocidas del péptido biomarcador utilizado para construir una curva patrón. La curva patrón se utiliza a continuación para estimar la concentración del biomarcador en una muestra clínica que se procesa de una manera idéntica. El resultado puede ser utilizado tal cual o combinado con otros marcadores como se describe a continuación.

B. Inmunoanálisis multiplexado para el péptido o la proteína: Cuando se necesita la detección de múltiples biomarcadores a partir de una sola muestra, puede ser más económico y conveniente llevar a cabo un análisis

multiplexado. Para cada analito en cuestión, es necesario un par de anticuerpos específicos y una micropartícula teñida de forma única para su uso en un analizador Luminex 100™. Cada anticuerpo de captura de la pareja se aplica como recubrimiento individualmente sobre una micropartícula única. El otro anticuerpo del par se conjuga con un fluoróforo tal como rFicoeritrina. Las micropartículas se agrupan y se diluyen a una concentración de aproximadamente 1000 partículas únicas por microlitro lo 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 productos conjugados se agrupan y se ajustan a una concentración final de aproximadamente 1 a 10 nM cada uno en el diluyente de micropartículas. Para llevar a cabo el análisis, una alícuota de muestra que se sospecha que contiene uno o más de los analitos se coloca en un pocillo de incubación seguido de la mitad del volumen de micropartículas agrupadas. La suspensión se incuba durante 30 minutos seguido de la adición de la mitad del volumen de solución de producto conjugado agrupado. Después de una incubación adicional de 30 minutos, la reacción se diluye mediante la adición de dos volúmenes de solución tamponada que contiene una sal y tensioactivo. La suspensión se mezcla y se aspira un volumen de aproximadamente dos veces el de la muestra mediante el aparato Luminex 100™ para su análisis. Opcionalmente, las micropartículas pueden lavarse después de cada incubación y después se resuspenden para su análisis. La fluorescencia de cada partícula individual se mide a 3 longitudes de onda; dos se utilizan para identificar la partícula y su analito asociado y la tercera se utiliza para cuantificar la cantidad de analito unido a la partícula. Al menos se miden 100 micropartículas de cada tipo y se calcula la mediana de la fluorescencia para cada analito. La cantidad de analito en la muestra se calcula por comparación con una curva patrón generada realizando el mismo análisis en una serie de muestras que contienen cantidades conocidas del péptido o proteína y trazando de la mediana de la fluorescencia de las muestras conocidas frente a la concentración conocida. Una muestra desconocida se clasifica que es cáncer o negativa para cáncer basándose en la concentración de analito (ya sea elevada o deprimida) con respecto a especímenes de cáncer o negativos para cáncer conocidos utilizando modelos tales como el Método de División y Puntuación el Método0 de División y Puntuación Ponderada como en el Ejemplo 7.

Por ejemplo, un paciente puede ser sometido a ensayo para determinar la probabilidad de tener cáncer de pulmón utilizando el panel de inmunoanálisis (IA) de 8 de la Tabla 18 y el Método de División y Puntuación. Después de obtener una muestra de análisis del paciente, la cantidad de cada uno de los 8 biomarcadores en la muestra de análisis del paciente (es decir, suero) es cuantificada y la cantidad de cada uno de los biomarcadores se compara a continuación con el punto de división predeterminado correspondiente (valor del punto de corte predeterminado) para el biomarcador, tal como se enumerado 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 tiene una cantidad que es superior a su correspondiente punto de división predeterminado (límite predeterminado), se puede asignar una puntuación de 1. Para cada biomarcador que tiene una cantidad que es menor que o igual a su correspondiente punto de división predeterminado (valor del punto de corte predeterminado), se puede asignar una puntuación de 0. La puntuación para cada uno de los 8 biomarcadores se combina a continuación matemáticamente (es decir, añadiendo cada una de las puntuaciones de los biomarcadores en conjunto) para llegar a la 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 de 8 IA de la Tabla 25a, a saber 1. Una puntuación del panel mayor que 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 un estudio de población anterior, este panel ha demostrado una especificidad de 30%, una tasa de falsos positivos de 70% y una sensibilidad de 90%. Un resultado positivo del panel para el paciente tiene una probabilidad de 70% de ser falsamente positivo. Adicionalmente, 90% de los pacientes con cáncer de pulmón tendrá un resultado positivo del panel. Por lo tanto, el paciente que tiene un resultado positivo del panel puede ser remitido para más pruebas por un indicio o sospecha de cáncer de pulmón.

