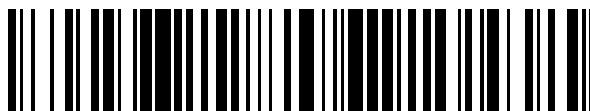


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 704 888**

51 Int. Cl.:

G01N 33/574 (2006.01)

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **11.11.2014 PCT/GB2014/053340**

87 Fecha y número de publicación internacional: **14.05.2015 WO15067969**

96 Fecha de presentación y número de la solicitud europea: **11.11.2014 E 14815406 (5)**

97 Fecha y número de publicación de la concesión europea: **10.10.2018 EP 3069143**

54 Título: **Método, matriz y uso de los mismos para determinar cáncer de páncreas**

30 Prioridad:

11.11.2013 GB 201319878

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

20.03.2019

73 Titular/es:

IMMUNOVIA AB (100.0%)

Medicon Village

223 81 Lund, SE

72 Inventor/es:

**BORREBAECK, CARL ARNE KRISTER;
WINGREN, CHRISTER LARS BERTIL y
GERDTSSON, SANDSTRÖM ANNA**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 704 888 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

Método, matriz y uso de los mismos para determinar cáncer de páncreas

5 Campo de la invención

La presente invención se refiere a métodos para detectar cáncer de páncreas, y a biomarcadores y matrices para su uso para ello.

10 Antecedentes

El adenocarcinoma ductal pancreático o cáncer de páncreas (PC) es la 4ª causa más común de muerte relacionada con el cáncer, con el resultado de casi tantas muertes como el cáncer de mama en los Estados Unidos, a pesar de que su incidencia es 10 veces menor [1]. El mal pronóstico se debe sobre todo a la incapacidad para detectar PC en los primeros estadios, a pesar incluso de que los datos corroboran que se tardan más de cinco años desde el inicio del tumor hasta que adquiere la capacidad metastásica [2], lo cual demuestra claramente una ventana de oportunidad para la detección temprana si se dispusiera de marcadores. Hoy en día, suele detectarse el cáncer cuando ha progresado hasta una enfermedad avanzada en la que los tumores no se pueden operar y se ha producido metástasis ya [1, 3]. Sin embargo, incluso aunque la supervivencia de cinco años de los tumores extirpados grandes sea solamente 10-20 % [4, 5], aumenta a 30-60 % si se pueden extirpar los tumores de <20 mm [6,7] y a >75 % si el tamaño en la extirpación es <10 mm [8]. La detección tardía se debe a síntomas clínicos inespecíficos, así como a la falta de tecnologías y marcadores sensibles para un diagnóstico temprano. Cabe destacar que los estudios indican que los tumores pancreáticos podrían extirparse hasta seis meses antes del diagnóstico clínico en una etapa asintomática [9,10].

El marcador más valorado hasta la fecha para el PC, CA 19-9, adolece de una escasa especificidad, con niveles elevados también en pancreatitis y otros cánceres y una completa ausencia de tumores de Lewis a y b negativos. En consecuencia, el uso de CA19-9 para detectar cáncer de páncreas se ha desaconsejado [11]. A pesar de los muchos esfuerzos, no se ha demostrado que ningún otro biomarcador simple supere CA19-9 y en los últimos años la tendencia en este campo se ha desplazado hacia paneles de marcadores multiplexados para una mayor sensibilidad y especificidad [12]. Los paneles han consistido principalmente en proteínas de la sangre altamente abundantes, reactivos en fase aguda frecuentemente (p.ej. CRP y SAA) marcadores de tumor conocidos (p.ej. CA242, CA125 y CEA), moléculas de adhesión (p.ej. ICAM-1 y ADCAM), proteínas relacionadas con la degradación de la matriz extracelular (p.ej. MMP y TIMP1), lipoproteínas (p.ej. Apo-C1, Apo-A2), y otras varias, sobre todo en combinación con CA19-9 [15-20]. A pesar de los informes sobre la alta sensibilidad para PC frente a los controles sanos o afecciones pancreáticas benignas, ninguno de estos paneles ha sido validado todavía para uso clínico.

Además de los marcadores ya mencionados, varias proteínas inmunorreguladoras pueden ser de interés como biomarcadores de tumor, ya que la asociación entre cáncer e inflamación sigue sin dilucidarse [21]. Teniendo en cuenta el efecto sistémico, así como la multitud de funciones de muchas de las proteínas inmunorreguladoras, generalmente se considera que paneles reducidos de 2 a 5 marcadores no serán suficientemente específicos para cáncer de páncreas, en particular, cuando se trata de discriminar cáncer de páncreas de pancreatitis y otras afecciones que presentan síntomas similares. Sin embargo, los estudios anteriores han demostrado que un aumento del número de estos analitos (≤ 25) puede proporcionar inmunofirmas de enfermedad muy específicas que reflejan la respuesta sistémica a la enfermedad [22-24].

Li et al., 2008, Chinese J., Cancer Prevention & Treatment, 15(15): 1166-1168 se refiere a un método para determinar la presencia de cáncer de páncreas en un individuo que comprende o que consiste en las etapas de (a) proporcionar una muestra para su análisis desde el individuo y (b) determinar una firma de biomarcador de la muestra de ensayo para medir la expresión en la muestra de ensayo de dos o más biomarcadores seleccionados del grupo definido en la Tabla III, en donde la expresión en la muestra de ensayo de dos o más biomarcadores seleccionados del grupo que se define en la Tabla III es indicativa del individuo que tiene cáncer de páncreas.

La patente estadounidense US 2007/212738 A1 (Haley et al.) se refiere a métodos de diagnóstico y pronóstico para predecir la eficacia del tratamiento de un paciente de cáncer con un inhibidor de quinasa EGFR.

La patente internacional WO 2006/110581 A2 (Chiron Corp. (US) & Sagres Discovery Inc.) se refiere a métodos para detectar cáncer o la probabilidad de desarrollar cáncer sobre la base de la presencia o la ausencia de la expresión de ciertos genes o proteínas codificados por dichos genes.

Fujiwara et al., 2005, Biochem. & Biophys. Res. Com., 338(2):1256-1262 divulga la observación de que la proteína LTBR de longitud completa (tipo silvestre) manifiesta actividad de transformación.

Wingren et al., 2012, Cancer Res., 72(10):2481-2490 divulga una firma de biomarcador en suero multiplexado prevalidado para el diagnóstico de cáncer de páncreas que puede mejorar el diagnóstico y prevención de enfermedades premalignas y en la identificación sistemática de individuos en alto riesgo.

Ingvarsson et al., 2008, *Proteomics*, 8(11):2211-2219 se refiere a la identificación de una firma de proteína basada en 19 analitos no redundantes que discrimina entre pacientes con cáncer y sujetos sanos.

Pauly et al., 2013, *Journal of Proteome Research*, 12(12): 5943-5953 demuestra que se podrían extraer proteínas de material de tejido impregnado en parafina fijado en formalina (FFPE), marcado y perfilado posteriormente de una manera multiplexada, sensible y reproducible utilizando micromatrices de anticuerpo scFv recombinantes.

Cho et al., 2008, *Gastroenterology*. 134(4): A-699 divulga la expresión de SOX11 y SOX4 en tumor de páncreas pseudopapilar sólido, pero no en páncreas normal o en adenocarcinoma pancreático.

La patente internacional WO 2012/120288 (Immunovia AB) se refiere a un método para determinar la presencia de cáncer pancreático en un individuo que comprende o que consiste en las etapas de (a) proporcionar una muestra para su análisis de un individuo y (b) determinar una firma de biomarcador de la muestra de ensayo midiendo la expresión en la muestra de ensayo de dos o más biomarcadores seleccionados del grupo definido en la Tabla III, en donde la expresión en la muestra de ensayo de dos o más biomarcadores seleccionados del grupo definido en la Tabla III es indicativo del individuo que tiene cáncer de páncreas.

No obstante, el análisis del proteoma inmunorregulador está asociado a varios retos. En primer lugar, la concentración en suero de las proteínas de interés despliega una inmensa gama dinámica, desde tan alta como microgramos a tan baja como picograma por ml, lo cual complica su detección simultánea utilizando metodologías proteómicas convencionales [25, 26]. En segundo lugar, parece haber un consenso en cuanto a que es más probable que se encuentren marcadores de cáncer prometedores entre las proteínas a menudo de peso molecular bajo y las menos abundantes [27,28]. En tercer lugar, es de esperar que los cambios asociados a la enfermedad en los niveles de suero de estos analitos poco abundantes sean pequeños y por tanto, se necesitará un significativo número de muestras clínicas para una estadística adecuada [29]. Para estos fines, los autores de la invención han diseñado micromatrices de anticuerpos recombinantes multiplexados con cerca de 300 anticuerpos scFv dirigidos principalmente a proteínas inmunoregulatorias [30]. Con estas matrices, se puede medir la expresión de proteína en cientos de muestras de una manera altamente productiva y reproductiva.

Sumario de la invención

La invención se define según las reivindicaciones del presente documento.

En este estudio multicéntrico, se analizaron muestras de suero de 338 individuos de pacientes con cáncer de páncreas, enfermedad pancreática benigna, así como controles normales, en las micromatrices de anticuerpo diseñadas dentro del laboratorio de los autores de la invención. Para definir las firmas de marcador más discriminatorias, se aplicó un procedimiento de eliminación retroactiva iterativa, sobre la base de análisis mecánico de vector de soporte y diseñado para predecir la óptima combinación de anticuerpos [31]. Para este fin, se identificaron inmunofirmas 25-plexada para discriminar PC de controles benignos y sanos en grupos de entrenamiento, prevalados en grupos de ensayo independientes y se compararon con firmas derivadas de análisis de expresión de proteína diferencial. Por otra parte, se pudo aplicar un perfil de proteína para estratificar muestras de suero sobre la base del emplazamiento original del tumor en el páncreas, lo cual hasta donde llega el conocimiento de los autores de la invención no se ha realizado anteriormente con proteómica. En conjunto, estos hallazgos añaden una importante información al rompecabezas del proteoma del cáncer de páncreas que podría tener como resultado final biomarcadores multiplexados que proporcionan un beneficio a miles de pacientes.

Por consiguiente, un primer aspecto de la invención proporciona un método para determinar la presencia de cáncer de páncreas en un individuo que comprende o que consiste en las etapas de:

- a) proporcionar una muestra para su análisis obtenida de un individuo;
- b) determinar la firma de biomarcador de la muestra de ensayo midiendo la expresión de la muestra de ensayo de dos o más biomarcadores seleccionados del grupo definido en la Tabla A;

en donde la etapa (b) comprende la medición de la expresión en la muestra de ensayo de HADH2, y en donde la expresión en la muestra de ensayo de dos o más biomarcadores seleccionados del grupo definido en la Tabla A es indicativo de la presencia de cáncer de páncreas.

Dentro de las expresiones "muestra para su análisis", "muestra de ensayo" o "muestra de control", se incluye una muestra de tejido o de fluido que se ha extraído o derivado previamente de un individuo. Preferentemente, la muestra para su análisis es de un mamífero. El mamífero puede ser un animal doméstico o de granja. Preferentemente, el mamífero es una rata, un ratón, una cobaya, un gato, un perro, un caballo o un primate. Lo más preferente el mamífero es un ser humano. Preferentemente, la muestra es una muestra de célula o tejido (o un derivado de los mismos) que comprende o que consiste en plasma, células del plasma, suero, células de tejido, siendo igualmente preferente, proteínas o ácidos nucleicos derivados de una muestra de tejido o célula. Preferentemente, las muestras de ensayo y de control se derivan de la misma especie.

En una realización, la muestra de tejido es tejido pancreático. En una realización, la muestra de célula es una muestra de células pancreáticas.

“Expresión” significa según los autores de la invención el nivel o la cantidad de un producto génico como por ejemplo ARNm o proteína.

Los métodos para detectar y/o medir la concentración de la proteína y/o ácido nucleico son conocidos entre las personas expertas en la materia, véase por ejemplo, Sambrook y Russell, 2001, Cold Spring Harbor Laboratory Press.

“Biomarcador” significa una molécula biológica natural, o un componente o fragmento de la misma, cuya medición puede proporcionar información útil en el pronóstico de cáncer de páncreas. Por ejemplo, el biomarcador puede ser una proteína natural o una fracción de carbohidrato, o un componente antigénico o fragmento del mismo.

En una realización, el método comprende o consiste en las etapas (a) y (b) y además las etapas de:

- c) proporcionar una muestra de control obtenida de un individuo no afligido por cáncer de páncreas;
- d) determinar una firma de biomarcador de la muestra de control midiendo la expresión de la muestra de control de dos o más biomarcadores medidos en la etapa (b);

en donde la presencia de cáncer de páncreas se identifica cuando la expresión en la muestra de ensayo de dos o más biomarcadores medidos en la etapa (b) es diferente de la expresión en la muestra de control de dos o más biomarcadores medidos en la etapa (d).

La una o más muestras de control pueden ser de un individuo sano (es decir, un individuo que no está afligido por ninguna enfermedad ni afección), un individuo afligido por una enfermedad o afección no pancreática o un individuo afligido por una enfermedad o afección pancreática benigna (por ejemplo, pancreatitis aguda o crónica).

En otra realización, el método comprende o consiste en las etapas (a) y (b) y, opcionalmente, las etapas (c) y (d) y las etapas adicionales de:

- e) proporcionar una o más muestras de control obtenidas de un individuo afligido por cáncer de páncreas;
- f) determinar una firma de biomarcador de la muestra de control midiendo la expresión en la muestra de control de dos o más biomarcadores medidos en la etapa (b);

en donde la presencia del cáncer de páncreas se identifica cuando la expresión en la muestra de ensayo de dos o más biomarcadores medidos en la etapa (b) se corresponde con la expresión en la muestra de control de dos o más biomarcadores medidos en la etapa (f).

En una realización, se utiliza un valor patrón o de referencia en lugar, o además de, el uno o más controles positivos o negativos. Por tanto, los valores patrón o de referencia pueden determinarse en procedimientos por separado a partir de los valores de ensayo.

En una realización, el método es para determinar la presencia de cáncer de páncreas originado en la cabeza del páncreas o (ii) el cuerpo o la cola del páncreas. En esta realización, la etapa (e) puede comprender (i) proporcionar una o más muestras de control de un individuo afligido con cáncer de páncreas originado en la cabeza del páncreas y/o (ii) proporcionar una o más muestras de control de un individuo afligido con cáncer de páncreas originado en el cuerpo o la cola del páncreas.

En una realización, la presencia de cáncer de páncreas originado en la cabeza del páncreas se define en el caso de que la expresión en la muestra (e)(i) (cuando está presente) de dos o más biomarcadores medidos en la etapa (b) se corresponde con la expresión en la muestra de control de dos o más biomarcadores medidos en la etapa (f) y/o la expresión en la muestra de ensayo (e) (ii) (cuando está presente) de dos o más biomarcadores medidos en la etapa (b) es diferente de la expresión en la muestra de control de dos o más biomarcadores medidos en la etapa (f).

En una realización, la presencia de cáncer de páncreas originado en el cuerpo o la cola del páncreas se identifica en el caso de que la expresión en la muestra de ensayo (e) (i) (cuando está presente) de dos o más biomarcadores medidos en la etapa (b) es diferente de la expresión en la muestra de control de dos o más biomarcadores medidos en la etapa (f) y/o la expresión en la muestra de ensayo (e)(ii) (cuando está presente) de dos o más biomarcadores medidos en la etapa (b) corresponde a la expresión en la muestra de control de dos o más biomarcadores medidos en la etapa (f).

Se entiende por “que corresponde a la expresión en la muestra de control” que la expresión de dos o más biomarcadores en la muestra que se analiza es igual o similar a la expresión de los dos o más biomarcadores de la muestra de control positivo. Preferentemente, la expresión de dos o más biomarcadores en la muestra que se va a analizar es idéntica a la expresión de dos o más biomarcadores de la muestra de control positivo.

La expresión diferencial (regulación al alza o regulación a la baja) de biomarcadores, o la falta de los mismos, puede determinarse a través de cualquier medio adecuado conocido entre las personas expertas en la materia. La expresión diferencial se determina para un valor p como mínimo de menos de 0,05 ($p = < 0,05$), por ejemplo, al menos $<0,04$, $<0,03$, $<0,02$, $<0,01$, $<0,009$, $<0,005$, $<0,001$, $<0,0001$, $<0,00001$ o al menos $<0,000001$.

5 Preferentemente, la expresión diferencial se determina utilizando una máquina de vector de soporte (MVS). Preferentemente, la MVS es una MVS como la que se describe más adelante. Lo más preferente, la MVS es la MVS descrita en la Tabla B más adelante.

10 Tal como apreciarán las personas expertas en la materia, la expresión diferencial se puede relacionar con un solo biomarcador o con varios biomarcadores considerados en combinación (es decir, una *firma* de biomarcador). Por lo tanto, un valor p puede asociarse con un solo biomarcador o con un grupo de biomarcadores. De hecho, las proteínas que tienen un valor p de expresión diferencial superior a 0,05 cuando se consideran individualmente pueden seguir siendo útiles no obstante como biomarcadores de acuerdo con los métodos de la invención cuando se consideran sus niveles de expresión en combinación con otro u otros biomarcadores.

15 Tal como se ilustra en los ejemplos adjuntos, la expresión de ciertas proteínas en una muestra de ensayo de tejido, sangre, suero, o plasma puede ser indicativa de cáncer de páncreas en un individuo. Por ejemplo, la expresión relativa de ciertas proteínas de suero en una sola muestra de ensayo puede ser indicativa de la presencia de cáncer de páncreas en un individuo.

20 La etapa (b) puede comprender además la medición de la expresión de 1 o más biomarcadores a partir de los biomarcadores enumerados en la Tabla A (i). La etapa (b) puede comprender o consistir en la medición de la expresión de cada uno de los biomarcadores enumerados en la Tabla A(i), por ejemplo, al menos 2 de los biomarcadores enumerados en la Tabla A(i). Por tanto, la etapa (b) puede comprender o consistir en la medición de la expresión de todos los biomarcadores enumerados en la Tabla A(i).

25 La etapa (b) puede comprender además medir la expresión de 1 o más biomarcadores de los biomarcadores enumerados en Tabla A(ii), por ejemplo al menos 2, 3, 4, 5, 6, 7, 8 o 9 de los biomarcadores enumerados en la Tabla A(ii). Por lo tanto, la etapa (b) puede comprender medir la expresión de todos los biomarcadores enumerados en la Tabla A(ii).

30 La etapa (b) puede comprender además medir la expresión de 1 o más biomarcadores de los biomarcadores enumerados en Tabla A(iii), por ejemplo al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 o 17 de los biomarcadores enumerados en la Tabla A(iii). Por lo tanto, la etapa (b) puede comprender medir la expresión de todos los biomarcadores enumerados en la Tabla A(iii).

35 La etapa (b) puede comprender además medir la expresión de 1 o más biomarcadores de los biomarcadores enumerados en Tabla A(iv), por ejemplo al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 o 22, 23 o 24 de los biomarcadores enumerados en la Tabla A(iv). Por lo tanto, la etapa (b) puede comprender medir la expresión de todos los biomarcadores enumerados en la Tabla A(iv).

40 La etapa (b) puede comprender además medir la expresión de 1 o más biomarcadores de los biomarcadores enumerados en Tabla A(v), por ejemplo al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 45 53, 54, 55 o 56 de los biomarcadores enumerados en la Tabla A(v). Por lo tanto, la etapa (b) puede comprender medir la expresión de todos los biomarcadores enumerados en la Tabla A(v).

50 La etapa (b) puede comprender además medir la expresión de 1 o más biomarcadores de los biomarcadores enumerados en Tabla A(vi). Por lo tanto, la etapa (b) puede comprender medir la expresión de todos los biomarcadores enumerados en la Tabla A(vi).

55 La etapa (b) puede comprender además medir la expresión de 1 o más biomarcadores de los biomarcadores enumerados en la Tabla A, por ejemplo al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109 o 110 de los biomarcadores enumerados en la Tabla A. Por lo tanto, la etapa (b) puede comprender medir la expresión de todos los biomarcadores enumerados en la Tabla A.

60 Cuando el método es para determinar la presencia de adenocarcinoma pancreático la etapa (b) comprende preferentemente medir la expresión de HADH2 y:

1 o más biomarcadores de los biomarcadores enumerados en Tabla A(i), por ejemplo al menos 2 de los biomarcadores enumerados en la Tabla A(i);

1 o más biomarcadores de los biomarcadores enumerados en Tabla A(ii), por ejemplo al menos 2, 3, 4, 5, 6, 7, 8

o 9 de los biomarcadores enumerados en la Tabla A(ii);

1 o más biomarcadores de los biomarcadores enumerados en Tabla A(iii), por ejemplo al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 o 17 de los biomarcadores enumerados en la Tabla A(iii);

1 o más biomarcadores de los biomarcadores enumerados en Tabla A(iv), por ejemplo al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 o 24 de los biomarcadores enumerados en la Tabla A(iv); y/o

1 o más biomarcadores de los biomarcadores enumerados en Tabla A(v), por ejemplo al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 o 56 de los biomarcadores enumerados en la Tabla A(v).

1 o más biomarcadores de los biomarcadores enumerados en Tabla A(vi). Por lo tanto, la etapa (b) puede comprender o consistir en medir la expresión de todos los biomarcadores enumerados en la Tabla A(vi).

Cuando el método es para determinar la presencia de adenocarcinoma pancreático originado en (i) la cabeza del páncreas o (ii) el cuerpo o la cola del páncreas, la etapa (b) comprende preferentemente medir la expresión de HADH2 y:

1 o más biomarcadores de los biomarcadores enumerados en Tabla A(ii), por ejemplo al menos 2, 3, 4, 5, 6, 7, 8 o 9 de los biomarcadores enumerados en la Tabla A(ii);

1 o más biomarcadores de los biomarcadores enumerados en Tabla A(iv), por ejemplo al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 13 o 24 de los biomarcadores enumerados en la Tabla A(iv); y/o

1 o más biomarcadores enumerados en la Tabla A(vi).

Cuando se hace referencia a una patología "normal", se pretende incluir los individuos no afligidos con una pancreatitis crónica (ChP) o una pancreatitis inflamatoria aguda (AIP). Preferentemente, los individuos no están afligidos de ninguna enfermedad o trastorno pancreático. Lo más preferente, los individuos son individuos sanos, es decir, no están afligidos de ninguna enfermedad o trastorno.

En los métodos de la invención, la etapa (b) comprende medir la expresión en la muestra de ensayo de HADH2 y al menos un biomarcador adicional seleccionado del grupo que se define en la Tabla A (i), (ii) o (iii).

Por lo tanto, en una realización, la etapa (b) comprende la medición de la expresión de factor de transcripción SOX-11. En otra realización adicional, la etapa (b) comprende medir la expresión de integrina alfa-10. En una realización adicional, la etapa (b) comprende medir la expresión de EDNR. En una realización adicional, la etapa (b) comprende medir la expresión de EPFR. En una realización adicional, la etapa (b) comprende medir la expresión de LSADHR. En una realización adicional, la etapa (b) comprende medir la expresión de SEAHLR. En una realización adicional, la etapa (b) comprende medir la expresión de AQHQWDGLLSYQDSLS. En una realización adicional, la etapa (b) comprende medir la expresión de WTRNSNMNYWLIIRL. En una realización adicional, la etapa (b) comprende medir la expresión de WDSR. En una realización adicional, la etapa (b) comprende medir la expresión de DFAEDK. En una realización adicional, la etapa (b) comprende medir la expresión de proteína FASN. En una realización adicional, la etapa (b) comprende medir la expresión de proteína GAK. En una realización adicional, la etapa (b) comprende medir la expresión de LNVWGK. En una realización adicional, la etapa (b) comprende medir la expresión de LTEFAK. En una realización adicional, la etapa (b) comprende medir la expresión de LYEIAR. En una realización adicional, la etapa (b) comprende medir la expresión de proteína tirosina quinasa asociada a megacariocito. En una realización adicional, la etapa (b) comprende medir la expresión de proteína 3 relacionada con proteína de unión a oxisterol 3. En una realización adicional, la etapa (b) comprende medir la expresión de QEASFK. En una realización adicional, la etapa (b) comprende medir la expresión de SSAYSR. En una realización adicional, la etapa (b) comprende medir la expresión de QEASFK. En una realización adicional, la etapa (b) comprende medir la expresión de TEEQLK. En una realización adicional, la etapa (b) comprende medir la expresión de TLYVGK. En una realización adicional, la etapa (b) comprende medir la expresión de FLLMQYGGMDEHAR. En una realización adicional, la etapa (b) comprende medir la expresión de GIVKLYEDEG. En una realización adicional, la etapa (b) comprende medir la expresión de GIVKLYEDEG. En una realización adicional, la etapa (b) comprende medir la expresión del miembro 3 de la superfamilia de receptor de factor de necrosis de tumor. En una realización adicional, la etapa (b) comprende medir la expresión de proteína tirosina quinasa SYK. En una realización adicional, la etapa (b) comprende medir la expresión de angiomotina. En una realización adicional, la etapa (b) comprende medir la expresión de quimiocina 2 de motivo C-C. En una realización adicional, la etapa (b) comprende medir la expresión de quimiocina 5 de motivo C-C. En una realización adicional, la etapa (b) comprende medir la expresión de ligando CD40. En una realización adicional, la etapa (b) comprende medir la expresión de péptido 1 de tipo glucagón. En una realización adicional, la etapa (b) comprende medir la expresión de inmunoglobulina M. En una realización

adicional, la etapa (b) comprende medir la expresión de interleucina1-alfa. En una realización adicional, la etapa (b) comprende medir la expresión de proteína antagonista de receptor de interleucina-1. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-11. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-12. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-16. En una realización adicional, la etapa (b) comprende medir la expresión de interleucina-18. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-2. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-3. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-4. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-6. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-7. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-9. En una realización adicional, la etapa (b) comprende medir la expresión de Lewis x. En una realización adicional, la etapa (b) comprende medir la expresión de linfotóxina-alfa. En una realización adicional, la etapa (b) comprende medir la expresión de factor de crecimiento de transformación beta-1. En una realización adicional, la etapa (b) comprende medir la expresión de factor de crecimiento endotelial vascular. En una realización adicional, la etapa (b) comprende medir la expresión de sistema visual homeobox 2. En una realización adicional, la etapa (b) comprende medir la expresión de HLA-DR/DP. En una realización adicional, la etapa (b) comprende medir la expresión de apolipoproteína A1. En una realización adicional, la etapa (b) comprende medir la expresión de apolipoproteína A4. En una realización adicional, la etapa (b) comprende medir la expresión de apolipoproteína B-100. En una realización adicional, la etapa (b) comprende medir la expresión de ATP sintasa subunidad beta, mitocondrial. En una realización adicional, la etapa (b) comprende medir la expresión de beta-galactosidasa. En una realización adicional, la etapa (b) comprende medir la expresión de cathepsina W. En una realización adicional, la etapa (b) comprende medir la expresión de quimiocina 13 de motivo C-C. En una realización adicional, la etapa (b) comprende medir la expresión de quimiocina 7 de motivo C-C. En una realización adicional, la etapa (b) comprende medir la expresión de proteína CD40. En una realización adicional, la etapa (b) comprende medir la expresión de Complemento C1q. En una realización adicional, la etapa (b) comprende medir la expresión de Complemento C1s. En una realización adicional, la etapa (b) comprende medir la expresión de Complemento C3. En una realización adicional, la etapa (b) comprende medir la expresión de Complemento C4. En una realización adicional, la etapa (b) comprende medir la expresión de Complemento C5. En una realización adicional, la etapa (b) comprende medir la expresión de Complemento factor B. En una realización adicional, la etapa (b) comprende medir la expresión de quinasas dependientes de ciclina 2. En una realización adicional, la etapa (b) comprende medir la expresión de Cistatina-C. En una realización adicional, la etapa (b) comprende medir la expresión de Eotaxina. En una realización adicional, la etapa (b) comprende medir la expresión de receptor de factor de crecimiento epidérmico. En una realización adicional, la etapa (b) comprende medir la expresión de receptor de péptido tipo glucagón 1. En una realización adicional, la etapa (b) comprende medir la expresión de factor de estimulación de colonia de granulocitos-macrófagos. En una realización adicional, la etapa (b) comprende medir la expresión de Integrina alfa-11. En una realización adicional, la etapa (b) comprende medir la expresión de molécula de adhesión intercelular 1. En una realización adicional, la etapa (b) comprende medir la expresión de Interferón gamma. En una realización adicional, la etapa (b) comprende medir la expresión de interleucina-1 beta. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-10. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-13. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-5. En una realización adicional, la etapa (b) comprende medir la expresión de Interleucina-8. En una realización adicional, la etapa (b) comprende medir la expresión de queratina, tipo I citoesquelética 19. En una realización adicional, la etapa (b) comprende medir la expresión de Leptina. En una realización adicional, la etapa (b) comprende medir la expresión de Lumican. En una realización adicional, la etapa (b) comprende medir la expresión de proteína quinasa 1 activada con mitógeno. En una realización adicional, la etapa (b) comprende medir la expresión de proteína quinasa 8 activada con mitógeno. En una realización adicional, la etapa (b) comprende medir la expresión de Mucina-1. En una realización adicional, la etapa (b) comprende medir la expresión de Miomesina-2. En una realización adicional, la etapa (b) comprende medir la expresión de Osteopontina. En una realización adicional, la etapa (b) comprende medir la expresión de subunidad alfa reguladora de fosfatidilinositol 3-quinasa. En una realización adicional, la etapa (b) comprende medir la expresión de inhibidor C1 de proteasa de plasma. En una realización adicional, la etapa (b) comprende medir la expresión de Properdina. En una realización adicional, la etapa (b) comprende medir la expresión de antígeno específico de próstata. En una realización adicional, la etapa (b) comprende medir la expresión de receptor erbB-2 de proteína tirosina quinasa. En una realización adicional, la etapa (b) comprende medir la expresión de regulador de transcritos no codificantes 3B. En una realización adicional, la etapa (b) comprende medir la expresión de proteína ribosómica S6 quinasa alfa-2. En una realización adicional, la etapa (b) comprende medir la expresión de Sialyl Lewis x. En una realización adicional, la etapa (b) comprende medir la expresión de proteína 2 adaptadora de transducción de señal. En una realización adicional, la etapa (b) comprende medir la expresión de enzima UBC9 conjugada con SUMO. En una realización adicional, la etapa (b) comprende medir la expresión del miembro 9 de la familia de dominio TBC. En una realización adicional, la etapa (b) comprende medir la expresión de péptido de transmembrana. En una realización adicional, la etapa (b) comprende medir la expresión de factor alfa de necrosis de tumor. En una realización adicional, la etapa (b) comprende medir la expresión de miembro 14 de la superfamilia de receptor de factor de necrosis de tumor. En una realización adicional, la etapa (b) comprende medir la expresión de proteína tirosina quinasa BTK. En una realización adicional, la etapa (b) comprende medir la expresión de proteína tirosina quinasa JAK3. En una realización adicional, la etapa (b) comprende medir la expresión de proteína tirosina fosfatasa tipo no receptor 1. En una realización adicional, la etapa (b) comprende medir la expresión de isozima L5 hidrolasa carboxilo terminal ubiquitina. En una realización adicional, la etapa (b) comprende medir la

expresión de enzima E2 C conjugada con ubiquitina. En una realización adicional, la etapa (b) comprende medir la expresión de FIQTDK.

“Péptido de transmembrana” o “péptido TM” significa un péptido derivado de una proteína 10TM, para la que tiene especificidad la construcción de anticuerpo scFv de SEQ ID NO: 1 a continuación (en donde las secuencias CDR se indican en **negrita**, texto en *cursiva*):

MAEVQLLESGGGLVQPGGSLRLSCAASGFT**FSSYGFHWVRQAPGK**GLEWV**SLISWDG**
GSTYYADSVKGRFTISRDN SKNTLYLQMNSLRAEDTAVYYCARGTWFD PWGQGTLVT
 VSSGGGGSGGGGSGGGGSGQSVLTQPPSASGTPGQRVTISCS**GSSSNIGNNAV**VMWYQ

QLPGTAPKLLIY**RNNQRPS**GV PDRFSGSKSGTSASLAISGLRSEDEADYY**CAAWDDSL**
SWVFGGGTKLTVLG

[SEQ ID NO: 1]

Por lo tanto, puede utilizarse este scFv o cualquier anticuerpo, o fragmento de unión a antígeno del mismo, que compita con este scFv por la unión con la proteína 10TM. Por ejemplo, el anticuerpo, o fragmento de unión a antígeno del mismo, que compita con las mismas CDR que las presentes en SEQ ID NO: 1.

Las personas expertas en la materia podrán apreciar que dicho anticuerpo puede producirse con una etiqueta de afinidad (p.ej., en el término C) para fines de purificación. Por ejemplo, se puede utilizar una etiqueta de afinidad de SEQ ID NO: 2, a continuación:

DYKDHDGDYKDHDIDYKDDDDKAAAHHHHHH
 [SEQ ID NO: 2]

En una realización adicional, los dos o más marcadores medidos en la etapa (b) comprenden HADH2 y uno o más biomarcadores seleccionados del grupo que consiste en: Angiotina, Apo-A1, Apo-A4, ATP-5B, BTK, inh C1, C1q, C3, C5, CD40, CD40L, Cistatina C, Eotaxina, Factor B, GAK, GM-CSF, IL-11, IL-13, IL-3, IL-4, IL-6, IL-8, IL-9, KSYK-1, LDL, MAPK1, MCP-1, PTP-1B, Sialyl Lewis x, Sox11A, TGFbeta1, TNF-alfa, TNFRSF3, UCHL5 y UPF3B.

En esta realización, el método puede servir para discriminar entre cáncer de páncreas (CP) y cáncer no pancreático (NC) y/o afecciones pancreáticas benignas (BC).

En una realización adicional, los dos o más biomarcadores medidos en la etapa (b) comprenden HADH2 y uno o más biomarcadores seleccionados del grupo que consiste en: Angiotina, ATP-5B, inh C1, C1q, C3, C5, CD40, Cistatina C, Eotaxina, Factor B, GAK, IL-11, IL-13, IL-6, IL-8, LDL y TNF-alfa.

En esta realización, el método puede servir para discriminar entre cáncer pancreático (CP) y cáncer no pancreático (NC).

En una realización adicional, los dos o más biomarcadores medidos en la etapa (b) comprenden HADH2 y uno o más biomarcadores seleccionados del grupo que consiste en: Apo-A1, Apo-A4, BTK, C1 inh, C5, CD40L, CIMS, Factor B, GM-CSF, HADH2, IL-3, IL-4, IL-4, IL-9, KSYK-1, MAPK1, MCP-1, PTP-1B, Sialyl Lewis x, Sox11A, TGF-beta1, TNFRSF3, UCHL5 y UPF3B.

En esta realización, el método puede servir para discriminar entre cáncer de páncreas (PC), y afecciones pancreáticas benignas (BC).

En una realización adicional, el método comprende la etapas (a) y (b), opcionalmente, la etapas (c) y (d), opcionalmente, la etapas (e) y (f), y la etapa adicional de:
 g) determinar la presencia de cáncer de páncreas sobre la base de la expresión de los dos o más biomarcadores medidos en la etapa (b).

En una realización adicional, el método se basa en las tendencias (regulación al alza y a la baja) que se identifican en la Tabla 2.

En una realización adicional, la etapa (b) y cuando están presentes las etapas (d) y (f) se realizan poniendo en contacto la muestra que se va a analizar con la fracción de unión para los dos o más biomarcadores medidos en la etapa (b), por ejemplo, una primera fracción de unión, tal como se define a continuación.

En una realización adicional, el método comprende o consiste en el uso *in vitro* de los biomarcadores enumerados

en la Tabla 3.

Generalmente, el diagnóstico se realiza con una ROC AUC de al menos 0,55, por ejemplo con una ROC AUC de al menos 0,60, 0,65, 0,70, 0,75, 0,80, 0,85, 0,90, 0,95, 0,96, 0,97, 0,98, 0,99 o con una ROC AUC de 1,00. Preferentemente, el diagnóstico se realiza con una ROC AUC de al menos 0,85, y lo más preferente con una ROC AUC de 1.

Normalmente, el diagnóstico se realiza utilizando una máquina de vector de soporte (MVS), como por ejemplo las que están disponibles por <http://cran.r-project.org/web/packages/e1071/index.html> (p.ej. e1071 1.5-24). Sin embargo, puede utilizarse cualquier otro medio adecuado.

Las máquinas de vector de soporte (MVS) son un conjunto de métodos de aprendizaje supervisados para clasificación y regresión. Dado un conjunto de ejemplos de entrenamiento, cada uno de ellos marcado como perteneciente a una o dos categorías, un algoritmo de entrenamiento de MVS construye un modelo que predice si un nuevo ejemplo entra en una categoría u otra. Intuitivamente, un modelo de MVS es una representación de los ejemplos ya que apunta en el espacio, cartografiado para que los ejemplos de las categorías por separado se dividan con un hueco claro que es lo más ancho posible. Se cartografían nuevos ejemplos después en ese mismo espacio y se predice que pertenecen a una categoría sobre la base de en qué lado del hueco entran.

Más formalmente, una máquina de vector de soporte construye un hiperplano o un conjunto de hiperplanos en un espacio dimensional alto o infinito, que se puede utilizar para clasificación, regresión u otras tareas. Intuitivamente, se consigue una buena separación mediante el hiperplano que tiene la distancia máxima para los puntos de datos de entrenamiento más cercanos de cualquier clase (llamado margen funcional) ya que, en general, cuanto mayor es el margen, menor es el error de generalización del clasificador. Para más información sobre las MVS, véase por ejemplo, Burges, 1998, *Data Mining y Knowledge Discovery*, 2:121-167.

Las MVS pueden "entrenarse" antes de realizar los métodos de la invención utilizando perfiles de biomarcador de individuos con una patología conocida (por ejemplo individuos que se sabe que tienen cáncer de páncreas, individuos que se sabe que tienen pancreatitis inflamatoria aguda, individuos que se sabe que tienen pancreatitis crónica o individuos que se sabe que están sanos). Al introducir en el ciclo las muestras de entrenamiento, la MVS es capaz de aprender que los perfiles de biomarcador están asociados con cáncer de páncreas. Una vez completado el proceso de entrenamiento, la SVM es capaz entonces de distinguir si una muestra de biomarcador analizada es de un individuo con cáncer de páncreas o no.

Sin embargo, este procedimiento de entrenamiento puede obviarse pre-programando la SMV con los parámetros de entrenamiento necesarios. Por ejemplo, pueden realizarse diagnósticos de acuerdo con parámetros de SVM conocidos utilizando el algoritmo de SVM detallado en la Tabla B, sobre la base de la medición de cualquiera o todos los biomarcadores enumerados en la Tabla A.

Las personas expertas en la materia podrán apreciar que pueden determinarse parámetros de MVS adecuados para cualquier combinación de los biomarcadores enumerados en la Tabla A entrenando una máquina MVS con la selección de datos apropiada (es decir, las mediciones de biomarcador de individuos que se sabe que tienen una patología del páncreas conocida). Alternativamente, los datos de la Tabla 2 y 3 pueden utilizarse para determinar una patología asociada a un cáncer de páncreas en particular, de acuerdo con cualquier otro método estadístico conocido en la técnica.

Preferentemente, el método divulgado tiene una precisión de al menos 60 %, por ejemplo 61 %, 62 %, 63 %, 64 %, 65 %, 66 %, 67 %, 68 %, 69 %, 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de precisión.

Preferentemente, el método divulgado tiene una sensibilidad de al menos 60 %, por ejemplo 61 %, 62 %, 63 %, 64 %, 65 %, 66 %, 67 %, 68 %, 69 %, 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de sensibilidad.

Preferentemente, el método divulgado tiene una especificidad de al menos 60 %, por ejemplo 61 %, 62 %, 63 %, 64 %, 65 %, 66 %, 67 %, 68 %, 69 %, 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de especificidad.

Por "Precisión", los autores de la invención se refieren a la proporción de resultados correctos de un método, por "sensibilidad" se refieren a la proporción de todas las muestras PaC positivas que se clasifican correctamente como positivos y por "especificidad" se refieren a la proporción de todas las muestras de PaC negativas que se han clasificado correctamente como negativas.

En una realización adicional, el individuo no afligido por cáncer de páncreas no está afligido por cáncer de páncreas (PaC), pancreatitis crónica (ChP) o pancreatitis inflamatoria aguda (AIP). Más preferentemente, el individuo sano no está afligido por ninguna enfermedad o afección pancreática. Incluso más preferentemente, el individuo que no está afligido con cáncer de páncreas no está afligido con ninguna otra enfermedad o afección. Lo más preferente, el individuo que no está afligido con cáncer de páncreas es un individuo sano. Dentro de "individuo sano" se incluye a los individuos que la persona experta en la materia considera como físicamente vigoroso y sin una enfermedad física.

Sin embargo, en otra realización, el individuo que no está afligido por cáncer de páncreas está afligido por pancreatitis crónica. En otra realización más, el individuo que no está afligido por cáncer de páncreas está afligido por pancreatitis inflamatoria aguda.

Tal como se ha mencionado anteriormente, el método de la presente invención sirve para determinar la presencia de cáncer de páncreas en un individuo. En una realización alternativa o adicional, el cáncer de páncreas se selecciona del grupo que consiste en adenocarcinoma, carcinoma adenoescamoso, carcinoma de célula en anillo de signet, carcinoma hepatoide, carcinoma coloide, carcinoma no diferenciado y carcinomas no diferenciados con células gigantes de tipo osteoclasto. Preferentemente, el cáncer de páncreas es un adenocarcinoma pancreático. Más preferentemente, el cáncer de páncreas es un adenocarcinoma ductal pancreático, también conocido como cáncer de páncreas exocrino.

En otra realización más, la etapa (b), (d) y/o la etapa (f) se realizan utilizando primeros agentes de unión capaces de unirse a dos o más biomarcadores (es decir, utilizando dos o más primeros agentes de unión, en donde cada agente de unión es capaz de unirse específicamente a uno de los dos o más biomarcadores). Las personas expertas en la materia apreciarán que el primer agente de unión puede comprender o consistir en una sola especie con especificidad para uno de los biomarcadores de proteína o una pluralidad de especies diferentes, cada una de ellas para un biomarcador de proteína diferente.

Los agentes de unión adecuados (también denominados moléculas de unión) se pueden seleccionar de una biblioteca, sobre la base de su capacidad para unirse a un motivo dado, tal como se explica más adelante.

A menos un tipo de los agentes de unión, y más normalmente todos los tipos, pueden comprender o consistir en un anticuerpo o fragmento de unión a antígeno del mismo, o una variante del mismo.

Los métodos para la producción y uso de anticuerpos son conocidos dentro de la técnica, como por ejemplo, véase *Antibodies: A Laboratory Manual*, 1988, Harlow & Lane, Cold Spring Harbor Press, ISBN-13: 978-0879693145, *Using Antibodies: A Laboratory Manual*, 1998, Harlow & Lane, Cold Spring Harbor Press, ISBN-13: 978-0879695446 y *Making y Using Antibodies: A Practical Handbook*, 2006, Howard & Kaser, CRC Press, ISBN-13: 978-0849335280.

Por tanto, un fragmento puede contener uno o más de los dominios variable pesado (V_H) o variable ligero (V_L). Por ejemplo, el término fragmento de anticuerpo incluye moléculas de tipo Fab (Better et al (1988) *Science* **240**, 1041); moléculas F_v (Skerra et al (1988) *Science* **240**, 1038); moléculas F_v monocaténarias ($ScFv$) en donde los dominios de pareja V_H y V_L están unidos a través de un oligopéptido flexible (Bird et al (1988) *Science* **242**, 423; Huston et al (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5879) y anticuerpos de dominio simple (dAbs) que comprenden dominios V aislados (Ward et al (1989) *Nature* **341**, 544).

El término "variante de anticuerpo" incluye cualquier anticuerpo sintético, anticuerpos recombinantes o híbridos de anticuerpo, como por ejemplo, pero sin limitarse a ellos, una molécula de anticuerpo monocaténaria producida por despliegue de fagos de las regiones variables y/o constantes de cadena ligera y/o pesada de inmunoglobulina, u otra molécula inmunointeractiva capaz de unirse a un antígeno en un formato de inmunoensayo conocido entre los expertos en la materia.