A modo de ejemplo adicional, de nuevo utilizando el panel 8 IA y el Método de División y Puntuación Ponderada, después de obtener una muestra de ensayo de un paciente, se cuantifica la cantidad de cada uno de los 8 biomarcadores en la muestra de análisis del paciente (es decir, suero) y la cantidad de cada uno de los biomarcadores se compara a continuación con los puntos de división predeterminados (puntos de corte predeterminados) tales como los puntos de división 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 3 puntos de división predeterminados (puntos de corte predeterminados). Por lo tanto, se pueden asignar 4 puntuaciones posibles para cada biomarcador. La puntuación para cada uno de los 8 biomarcadores se combinan a continuación matemáticamente (es decir, añadiendo cada una de las puntuaciones de los biomarcadores en conjunto) para llegar a la puntuación total para el paciente. La puntuación total se convierte después en la puntuación del panel. La puntuación del panel se puede comparar con el umbral predeterminado (o puntuación total predeterminada) para el modelo de 8 IA, que se calculó que era 11,2. Una puntuación del panel de pacientes mayor que 11,2 sería un resultado positivo. Una puntuación del panel de pacientes menor de o igual a 11,2 sería un resultado negativo. En un estudio de población anterior, este panel ha demostrado una especificidad de 34%, una tasa de falsos positivos de 66% y una sensibilidad de 90%. El resultado positivo del panel tiene 66% de probabilidad de ser falsamente positivo.

Adicionalmente, 90% de los pacientes con cáncer de pulmón tienen un resultado positivo del panel. Por lo tanto, el paciente que tiene un resultado positivo del panel puede ser remitido para más pruebas por un indicio o sospecha de cáncer de pulmón.

C. Inmunoanálisis de espectrometría de masas: La preparación de las muestras para la espectrometría de masas también pueden utilizar métodos inmunológicos, así como métodos cromatográficos o electroforéticos. La micropartículas superparamagnéticas recubiertas con anticuerpos específicos para un biomarcador peptídico se ajustan a una concentración de aproximadamente 0,1% p/v en una solución tampón que contiene sal. Una alícuota de la muestra de suero del paciente se mezcla con un volumen igual de micropartículas recubiertas con anticuerpo y el doble de volumen de diluyente. Después de una incubación, las micropartículas se lavan con un tampón de lavado que contiene una sal de tamponamiento y, opcionalmente, sal y agentes tensioactivos. Las micropartículas se lavan a continuación con agua desionizada. El analito inmunopurificado se eluye de las micropartículas mediante la adición de un volumen de acetonitrilo acuoso que contiene ácido trifluoroacético. La muestra se mezcla a continuación con un volumen igual de solución de matriz de ácido sinapínico y un pequeño volumen (aproximadamente 1 a 3 microlitros) se aplica a una diana MALDI para el análisis de masas de tiempo de vuelo. La corriente de iones a m/z deseada se compara con la corriente de iones derivadas de una muestra que contiene una cantidad conocida del biomarcador peptídico que ha sido procesada de una manera idéntica.

Debe tenerse en cuenta que la corriente de iones está directamente relacionada con la concentración y la corriente de iones (o intensidad) a un valor m/z (o ROI) particular puede ser convertida en concentración si se desea. Tales concentraciones o intensidades se pueden utilizar como entrada en cualquiera de los algoritmos de construcción de modelos descritos en el Ejemplo 7.

D. espectrometría de masas mediante ROI: Una muestra de sangre se obtiene de un paciente y se deja coagular para formar una muestra de suero. La muestra se prepara para el análisis de espectrometría de masas SELDI y se carga en un Chip de Proteína en un Bioprocesador y se trata como se proporciona en el Ejemplo 2. El Chip de Proteína se carga en un espectrómetro de masas de tiempo de vuelo MALDI Ciphergen 4000 y se analiza como en el Ejemplo 3. Cada espectro se somete a ensayo de aceptación mediante análisis multivariante. Por ejemplo, se calculan la corriente de iones total y el ángulo de contraste espectral (entre la muestra desconocida y una población de referencia conocida). A continuación se determina la distancia de Mahalanobis. Para el espectro cuya distancia de Mahalanobis es menor que el valor crítico establecido, el espectro es calificado. Para el espectro cuya distancia de Mahalanobis es mayor que el valor crítico establecido, el espectro se excluye del análisis adicional y la muestra debe ser re-analizada. Después de la calificación, el espectro de masas se normaliza.

El espectro de masas resultante se evalúa mediante la medición de la corriente de iones en las regiones de interés adecuada para el modelo de análisis de datos elegido. Basándose en los resultados del análisis, se considera que el paciente tiene riesgo de o tienen una alta probabilidad de tener cáncer de pulmón y debe ser conducido a procedimientos de diagnóstico adicionales.

Para el uso del Método de División y Puntuación, se miden para el paciente las intensidades en las ROI en los valores de m/z proporcionados en la Tabla 5. El resultado para el paciente se obtuvo observando si los valores del paciente están en el lado cáncer o el lado negativo para cáncer de los valores de los puntos de división promedio proporcionados en la Tabla 6. Se asigna una puntuación de 1 para cada valor de ROI que se encontraba que estaba en el lado del cáncer del punto de división. Las puntuaciones de 3 y superiores indican que el paciente tiene riesgo elevado de cáncer y debe ser remitido a procedimientos de diagnóstico adicionales.