Se puede encontrar una revisión general de las técnicas relacionadas con la síntesis de fragmentos de anticuerpo que retienen sus sitios de unión específicos en Winter & Milstein (1991) *Nature* **349**, 293-299.

En los métodos de la invención pueden utilizarse bibliotecas moleculares, como las bibliotecas de anticuerpos (Clackson et al, 1991, *Nature* **352**, 624-628; Marks et al, 1991, *J Mol Biol* **222**(3): 581-97), bibliotecas de péptidos (Smith, 1985, *Science* **228** (4705): 1315-7), bibliotecas de ADNc expresado (Santi et al (2000) *J Mol Biol* **296**(2): 497-508), bibliotecas de otros andamios distintos al marco de anticuerpo como afficuerpos (Gunneriusson et al, 1999, *Appl Environ Microbiol* **65**(9): 4134-40) o bibliotecas basadas en aptámeros (Kenan et al, 1999, *Methods Mol Biol* **118**, 217-31) como fuente para la selección de moléculas de unión que son específicas para un motivo dado.

Las bibliotecas moleculares pueden expresarse *in vivo* en células procariotas (Clackson et al, 1991, *op. cit.*; Marks et al, 1991, *op. cit.*) o eucariotas (Kieck et al, 1999, *Proc Natl Acad Sci EE.UU.*, **96**(10):5651-6) o pueden expresarse *in vitro* sin participación de células (Hanes & Pluckthun, 1997, *Proc Natl Acad Sci EE.UU.* **94**(10):4937-42; He & Taussig, 1997, *Nucleic Acids Res* **25**(24):5132-4; Nemoto et al, 1997, *FEBS Lett*, **414**(2):405-8).

En los casos en los que se utilizan bibliotecas a base de proteínas, generalmente, los genes que codifican las bibliotecas de moléculas de unión posibles están empaquetados en virus y la molécula de unión posible se despliega en la superficie del virus (Clackson *et al*, 1991, *op. cit.*; Marks *et al*, 1991, *op. cit.*; Smith, 1985, *op. cit.*).

El sistema cuyo uso está más extendido hoy en día es bacteriofago que despliega fragmentos de anticuerpo en su superficie, expresándose los fragmentos de anticuerpo como una fusión de la proteína de revestimiento menor del bacteriofago (Clackson *et al*, 1991, *op. cit.*; Marks *et al*, 1991, *op. cit.*). Sin embargo, también se han empleado otros sistemas para despliegue utilizando otros virus (EP 39578), bacterias (Gunneriusson *et al*, 1999, *op. cit.*; Daugherty *et al*, 1998, *Protein Eng* 11(9):825-32; Daugherty *et al*, 1999, *Protein Eng* 12(7):613-21) y levadura (Shusta *et al*, 1999, *J Mol Biol* 292(5):949-56).

Además, los sistemas de despliegue se han desarrollado utilizando la unión del producto de polipéptido con su ARNm codificante en lo que se conoce como sistemas de despliegue (Hanes & Pluckthun, 1997, *op. cit.*; He & Taussig, 1997, *op. cit.*; Nemoto *et al*, 1997, *op. cit.*), o alternativamente, la unión del producto de polipéptido con el ADN codificante (véase patente estadounidense No. 5.856.090 y patente internacional WO 98/37186).

Cuando se seleccionan las posibles moléculas de unión de bibliotecas, normalmente se emplean uno o más péptidos seleccionadores que tienen motivos definidos. Se pueden utilizar restos de aminoácido que proporcionan estructura, menor flexibilidad en el péptido o cadenas laterales cargadas, polares o hidrófobas que permiten la interacción con la molécula de unión, en el diseño de los motivos de los péptidos seleccionadores.

Por ejemplo:

- (i) Prolina puede estabilizar una estructura de péptido cuando su cadena lateral está unida tanto al carbono alfa como al nitrógeno;
- (ii) Fenilalanina, tirosina, triptófano tienen cadenas laterales aromáticas y son altamente hidrófobos, mientras que leucina e isoleucina tienen cadenas laterales alifáticas y también son hidrófobos;
- (iii) Lisina, arginina e histidina tienen cadenas laterales básicas y estarán positivamente cargadas a un pH neutro, mientras que aspartato y glutamato tienen cadenas laterales ácidas y estarán cargados negativamente a un pH neutro.
- (iv) Asparagina y glutamina son neutros a un pH neutro pero contienen un grupo amida que puede participar en las uniones de hidrógeno;
- (v) Las cadenas laterales de serina, treonina y tirosina contienen grupos hidroxilo, que pueden participar en enlaces de hidrógeno.

Normalmente, la selección de los agentes de unión puede implicar el uso de un conjunto de tecnologías y sistemas de matriz para analizar la unión por puntos que corresponde a los tipos de moléculas de unión.

En una realización alternativa o adicional, el (los) primer(os) agente(s) de unión está(n) inmovilizados en una superficie (p.ej., en una matriz o placa de varios pocillos).

Los dominios variable pesado (V_H) y variable ligero (V_L) del anticuerpo participan en el reconocimiento de antígeno, un hecho reconocido primero a través de experimentos de digestión de proteasa. Se observó una posterior confirmación por "humanización" de anticuerpos de roedores. Los dominios variables de origen roedor pueden fusionarse con dominios constantes de origen humano de manera que el anticuerpo resultante retenga la especificidad antigénica del anticuerpo parental roedor (Morrison *et al* (1984) *Proc. Natl. Acad. Sci. EE.UU.* **81**, 6851-6855).

Se sabe que la especificidad antigénica se confiere mediante dominios variables y es independiente de los dominios constantes por los experimentos que implican la expresión bacteriana de fragmentos de anticuerpo, que contienen uno o más dominios variables. Estas moléculas incluyen moléculas de tipo Fab (Better *et al* (1988) *Science* **240**, 1041); moléculas Fv (Skerra *et al* (1988) *Science* **240**, 1038); moléculas Fv monocatenarias (ScFv) en donde los dominios emparejados V_H y V_L están unidos a través de un oligopéptido flexible (Bird *et al* (1988) *Science* **242**, 423; Huston *et al* (1988) *Proc. Natl. Acad. Sci. EE.UU.*, **85**, 5879) y anticuerpos de dominio simple (dAbs) que comprenden dominios V aislados (Ward *et al* (1989) *Nature* **341**, 544). Una revisión general de las técnicas relacionadas con la síntesis de fragmentos de anticuerpos que retienen sus sitios de unión específicos se puede encontrar en Winter & Milstein (1991) *Nature* **349**, 293-299.

Por "moléculas ScFv" los autores de la invención se refieren a moléculas en las que los dominios emparejados V_H y V_L están unidos a través de un oligopéptido flexible.

Las ventajas de emplear fragmentos de anticuerpo en lugar de anticuerpos enteros son múltiples. El pequeño tamaño de los fragmentos puede conllevar propiedades farmacológicas mejoradas, como una mejor penetración de tejido sólido. Las funciones de efector de anticuerpos enteros, como la unión a complemento, se eliminan. Los fragmentos de anticuerpo Fab, Fv, ScFv y dAb pueden expresarse y secretarse en *El coli*, y desde él, permitiendo así la fácil producción de grandes cantidades de dichos fragmentos.

Los anticuerpos enteros, y los fragmentos F(ab')₂ son "bivalentes". Se entiende por "bivalente" que dichos anticuerpos y fragmentos F(ab')₂ tienen dos sitios de combinación de antígeno. En contraste, los fragmentos Fab, Fv, ScFv y dAb son monovalentes, y tienen solamente un sitio de combinación de antígeno.

- 5 Los anticuerpos pueden ser monoclonales o policlonales. Los anticuerpos monoclonales adecuados pueden prepararse a través de técnicas conocidas, como por ejemplo, las divulgadas en "Monoclonal Antibodies: A manual of techniques", H Zola (CRC Press, 1988) , así como en "Monoclonal Hybridoma Antibodies: Techniques and applications", J G R Hurrell (CRC Press, 1982).

- 10 En una realización alternativa o adicional, el primer agente de unión se inmoviliza en una superficie (p.ej., en una matriz o placa de varios pocillos).

Por lo tanto, el primer agente de unión puede comprender o consistir en un anticuerpo o un fragmento de unión a antígeno del mismo. Preferentemente, el anticuerpo o fragmento de unión a antígeno del mismo es un anticuerpo recombinante o un fragmento de unión a antígeno del mismo. El anticuerpo o fragmento de unión a antígeno del mismo se puede seleccionar del grupo que consiste en: scFv, Fab y un dominio de unión de una molécula de inmunoglobulina.

- 15 El primer agente de unión puede inmovilizarse en una superficie.

- 20 Los dos o más biomarcadores en la muestra de ensayo pueden etiquetarse con una fracción detectable. Se incluye en el significado de "fracción detectable" que la fracción es aquella que puede detectarse y determinarse la cantidad relativa y/o emplazamiento de la fracción (por ejemplo, el emplazamiento en una matriz).

- 25 Las fracciones detectables adecuadas son muy conocidas en la técnica.

Por lo tanto, la fracción detectable puede ser una fracción fluorescente y/o luminiscente y/o quimioluminiscente que puede detectarse cuando se expone en condiciones específicas. Por ejemplo, es posible que sea necesario exponer una fracción fluorescente a radiación (es decir luz) a una longitud de onda y una intensidad específicas para causar la excitación de la fracción fluorescente, permitiendo así que emita fluorescencia detectable a una longitud de onda específica que se pueda detectar.

- 30 Alternativamente, la fracción detectable puede ser una enzima que es capaz de convertir un sustrato (preferentemente no detectable) en un producto detectable que se puede visualizar y/o detectar. Entre los ejemplos de enzimas adecuadas se incluyen las explicadas con más detalle más adelante en relación con los ensayos ELISA.

- 35 Alternativamente, la fracción detectable puede ser un átomo radioactivo que es útil en la obtención de imágenes. Entre los átomos radioactivos adecuados se incluyen Tc^{99m} e I¹²³ para estudios escintográficos. Otras fracciones detectables incluyen por ejemplo, marcadores de espín para imagen por resonancia magnética (IRM) como I¹²³ de nuevo, I¹³¹, In¹¹¹, F¹⁹, C¹³, N¹⁵, O¹⁷, gadolinio, magnesio o hierro. Claramente, el agente que se va a detectar (por ejemplo, el uno o más biomarcadores en la muestra de ensayo y/o la muestra de control descrita en el presente documento y/o la molécula de anticuerpo para su uso en la detección de una proteína seleccionada) debe tener suficientes isótopos atómicos apropiados para poder detectar fácilmente la fracción detectable.

- 40 Los radio marcadores y otros marcadores pueden incorporarse en los agentes utilizados en la invención (es decir, las proteínas presentes en las muestras de los métodos de la invención y/o los agentes de unión utilizados en la invención) de maneras conocidas. Por ejemplo, si la fracción de unión es un polipéptido, puede biosintetizarse o se puede sintetizar por síntesis de aminoácidos química empleando precursores de aminoácido adecuados lo cual implica por ejemplo, flúor-19 en lugar de hidrógeno. Se pueden unir marcadores como Tc^{99m}, I¹²³, Rh¹⁸⁶, Rh¹⁸⁸ e In¹¹¹, por ejemplo, a través de restos de cisteína en la fracción de unión. Se puede unir Itrio-90 a través de un resto de lisina. Puede emplearse el método IODOGEN (Fraker et al (1978) Biochem. Biophys. Res. Comm. 80, 49-57) para incorporar I¹²³. La referencia ("Monoclonal Antibodies in Immunoscintigraphy", J-F Chatal, CRC Press, 1989) describe otros métodos en detalle. Los métodos para conjugar otras fracciones detectables (como por ejemplo fracciones enzimáticas, fluorescentes, luminiscentes, quimioluminiscentes o radiactivas) a proteínas son muy conocidos en la técnica.

- 45 Preferentemente, los dos o más biomarcadores en la(s) muestra(s) de control están marcados con una fracción detectable. La fracción detectable puede seleccionarse del grupo que consiste en una fracción fluorescente; una fracción luminiscente; una fracción quimioluminiscente; una fracción radioactiva; una fracción enzimática. Sin embargo, es preferente que la fracción detectable sea biotina.

- 50 En una realización adicional la etapa (b), (d) y/o la etapa (f) se realiza empleando un ensayo que comprende segundos agentes de unión capaces de unirse a los dos o más biomarcadores, comprendiendo el segundo agente de unión una fracción detectable. Preferentemente, el segundo agente de unión comprende o consiste en un anticuerpo o fragmento de unión a antígeno del mismo. Preferentemente, el anticuerpo o fragmento de unión a antígeno del mismo es un anticuerpo recombinante o fragmento de unión a antígeno del mismo. Lo más preferente,

el anticuerpo o fragmento de unión a antígeno del mismo se selecciona del grupo que consiste en: scFv, Fab y un dominio de unión de una molécula de inmunoglobulina. En una realización alternativa o adicional, la fracción detectable se selecciona del grupo que consiste en: una fracción fluorescente; una fracción luminiscente; una fracción quimioluminiscente; una fracción radioactiva y una fracción enzimática. Preferentemente, la fracción detectable es una fracción fluorescente (por ejemplo un colorante Alexa Fluor, p.ej., Alexa647).

En un caso alternativo o adicional, el método del primer aspecto de la invención puede llevarse a cabo por medio de un ensayo ELISA (Ensayo inmunoabsorbente ligado a enzima). Los ensayos preferentes para detectar proteínas en suero o plasma incluyen ensayo inmunoabsorbente ligado a enzima (ELISA), radioinmunoensayo (RIA), ensayos inmunoradiométricos (IRMA) y ensayos inmunoenzimáticos (IEMA), incluyendo ensayos sándwich utilizando anticuerpos monoclonales y/o policlonales. En David et al en las patentes estadounidenses Nos. 4.376.110 y 4.486.530, se describen ejemplos de ensayos sándwich. Se puede utilizar la tinción con células de anticuerpos en portaobjetos en los métodos perfectamente conocidos de los análisis de diagnóstico de laboratorio citológicos, tal como conocen las personas expertas en la materia.

Normalmente, el ensayo en ELISA (ensayo inmunoabsorbente ligado a enzimas) que implica normalmente el uso de enzimas que dan un producto de reacción con color, normalmente en ensayos en fase sólida. Está muy extendido el uso de enzimas como peroxidasa de rábano picante rusticano y fosfatasa. Una forma de amplificar la reacción de fosfatasa es con el uso de NADP como sustrato para generar NAD que actúa ahora como una coenzima para un segundo sistema de enzima. La pirofosfatasa de *Escherichia coli* proporciona un buen conjugado ya que la enzima no está presente en tejidos y es estable y da un buen color de reacción. También pueden utilizarse sistemas quimioluminiscentes basados en enzimas como luciferasa.

Los métodos ELISA son muy conocidos en la técnica, véase por ejemplo la Guía ELISA (Methods in Molecular Biology), 2000, Crowther, Humana Press, ISBN-13: 978-0896037281.

La conjugación con la vitamina biotina se utiliza frecuentemente ya que se puede detectar fácilmente a través de su detección con avidina ligada a enzima o estreptavidina con la que se une con gran especificidad y afinidad.

Sin embargo, se realiza alternativamente la etapa (b), (d) y/o la etapa (f) utilizando una matriz. Las matrices en sí son conocidas dentro de la técnica. Normalmente, se forman de una estructura lineal o bidimensional que tiene regiones separadas por un espacio (es decir, diferenciadas) ("puntos") que tienen cada una de ellas un área finita formada en la superficie de un soporte sólido. Una matriz también puede ser una estructura de perla en la que se puede identificar cada perla a través de un código molecular o un código de color o se puede identificar en un flujo continuo. El análisis puede realizarse también secuencialmente pasándose la muestra sobre una serie de puntos que adsorben cada uno de ellos la clase de moléculas de la solución. El soporte sólido es normalmente vidrio o un polímero, siendo los polímeros más comúnmente utilizados celulosa, poliácridamida, nilón, poliestireno, policloruro de vinilo, polipropileno. Los soportes sólidos pueden presentarse en forma de tubos, perlas, discos, copos de silicio, microplacas, membrana de polidifluoruro de vinilideno (PVDF), membrana de nitrocelulosa, membrana de nilón, otras membranas porosas, membrana no porosa (p.ej., plástico, polímero, perspex, silicio, entre otros), una pluralidad de pasadores poliméricos o una pluralidad de pocillos de microtitulación, o cualquier otra superficie adecuada para inmovilizar proteínas, polinucleótidos y otras moléculas adecuadas y/o llevar a cabo un inmunoensayo. Los procesos de unión son muy conocidos dentro de la técnica y consisten generalmente en reticular, unir covalentemente o adsorber físicamente una molécula de proteína, polinucleótido o similar a un soporte sólido. Al utilizar técnicas muy conocidas, como impresión con contacto o sin contacto, enmascaramiento o fotolitografía, se puede definir el emplazamiento de cada punto. Para una revisión véase Jenkins, R.E., Pennington, S.R. (2001, Proteomics, 2, 13-29) y Lal et al (2002, Drug Discov Today 15; 7(18 Suppl):S143-9).

Normalmente, la matriz es una micromatriz. El significado de "micromatriz" incluye una matriz de regiones que tienen una densidad de regiones diferenciadas de menos de aproximadamente 100/cm² y, preferentemente, al menos aproximadamente 1000/cm². Las regiones en una micromatriz tienen dimensiones típicas, p.ej., diámetros, en el intervalo entre aproximadamente 10-250 µm y están separadas de otras regiones en la matriz por aproximadamente la misma distancia. La matriz puede ser también una micromatriz o una nanomatriz.

Una vez identificadas y aisladas las moléculas de unión adecuadas (tal como se ha explicado), la persona experta en la materia puede fabricar una matriz aplicando los métodos conocidos en la técnica de biología molecular.

Por lo tanto, la matriz puede ser, la matriz es, una matriz a base de perlas, o una matriz de base superficial. Preferentemente, la matriz se selecciona del grupo que consiste en una macromatriz, una micromatriz y una nanomatriz.

En una realización adicional, el método de acuerdo con el primer aspecto de la invención comprende:

- (i) marcar los biomarcadores presentes en la muestra con biotina;
- (ii) contacto de las proteínas marcadas con biotina con una matriz que comprende una pluralidad de scFv inmovilizados en emplazamientos diferenciados en su superficie, teniendo scFv especificidad para dos o más de

las proteínas de la Tabla III, incluyendo HADH2;

(iii) contacto de scFV inmovilizado con un conjugado de estreptavidina que comprende un colorante fluorescente;

y

(iv) detección de la presencia del colorante en emplazamientos diferenciados en la superficie de matriz

5 en donde la expresión del colorante en la superficie de la matriz es indicativo de la expresión de un biomarcador de la Tabla A en la muestra.

10 En una realización alternativa, la etapa (b), (d) y/o (f) comprende medir la expresión de una molécula de ácido nucleico que codifica los dos o más biomarcadores. Preferentemente la molécula de ácido nucleico es una molécula de ADNc o una molécula de ARNm. Lo más preferente, la molécula de ácido nucleico es una molécula de ARNm.

15 Por lo tanto la expresión de los dos o más biomarcadores en la etapa (b), (d) y/o (f) puede realizarse utilizando un método seleccionado del grupo que consiste en hibridación Southern, hibridación Northern, reacción en cadena de la polimerasa (PCR), PCR de transcriptasa inversa (RT-PCR), PCR en tiempo real cuantitativa (qRT-PCR), nanomatriz, micromatriz, macromatriz, autorradiografía e hibridación *in situ*. Preferentemente, la expresión del uno o más biomarcadores en la etapa (b) se determina utilizando una micromatriz de ADN.

20 En una realización alternativa o adicional, la medición de la expresión de los dos o más biomarcadores en la etapa (b), (d) y/o (f) se lleva a cabo utilizando dos o más fracciones de unión, capaz cada una de ellas de unirse selectivamente a una molécula que codifica un de los biomarcadores identificados en la Tabla A.

25 En otra realización más, la una o más fracciones de unión comprende o consiste cada una de ellas en una molécula de ácido nucleico. Por tanto, cada una o más de las fracciones de unión pueden comprender o consistir en ADN, ARN, PNA, LNA, GNA, TNA o PMO. Sin embargo, preferentemente cada una o más de las fracciones de unión comprende o consiste en ADN.

30 Preferentemente, la una o más de las fracciones de unión tienen de 5 a 100 nucleótidos de longitud. Más preferentemente, la una o más moléculas de ácido nucleico es de 15 a 35 nucleótidos de longitud. Más preferentemente, la fracción de unión comprende una fracción detectable.

35 En una realización adicional, la fracción detectable se selecciona del grupo que consiste en: una fracción fluorescente; una fracción luminiscente; una fracción quimioluminiscente; una fracción radioactiva (por ejemplo, un átomo radioactivo) y una fracción enzimática. Preferentemente, la fracción detectable comprende o consiste en un átomo radioactivo. El átomo radioactivo puede seleccionarse del grupo que consiste en tecnecio-99m, yodo-123, yodo-125, yodo-131, indio-111, flúor-19, carbono-13, nitrógeno-15, oxígeno-17, fósforo-32, azufre-35, deuterio tritio, renio-186, renio-188 e itrio-90.

40 Sin embargo, la fracción detectable de la fracción de unión puede ser una fracción fluorescente (por ejemplo, un colorante Alexa Fluor, *p.ej.* Alexa647).

45 En una realización adicional, la mezcla que se proporciona en la etapa (b), (d) y/o (f) se selecciona del grupo que consiste en sangre no fraccionada, plasma, suero, fluido de tejido, tejido pancreático, jugo pancreático, bilis y orina. Preferentemente, la muestra proporcionada en la etapa (b), (d) y/o (f) se selecciona del grupo que consiste en sangre no fraccionada, plasma y suero. Más preferentemente, la muestra proporcionada en la etapa (b), (d) y/o (f) es plasma. En otra realización preferente, la muestra proporcionada en la etapa (b), (d) y/o (f) es suero.

50 En una realización alternativa, el método del primer aspecto de la invención es para diferenciar cáncer de cuerpo y/o cola del páncreas del cáncer de la cabeza del páncreas que comprende o consiste en las etapas de:

a) proporcionar una muestra para su análisis obtenida de un individuo;

b) determinar una firma de biomarcador de la muestra de ensayo midiendo la expresión en la muestra de ensayo de al menos dos biomarcadores seleccionados del grupo definido en la Tabla A (i), (ii), (iv) y/o (vi), incluyendo HADH2;

55 en donde la expresión en la muestra de ensayo del al menos un biomarcador seleccionado del grupo definido en la Tabla A (i), (ii) y/o (iii) es indicativo de la presencia de cáncer del cuerpo y/o la cola del páncreas o cáncer de la cabeza del páncreas. La muestra de ensayo puede haberse identificado ya como una muestra de cáncer pancreático (por ejemplo, aplicando el método que se ha descrito). Alternativamente, la muestra de ensayo puede no haber sido identificada como muestra de cáncer pancreático. Las muestras de control positivo y negativo se pueden seleccionar apropiadamente.

65 En una realización alternativa, se trata apropiadamente al paciente tras la identificación del cáncer de páncreas. Por ejemplo, pueden extirparse quirúrgicamente el (los) tumor(es), tratarse por quimioterapia (es decir fármacos anti-neoplásicos) y/o tratarse con radioterapia. Por tanto, la presente invención incluye un método de tratamiento de una persona que tiene cáncer de páncreas, en donde se identifica al paciente por tener cáncer de páncreas aplicando el

método del primer aspecto de la invención.

Se describe asimismo en el presente documento una matriz para determinar la presencia de cáncer de páncreas en un individuo, tal como se define en las reivindicaciones del presente documento.

Las matrices adecuadas para su uso en los métodos de la invención se han explicado anteriormente.

Preferentemente, los dos o más agentes de unión son capaces de unirse a todos los biomarcadores definidos en la Tabla A.

Un segundo aspecto de la presente invención proporciona el uso de dos o más biomarcadores (incluyendo HADH2) seleccionados del grupo que se define en el primer aspecto de la invención como marcador de diagnóstico para determinar la presencia de cáncer de páncreas en un individuo *in vitro*, tal como se define en las reivindicaciones del presente documento. Preferentemente, todas las proteínas definidas en la Tabla III se utilizan como marcadores de diagnóstico para determinar la presencia de cáncer de páncreas en un individuo.

Asimismo, se describe en el presente documento un kit para determinar la presencia de cáncer de páncreas que comprende:

A) uno o más agentes de unión o una matriz tal como se describe en el presente documento; y

B) instrucciones para realizar el método de acuerdo con el primer aspecto de la invención.

Se describe asimismo en el presente documento el uso de uno o más agentes/fracciones de unión seleccionadas del grupo que se define en el primer aspecto de la invención para determinar la presencia de cáncer de páncreas en un individuo. Preferentemente, los agentes/fracciones de unión para todas las proteínas definidas en la Tabla III se utilizan como marcadores de diagnóstico para determinar la presencia de cáncer de páncreas en un individuo.

A continuación, se describirán ejemplos no exhaustivos preferentes que representan ciertos aspectos de la invención haciendo referencia a las siguientes Tablas y Figuras:

Figura 1. PCA con muestras coloreadas de acuerdo con el diagnóstico (rojo = PC, amarillo = BC, azul = NC). Los datos se han filtrado a $p = 1e-10$ (63 anticuerpos).

Figura 2. Análisis de eliminación retroactiva. Se registró en gráfico el error Kullback-Leibler (K-L) después de cada ronda de eliminación de anticuerpo en el primer grupo de entrenamiento para para **(A)** PC frente a NC y **(B)** PC frente a BC. **(C)** Boxplot de los valores AUC generados a partir de modelos MVS de 25 anticuerpos en diez pares diferentes de grupos de ensayo/entrenamiento.

Figura 3. PCA con muestras coloreadas de acuerdo con el emplazamiento del tumor (rojo = tumores de cabeza, amarillo = tumores de cuerpo y cola, azul = controles normales). Se han filtrado los datos para $p = 1e-10$ (46 anticuerpos).

Figura 4. Imagen representativa de un portaobjetos de micromatriz con 13 submatrices representadas con A1-6 y B1-7, tras la hibridación y exploración de la muestra. Se aumenta la submatriz B4, presentando la disposición de matriz con 33 x 31 puntos. La matriz consiste en tres segmentos separados por filas impresas de BSA marcado (fila 1, 11, 21, 31). Se imprimió cada uno de los anticuerpos en tres puntos replicados, uno en cada segmento, distribuidos para diferentes posiciones dentro de cada segmento.

Figura 5. Reprocesamiento de los datos presentados como gráficos PCA. A) Datos en bruto coloreados de acuerdo con las cinco rondas (días) de análisis, representadas R1-R5. B) Datos normalizados que presentan que se ha eliminado la agregación sobre la base de las rondas de análisis.

Figura 6. Análisis de ruta de enfermedades según se identifica mediante los biomarcadores, sobre la base de los datos de un grupo completo de anticuerpos que corresponde a los valores p y número de veces de cambio de PC frente a NC.

EJEMPLOS

Materiales y métodos

Muestras

Después de firmar el consentimiento informado, se recogieron 338 muestras de suero de cinco sitios diferentes en España, de pacientes con cáncer de páncreas (PC, $n=156$), controles benignos (BC, $n=152$), y controles normales (NC, $n=30$) (Tabla 1). Se marcó todo el conjunto de muestras en una sola ocasión utilizando un protocolo optimizado previamente [32, 33]. Brevemente, se diluyeron muestras en bruto 1:45 en PBS, con el resultado de una

concentración de proteína aproximada de 2 mg/ml, y marcadas con un exceso molar 15:1 de biotina a proteína utilizando 0,6 mM EZ-Link Sulfo-NHS-LC-Biotin (Pierce, Rockford, IL, EE.UU.). Se eliminó la biotina no unida por diálisis contra PBS durante 72 horas. Se repartieron partes alícuotas de las muestras marcadas y se almacenaron a -20 °C.

5

Anticuerpos

Se utilizaron en total, 293 anticuerpos scFv recombinantes humanos dirigidos contra 98 antígenos de suero conocidos y 31 motivos de 4-6 aminoácidos (representados en el presente documento CIMS 1-31) [34], como contenido de las micromatrices de anticuerpo (véase Tabla 4 para la lista completa de anticuerpos utilizados). Se produjeron los anticuerpos en 15 ml de cultivos de *E. coli* y se purificaron del periplasma en 300 µl utilizando un sistema de purificación de proteínas MagneHis (Promega, Madison, WI, EE.UU.) y un robot KingFisher96 (Thermo Fisher Scientific, Waltham, MA, EE.UU.). Se intercambió el tampón de elución para PBS utilizando placas de espín desaladas de 96 pocillos Zeba (Pierce). Se midió el rendimiento de proteína utilizando NanoDrop (Thermo Scientific, Wilmington, DE, EE.UU.) y se comprobó la pureza utilizando SDS-PAGE 10 % (Invitrogen, Carlsbad, CA, EE.UU.).

15

Micromatrices de anticuerpo

Se produjeron las micromatrices de anticuerpo en portaobjetos MaxiSorp negros (NUNC, Roskilde, Dinamarca) utilizando una impresora sin contacto (SciFlexarrayer S11, Scienion, Berlín, Alemania). Se imprimieron trece submatrices idénticas en cada portaobjetos consistiendo cada matriz en 31 x 33 puntos, 130 µm de diámetro, con 200 µm de distancia del centro de punto a punto. Cada submatriz consistió en tres segmentos separados por filas impresas de BSA marcado (Fig. 4). Se imprimió cada anticuerpo en tres replicados, uno en cada segmento de submatriz, situado en diferentes posiciones para asegurar una reproducibilidad adecuada. Se imprimieron ocho portaobjetos, es decir 104 submatrices cada día durante cinco días. Se realizó la impresión durante toda la noche y se utilizaron inmediatamente las matrices para análisis de la matriz al día siguiente.

25

Se montó cada portaobjetos en una junta de hibridación (Schott, Jena, Alemania) y se bloqueó con PBSMT (1 % (w/v) leche, 1 % (v/v) Tween-20 en PBS) durante una hora. Al mismo tiempo, se descongelaron partes alícuotas de muestras marcadas sobre hielo y se diluyeron 1:10 en PBSMT, en placas de 96 pocillos. Se lavaron los portaobjetos con PBST (0,05 % (v/v) Tween-20 en PBS) cuatro veces antes de añadir las muestras desde las placas de dilución. Se incubaron las muestras durante dos horas en una mesa basculante, se lavaron cuatro veces con PBST, se incubaron con 1 µg/ml de estreptavidina-Alexa en PBSMT durante una hora en una mesa basculante y se volvieron a lavar cuatro veces en PBST. Finalmente, se desmontaron los portaobjetos desde las cámaras de hibridación, se sumergieron rápidamente en dH₂O y se secaron bajo una corriente de N₂. Se exploraron inmediatamente los portaobjetos en un escáner de micromatriz confocal (PerkinElmer Life y Analytical Sciences, Wellesley, MA, EE.UU.) a una resolución de 10 µm, utilizando 60 % de ganancia de PMT y 90 % polvo de láser. Se cuantificaron las intensidades de señal utilizando el software ScanArray Express versión 4.0 (PerkinElmer Life and Analytical Sciences), utilizando la opción círculo fijo. Se utilizaron los valores de intensidad con substracción del fondo local para el análisis de datos.

30

35

40

Pre-procesamiento de datos

Se utilizó un valor promedio de tres puntos replicados, a no ser que ninguno de los replicados CV excediera 15 % desde el valor medio, en cuyo caso se descartó, utilizándose en su lugar el valor promedio de los dos replicados restantes. El CV replicado promedio fue 8,3 % ($\pm 5,5$ %). Se calculó la aplicación de un CV límite de 15 %, 70 % de los valores de los datos desde los tres puntos replicados y el 30 % restante desde dos replicados.

45

Para la evaluación de las estrategias de normalización y el análisis inicial de la varianza, se visualizaron los datos utilizando el análisis de componente principal 3D (Qlucore, Lund, Suecia). Se consideró una muestra (pancreatitis crónica) como atípica (apenas se obtuvieron señales) y se excluyó de los datos principales. El análisis de componente principal sobre datos en bruto 10 log demostró que no se pudieron encontrar diferencias entre i) el posicionamiento de la submatriz de muestra sobre el portaobjetos, ii) el género del paciente, iii) la edad del paciente y iv) el centro de recogida de la muestra de suero. Se pudieron observar diferencias sistemáticas menores entre los días del análisis (rondas 1-5, probablemente debido a pequeñas diferencias en la humedad durante la impresión de la matriz, en particular durante el día 1, véase Fig. 4), lo cual se neutralizó por normalización. Se normalizaron los datos en dos etapas. En primer lugar, se eliminaron las diferencias entre las rondas (días) del análisis utilizando una estrategia de media del grupo de sustrato. Se calculó la intensidad promedio de cada anticuerpo dentro de cada ronda de análisis y se sustrajo de los valores simples, centrando así los datos al centro. Se añadió la señal media global de cada anticuerpo a cada uno de los puntos de los datos correspondientes para evitar valores negativos. En segundo lugar, se gestionaron las diferencias entre una matriz y otra (p.ej., las diferencias de fluorescencia de fondo de la muestra inherentes (véase Fig. 4)) calculando un factor de escala para cada submatriz, sobre la base del 20 % de los anticuerpos con el CV más bajo, tal como se ha descrito anteriormente [23, 35]. La normalización de los datos se visualizó en gráficos PCA (Fig. 5).

55

60

65

Análisis de los datos

El análisis se basó en comparaciones de dos grupos de PC frente a NC y PC frente a BC utilizando PCA, agrupación jerárquica, prueba t de Student y número de veces de cambio, así como ANOVA multigrupo (PC, NC y CP) (Qlucore). Se llevó a cabo el análisis de máquina de vector de soporte (MVS) en R (www.r-project.org) utilizando un núcleo lineal con el coste de limitaciones establecidas en uno (valor por defecto). Se dividieron aleatoriamente los datos en grupos de ensayo y de entrenamiento, con dos terceras partes de las muestras de cada grupo en el grupo de entrenamiento y el tercio restante de las muestras en el grupo de ensayo. Se aplicó el algoritmo de eliminación retroactivo utilizando datos del grupo de entrenamiento, excluyendo un anticuerpo cada vez y eliminando interactivamente el anticuerpo que se eliminó cuando se obtuvo la divergencia de Kullback-Leibler mínima en el análisis de clasificación, tal como se ha descrito previamente [31]. Se utilizaron los últimos 25 anticuerpos para su eliminación para construir un modelo de clasificación en los datos de entrenamiento que se aplicó directamente en el grupo de ensayo correspondiente. El área bajo la curva ROC (AUC) fue utilizada como medición de la precisión de la firma en el grupo de ensayo. Se repitió este procedimiento diez veces, en diez pares de grupos de ensayo y de entrenamiento generados aleatoriamente diferentes. Finalmente, se dio a cada anticuerpo una puntuación sobre la base del orden de eliminación en diez grupos de ensayo. Se calculó la puntuación a partir de la resistencia promedio en el proceso de eliminación (primer anticuerpo para su eliminación = 1, último anticuerpo para su eliminación = 293). Se calcularon las sensibilidades (SN) y especificidades (SP) a partir del umbral óptimo de los valores de predicción MVS, definidos en el presente documento como la suma máxima de SN + SP. Finalmente se llevaron a cabo los análisis de rutas utilizando MetaCore (Thomson Reuters, Nueva York, NY, EE.UU.).

Resultados

Análisis de expresión de proteína diferencial

El análisis de la varianza reveló un gran número de anticuerpos con patrones de señalización diferencial fuerte entre cáncer y controles en las muestras de suero. De hecho, el multigrupo ANOVA demostró que el 75 % de los anticuerpos presentó un nivel de significancia de $p < 0,001$ en los datos del suero. El análisis de los componentes principales demostró que las muestras de cáncer se diferenciaban claramente más con respecto a los normales que con respecto a los controles benignos (Fig. 1).

Para cada comparación de subgrupo, en la Tabla 2 se muestran los 25 anticuerpos que presentaron los valores p más bajos. La comparación PC frente a NC reveló analitos fuertemente regulados al alza y a la baja, mientras que el análisis PC frente a BC identificó analitos con un número de cambios más modesto y que se regularon al alza casi exclusivamente en PC frente a BC (Tabla 2). Los analitos que presentaron el nivel máximo de expresión diferencial fueron GAK, IL-6, LDL, y MAPK8 ($p \leq 4 \cdot 10^{-20}$) para PC frente a NC, mientras que se identificó un grupo de proteínas significativamente diferente para PC frente a BC, presentando Cistatina C, IL-13 y IL-1 α los valores p más bajos ($p \leq 3 \cdot 10^{-10}$).

Un análisis de las rutas basado en todo el grupo de anticuerpos ($n = 293$) con los valores p y el número de veces de cambios correspondientes para PC frente a NC, indicó enfermedades relacionadas con la producción de insulina (hiperinsulinismo y resistencia a la insulina) como resultados superiores en la búsqueda de enfermedades mediante biomarcadores (Fig. 6). Las afecciones asociadas con el metabolismo, como trastornos del metabolismo de la glucosa, obesidad y sobrepeso se relacionaron también significativamente, así como las redes de biomarcador de núcleo de diabetes de tipo I y II, enfermedad de Crohn, hepatitis, afecciones autoinmunes e infecciosas, y diversos tipos de neoplasmas (incluyendo pancreáticas). La mayoría de estas enfermedades identificadas por análisis de rutas han sido asociadas así o correlacionadas sintomáticamente con PC.

Firmas para clasificación

A continuación, se subdividieron los datos aleatoriamente en grupos de ensayo y de entrenamiento. Se utilizaron los grupos de entrenamiento para identificar la combinación más discriminadora de anticuerpos aplicando un algoritmo de eliminación retroactivo basado en análisis MVS, excluyendo los anticuerpos uno por uno. La clasificación de PC y NC fue altamente precisa, tal como se dedujo por las pequeñas divergencias Kullback-Leibler (K-L) ($\leq 33,2$) a través del proceso de eliminación Fig. 2A). Se alcanzó una K-L mínima diferenciadas (12,0) cuando se dejaron únicamente siete anticuerpos en el proceso de eliminación. Este panel de proteínas de 7 plexadas que incluye IL-6, Cistatina C, IL-8, IL-11, inhibidor C1, Eotaxina, y HADH2, presentó una sensibilidad (SN) y especificidad (SP) de 100 % y 96 %, respectivamente, en el grupo de ensayos correspondiente de muestras no observadas previamente, lo cual demostró la posibilidad de combinar varios analitos para una clasificación de próxima a perfecta de PC y NC. En contraste, los valores K-L fueron significativamente más altos ($\leq 181,3$) cuando se comparó PC con BC. En este punto, el valor K-L mínimo (50,0) no fue tan diferenciado y supuso que fuera necesario un panel de anticuerpos mucho mayor para la diferenciación óptima de PC y BC (Fig. 2B). Para cada comparación de subgrupo, se repitió el procedimiento hasta que se hubieron utilizado diez grupos de entrenamiento generados aleatoriamente diferentes para eliminación retroactiva. Las curvas K-L fueron altamente similares a las presentadas en la Fig. 2 lo cual indica que sería necesario un promedio de 67 anticuerpos para una clasificación óptima de PC frente a BC (Fig. 2B).

Para generar y evaluar las firmas de un número de anticuerpos viables, se seleccionaron 25 anticuerpos principales de cada proceso de eliminación y se utilizaron para construir modelos de clasificación MVS en los grupos de entrenamiento. Los valores AUC generados por los modelos en los grupos de ensayo correspondientes sirvieron como medición de la precisión de clasificación (Fig. 2C). Cada firma de 25 anticuerpos generados pudo predecir PC de NC con una alta precisión (AUC promedio 0,98). La sensibilidad y especificidad de las diez firmas osciló entre 90 % SN y 85 % SP, y 100 % SN y SP, con un promedio de SN y SP de 95 %. En contraste, PC fue más difícil de predecir a partir de BC (AUC promedio 0,67) con 76 % SN y 67 % SP para la firma de mejor rendimiento.

Finalmente, se le dio a cada anticuerpo una puntuación, correspondiente a su resistencia promedio en los procesos de eliminación (Tabla 3). Dentro del análisis PC frente a NC, las diez firmas fueron altamente similares. Por ejemplo, el anticuerpo superior, dirigido a IL-11, tuvo una puntuación total de 291,4 de 293 eliminaciones, siendo el último anticuerpo en ser eliminado 4 de 10 veces. En todos, los 25 anticuerpos de máxima puntuación para PC frente NC representaron los 20 analitos no redundantes, comprendidos entre citoquinas y quimiocinas (IL-11, IL-6, IL-13, IL-8, TNF- α , y Eotaxina), componentes complemento (inhibidor C1, C1q, C5, y Factor B), enzimas (HADH2, GAK, y ATP-5B) y más. Un grupo altamente diferente de proteínas apareció como marcadores principales para PC frente BC, con MAPK1, TNFRSF3, UCHL5, IL-4, Apo-A1, Apo-A4, ligando CD40, KSYK y otros entre los analitos puntuados principales (Tabla 3).

Aunque las firmas basadas en la puntuación de anticuerpo (Tabla 3) fueron diferentes a las derivadas de los valores p (Tabla 2), se observó una significativa superposición, en particular para las firmas PC frente a NC, en donde 8 de cada 25 anticuerpos (IL-11, IL-6, IL-13, HADH2, LDL, GAK, C1q y TNF- α) aparecieron en ambas firmas. Todos los 25 anticuerpos principales puntuados para PC frente a NC se expresaron de hecho de forma significativamente diferencial incluso aunque no todos estuvieran presentes en la lista de valor p de 25 principales.

Localización de tumor

Basándose en los perfiles de proteína en suero, se pudieron estratificar las muestras de cáncer también de acuerdo con el origen del tumor en el páncreas. El análisis del componente principal demostró que las muestras de los tumores localizados en el cuerpo o en la cola del páncreas se agrupaban en cierto grado más próximas que los controles normales que las muestras de tumores localizados en la cabeza del páncreas, y que se pudo observar la separación de las muestras de cáncer sobre la base de la localización del tumor (cabeza frente a cuerpo/cola) (Fig. 3). El análisis de expresión de proteína diferencial reveló una extensa lista de analitos con niveles significativamente diferentes en las muestras del suero de tumores de la cabeza y el cuerpo/cola, presentando un 39 % de los anticuerpos valores $p < 0,001$ y que presentaron casi exclusivamente niveles regulados al alza en los tumores de cabeza en comparación con los tumores de cuerpo y cola (Tabla 5).

Conclusión

Este estudio representa uno de los análisis multicéntrico más extenso de paneles de biomarcadores para predecir cáncer de páncreas que se ha llevado a cabo hasta ahora. Hasta donde llega el conocimiento de los autores de la invención, se trata del mayor análisis multiplexado de analitos de suero en dicha cohorte de cáncer de páncreas grande (>300 muestras) utilizando proteómica de afinidad. Se llevó a cabo el análisis utilizando micromatrices de anticuerpo recombinantes del laboratorio, una plataforma que ha ido avanzando a lo largo de varios años [32, 36], y que ahora incluye cerca de 300 anticuerpos, seleccionados rigurosamente contra una gama de dianas principalmente del proteoma inmunorregulador. Con el empleo de nuevos protocolos de alta producción y purificación, se generaron fácilmente estos anticuerpos en menos de una semana y se imprimieron rápidamente en tres puntos replicados en la escala de picolitros en micromatrices planas. En la configuración actual, se pueden analizar más de cien muestras en paralelo al día y por estación de trabajo utilizando solamente volúmenes diminutos (<1 μ l) de suero sin diluir, lo cual permite el análisis de grandes cohortes de muestras en el curso de unos días.