Un experto en la técnica apreciará fácilmente que la presente invención está bien adaptada para llevar a cabo los objetos y obtener los fines y ventajas mencionados, así como aquellos inherentes a la misma. Las composiciones, formulaciones, métodos, procedimientos, tratamientos, moléculas, compuestos específicos descritos en la presente son actualmente representativos de realizaciones preferidas, son ilustrativos y no se pretende que sean limitaciones del alcance de la invención. Será fácilmente evidente para un experto en la técnica que se puede realizar en la invención descrita en la presente memoria diversas sustituciones y modificaciones sin apartarse del alcance de la invención como se define en las reivindicaciones adjuntas.

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

LISTA DE SECUENCIAS

<110> Abbott Laboratories

5 <120> Métodos y Combinaciones de Marcadores Para el Escrutinio de la Predisposición al Cáncer de Pulmón

<130> 8064EP01

<160> 10

10 <170> PatentIn Versión 3.2

<210> 1

<211> 296

15 <212> PRT

<213> Homo sapiens

<400> 1

```

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
20

```


ES 2 606 015 T3

	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
 5 <212> PRT
 <213> Homo sapiens
 <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	
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	

10

ES 2 606 015 T3

Leu 145 Leu Asp Trp Leu 150 Leu Glu Val Cys Glu Val 155 Tyr Thr Leu His Arg 160
 Glu Thr Phe Tyr Leu 165 Ala Gln Asp Phe Phe 170 Asp Arg Phe Met Leu Thr 175
 Gln Lys Asp Ile 180 Asn Lys Asn Met Leu 185 Gln Leu Ile Gly Ile Thr Ser 190
 Leu Phe Ile 195 Ala Ser Lys Leu Glu 200 Glu Ile Tyr Ala Pro 205 Lys Leu Gln
 Glu Phe 210 Ala Tyr Val Thr Asp 215 Gly Ala Cys Ser Glu 220 Glu Asp Ile Leu
 Arg Met Glu Leu Ile Ile 230 Leu Lys Ala Leu Lys 235 Trp Glu Leu Cys Pro 240
 Val Thr Ile Ile Ser 245 Trp Leu Asn Leu Phe 250 Leu Gln Val Asp Ala Leu 255
 Lys Asp Ala Pro 260 Lys Val Leu Leu Pro 265 Gln Tyr Ser Gln Glu Thr Phe 270
 Ile Gln Ile 275 Ala Gln Leu Leu Asp 280 Leu Cys Ile Leu Ala 285 Ile Asp Ser
 Leu Glu Phe Gln Tyr Arg Ile 295 Leu Thr Ala Ala 300 Ala Leu Cys His Phe
 Thr Ser Ile Glu Val Val 310 Lys Lys Ala Ser Gly 315 Leu Glu Trp Asp Ser 320
 Ile Ser Glu Cys Val 325 Asp Trp Met Val Pro 330 Phe Val Asn Val Val 335 Lys
 Ser Thr Ser Pro 340 Val Lys Leu Lys Thr 345 Phe Lys Lys Ile Pro Met Glu 350
 Asp Arg His 355 Asn Ile Gln Thr His 360 Thr Asn Tyr Leu Ala 365 Met Leu Cys
 Met Ile 370 Ser Ser His Val

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

<400> 3

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

ES 2 606 015 T3

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

5 <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

10 <210> 5
<211> 19
<212> PRT
15 <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

20 <210> 6
<211> 103
<212> PRT
25 <213> Homo sapiens

<400> 6

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

ES 2 606 015 T3

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

$\langle 400 \rangle$ 7

Ser 1	Phe	Phe	Ser	Phe 5	Leu	Gly	Glu	Ala	Phe 10	Asp	Gly	Ala	Arg	Asp 15	Met
Trp	Arg	Ala	Tyr 20	Ser	Asp	Met	Arg	Glu 25	Ala	Asn	Tyr	Ile	Gly 30	Ser	Asp
Lys	Tyr	Phe 35	His	Ala	Arg	Gly	Asn 40	Tyr	Asp	Ala	Ala	Lys 45	Arg	Gly	Pro
Gly	Gly 50	Ala	Trp	Ala	Ala	Glu 55	Val	Ile	Ser	Asp	Ala 60	Arg	Glu	Asn	Ile
Gln 65	Arg	Phe	Phe	Gly	His 70	Gly	Ala	Glu	Asp	Ser 75	Leu	Ala	Asp	Gln	Ala 80
Ala	Asn	Glu	Trp	Gly 85	Arg	Ser	Gly	Lys	Asp 90	Pro	Asn	His	Phe	Arg 95	Pro
Ala	Gly	Leu	Pro 100	Glu	Lys	Tyr									

<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

ES 2 606 015 T3

<211> 42
 <212> PRT
 <213> Homo sapiens

5 <400> 9

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

10 <210> 10
 <211> 24
 <212> PRT
 <213> Homo sapiens

15 <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

1

REIVINDICACIONES

1. Un método para ayudar en 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;
- b. comparar la cantidad de cada biomarcador en el panel con un punto corte predeterminado de 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 llegar a una puntuación total de 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 puntuación total, en donde los biomarcadores cuantificados son al menos Acn9459 y TFA9133, en donde Acn9459 tiene la secuencia de aminoácidos expuesta en el SEQ ID NO: 8, y en donde TFA9133 tiene la secuencia de aminoácidos expuesta en el SEQ ID NO: 8 que ha perdido un ácido siálico ~~del~~ radical glicosilado Thr74.