En este estudio, se trazó el perfil de 338 muestras de suero en estas matrices, comparando cáncer de páncreas con controles normal y benigno. Utilizando ensayos altamente multiplexados como el actual, es probable que aparezca un cierto nivel de correlación, en particular cuando se enfoca a proteínas altamente interconectadas como las del sistema inmunitario. Incluso aunque la potencia discriminativa de anticuerpos individuales, representados por valores p únicos podría ser de interés, otros enfoques estadísticos pueden ser más precisos para identificar la combinación óptima de anticuerpos [37]. En este punto, se utilizó un modelo supervisado basado en el análisis de máquina de vector de soporte. Se dividieron los datos en grupos de entrenamiento de los que se identificaron las firmas de biomarcador mediante un algoritmo de eliminación retroactiva iterativa y grupos de validación complementarios en los que se analizó la potencia de clasificación de las firmas.

Al adoptar este enfoque, se derivaron listas de proteínas de interés tanto de los patrones de expresión diferencial, representados por valores p, como del análisis de eliminación retroactiva. Éste último demostró que solamente de 4 a 10 anticuerpos serían suficientes para una discriminación próxima a perfecta de PC frente a NC, resultados que son muy alentadores y que confirman lo que han observado los autores de la invención previamente en cuanto a que PC puede discriminarse fácilmente de controles sanos mediante ensayo del proteoma inmunorregulador [23, 24]. Sin embargo, cuando se comparó PC con BC (principalmente pancreatitis crónica), hasta 67 anticuerpos parecieron

requerir una potencia de clasificación óptima. Esto puede explicarse en parte por el tipo de proteínas medidas; no fue probable que los analitos inmunorreguladores presentaran una especificidad de enfermedad manejable, ya que el sistema inmune está altamente afectado en cualquier afección. De hecho el análisis de rutas llevado a cabo en este estudio confirmó el impacto sistémico altamente similar de PC y varias otras afecciones, como hiperinsulinismo, resistencia a la insulina y enfermedades metabólicas, así como autoinmunidad e infecciones, lo cual demuestra una vez más el reto que supone distinguir PC de afecciones benignas sintomáticamente relacionadas. La identificación de una inmunofirma relevante para PC es por tanto un acto de equilibrio. Es deseable un pequeño panel ya que cada agente de unión se añade al coste y la complejidad del ensayo, y con todo la firma ha de ser lo suficientemente grande como para constituir una huella sensible y específica de la enfermedad. En esta fase del descubrimiento, los autores de la invención llegaron a la conclusión de que 25 analitos serían un punto de partida viable y los estudios de seguimiento servirán para dilucidar si estas firmas pueden condensarse en paneles de biomarcadores más reducidos al mismo tiempo que presentan la sensibilidad y especificidad requerida para el inmunoensayo de diagnóstico.

A pesar de adoptar dos estrategias muy diferentes para la identificación de firmas, siguió habiendo un gran solapamiento entre las firmas derivadas de los diferentes análisis, es decir, los marcadores que obtuvieron las puntuaciones totales de eliminación retroactiva más altas se expresaron también muy diferenciadamente también de forma general. Por ejemplo, IL-11, IL-6, inhibidor C1, IL-13, HADH2, LDL, GAK, C1q y TNF- α aparecieron en ambas firmas y demostraron así tanto valores p bajos como puntuaciones de eliminación retroactiva altas para PC frente a NC. Para PC frente a BC las dos firmas se solaparon con, Apo-A4, BTK, TGF- β 1, MCP-1, y UPF3B. Cabe señalar que una serie de marcadores presentó un potencial para PC tanto cuando se compararon con NC como con BC. Varios de ellos, incluyendo el inhibidor C1, C5, Factor B, IL-13, MCP-1, y TNF- α , han sido asociados con PC en estudios anteriores por los autores de la invención [23, 24] y otros [38-40], mientras que HADH2, una acetil CoA dehidrogenasa, hasta donde llega el conocimiento con los que cuentan los autores de la invención, por el momento no se ha notificado para PC. Por otra parte, otras proteínas que no han sido medidas anteriormente, también presentaron potencial para diferenciación de PC. Por ejemplo, GAK, una serina/treonina quinasa, se reguló fuertemente a la baja en PC frente a NC y también apareció en firmas de eliminación retroactiva para PC frente a NC. Por otra parte, se incluyeron TNFRSF3 (TNF- β receptor) y UPF3B, una proteína reguladora de ARNm en las firmas de valor p y anticuerpo para PC frente a BC. Finalmente, MAPK1 (ERK2), una quinasa de la ruta de señalización MAPK/ERK que según se ha demostrado se desregula en PC [41] fue la proteína con la puntuación máxima en las firmas de eliminación retroactiva para PC frente a BC.

Se estratificaron también las muestras de suero sobre la base del emplazamiento del tumor en el páncreas, lo cual, hasta donde llega el conocimiento que tienen los autores de la invención, no se ha realizado anteriormente con proteómica. Se demostró que las muestras, hasta cierto grado, se pudieron separar sobre la base del origen del tumor y que la parte predominante de las muestras de pacientes con tumores de cola y cuerpo se agrupaban ligeramente más próximos que los controles normales que el suero de pacientes con tumores de cabeza pancreática. Estos resultados demuestran pues las discrepancias en el impacto sistémico del cáncer sobre la base de su origen en el páncreas. El inmenso número de proteínas inmunorreguladoras de suero encontradas en muestras reguladas al alza de tumores de cabeza en comparación con las muestras de tumores de cuerpo/cola puede reflejar un impacto sistémico más profundo de tumores localizados en la cabeza pancreática, que a menudo invaden los vasos sanguíneos mesentéricos circundantes que conectan el páncreas con el duodeno [3]. Estos hallazgos que indican que el perfil del suero en las micromatrices de anticuerpo podría aplicarse para distinguir entre tumores de la cabeza y el cuerpo/cola del páncreas, lo cual podría favorecer posiblemente la decisión sobre el tratamiento del tumor.

A pesar de ser la 4^o causa de cáncer más letal, la incidencia de PC es baja, ~11 de cada 100.000 individuos en Estados Unidos [42]. Desde un punto de vista económico-sanitario, la baja incidencia hace difícil la motivación de la detección sistemática de PC en la población en general. Sobre la base de las sensibilidades y especificidades que han presentado anteriormente los autores de la invención [24], un estudio reciente confirma sin embargo, la rentabilidad que supone identificar sistemáticamente a individuos de alto riesgo para PC [43]. Los factores de riesgo para PC incluyen no solamente pancreatitis y neoplasmas benignos, sino también, p.ej., el cáncer de páncreas familiar, pancreatitis hereditaria, mutaciones BRCA, síndrome Peutz-Jeghers, diabetes melitus, infección por *Helicobacter pylori* [44]. La siguiente etapa será explorar las firmas identificadas en el presente estudio, para el diagnóstico de PC entre individuos con un mayor riesgo de PC que representan una población diana pertinente para el diagnóstico por inmunoensayo de PC.

En conclusión, este estudio multicéntrico multiplexado extensivamente que revela inmuno-firmas asociadas con cáncer de páncreas, presentando sensibilidades, especificidades en el intervalo de 90-100 %, demuestra claramente la susceptibilidad de aplicación para el diagnóstico de PC y también indica el potencial de micro matrices de anticuerpo recombinante para estratificar muestras de suero basándose en el emplazamiento del tumor en el páncreas.

Referencias

[1] Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA: a cancer journal for clinicians. 2012; 62:10-29.

- [2] Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*. 2010; 467:1114-7.
- [3] Hidalgo M. Pancreatic cancer. *The New England journal of medicine*. 2010; 362:1605-17.
- [4] Conlon KC, Klimstra DS, Brennan MF. Long-term survival after curative resection for pancreatic ductal adenocarcinoma. Clinicopathologic analysis of 5-year survivors. *Annals of surgery*. 1996; 223:273-9.
- [5] Sohn TA, Yeo CJ, Cameron JL, Koniaris L, Kaushal S, Abrams RA, et al. Resected adenocarcinoma of the pancreas-616 patients: results, outcomes, and prognostic indicators. *Journal of gastrointestinal surgery: official journal of the Society for Surgery of the Alimentary Tract*. 2000; 4:567-79.
- [6] Furukawa H, Okada S, Saisho H, Ariyama J, Karasawa E, Nakaizumi A, et al. Clinicopathologic features of small pancreatic adenocarcinoma. A collective study. *Cancer*. 1996; 78:986-90.
- [7] Shimizu Y, Yasui K, Matsueda K, Yanagisawa A, Yamao K. Small carcinoma of the pancreas is curable: new computed tomography finding, pathological study and postoperative results from a single institute. *Journal of gastroenterology and hepatology*. 2005; 20: 1591-4.
- [8] Ishikawa O, Ohigashi H, Imaoka S, Nakaizumi A, Uehara H, Kitamura T, et al. Minute carcinoma of the pancreas measuring 1 cm or less in diameter--collective review of Japanese case reports. *Hepato-gastroenterology*. 1999; 46:8-15.
- [9] Gangi S, Fletcher JG, Nathan MA, Christensen JA, Harmsen WS, Crownhart BS, et al. Time interval between abnormalities seen on CT and the clinical diagnosis of pancreatic cancer: retrospective review of CT scans obtained before diagnosis. *AJR American journal of roentgenology*. 2004; 182:897-903.
- [10] Pelaez-Luna M, Takahashi N, Fletcher JG, Chari ST. Resectability of presymptomatic pancreatic cancer and its relationship to onset of diabetes: a retrospective review of CT scans and fasting glucose values prior to diagnosis. *The American journal of gastroenterology*. 2007; 102:2157-63.
- [11] Locker GY, Hamilton S, Harris J, Jessup JM, Kemeny N, Macdonald JS, et al. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2006; 24:5313-27.
- [12] Koopmann J, Rosenzweig CN, Zhang Z, Canto MI, Brown DA, Hunter M, et al. Serum markers in patients with resectable pancreatic adenocarcinoma: macrophage inhibitory cytokine 1 versus CA19-9. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2006; 12:442-6.
- [13] Brand RE, Nolen BM, Zeh HJ, Allen PJ, Eloubeidi MA, Goldberg M, et al. Serum biomarker panels for the detection of pancreatic cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2011; 17:805-16.
- [14] Bunger S, Laubert T, Roblick UJ, Habermann JK. Serum biomarkers for improved diagnostic of pancreatic cancer: a current overview. *Journal of cancer research and clinical oncology*. 2011; 137:375-89.
- [15] Jiang JT, Wu CP, Deng HF, Lu MY, Wu J, Zhang HY, et al. Serum level of TSGF, CA242 and CA19-9 in pancreatic cancer. *World journal of gastroenterology: WJG*. 2004; 10: 1675-7.
- [16] Duraker N, Hot S, Polat Y, Hobek A, Gender N, Urhan N. CEA, CA 19-9, and CA 125 in the differential diagnosis of benign and malignant pancreatic diseases with or without jaundice. *Journal of surgical oncology*. 2007;95:142-7.
- [17] Ni XG, Bai XF, Mao YL, Shao YF, Wu JX, Shan Y, et al. The clinical value of serum CEA, CA19-9, and CA242 in the diagnosis and prognosis of pancreatic cancer. *European journal of surgical oncology: the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology*. 2005; 31:164-9.
- [18] Faca VM, Song KS, Wang H, Zhang Q, Krasnoselsky AL, Newcomb LF, et al. A mouse to human search for plasma proteome changes associated with pancreatic tumor development. *PLoS medicine*. 2008; 5:e123.
- [19] Firpo MA, Gay DZ, Granger SR, Scaife CL, DiSario JA, Boucher KM, et al. Improved diagnosis of pancreatic adenocarcinoma using haptoglobin and serum amyloid A in a panel screen. *World journal of surgery*. 2009; 33:716-22.
- [20] Orzechowski R, Hamelinck D, Li L, Gliwa E, vanBrocklin M, Marrero JA, et al. Antibody microarray profiling reveals individual and combined serum proteins associated with pancreatic cancer. *Cancer research*. 2005; 65:11193-202.
- [21] Coussens LM, Werb Z. Inflammation and cancer. *Nature*. 2002; 420:860-7.
- [22] Carlsson A, Wuttge DM, Ingvarsson J, Bengtsson AA, Sturfelt G, Borrebaeck CA, et al. Serum protein profiling of systemic lupus erythematosus and systemic sclerosis using recombinant antibody microarrays. *Molecular & cellular proteomics : MCP*. 2011; 10:M110 005033.
- [23] Ingvarsson J, Wingren C, Carlsson A, Ellmark P, Wahren B, Engstrom G, et al. Detection of pancreatic cancer using antibody microarray-based serum protein profiling. *Proteomics*. 2008; 8:2211-9.
- [24] Wingren C, Sandstrom A, Segersvard R, Carlsson A, Andersson R, Lohr M, et al. Identification of serum biomarker signatures associated with pancreatic cancer. *Cancer research*. 2012; 72:2481-90.
- [25] Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Molecular & cellular proteomics : MCP*. 2002; 1:845-67.
- [26] Zhang H, Liu AY, Loriaux P, Wollscheid B, Zhou Y, Watts JD, et al. Mass spectrometric detection of tissue proteins in plasma. *Molecular & cellular proteomics : MCP*. 2007; 6:64-71.
- [27] Surinova S, Schiess R, Huttenhain R, Cerciello F, Wollscheid B, Aebersold R. On the development of plasma protein biomarkers. *Journal of proteome research*. 2011; 10:5-16.
- [28] Haab BB, Geierstanger BH, Michailidis G, Vitzthum F, Forrester S, Okon R, et al. Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens: systematic variation between sample types and calibration of mass spectrometry data. *Proteomics*. 2005; 5:3278-91.

- [29] Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Statistics in medicine*. 2002; 21:835-52.
- [30] Borrebaeck CA, Wingren C. Design of high-density antibody microarrays for disease proteomics: key technological issues. *Journal of proteomics*. 2009; 72:928-35.
- [31] Carlsson A, Wingren C, Kristensson M, Rose C, Ferno M, Olsson H, et al. Molecular serum portraits in patients with primary breast cancer predict the development of distant metastases. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:14252-7.
- [32] Wingren C, Ingvarsson J, Dexlin L, Szul D, Borrebaeck CA. Design of recombinant antibody microarrays for complex proteome analysis: choice of sample labeling-tag and solid support. *Proteomics*. 2007; 7:3055-65.
- [33] Carlsson A, Persson O, Ingvarsson J, Widegren B, Salford L, Borrebaeck CA, et al. Plasma proteome profiling reveals biomarker patterns associated with prognosis and therapy selection in glioblastoma multiforme patients. *Proteomics Clinical applications*. 2010; 4:591-602.
- [34] Olsson N, Wallin S, James P, Borrebaeck CA, Wingren C. Epitope-specificity of recombinant antibodies reveals promiscuous peptide-binding properties. *Protein science: a publication of the Protein Society*. 2012; 21:1897-910.
- [35] Carlsson A, Wingren C, Ingvarsson J, Ellmark P, Baldertorp B, Ferno M, et al. Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays. *Eur J Cancer*. 2008; 44:472-80.
- [36] Ingvarsson J, Larsson A, Sjöholm AG, Truedsson L, Jansson B, Borrebaeck CA, et al. Design of recombinant antibody microarrays for serum protein profiling: targeting of complement proteins. *Journal of proteome research*. 2007; 6:3527-36.
- [37] Quackenbush J. Computational analysis of microarray data. *Nature reviews Genetics*. 2001; 2:418-27.
- [38] Ma Y, Hwang RF, Logsdon CD, Ullrich SE. Dynamic mast cell-stromal cell interactions promote growth of pancreatic cancer. *Cancer research*. 2013; 73:3927-37.
- [39] Kudo-Saito C, Shirako H, Ohike M, Tsukamoto N, Kawakami Y. CCL2 is critical for immunosuppression to promote cancer metastasis. *Clinical & experimental metastasis*. 2013; 30:393-405.
- [40] McDade TP, Perugini RA, Vittimberga FJ, Jr., Carrigan RC, Callery MP. Salicylates inhibit NF-kappaB activation and enhance TNF-alpha-induced apoptosis in human pancreatic cancer cells. *The Journal of surgical research*. 1999; 83:56-61.
- [41] Arlt A, Muerkoster SS, Schafer H. Targeting apoptosis pathways in pancreatic cancer. *Cancer letters*. 2013; 332:346-58.
- [42] Shaib YH, Davila JA, El-Serag HB. The epidemiology of pancreatic cancer in the United States: changes below the surface. *Alimentary pharmacology & therapeutics*. 2006; 24:87-94.
- [43] Ghatnekar O, Andersson R, Svensson M, Persson U, Ringdahl U, Zeilon P, et al. Modelling the benefits of early diagnosis of pancreatic cancer using a biomarker signature. *International journal of cancer Journal international du cancer*. 2013.
- [44] Konstantinou F, Syrigos KN, Saif MW. Pancreatic cancer: what about screening and detection? *JOP : Journal of the pancreas*. 2013; 14:312-5.

Tabla 1. Demografía de muestras de suero

Diagnóstico	Subgrupo	Género (M/F)	Edad (media ± sd)	No. de muestras
Cáncer de páncreas (CP)		92/64	66 ± 13	156
	Cuerpo			16
	Cabeza			97
	Cola			10
	Otros			16
	Sin especificar			17
Controles benignos (BC)		117/35	52 ± 14	152
	Pancreatitis aguda			33
	Pancreatitis crónica			110
	Neoplasma de Langerhan			3
	Neoplasma pancreático			6
Controles normales (NC)		20/20	62 ± 14	30
Total				338

Tabla 2. 25 Anticuerpos principales basados en la significancia del análisis de expresión de proteína diferencial. Entre paréntesis se presenta el sufijo del clon de anticuerpo individual (para los analitos dirigidos por varios clones de anticuerpo). FC = Número de veces del cambio

PC-NC-BC ANOVA		prueba tPC-NC		prueba t PC-BC			
Anticuerpo	valor p	Anticuerpo	valor p	FC	Anticuerpo	valor p	FC
GAK (3)	4,14E-48	GAK (3)	3,44E-30	0,68	Cistatina C (3)	3,80E-11	1,08
IL-6 (7)	4,00E-43	IL-6 (7)	1,11E-27	0,71	IL-13 (3)	1,09E-10	1,13
GAK (2)	3,25E-36	GAK (1)	3,25E-26	0,77	IL-1α (1)	3,08E-10	1,11
IL-11 (2)	1,78E-32	LDL (2)	1,15E-22	0,62	Ag superficial X	5,57E-09	1,10
LDL (2)	6,09E-32	MAPK8 (1)	4,24E-20	0,81	BTK(2)	6,06E-09	1,05

TNF- α (3)	1,89E-31	IL-11 (2)	4,56E-18	0,69	Cistatina C (4)	6,60E-09	1,05
Procatepsina W	4,88E-27	TNF- α (3)	2,93E-16	0,76	CIMS(26)	6,79E-09	1,10
IL-13 (3)	8,15E-27	HADH2 (3)	2,49E-15	0,89	CD40 (1)	1,02E-08	1,09
MAPK8 (1)	3,02E-22	Procatepsina W	3,21E-14	0,79	TNFRSF3 (2)	1,29E-08	1,05
IL-1 α (1)	5,83E-20	TNFRSF3 (1)	5,08E-13	0,91	ORP-3 (2)	2,16E-08	1,07
IL-13 (2)	3,36E-19	VEGF (3)	1,52E-12	1,24	Apo-A4 (3)	2,66E-08	1,07
TNFRSF3 (1)	8,51E-16	IL-1ra (1)	3,24E-12	1,16	UPF3B (2)	2,84E-08	1,04
IL-18 (2)	5,33E-15	C1 inh (1)	1,20E-11	1,22	MUC-1 (1)	3,49E-08	1,07
IL-1ra (1)	9,67E-15	IL-13 (3)	2,73E-11	0,77	TNF- α (3)	4,31E-08	1,08
HADH2 (3)	1,06E-14	C1q	4,12E-11	1,11	CIMS (16)	4,67E-08	1,08
CD40 (1)	1,48E-14	IL-16 (3)	5,19E-11	1,15	ATP-5B (1)	5,68E-08	1,06
Cistatina C (4)	2,36E-14	VEGF (1)	5,43E-11	1,19	CIMS (12)	6,00E-08	1,06
IL-4 (3)	3,12E-14	CD40L	6,21E-11	1,13	IL-13 (1)	6,31E-08	1,08
CIMS (18)	6,91E-14	IL-4 (3)	8,82E-11	1,24	MCP-1 (4)	9,64E-08	1,03
CIMS (16)	7,21E-14	Sialyl Lewis x	9,47E-11	1,10	CIMS (1)	1,14E-07	1,08
VEGF (3)	8,52E-14	IL-18 (2)	1,12E-10	1,20	CD40 (3)	1,28E-07	1,06
FASN (3)	1,18E-13	CIMS (18)	1,43E-10	1,16	Procatepsina W	1,43E-07	1,07
TGF- β 1 (2)	2,59E-13	CIMS (23)	2,64E-10	1,13	Digoxina	1,55E-07	1,08
CIMS (26)	2,61E-13	MCP-1 (2)	3,08E-10	1,16	TGF- β 1 (2)	1,88E-07	1,09
CIMS (25)	6,47E-13	CIMS (25)	3,44E-10	1,24	CIMS (24)	3,32E-07	1,08

Tabla 3. 25 anticuerpos principales basados en la puntuación del anticuerpo a partir de diez iteraciones de eliminación retroactiva. Se presentan en entre paréntesis el sufijo del clon de anticuerpo individual (para análisis dirigidos por varios clones de anticuerpos)

5

PC frente a NC	Puntuación	PC frente a BC	Puntuación
IL-11 (2)	291,4	MAPK1 (3)	276,5
IL-6 (7)	288,1	C5 (2)	273
Cistatina C (1)	286,9	TNFRSF3 (1)	265,5
inh C1 (3)	279,2	TNFRSF3 (2)	260,9
Angiomotina (1)	276	UCHL5	259,6
IL-13 (2)	272,9	IL-4 (3)	258,7
IL-13 (3)	270,7	Factor B (3)	258
CD40 (1)	270,6	Apo-A4 (3)	257,5
HADH2 (3)	270,4	KSYK-1	255,2
HADH2 (4)	269,7	Sox11A	253,1
inh.C1 (4)	269,4	CD40L	252,2
inh.C1 (2)	269,2	Apo-A1 (1)	251,4
LDL (2)	268,1	CIMS (13)	250,1
GAK (3)	268	BTK (2)	246,1
C3 (1)	266,1	GM-CSF (5)	245
CIMS (5)	264,3	TGF- β 1 (2)	239,5
C1q	261,1	PTP-1B (2)	237,2
CD40 (4)	259,6	MCP-1 (7)	235,1
IL-8 (2)	259,4	UPF3B (1)	232,5
C5 (2)	258,5	C1 inh, (4)	228,3
ATP-5B (3)	257,1	Sialyl Lewis x	227,6
Factor B (4)	256,2	IL-3 (1)	225,8
CIMS (10)	253,6	IL-9 (2)	224,2
TNF- α (3)	253,5	HADH2 (2)	222,7
Eotaxina (3)	248,4	IL-4 (4)	222,4