2. El método de la reivindicación 1, en donde los dos o más biomarcadores se seleccionan del grupo de anticuerpos, antígenos y regiones de interés.

3. El método de la reivindicación 1, que comprende adicionalmente la obtención de al menos un parámetro biométrico del sujeto.

4. El método de la reivindicación 3, en donde el al menos un parámetro biométrico se basa en el historial de tabaquismo del sujeto.

5. El método de la reivindicación 3, que comprende adicionalmente la etapa de comparar el al menos un parámetro biométrico frente a un punto de corte predeterminado para cada uno de dichos parámetros biométricos y asignar una puntuación a cada parámetro biométrico basándose en dicha comparación, combinando la puntuación asignada a cada parámetro biométrico con la puntuación asignada a cada biomarcador cuantificado en la etapa b para llegar a una puntuación total de dicho sujeto en 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 en la etapa e.

6. El método de la reivindicación 1, en donde los biomarcadores adicionales se seleccionan del grupo que consiste en anticuerpo anti-p53, anticuerpo anti-TMP21, anticuerpo anti-dominio de NPC1L1C, anticuerpo anti-TMOD1, anticuerpo anti-CAMK1, anticuerpo anti-RGS1, anticuerpo anti-PACSIN1, anticuerpo anti-RCV1, anticuerpo anti-MAPKAPK3, al menos un anticuerpo contra ciclina E2 inmunorreactiva, citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA15-3, SCC, proGRP, CA19-9, amiloide A sérico, alfa-1-anti-tripsina, apolipoproteína CIII, Pub11597, Pub4789, y TFA2759, en donde Pub11597 tiene la secuencia de aminoácidos expuesta en el SEQ ID NO: 7. en donde Pub4789 tiene la secuencia de aminoácidos expuesta en el SEQ ID NO: 9. y en donde TFA2759 tiene la secuencia de aminoácidos expuesta en el SEQ ID NO: 10.

7. El método de la reivindicación 3, en donde el parámetro biométrico se selecciona del grupo que consiste en: el historial de tabaquismo, la edad, la exposición a carcinógenos y el género del sujeto.

8. El método de la reivindicación 7, en donde el parámetro biométrico es los paquetes-año de tabaquismo.

9. El método de la reivindicación 1, en donde el panel comprende al menos un antígeno, al menos un anticuerpo o una combinación de al menos un antígeno y al menos un anticuerpo.

10. El método de la reivindicación 9, en donde el al menos un anticuerpo se selecciona del grupo que consiste en: anti-p53, anti-TMP21, anti-dominio de NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSIN1, anti-RCV1, anti-MAPKAPK3 y al menos un anticuerpo contra ciclina E2 inmunorreactiva.

11. El método de la reivindicación 9, en donde el al menos un antígeno se selecciona del grupo que consiste en: citoqueratina 8, citoqueratina 19, citoqueratina 18, CEA, CA125, CA19-9, CA15-3, SCC, proGRP, el amiloide A sérico, alfa-1-antitripsina y apolipoproteína CIII.

12. El método de la reivindicación 1, en donde el panel comprende al menos una región de interés seleccionada del grupo que consiste en: Pub11597, Pub4789, y TFA2759, en donde Pub11597 tiene la secuencia de aminoácidos expuesta en el SEQ ID NO: 7. en donde Pub4789 tiene la secuencia de aminoácidos expuesta en el SEQ ID NO: 9. y en donde TFA2759 tiene la secuencia de aminoácidos expuesta en el SEQ ID NO: 10.

13. El método de la reivindicación 1, en donde la DFI de los biomarcadores con respecto al cáncer de pulmón es de menos de 0,4.

14. El método de la reivindicación 1, en donde los puntos de corte predeterminados se basan en las curvas ROC.

15. El método de la reivindicación 1, en donde la etapa b. comprende comparar la cantidad de cada biomarcador en el panel con varios puntos 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 a cada biomarcador determinada en la etapa b. para llegar a una puntuación total de dicho sujeto, la etapa d. comprende comparar la puntuación total determinada en la etapa c. con varias puntuaciones totales predeterminadas, y la etapa e. comprende determinar si dicho sujeto tiene cáncer de pulmón basándose en la puntuación total.

16. El método de la reivindicación 6, en donde los biomarcadores cuantificados son anticuerpo anti-p53, citoqueratina 19, CEA, Acn9459, Pub11597, Pub4789, TFA2759, y TFA9133.