Tabla 4. Antígenos dirigidos sobre la micro matriz de anticuerpo

Proteína	Nombre completo	No. de clones de anticuerpo
Angiomotina	Angiomotina	2
Apo-A1	Apolipoproteína A1	3
Apo-A4	Apolipoproteína A4	3
ATP-5B,	ATP sintasa subunidad beta mitocondrial	3
β -galactosidasa	Beta-galactosidasa	1
BTK	Tirosina-proteína quinasa BTK	4
C1	Inhibidor C1 Inhibidor Plasma proteasa	4
C1q	Complemento C1q	1
C1s	Complemento C1s	1
C3	Complemento C3	6

C4	Complemento C4	4
C5	Complemento C5	3
CD40	CD40 proteína	4
CD40L	Ligando CD40	1
CDK-2	Quinasa dependiente de cdc2	2
CHX10	Sistema visual homeobox 2	3
CT17	Subunidad toxina cólera B	1
Cistatina C	Cistatina C	4
Digoxina	Digoxina	1
DUSP9	Especificidad dual proteína fosfatasa 9	1
EGFR	Receptor de factor de crecimiento epidérmico (Cetuximab (Erbtux) anticuerpo)	1
Eotaxina	Eotaxina	3
Factor B	Complemento factor B	4
FASN	Proteína FASN	4
GAK	Proteína GAK	3
GLP-1	Péptido-1 tipo Glucagon	1
GLP-1R	Receptor Péptido 1 tipo Glucagon	1
GM-CSF	Factor estimulador de colonias de macrófagos-granulocito	6
HADH2	Proteína HADH2	4
Her2/ErbB-2	Receptor tirosina-proteína quinasa erbB-2	4
HLA-DR/DP	HLA-DR/DP	1
ICAM-1	Molécula de adhesión intercelular 1	1
IFN-γ	gamma interferón	3
IgM	Inmunoglobulina M	5
IL-10	Interleucina-10	3
IL-11	Interleucina-11	3
IL-12	Interleucina-12	4
IL-13	Interleucina-13	3
IL-16	Interleucina-16	3
IL-18	Interleucina-18	3
IL-1α	Interleucina-1 alfa	3
IL-1β	Interleucina-1 beta	3
IL-1ra	Interleucina-1 proteína antagonista de receptor	3
IL-2	Interleucina-2	3
IL-3	Interleucina-3	3
IL-4	Interleucina-4	4
IL-5	Interleucina-5	3
IL-6	Interleucina-6	8
IL-7	Interleucina-7	2
IL-8	Interleucina-8	3
IL-9	Interleucina-9	3
Integrina α-10	Integrina alfa-10	1
Integrina α-11	Integrina alfa-11	1
JAK3	Tirosina-proteína quinasa JAK3	1
Queratina 19	Queratina tipo citoesquelético 19	3
KSYK	Tirosina-proteína quinasa SYK	2
LDL	Apolipoproteína B-100	2
Leptina	Leptina	1
Lewis x	Lewis x	2
Lewis y	Lewis y	1
Lumican	Lumican	1
MAPK1 4	proteína quinasa 1 activada con mitógeno	1
MAPK8	proteína quinasa activada con mitógeno 8	3
MATK	Proteína tirosina- quinasa asociada a Megacariocito	3
MCP-1	motivo C-C quimiocina 2	9
MCP-3	Motivo C-C quimiocina 7	3
MCP-4	Motivo C-C quimiocina 13	3
MUC-1	Mucina-1	6
Myomesin-2	Myomesina-2	2
ORP-3	Proteína 3 relacionada con proteína de unión a oxiterol	2
Osteopontina	Osteopontina	3
P85A	Fosfatidilinositol 3-subunidad alfa reguladora de quinasa	3
PKB gamma	RAC-gamma serina/treonina-proteína quinasa	2
Procatepsina W	Catepsina W	1
Properdina	Properdina	1

PSA	Antígeno específico de próstata	1
PTK-6	Proteína-tirosina quinasa 6	1
PTP-1B	Tirosina-proteína fosfatasa no receptor tipo 1	3
RANTES	motivo C-C quimiocina 5	3
RPS6KA2	Proteína ribosómica S6 quinasa alfa-2	3
Sialyl Lewis x	Sialyl Lewis x	1
Sox11A	Factor de transcripción SOX-11	1
STAP2	Proteína 2 adaptador de transducción de señal	4
STAT1	Transductor de señal y activador de transcripción 1-alfa/beta	2
Surface Ag X	Ag X superficial	1
TBC1D9	Miembro 9 de la familia de dominio TBC1	3
TENS4	Tensina-4	1
TGF-β1	factor beta-1 de crecimiento de transformación	3
Péptido TM	Péptido de transmembrana	1
TNF-α	Factor de necrosis de tumor	3
TNF-β	Linfotoxina-alfa	4
TNFRSF14	Miembro 14 de la superfamilia de factor de necrosis de tumor	2
TNFRSF3	Miembro 3 de la superfamilia de factor de necrosis de tumor	3
UBC9	SUMO- UBC9 enzima de conjugación	3
UBE2C	Ubiquitina- E2 C enzima de conjugación	2
UCHL5	Ubiquitina hidrolasa isozima carboxilo-terminal L5	1
UPF3B	Regulador de transcritos no codificantes 3B	2
VEGF	Factor de crecimiento endotelial vascular	4

Tabla 5. Análisis de expresión de proteína diferencial de muestras de suero extraídas de pacientes con tumores de páncreas en distintos emplazamientos. Los resultados se muestran para tumores cuerpo+ cola frente a tumores de cabeza, para los 40 anticuerpos principales ($p < 5,10^{-5}$). FC = número de veces de cambio.

<i>Anticuerpo</i>	<i>val p</i>	<i>FC</i>
IL-1α (2)	4,32E-07	0,874447
CIMS (16)	1,28E-06	0,874676
VEGF (1)	1,47E-06	0,870962
TNF-β (2)	1,57E-06	0,872955
IL-11 (2)	2,15E-06	0,833144
CIMS (18)	3,09E-06	0,891206
CD40L	3,89E-06	0,917468
IL-3 (3)	4,19E-06	0,87497
CIMS (30)	4,88E-06	0,919455
IL-6 (1)	7,01E-06	0,873484
HLA-DR/DP	8,10E-06	0,879348
IL-2 (3)	8,34E-06	0,845869
Angiomotina (2)	8,45E-06	0,892806
Integrina α-10	9,07E-06	0,864779
IL-18 (3)	9,40E-06	0,899675
Sox11A	1,05E-05	0,898637
IL-7 (1)	1,18E-05	0,892816
MCP-1 (3)	1,20E-05	0,885187
Ag X superficial	1,20E-05	0,865919
IL-9 (1)	1,21E-05	0,872639
CIMS (20)	1,36E-05	0,873286
IL-12 (3)	1,67E-05	0,872587
Lewis x (1)	1,77E-05	0,898973
IgM (3)	1,91E-05	0,893885
IL-7 (2)	1,98E-05	0,893886
CIMS (25)	2,25E-05	0,851688
CIMS (2)	2,30E-05	0,88583
CIMS (6)	2,41E-05	0,858863
IL-16 (3)	2,57E-05	0,905756
GLP-1	2,78E-05	0,890237
CHX10 (3)	2,83E-05	0,889864
IL-4 (3)	3,02E-05	0,860565
VEGF (4)	3,20E-05	0,848107
IL-3 (2)	3,35E-05	0,871146
IL-2 (2)	3,89E-05	0,843949
IL-1ra (3)	3,93E-05	0,88975
RANTES (3)	4,51E-05	0,933006
CIMS (31)	4,80E-05	0,886723

CIMS (27)	490E-05	0,915073
TGF- β 1 (3)	4,91E-05	0,889071

TABLA A

				Diagnóstico	
ID de entrada	Nombre de proteína recomendada	Nombre corto	PC	PC (cuerpo y cola) frente a PC (cabeza)	
(i) Biomarcadores núcleo					
Q6IBS9	HADH2 proteína	HADH2	X		
	Miembro 3 superfamilia				
P36941	Receptor factor necrosis de tumor	TNFRSF3	X		
(ii) Biomarcadores preferentes (PC y BTvH)					
P35716	Factor de transcripción SOX-11	Sox11A	X	X	
O75578	Integrina alfa-10	Integrina α-10	X	X	
NA	EDFR		X	X	
NA	EPFR		X	X	
NA	LSADHR		X	X	
NA	SEHLR		X	X	
NA	AQQHQWDGLLSYQDSLS		X	X	
NA	WTRNSNMNYWLIIRL		X	X	
NA	WDSR		X	X	
(iii) Biomarcadores preferentes (PC)					
NA	DFAEDK		X		
Q6PJJ3	FASN proteína	FASN	X		
Q5U4P5	GAK proteína	GAK	X		
NA	LNWVGK		X		
NA	LTEFAK		X		
NA	LYEIAI		X		
P42679	Proteína tirosina quinasa asociada a megacariocito	MATK	X		
Q9H4L5	Proteína 3 relacionada con proteína de unión a oxisterol	ORP-3	X		
NA	QEASFK		X		
NA	SSAYSIR		X		
NA	QEASFK		X		
NA	TEEQLK		X		
NA	TLYVGK		X		
NA	FLLMQYGGMDEHAR		X		
NA	GIVKLYEDEG		X		
NA	GIVKLYEDEG		X		
P43405	Proteína tirosina- quinasa SYK	KSYK	X		
(iv) Biomarcadores opcionales (PC y BTvH)					
Q4VCS5	Angiomotina	Angiomotina	X	X	
P13500	Quimiocina 2 motivo C-C	MCP-1	X	X	
P13501	Quimiocina 5 motivo C-C	RANTES	X	X	
P29965	Ligando CD40	CD40L	X	X	
P01275	Péptido i-1 tipo Glucagon	GLP-1	X	X	
NA	Inmunoglobulina M	IgM	X	X	
P01583	Interleucina-1 alfa	IL-1a	X	X	
P18510	Proteína antagonista de receptor de Interleucina-1	IL-1ra	X	X	
P20809	Interleucina-11	IL-11	X	X	
P29459/60	Interleucina-12	IL-12	X	X	
Q14005	Interleucina-16	IL-16	X	X	
Q14116	Interleucina-18	IL-18	X	X	
P60568	Interleucina-2	IL-2	X	X	
P08700	Interleucina-3	IL-3	X	X	
P05112	Interleucina-4	IL-4	X	X	
P05231	Interleucina-6	IL-6	X	X	
P13232	Interleucina-7	IL-7	X	X	
P15248	Interleucina-9	IL-9	X	X	
NA	Lewis x	Lewis x	X	X	
P01374	Linfotoxina-alfa	TNF-b	X	X	
P01137	Factor de crecimiento de transformación beta-1	TGF-b1	X	X	

P15692	Factor de crecimiento endotelial vascular	VEGF	X	X
P58304	Sistema visual homeobox 2	CHX10	X	
P01903/P01911/ P79483/P13762/ Q30154/P20036/ P04440		HLA-DR/DP		X
(v) Biomarcadores opcionales (PC)				
P02647	Apolipoproteína A1	Apo-A1	X	
P06727	Apolipoproteína A4	Apo-A4	X	
P04114	Apolipoproteína B-100	ATP LDL	X	
P06576	Subunidad sintasa beta mitocondrial	ATP-5B	X	
P16278	Beta-galactosidasa	Bgalactosidasa	X	
P56202	Catepsina W	ProcathepsinaW	X	
Q99616	quimiocina 13 motivo C-C	MCP-4	X	
P80098	quimiocina 7 motivo C-C	MCP-3	X	
Q6P2H9	CD40 proteína	CD40	X	
P02745/6/7	Complemento C1q	C1q	X	
P09871	Complemento C1s	C1s	X	
P01024	Complemento C3	C3	X	
P0COL4/5	Complemento C4	C4	X	
P01031	Complemento C5	C5	X	
P00751	Complemento factor B	Factor B	X	
P24941	Cinasa 2 dependiente de ciclina	CDK-2	X	
P01034	Cistatina-C	Cistatina C	X	
P51671	Eotaxina	Eotaxina	X	
P00533	Receptor de factor de crecimiento epidérmico	EGFR	X	
P43220	Receptor péptido-1 tipo 1 Glucagon	GLP-1R	X	
P04141	Factor estimulador de colonia de granulocitos-macrófagos	GM-CSF	X	
Q9UKX5	Integrina alfa-11	Integrina a-11	X	
P05362	Molécula 1de adhesión Intercelular	ICAM-1	X	
P01579	Gamma Interferón	IFN-g	X	
P01584	Interleucina 1 beta	IL-lb	X	
P22301	Interleucina-10	IL-10	X	
P35225	Interleucina-13	IL-13	X	
P05113	Interleucina-5	IL-5	X	
P10145	Interleucina-8	IL-8	X	
P08727	Queratina, tipo I citoesqueletal 19	Keratin19	X	
P41159	Leptina	Leptina	X	
P51884	Lumican	Lumican	X	
P28482	Proteína quinasa1 activada con mitógeno	MAPK1	X	
P45983	Proteína quinasa8 activada con mitógeno	MAPK8	X	
P15941	Mucina-1	MUC-1	X	
P54296	Miomesina-2	Miomesina-2	X	
P10451	Osteopontina	Osteopontina	X	
P27986	Fosfatidil inositol 3-cinasa	P85A	X	
P05155	Subunidad alfa reguladora	inh C1	X	
P27918	Inhibidor C1 proteasa de plasma	Properdina	X	
P07288	Antígeno específico de próstata	PSA	X	
P04626	Receptor tirosina-proteína quinasa erbB-2	Her2/ErbB-2	X	
Q9BZ17	Regulador de transcritos 3B no codificante	UPF3B	X	
Q15349	Proteína S6 cinasa alfa-2 ribosónica	RPS6KA2	X	
NA	Sialyl Lewis x	Sialyl Lewis x	X	
Q9UGK3	Proteína 2 adaptadora transductora de señal	STAP2	X	
P63279	SUMO- UBC9 enzima de conjugación	UBC9	X	
Q6ZT07	Miembro 9 de la familia de dominioTBC1	TBC1D9	X	
NA	Péptido transmembrana	(péptido Tm)	X	
P01375	Factor alfa necrosis de tumor	TNF-a	X	
Q92956	Miembro 14 superfamilia de receptor de factor de necrosis de tumor	TNFRSF14	X	
Q06187	Proteína tirosina- quinasa BTK	BTK	X	

P52333	Proteína tirosina- quinasa JAK3	JAK3	X
P18031	Proteína tirosina-fosfatasa no-receptor tipo 1	PTP-1B	X
	Ubiquitina carboxilo-terminal		
Q9Y5K5	Hidrolasa isozyma L5	UCHL5	X
O00762	Ubiquitina- E2 C enzima de conjugación	UBE2C	X
(vi) Biomarcadores preferentes (BTvH)			
NA	FIQTDK		X

TABLA B - Programa SVM entrenado

Se obtuvieron los siguientes programas empleando el e1071 1.5-24 SVM, distribuido por <http://cran.r-project.org/web/packages/e1071/index.html>.

```

5  (i) Función de eliminación
# getWorstAb: Retorna el nombre de la sonda de "peor rendimiento"
getWorstAb <- function(errors, abNames)
{
10  return(abNames[order(errors, decreasing = F)[1]])
}
# testModels: Analiza todos los modelos con cada sonda eliminada de una vez
testModels <- function(models, elimData, averages, svmfac)
{
15  nsamples <- ncol(elimData)
d <- as.numeric(svmfac)-1
y <- numeric(nsamples)
E <- numeric(nsamples)
analytes <- nrow(elimData)
errors <- numeric(nrow(elimData))
20  for(k in 1 :analytes)
{
# Reemplazar sonda "k" con el promedio, pero almacenar valores originales
backup <- elimData[k,]
elimData[k,] <- averages[k]
25  # Hacer un bucle LOO sobre el conjunto de datos utilizando modelos ya calibrados
para (i en 1:nmuestras)
{
pred <- predict(models[[i]], t(elimData[,i]), decision.values=TRUE)
# Salvar valores de desión
30  y[i] <- as.numeric(attributes(pred)$decision.values)
}
# Calcular "probabilidades"
y = 1 -(1/(1 + exp(-y)))

35  # Calcular el error K-L para la clasificación con una prueba eliminada
para (i en 1:n muestras)
{
E[i] <- -(d[i]*log(y[i])+(1-d[i])*log(1-y[i]))
}
40  # Salvar el error
errors[k] <- sum(E)

# Volver a poner los datos de la sonda original
45  elimData[k,] <- backup
}
return( errors)
}

50  # getNewElimData: Elimina el anticuerpo correcto de los datos de entrenamiento
getNewElimData <- function(errors, elimData)
{
# Posición para el error mínimo
tasBort <- order(errors,decreasing = F)[1]
55  return(elimData[-tasBort,])
}

# getSmallestError: Retorna al error mínimo
getSmallestError <- function(errors)

```

```

{
  return(min(errors))
}

5  # getNewAverages: crea una nueva lista de promedios tras la eliminación de una sonda
   getNewAverages <-function(errors, averages)
   {
     # encuentra la posición del error mínimo
     tasBort <- order(errors, decreasing = F)[1]
10  return(averages[-tasBort])
   }

   # getRemovedAb: Retorna el nombre de la sonda que se va a eliminar
   getRemovedAb <- function(errors, abNames)
15  {
     return(abNames[order(errors, decreasing = T)[1 ]])
   }

20  backElim <- function(filename, group1, group2, log = FALSE)
   {

     # Incluye bibliotecas "e1071" y "tcltk"
     library(e1071)
25     library(tcltk)

     # Lectura de datos
     rawfile <- read.delim(filename)

30     # Lectura de grupos
     <- rawfile[,2]

     # Lectura de nombres de muestra
     samplenames <- as.character(rawfile[,1])
35     # Crear grupo de datos
     data <- t(rawfile[,c(1,2)])

     # Log de datos si se especifican
40     if (log){data <- log(data)/log(2)}

     # Obtener nombres de sonda
     ProteinNames <- read.delim(filename,header=FALSE)
     ProteinNames <- as.character(as.matrix(ProteinNames)[1,])
45     ProteinNames <- ProteinNames[-(1:2)]

     # Obtener número de sondas
     antal <- length(ProteinNames)

50     # Nombrar muestras y sondas
     rownames(data) <- ProteinNames
     colnames(data) <- samplenames

     # Crear subgrupos (booleano)
55     subset1 <- is.element(groups , strsplit(group1,"")[[1]])
     subset2 <- is.element(groups , strsplit(group2,"")[[1]])

     # Crear lista de grupos como factores
     svmfac <- factor(rep('rest',ncol(data )),levels=c(group1 ,group2,'rest'))
60     svmfac[subset1] <- group1
     svmfac[subset2] <- group2
     svmfac <- svmfac[subset1|subset2]

     # Crear vector para almacenamiento de errores K-L de cada longitud de firma
65     smallestErrorPerLength <- rep(NA,antal)

```

```

# Calcular la señal media para cada sonda de todas las muestras
averages <- apply(data, 1, mean)

5  # Crear vector para almacenamiento de nombres de sonda. La sonda que ha sido
# eliminada cuando el error estaba en el mínimo se almacenará aquí
abOrder <- rep(NA, antal)

# Crear un grupo de datos para hacer la eliminación en
10 elimData <- data[,subset1|subset2]

# Encontrar número de muestras
nsamples <- ncol(elimData)

15 # Nuevos subgrupos en elimData
subset1 <- svmfac==group1
subset2 <- svmfac==group2

print(paste(nsamples, "samples"),quote=F)
20 print(paste(" ",sum(subset1), "in", group1),quote=F)
print(paste(" ",sum(subset2), "in", group2),quote=F)
# Lista para almacenamiento de modelos SVM
models <- numeric(nsamples)

25 # Variable para almacenamiento del número de sondas eliminadas
borttagna <- 0

# Crear una barra de progreso
wrst <- 0
30 proc <- 0
m <- 0
for(i in 1:(antal-1)){
  m <- m+(antal-i)*sqrt(antal-i)
}
35 pb <- tkProgressBar(title = "Progress:", min = 0, max = m, width = 400)

# Actualizar barra de progreso
setTkProgressBar(pb, proc, label=paste("Eliminating first ab\t", round((proc)/(m)*100, 0), "% done"))

40 # Comprobar si los grupos se dan en el orden correcto
control <- as.numeric(svmfac)
if(sum(control[subset1]) > sum(control[subset2]))
{
  print("ERROR: Change the order of group1 and group2 in the data file!!!")
45 break
}

# Volver hacia atrás la eliminación hasta que solamente quedan dos sondas
for(j in 1 :(antal-1))
50 {

  # Entrenamiento de un modelo para cada combinación de muestras N-1 utilizando todas las sondas que quedan
  # en el grupo de datos elimData
  for (i in 1: nsamples)
55 {
    # Almacenar los modelos en un vector (modelos) que contienen las listas de modelo.
    models[i] <- list(svm(t(elimData[,-i]), svmfac[-i], kernel="linear"))
  }

60 # Todos los modelos necesarios para la validación cruzada LOO están ahora calibrados y almacenados en
"modelos".
# Los modelos están todos analizados en sus correspondientes muestras LOO en elimData.
# Cada sonda en elimData se elimina en una ronda LOO reemplazándola con el
# valor medio previamente calculado.
65 #
# Cuando todos los modelos están analizados en elimData con una sonda eliminada el error K-L