17. Un kit que comprende:

a. un primer reactivo que contiene al menos un anticuerpo para cuantificar la región de interés ACN9459, en donde ACN9459 tiene la secuencia de aminoácidos expuesta en el SEQ ID NO: 8;

b. un segundo reactivo que contiene al menos un anticuerpo cuantificar la región de interés TFA9133, en donde TFA9133 tiene la secuencia de aminoácidos expuesta en el SEQ ID NO: 8 que ha perdido un ácido siálico del radical glicosilado Thr74;

c. 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, CA15-3, SCC, CA19-9, ProGRP, amiloide A sérico, alfa-1-antitripsina y apolipoproteína CIII;

d. reactivos que contienen uno o más antígenos para cuantificar al menos un anticuerpo en una muestra de ensayo; en donde dichos anticuerpos son: anti-p53, anti-TMP21, anti-dominio NPC1L1C, anti-TMOD1, anti-CAMK1, anti-RGS1, anti-PACSIN1, anti-RCV1, anti-MAPKAPK3 y al menos un anticuerpo contra ciclina E2 inmunorreactiva; y

e. reactivos que contienen al menos un anticuerpo para cuantificar una o más regiones de interés seleccionadas del grupo que consiste en: Pub11597, Pub4789, y TFA2759, en donde Pub11597 tiene la secuencia de aminoácidos expuesta en el SEQ ID NO: 7, en donde Pub4789 tiene la secuencia de aminoácidos expuesta en el SEQ ID NO: 9. y en donde TFA2759 tiene la secuencia aminoácidos expuesta en el SEQ ID NO: 10.

18. El kit de la reivindicación 17 que comprende adicionalmente un péptido seleccionado del 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.

19. El kit de la reivindicación 17, en donde los uno o más antígenos de c. son citoqueratina 19 y CEA; en donde el al menos un anticuerpo de d. es anti-p53; y en donde las una o más regiones de interés de e. son Pub11597, Pub4789, y TFA2759.

Fig. 1

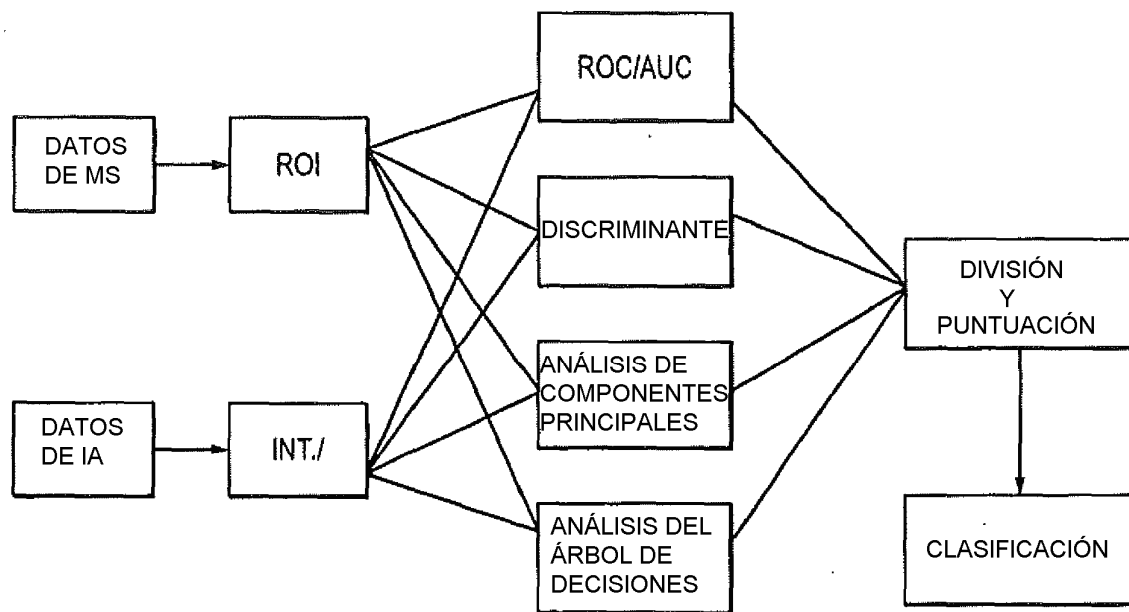
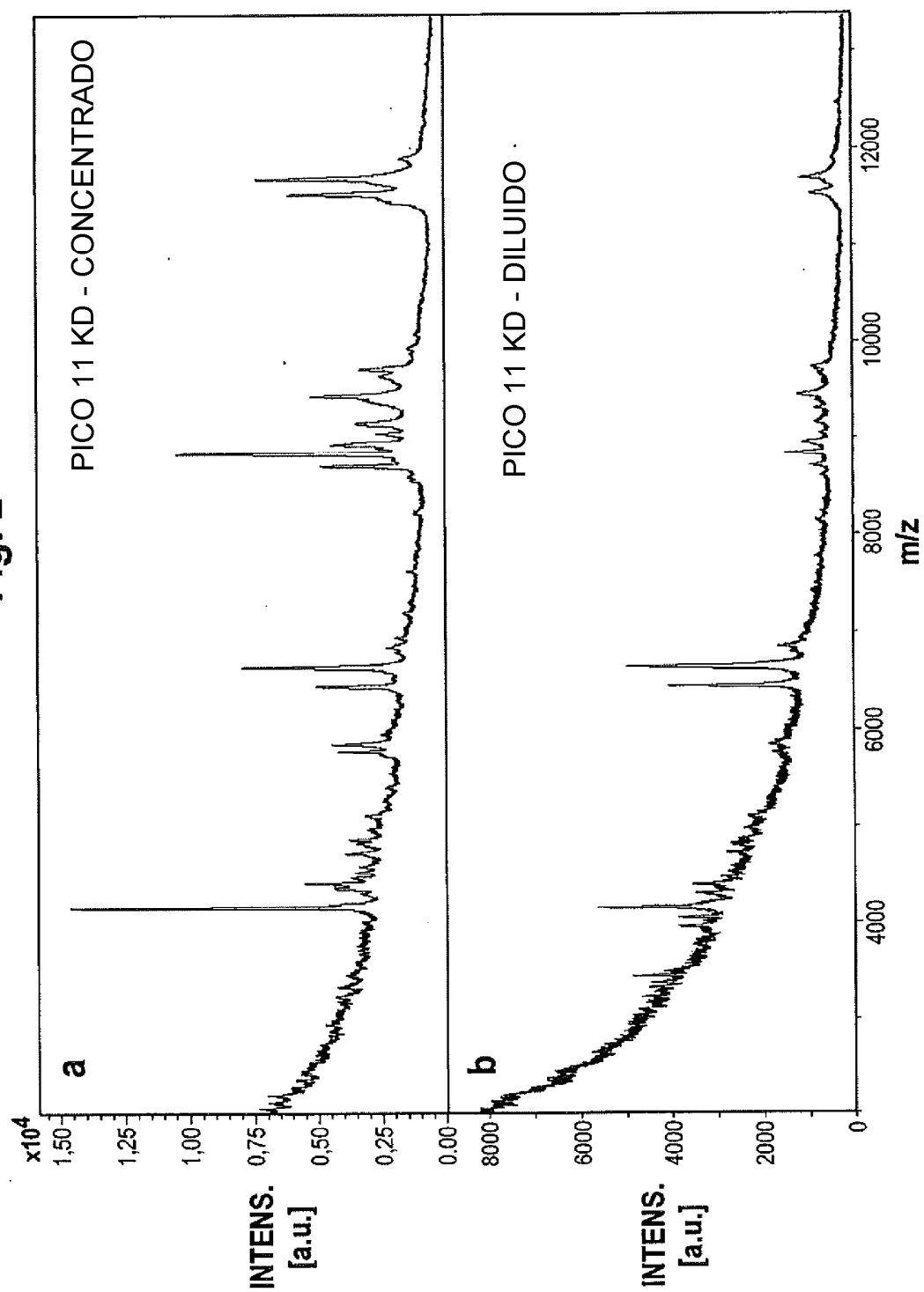


Fig. 2



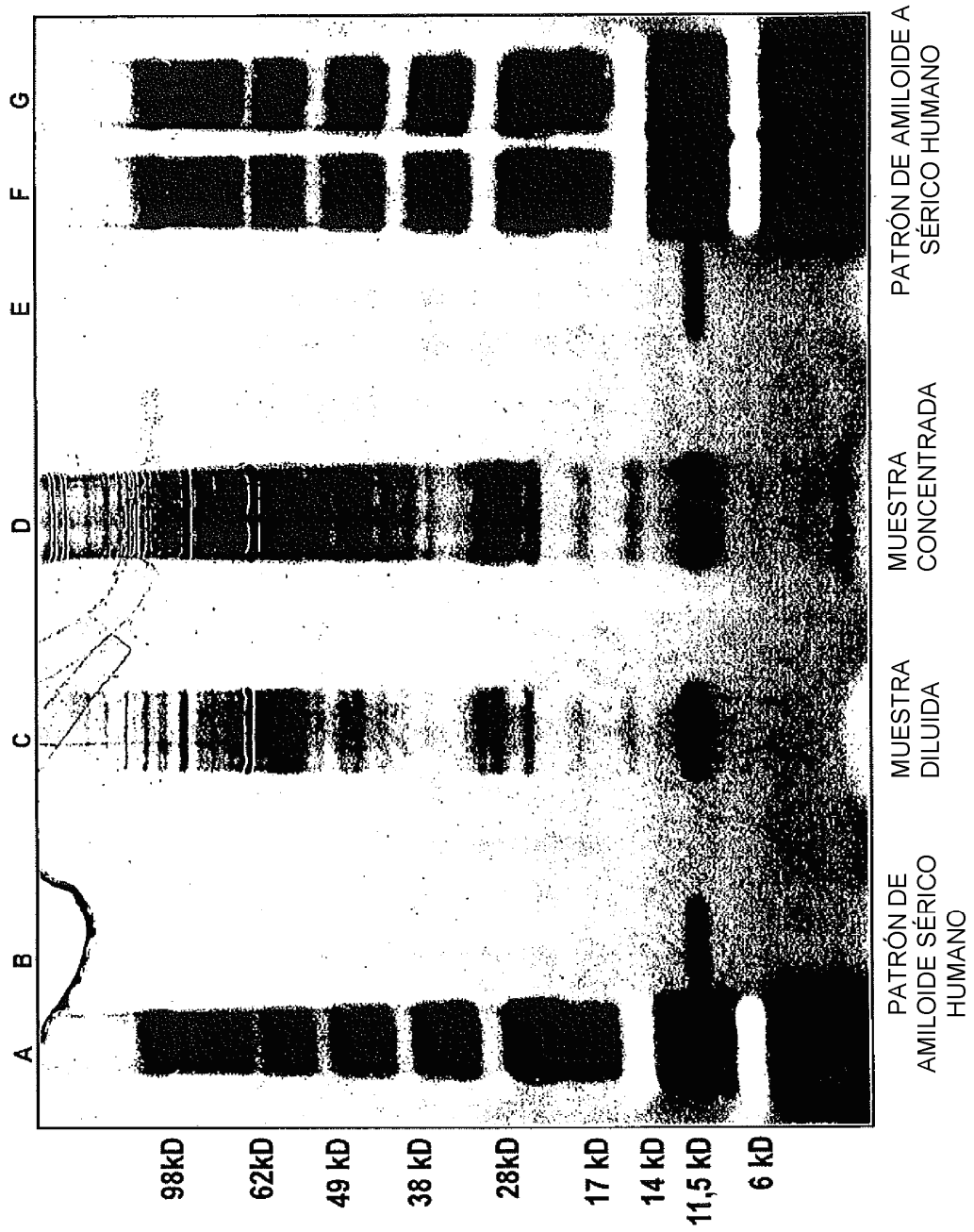


Fig. 3

Fig. 4a

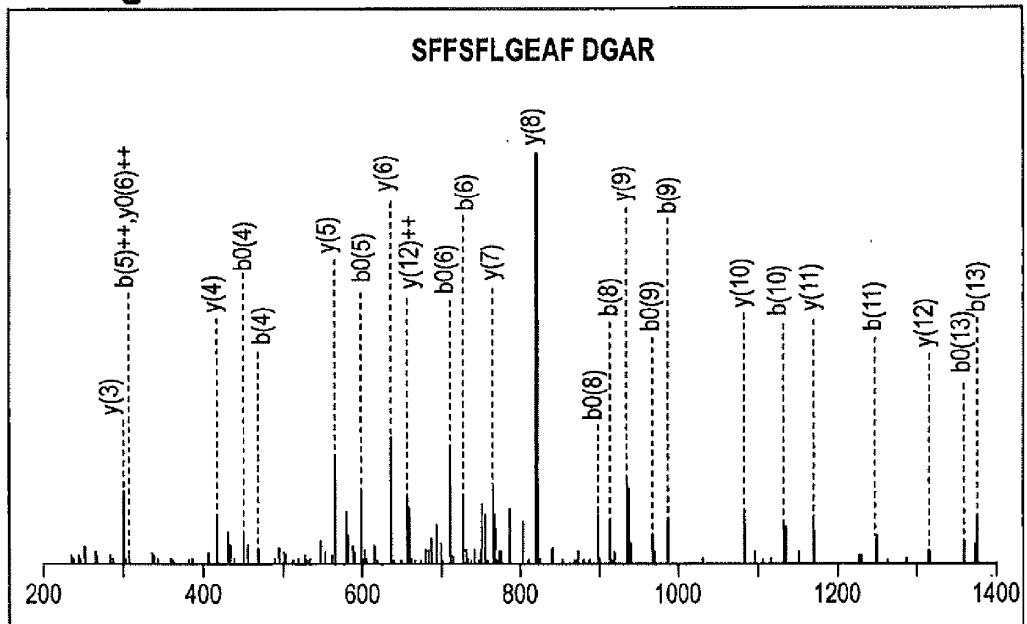


Fig. 4b