```

```

# se calcula y se almacena en "errors". Este procedimiento se repite hasta que todas las sondas
# han sido eliminadas una vez, con el resultado de una lista de errores K-L con la misma longitud
# como el número de sondas que quedan en el grupo de datos
# Crear una lista con los errores K-L para el número actual de sondas (antal + 1-j)
5 # cuando se ha eliminado cada sonda
errors <- testModels(models, elimData, averages, svmfac)

# Añadir el nombre de la sonda con el peor efecto en la clasificación a "abOrder"
wrst<-getWorstAb(errors, row.names(elimData))
10 abOrder[j] <- wrst

# Añadir el valor de error K-L más pequeño (que corresponde a la sonda añadida a "abOrder" mencionado)
smallestErrorPerLength[j] <- getSmallestError(errors)

15 # Eliminar la sonda de "averages"
averages <- getNewAverages(errors, averages)

# Eliminar la sonda de "elimData"
elimData <- getNewElimData(errors, elimData)
20

# Añadir 1 al número de sondas eliminadas
borttagna <- borttagna + 1

# Informar a usuario de que la sonda ha sido eliminada y qué hora es
25 print(paste(j, "probes eliminated @", Sys.time(), sep=""))

# Actualizar barra de progreso
proc <- proc + (antal-j)*sqrt(antal-j)
setTkProgressBar(pb, proc, label=paste(wrst,"eliminated\t", round((proc/m)*100, 0),"% done"))
30 }

# Añadir el nombre de la última sonda "uneliminated" a "abOrder"
abOrder[length(abOrder)] <- setdiff(ProteinNames, abOrder)

35 # Salvar el resultado al archivo con un número aleatorio adjunto al
# archivo para evitar reemplazar un ciclo antiguo.
filename <- paste("Backward elimination result(",rnorm(1)+1,")".txt",sep="")
write.table(cbind(smallestErrorPerLength,abOrder), file=filename, sep="\t", quote = F,row.names =
F)
40 plot(smallestErrorPerLength, type ="b", ylab = "K-L Error", xlab = "Eliminations")
}

```

(ii) Naive Bayesian

```

45 NBtrainer <- function(data, fac){
MeanVariancePval <- function(vec , fac){
vec1 <-vec[fac==levels(fac)[1]]
vec2 <- vec[fac==levels(fac)[2]]
50 if (sum(!is.na(vec1))<=2 | sum(!is.na(vec2))<=2){
return(c(NA,NA,NA,NA,NA))
}
mean1 <- mean(vec1 , na.rm=TRUE)
var1 <- var(vec1 , na.rm=TRUE)
mean2 <- mean(vec2 , na.rm=TRUE)
55 var2 <- var(vec2 , na.rm=TRUE)
if (var1==0 | var2==0){return(c(NA,NA,NA,NA,NA))}
pval <- t.test(vec1,vec2,var.equal=TRUE)$p.value
return(c(mean1 ,var1 ,mean2,var2,pval))
}
60 return(t(apply(data , 1 , MeanVariancePval , fac)))
}

NBpredicter <- function(testdata , NBtrained , topnumber=Inf, logfoldcut=0 , pcut =1){
if (topnumber==Inf){
65 indices <- !is.na(NBtrained[,5]) & NBtrained[,5]<=pcut & abs(NBtrained[,1]-NBtrained[,3])>=logfoldcut
}else{

```

```

preindices <- !is.na(NBtrained[,5]) & NBtrained[,5]<=pcut
abs(NBtrained[preindices,1]-NBtrained[preindices,3]) -> foldchange
cutfold <- sort(foldchange , decreasing=TRUE )[min(topnumber,length(foldchange))]
indices <- preindices & (abs(NBtrained[,1]-NBtrained[,3]) >= cutfold)
5 }
NBtrainedred <- matrix(NBtrained[indices,],ncol=ncol(NBtrained))
testdatared <- matrix(testdata[indices,], ncol=ncol(testdata))
singlegene <- function( genepred){
l1 <- -((genepred[6] - genepred[1])^2)/(2*genepred[2])-0.5*log(2*pi*genepred[2])
10 l2 <- -((genepred[6] - genepred[3])^2)/(2*genepred[4])-0.5*log(2*pi*genepred[4])
#print(genepred)
return(l1-l2)
}
NBvectorpredicter <- function(vec){
15 combined <- cbind(NBtrainedred , vec)
combined <- matrix(combined[!is.na(vec),], ncol=6)
return(sum(apply( combined , 1 , singlegene)))
}
return(apply(testdatared , 2 , NBvectorpredicter))
20 }

myROC <- function(numbers , fac){
n1 <- sum(fac==levels(fac)[1 ])
n2 <- sum(fac==levels(fac)[2])
25 wilcoxresult <- wilcox.test(numbers~fac , alternative="greater")
ROCarea <- as.numeric(wilcoxresult$statistic)/(n1*n2)
pval <- wilcoxresult$p.value
return(c(ROCarea,pval))
}
30

SensitivitySpecificity <- function(numbers, fac){
n1 <- sum(fac==levels(fac)[1])
n2 <- sum(fac==levels(fac)[2])
un <- sort(unique(numbers), decreasing=TRUE)
35 SenSpe <- function(x){
sen <- sum(numbers>=x & fac==levels(fac)[1])/n1
spe <- 1 - sum(numbers>=x & fac==levels(fac)[2])/n2
return(list(Sensitivity=sen,Specificity=spe))
}
40 return(t(sapply(un , SenSpe)))
}

NBloopreparer <- function(data , fac){
nsamples <- ncol(data)
45 ngenes <- nrow(data)
NBtrainedarray <- array(NA , dim=c(ngenes,5,nsamples))
for (i in 1:nsamples){ print(i)
NBtrainedarray[,i] <- NBtrainer(matrix(data[,i],ncol=nsamples-1),fac[-i])
}
50 return(NBtrainedarray)
}

NBleaveoneout <- function(NBtrainedarray , data , fac , topnumber=Inf, logfoldcut=0 , pcut=1){
nsamples <- ncol(data)
55 loglikelihoods <- rep(NA , nsamples)
for (i in 1:nsamples){
loglikelihoods[i]<- NBpredicter(matrix(data[,i],ncol=1),NBtrainedarray![,i],topnumber,logfoldcut,pcut)
}
return(loglikelihoods)
60 }

NBloocv <- function(NBtrainedarray , data , fac , topnumber=Inf, logfoldcut=0 , pcut=1){
n1 <- sum(fac==levels(fac)[1])
n2 <- sum(fac==levels(fac)[2])
65 SampleInformation <- paste(levels(fac)[1], " ",n1," ", levels(fac)[2], " ",n2,sep="")
loglikelihoods <- NBleaveoneout(NBtrainedarray , data,fac,topnumber,logfoldcut,pcut)

```

```

names <- colnames(data , do.NULL=FALSE)
orden <- order(loglikelihoods , decreasing=TRUE)
Samples <- data.frame(names[orden],loglikelihoods!orden],fac!orden])
ROCdata <- myROC(loglikelihoods,fac)
5 SenSpe <- SensitivitySpecificity(loglikelihoods,fac)
return (list(SampleInformation=SampleInformation,ROCarea=ROCdata[1],p.value=ROCdata[2],
             topnumber=topnumber,pcut=pcut,SenSpe <- SenSpe,samples=Samples))
}

10 NBtwooutpreparer <- function(data , fac){
  nsamples <- ncol(data)
  ngenes <- nrow(data)
  NBdoublearray <- array(NA , dim=c(ngenes,5,nsamples*(nsamples-1)/2))
  for (i in 2:nsamples){
    for (j in 1:(i-1)){ print(paste(i," ",j));
15     NBdoublearray[,,(i-1)*(i-2)/2+j] <- NBtrainer(matrix(data[,-c(i,j)],ncol=nsamples-2),fac[-c(i,j)])
    }
  }
  return(NBdoublearray)
20 }

NBmaximizer <- function(NBtrainedarray , data , fac){
  functomaximize <- function(pcut, topnumber){
    NBloocv(NBtrainedarray , data , fac , topnumber=top , pcut=pcut)$ROCarea
25 }
  rocmax <- 0
  pcutmax <- numeric(0)
  topmax <- numeric(0)
  pcutset <- c(1,0.05,0.01,0.005,0.001,0.0003,0.0005,0.0001)
30 topset <- c(1,2,5,10,20,50,100)
  for (pout in pcutset){
    for (top in topset){
      currentroc <- functomaximize(pcut,top); # print(paste(pcut," ",top," ",currentroc))
      if (currentroc >= rocmax){
35       rocmax <- currentroc
       pcutmax <- pcut
       topmax <- top
      }
    }
  }
40 }
  print(paste("Result ".pcutmax," ".topmax," ".rocmax))
  return(c(pcutmax,topmax))
}

45 NBtotalvalidation <- function(NBdoublearray , NBtrainedarray , data ,fac){
  n1 <- sum(fac==levels(fac)[1])
  n2 <- sum(fac==levels(fac)[2])
  nsamples <- n1+n2
  ngenes <- nrow(data)
50 SampleInformation <- paste(levels(fac)[1]," ",n1," ",levels(fac)[2]," ",n2,sep="")
  maxarray <- matrix(NA , nrow=nsamples , ncol=2)
  colnames(maxarray) <- c('pcut','topnumber')
  NormScore <- numeric(nsamples)
  loglikelihoods <-numeric(nsamples)
55 for (i in 1:nsamples){
  NBtemptrainedarray <- array(NA , dim=c(ngenes,5,nsamples-1))
  if (i >1){
    for (j in 1:(i-1)){
      NBtemptrainedarray[,,j] <- NBdoublearray[,,(i-1)*(i-2)/2+j]
60    }
  }
  if (i < nsamples){
    for (j in (i+1):nsamples){
      NBtemptrainedarray[,,j-1] <- NBdoublearray[,,(j-1)*(j-2)/2+i]
65    }
  }
}

```



```

maxarray[i,] <- NBmaximizer( NBtemptrainedarray , data[,i], fac[i])
temploglikelihoods <- NBpredictor( data, NBtrainedarray[,i] , pcut = maxarray[i,1] ,
topnumber=maxarray[i,2])
loglikelihoods[i] <- temploglikelihoods[i]
5   meanll <- mean(temploglikelihoods[-i])
    sdll <- sd(temploglikelihoods[-i])
    NormScore[i] <- (temploglikelihoods[i] - meanll)/sdll
  }
  names <- colnames(data , do.NULL=FALSE)
10  orden <- order( NormScore , decreasing=TRUE)
    Samples <- data.frame(names[orden],NormScore[orden],loglikelihoods[orden],fac[orden] ,maxarray[orden,])
    ROCdata <- myROC(NormScore,fac)
    SenSpe <- SensitivitySpecificity(NormScore,fac)
    return(list(SampleInformation=SampleInformation,ROCarea=ROCdata[1],p.value=ROCdata[2],
15      ,SenSpe <- SenSpe,samples=Samples))
  }

```

```

ROCplot <- function(clasRes , sensspecnumber=6){
Sensitivity <- as.numeric(clasRes[[sensspecnumber]][,1])
20  Specificity <- as.numeric(clasRes[[sensspecnumber]][,2])
    OneMinusSpecificity <- 1- Specificity
    ROCarea <- round(clasRes$ROC,digits=2)
    plot(OneMinusSpecificity , Sensitivity , type="l", xlab="1-specificity", ylab="sensitivity")
    title(paste("ROC area = ", ROCarea),font.main=1)
25  }

```

```

ROCplotReverse <- function(clasRes){
Sensitivity <- rev(as.numeric(clasRes[[4]][,2]))
Specificity <- rev(as.numeric(clasRes[[4]][,1]))
30  OneMinusSpecificity <- 1- Specificity
    ROCarea <- round(clasRes$ROC,digits=2)
    plot(OneMinusSpecificity , Sensitivity , type="l", xlab="1-specificity", ylab="sensitivity")
    title(paste("ROC area = ",ROCarea),font.main=1)
  }
35

```

(iii) ChP-PaC train1+test1 model

```

> str(svmtrain)
List of 29
40  $ call      : language svm.default(x = t(training[aprioriBoolean, ]), y = facTr, kernel = "linear")
    $ type     : num 0
    $ kernel   : num 0
    $ cost     : num 1
    $ degree   : num 3
45  $ gamma    : num 0.04
    $ coef0    : num 0
    $ nu       : num 0.5
    $ epsilon  : num 0.1
    $ sparse   : logi FALSE
50  $ scaled   : logi [1:25] TRUE TRUE TRUE TRUE TRUE TRUE ...
    $x.scale   :List of 2
    ..$ scaled:center: Named num [1:25] 3.14 3.42 3.28 3.33 2.66 ...
    .. - attr(*, "names")= chr [1:25] "ApoA1_001_C08" "APOA4_5" "C1e_032_A11" "C5.12"...
    ..$ scaled.scale : Named num [1:25] 0.117 0.157 0.183 0.109 0.112 ...
55  .. - attr(*, "names")= chr [1:25] "ApoA1_001_C08" "APOA4_5" "C1e_032_A11" "C5.12"...
    $ y.scale   : NULL
    $ nclasses  : int 2
    $ levels    : chr [1:3] "ChP" "PaC" "rest"
    $ tot.nSV   : int 97
60  $ nSV       : int [1:2] 49 48
    $ labels    : int [1:2] 12
    $ SV        : num [1:97, 1:25]-0.407 0.147-0.912-0.256 0.478 ...
    ..- attr(*, "dimnames")=List of 2
    ...$: chr [1:97] "R2S1B7" "R4S5A1" "R1S1B4" "R1S8A3"...
65  ..$: chr [1:25] "ApoA1_001_C08" "APOA4_5" "C1e_032_A11" "C5.12"...
    $ index     : int [1:97] 1 3 6 8 13 16 20 21 22 23 ...

```

```

$ rho      : num 0.0805
$ compprob : logi FALSE
$ probA    : NULL
$ probB    : NULL
5 $ sigma   : NULL
$ coefs    : num [1:97, 1] 1 1 0.0466 1 1 ...
$ na.action : NULL
$ fitted   : Factor w/ 3 levels "ChP","PaC","rest": 1 1 1 1 1 1 2 1 1...
attr(*, "names")= chr [1:204] "R2S1B7" "R1S2A6" "R4S5A1" "R3S5A2"...
10 $ decision.values: num [1:204, 1] 0.935 2.207 0.329 3.534 1.665 ...
..attr(*, "dimnames")=List of 2
.. ..$ : chr [1:204] "R2S1B7" "R1S2A6" "R4S5A1" "R3S5A2"...
.. ..$ : chr "ChP/PaC"
- attr(*, "class")= chr "svm"
15 >

```

(iv) N-PaC train1+test1 model

```

> str(svmtrain)
20 List of 29
  $ call      : language svm.default(x = t(training[aprioriBoolean, ]), y = facTr, kernel = "linear")
  $ type      : num 0
  $ kernel    : num 0
  $ cost      : num 1
25 $ degree    : num 3
  $ gamma     : num 0.04
  $ coef0     : num 0
  $ nu        : num 0.5
  $ epsilon   : num 0.1
30 $ sparse    : logi FALSE
  $ scaled    : logi [1:25] TRUE TRUE TRUE TRUE TRUE TRUE ...
  $ x.scale    : List of 2
..$ scaled:center: Named num [1:25] 3.53 3.37 3.11 3.26 3.72 ...
.. attr(*, "names")= chr [1:25] "Angiomotin.1" "APOA4_2" "BITM8.001.B04" "C1e_032_A11"...
35 ..$ scaled:scale : Named num [1:25] 0.135 0.186 0.235 0.192 0.177 ...
.. attr(*, "names")= chr [1:25] "Angiomotin.1" "APOA4_2" "BITM8.001.B04" "C1e_032_A11"...
  $ y.scale    : NULL
  $ nclasses   : int 2
  $ levels     : chr [1:3] "N" "PaC" "rest"
40 $ tot.nSV    : int 15
  $ nSV        : int [1:2] 8 7
  $ labels     : int [1:2] 1 2
  $ SV         : num [1:15, 1:25]-0.392-1.505 0.246 0.174 1.489 ...
..- attr(*, "dimnames")=List of 2
45 .. ..$ : chr [1:15] "R3S6A1" "R3S1A5" "R4S5A5" "R3S5A3"...
.. ..$ : chr [1:25] "Angiomotin.1" "APOA4_2" "BITM8.001.B04" "C1e_032_A11"...
  $ index      : int [1:15] 1 9 14 15 16 17 19 20 27 69...
  $ rho        : num 1.4
  $ compprob   : logi FALSE
50 $ probA      : NULL
  $ probB      : NULL
  $ sigma      : NULL
  $ coefs      : num [1:15, 1] 0.0642 0.0156 0.0414 0.0302 0.0184 ...
  $ na.action   : NULL
55 $ fitted     : Factor w/3 levels "N","PaC","rest": 1 1 1 1 1 1 1 1 1 ...
attr(*, "names")= chr [1:124] "R3S6A1" "R6S1A3" "R3S4B3" "R6S2B1"...
$ decision.values: num [1:124, 1] 1 1.88 1.2 1.67 2.61 ...
..- attr(*, "dimnames")=List of 2
.. ..$ : chr [1:124] "R3S6A1" "R6S1A3" "R3S4B3" "R6S2B1"...
60 .. ..$ : chr "N/PaC"
- attr(*, "class")= chr "svm"
>

```

(v) Script ChP-PaC train1+test1

```

65 apri<-c("IL-1a-145",

```

```

"      IL-4-55 ",
"      TGF-b1-34      ",
"      C-BTK-1",
"      C-UBE2C-1      ",
5  "      C-PTPN1-3      ",
"      CT17      ",
"      C5-12      ",
"      Sox11A_A6      ",
"      C-HADH2-8      ",
10  "      MCP1_001_A04      ",
"      C-TNFRSF3-1      ",
"      C-UCHL5-1      ",
"      TNF-a-89      ",
"      C-TNFRSF3-2      ",
15  "      Eotaxin-2      ",
"      MCP1_005_A11      ",
"      C1e_032_A11      ",
"      APOA4_5      ",
"      ApoA1_001_C08      ",
20  "      C-MK08-1      ",
"      C-FASN-6      ",
"      IL-16-1      ",
"      C-Keratin19-3      ",
"      C-GAK-5      ")

25  library(MASS)
    library(gplots)
    library(e1071)
    #source("C:/Program/R/R-2.8.1/library/NaiveBayesian") source(" NaiveBayesian")

30  filnamn<-"serum ChP-PaC trainl+testl.txt"
    rawfile <- read.delim(filnamn)
    samplenames <- as.character(rawfile[,1])
    groups <- rawfile[,2]
35  data <- t(rawfile[,c(1,2)])
    ProteinNames <- read.delim(filnamn,header=FALSE)
    ProteinNames <- as.character(as.matrix(ProteinNames)[1,])
    ProteinNames <- ProteinNames[-(1:2)]
    rownames(data) <- ProteinNames
40  colnames(data) <- samplenames

    group1 <- "ChP"
    group2 <- "PaC"

45  nTrainingSamples <- 204
    nTestSamples <-102

    training <- data[,1 :nTrainingSamples ]
    test <- data[(nTrainingSamples+1):(nTrainingSamples+nTestSamples)]

50  aprioriBoolean <- is.element(rownames(data), apri)
    facTr<- factor(rep("rest",ncol(training)),levels=c(group1, group2, "rest"))
    subset1Tr <- is.element(groups[1:nTrainingSamples], group1)
    subset2Tr <- is.element(groups[1:nTrainingSamples] , group2)
55  facTr[subset1Tr] <- group1
    facTr[subset2Tr] <- group2

    facTe <- factor(rep("rest",ncol(test)),levels=c(group1, group2, "rest"))
    subset1Te <- is.element(groups[(nTrainingSamples+1):(nTrainingSamples+nTestSamples)]
60  strsplit(group1,",")[[1]])
    subset2Te <- is.element(groups[(nTrainingSamples+1):(nTrainingSamples+nTestSamples)]
    strsplit(group2,",")[[1 ]])
    facTe[subset1Te] <- group1
    facTe[subset2Te] <- group2

65  svmtrain <- svm(t(training[aprioriBoolean,]) , facTr, kernel="linear")

```

```

pred <- predict(svmtrain, t(test[aprioriBoolean,]), decision.values = TRUE, probability = T)
res <- as.numeric(attributes(pred)$decision.values, probability = T)
facnames <- colnames(attributes(pred)$decision.values)[1]

```

```

5      ROCdata <- myROC(res,facTe)
      ROCdata[1]

      SenSpe <- SensitivitySpecificity(res,facTe)
      Sensitivity <- as.numeric(SenSpe[,1])
10     Specificity <- as.numeric(SenSpe[,2])
      omSpecificity <- 1-Specificity
      plot(omSpecificity, Sensitivity, ylab="Sensitivity", xlab="1-Specificity",type="l")
      mtext(side=1, line = -1.1, pastefROC AUC = ",signif(ROCdata[1], digits=2)))

15     ROCdata

      Sensitivity+Specificity
      max(Sensitivity+Specificity)

20     SenSpe

```

(vi) Script N-PaC train1+test1

```

25     apri<-c("C-IL6-5 ",
      "      CystC_002_A01 ",
      "      IL-8-10 ",
      "      IL-11-42",
      "      C1e_032_A11 ",
      "      Eotaxin-2 ",
30     "      C-HADH2-8 ",
      "      FB_009_B01 ",
      "      C5-9 ",
      "      CD40-44 ",
      "      C-CDK2-1 ",
35     "      FN3-001-B04 ",
      "      Angiomotin-1 ",
      "      LDL-2 ",
      "      C5_029_C08 ",
      "      C1q-4 ",
40     "      IL-1 a-122 ",
      "      C-CDK2-2 ",
      "      C1e_032_E05 ",
      "      APOA4_2 ",
      "      FB_009_E05 ",
45     "      C-P85A-4 ",
      "      C-BTK-1",
      "      BITM8-001-B04 ",
      "      IL-13-16")

50     library(MASS)
      library(gplots)
      library(e1071)
      #source("C:/Program/R/R-2.8.1/library/NaiveBayesian")
      source ("NaiveBayesian")

55     filnamn<-"serum N-PaC train 1+test1.txt"
      rawfile <- read.delim(filnamn)
      samplenames <- as.character(rawfile[,1])
      groups <- rawfile[,2]
60     data <- t(rawfile[,-c(1,2)])
      ProteinNames <- read.delim(filnamn,header=FALSE)
      ProteinNames <- as.character(as.matrix(ProteinNames)[1,])
      ProteinNames <- ProteinNames[-(1:2)]
      rownames(data) <- ProteinNames
65     colnames(data) <- samplenames