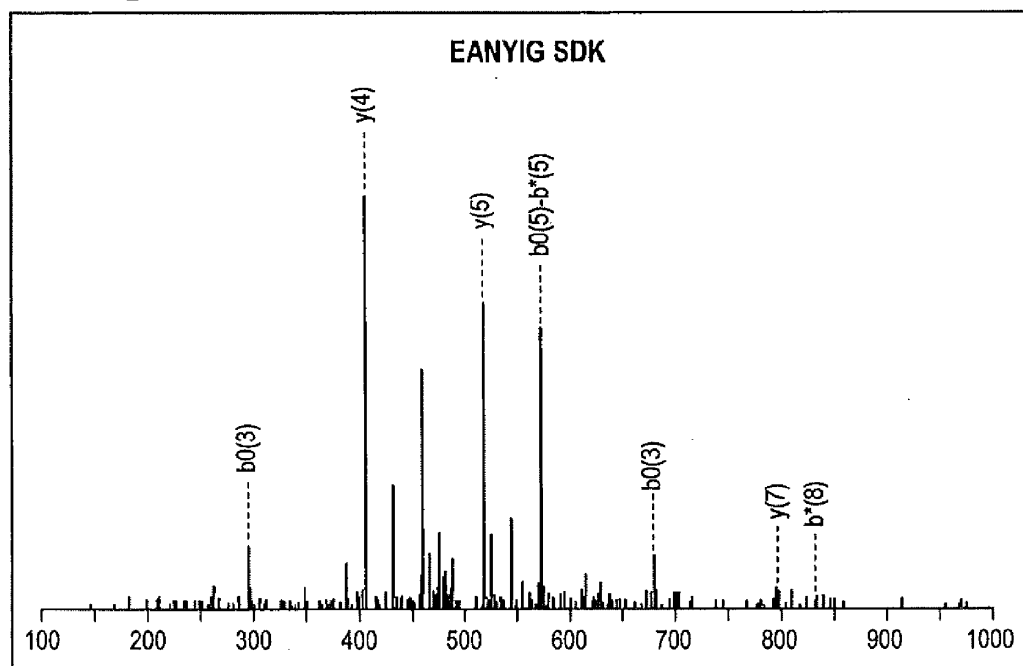


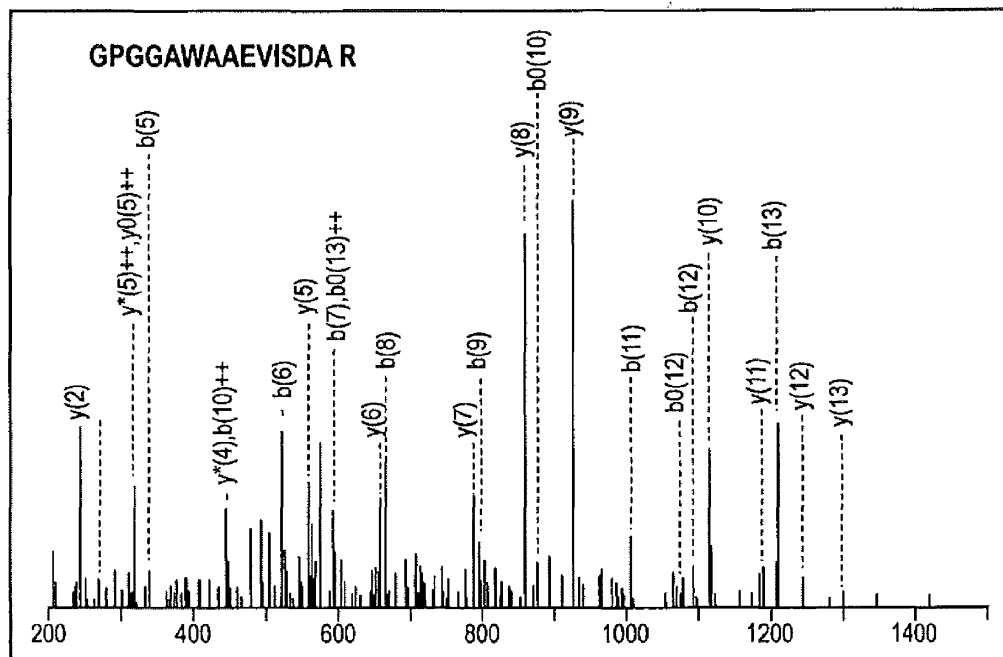
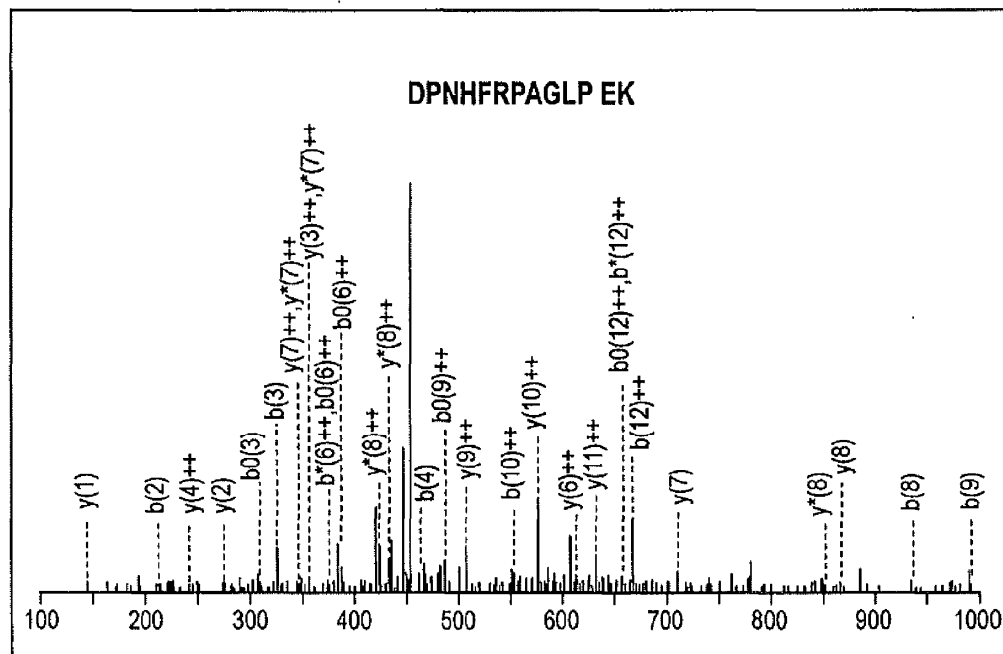
Fig. 4c**Fig. 4d**

Fig. 5