```

```

group1 <- "N"
group2 <- "PaC"

nTrainingSamples <- 124
5 nTestSamples <- 62

training <- data[,1:nTrainingSamples]
test <- data[(nTrainingSamples+1):(nTrainingSamples+nTestSamples)]

10 aprioriBoolean <- is.element(rownames(data), apri)
facTr <- factor(rep("rest",ncol(training)),levels=c(group1, group2, "rest"))
subset1Tr <- is.element(groups[1:nTrainingSamples], group1)
subset2Tr <- is.element(groups[1:nTrainingSamples], group2)
facTr[subset1Tr] <- group1
15 facTr[subset2Tr] <- group2

facTe <- factor(rep("rest",ncol(test)),levels=c(group1, group2, "rest"))
subset1Te <- is.element(groups[(nTrainingSamples+1):(nTrainingSamples+nTestSamples)]
,
strsplit(group1,"")[[1]])
20 subset2Te <- is.element(groups[(nTrainingSamples+1):(nTrainingSamples+nTestSamples)]
,
strsplit(group2,"")[[1]])
fade[subset1Te] <- group1
facTe[subset2Te] <- group2

25 svmtrain <- svm(t(training[aprioriBoolean,]), fadr, kernel="linear")
pred <- predict(svmtrain, t(test[aprioriBoolean,]), decision.values = TRUE, probability = T)
res <- as.numeric(attributes(pred)$decision.values, probability = T)
facnames <- colnames(attributes(pred)$decision.values)[1]
ROCdata <- my ROC (res,fade)
30 ROCdata[1]

SenSpe <- SensitivitySpecificity(res.facTe)
Sensitivity <- as.numeric(SenSpe[, 1 ])
Specificity <- as.numeric(SenSpe[,2])
35 omSpecificity <- 1-Specificity
plot(omSpecificity, Sensitivity, ylab="Sensitivity", xlab="1-Specificity",type="l")
mtext(side=1, line = -1.1, paste(ROC AUC = ",signif(ROCdata[1], digits=2)))

ROCdata

40 Sensitivity+Specificity
max(Sensitivity+Specificity)

SenSpe
45

```

Tabla C – Ejemplo de poder discriminatorio de biomarcadores y combinaciones de biomarcadores (PC frente a NC)

ROC-AUC	Firma de biomarcador
0,9	HADH2
0,83	TNFRSF3
0,93	HADH2+TNFRSF3
0,96	HADH2+TNFRSF3+GAK
0,95	HADH2+TNFRSF3+GAK+UPF3B
0,95	HADH2+TNFRSF3+GAK+UPF3B+Integrina a-10

REIVINDICACIONES

1. Un método para determinar la presencia de cáncer de páncreas en un individuo, que comprende o que consiste en las etapas de:
 - a) proporcionar una muestra para su análisis obtenida de un individuo;
 - b) determinar una firma de biomarcador de la muestra de ensayo midiendo la expresión de la muestra de ensayo de dos o más biomarcadores seleccionados del grupo definido en la Tabla A;
- en donde la etapa (b) comprende la medición de la expresión en la muestra de HADH2, y en donde la expresión en la muestra de ensayo de dos o más biomarcadores seleccionados del grupo definido en la Tabla A es indicativo de la presencia de cáncer de páncreas.
2. El método de acuerdo con la reivindicación 1, que comprende o consiste además en las etapas de:
 - c) proporcionar una muestra de control obtenida de un individuo no afligido por cáncer de páncreas;
 - d) determinar una firma de biomarcador de la muestra de control midiendo la expresión de la muestra de control de dos o más biomarcadores medidos en la etapa (b);
- en donde la presencia de cáncer de páncreas se identifica en caso de que la expresión en la muestra de ensayo de dos o más biomarcadores medidos en la etapa (b) sea diferente de la expresión en la muestra de control de dos o más biomarcadores medidos en la etapa (d).
3. El método de acuerdo con la reivindicación 2, que comprende o consiste además en las etapas de:
 - e) proporcionar una muestra de control obtenida de un individuo afligido por cáncer de páncreas;
 - f) determinar una firma de biomarcador de la muestra de control midiendo la expresión en la muestra de control de dos o más biomarcadores medidos en la etapa (b);
- en donde la presencia de cáncer de páncreas se identifica en el caso de que la expresión en la muestra de ensayo de dos o más biomarcadores medidos en la etapa (b) se corresponda con la expresión en la muestra de control de dos o más biomarcadores medidos en la etapa (f).
4. El método de acuerdo con una cualquiera de las reivindicaciones anteriores, en donde la etapa (b) comprende además medir la expresión de
 - (a) TNFRSF3;
 - (b) uno o más de los biomarcadores enumerados en la Tabla A(ii);
 - (c) uno o más biomarcadores de los biomarcadores enumerados en Tabla A(iii);
 - (d) uno o más biomarcadores de los biomarcadores enumerados en Tabla A(iv);
 - (e) uno o más biomarcadores de los biomarcadores enumerados en Tabla A(v) ; y/o
 - (f) uno o más biomarcadores de los biomarcadores enumerados en Tabla A(vi).
5. El método de acuerdo con una cualquiera de las reivindicaciones anteriores en donde la etapa (b) comprende o consiste en medir la expresión en la muestra de ensayo de todos los biomarcadores definidos en la Tabla A.
6. El método de acuerdo con una cualquiera de las reivindicaciones anteriores en donde el cáncer de páncreas se selecciona del grupo que consiste en adenocarcinoma, carcinoma adenoescamoso, carcinoma de célula en anillo de signet, carcinoma hepatoide, carcinoma coloide, carcinoma no diferenciado y carcinomas no diferenciados con células gigantes de tipo osteoclasto.
7. El método de acuerdo con una cualquiera de las reivindicaciones anteriores en donde la etapa (b), (d) y/o la etapa (f) se llevan a cabo utilizando primeros agentes de unión capaces de unirse a los dos o más biomarcadores, en donde cada uno de los primeros agentes de unión comprende o consiste en un anticuerpo o un fragmento de unión a antígeno del mismo.
8. El método de acuerdo con una cualquiera de las reivindicaciones anteriores en donde los dos o más biomarcadores en la muestra de ensayo y/o la muestra de control se marcan con una fracción detectable.
9. El método de acuerdo con una cualquiera de las reivindicaciones 7 u 8 en donde las etapas (b), (d) y/o la etapa (f) se llevan a cabo utilizando un ensayo que comprende segundos agentes de unión capaces de unirse a los dos o más biomarcadores, comprendiendo cada uno de los segundos agentes de unión una fracción detectable, en donde cada uno de los segundos agentes de unión comprende o consiste en un anticuerpo o un fragmento de unión a antígeno del mismo.
10. El método de acuerdo con una cualquiera de las reivindicaciones anteriores en donde el método comprende:

- (i) marcar biomarcadores presentes en la muestra con biotina;
 - (ii) poner en contacto las proteínas marcadas con biotina con una matriz que comprende una pluralidad de scFv inmovilizados en emplazamientos diferenciados en su superficie, teniendo scFv especificidad para una o más de las proteínas de la Tabla A;
 - 5 (iii) poner en contacto el scFv inmovilizado con un conjugado de estreptavidina que comprende un colorante fluorescente; y
 - (iv) detectar la presencia del colorante en emplazamientos diferenciados en la superficie de matriz
- en donde la expresión del colorante en la superficie de matriz es indicativa de la expresión de un biomarcador de la
- 10 Tabla A en la muestra.
11. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 6 en donde las etapas (b), (d) y/o (f) comprenden medir la expresión de una molécula de ácido nucleico que codifica dos o más biomarcadores.
- 15 12. El método de acuerdo con la reivindicación 11, en donde la medición de la expresión de los dos o más biomarcadores en las etapas (b), (d) y/o (f) se lleva a cabo utilizando dos o más fracciones de unión capaces cada una de ellas individualmente de unirse selectivamente a una molécula de ácido nucleico que codifica uno de los biomarcadores identificados en la Tabla A, en donde cada una de las una o más fracciones de unión comprende o
- 20 consiste en una molécula de ácido nucleico.
13. El método de acuerdo con una cualquiera de las reivindicaciones anteriores en donde la muestra proporcionada en las etapas (a), (c) y/o (e) se selecciona del grupo que consiste en sangre sin fraccionar, plasma, suero, fluido tisular, tejido pancreático, jugos pancreáticos, bilis y orina.
- 25 14. Uso de una matriz que comprende dos o más agentes de unión tal como se definen en las reivindicaciones 7 o 9 para determinar la presencia de cáncer de páncreas en un individuo, en donde la matriz comprende un agente de unión capaz de unirse a HADH2, en donde cada uno de los agentes de unión comprende o consiste en un anticuerpo o un fragmento de unión a antígeno del mismo.
- 30 15. Uso de dos o más biomarcadores seleccionados del grupo definido en la Tabla A como marcadores de diagnóstico para determinar *in vitro* la presencia de cáncer de páncreas en un individuo, en donde los dos o más biomarcadores incluyen HADH2.

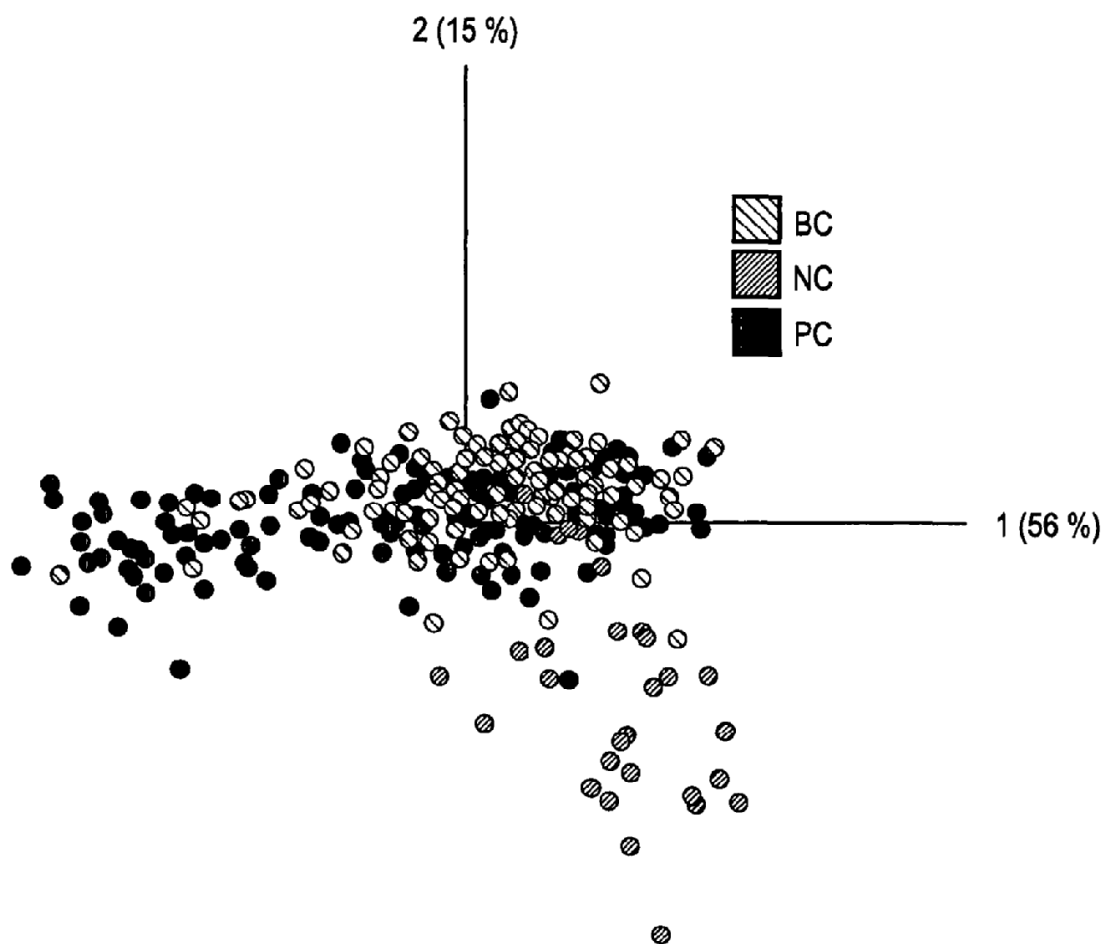


FIG. 1

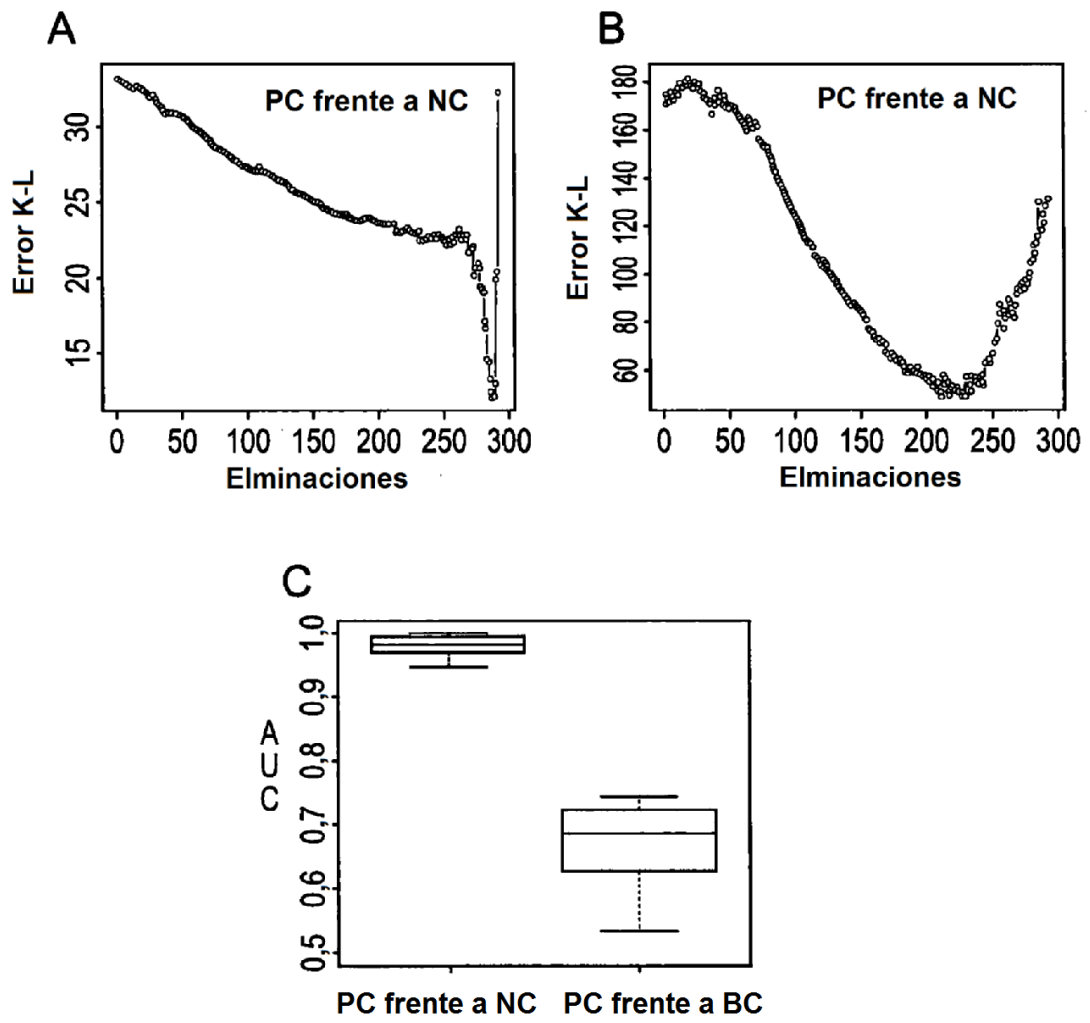


FIG. 2

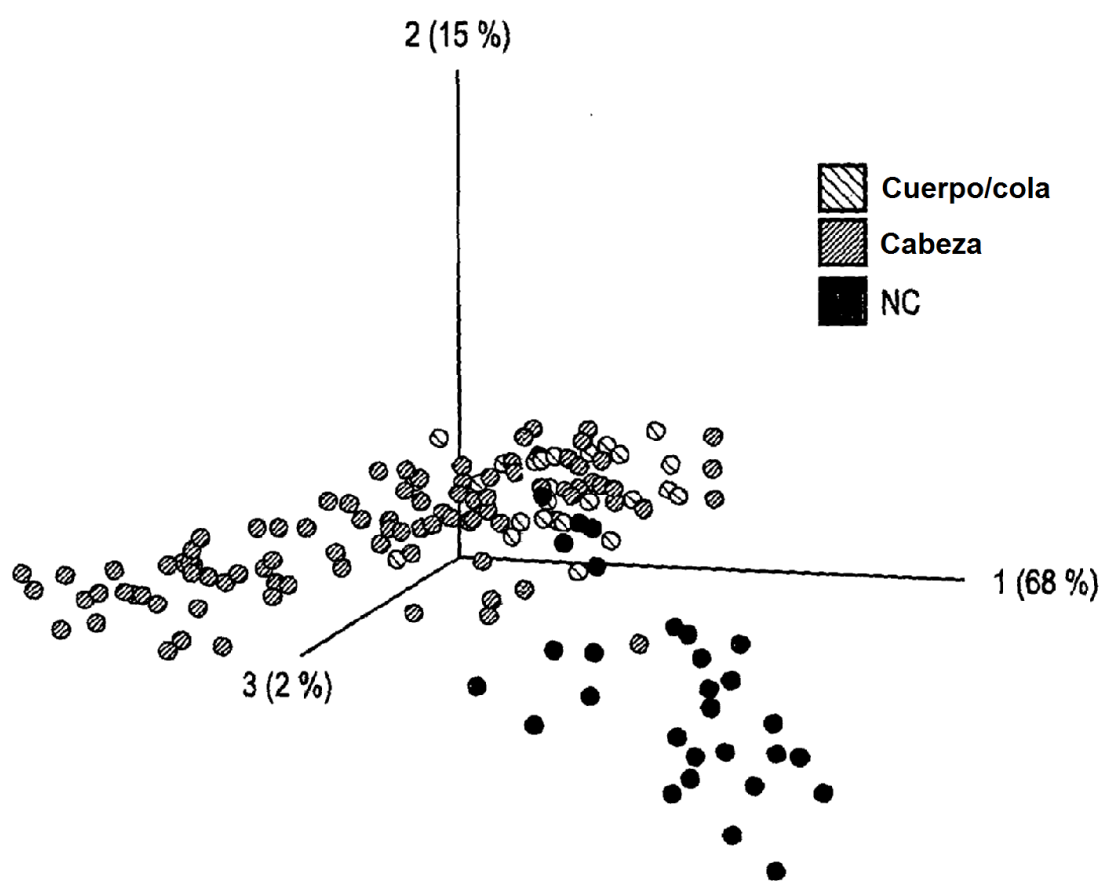


FIG. 3

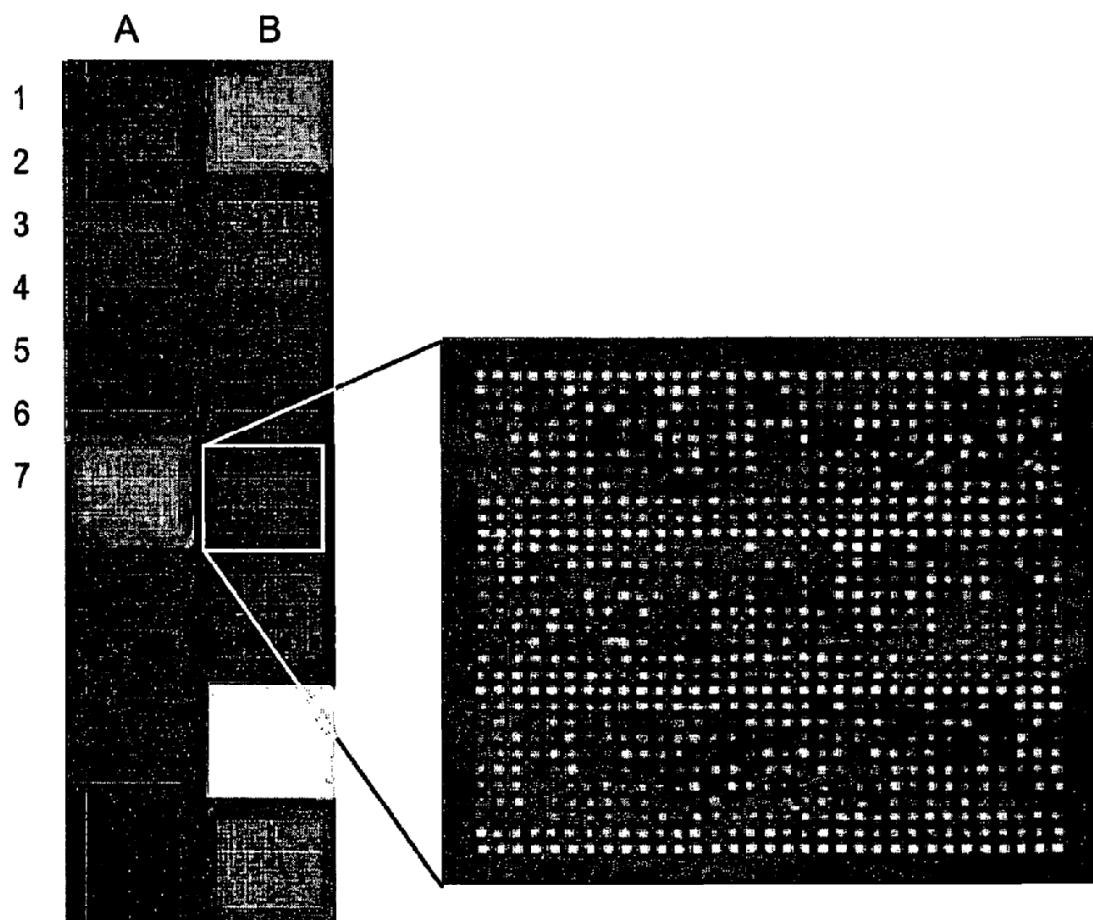


FIG. 4

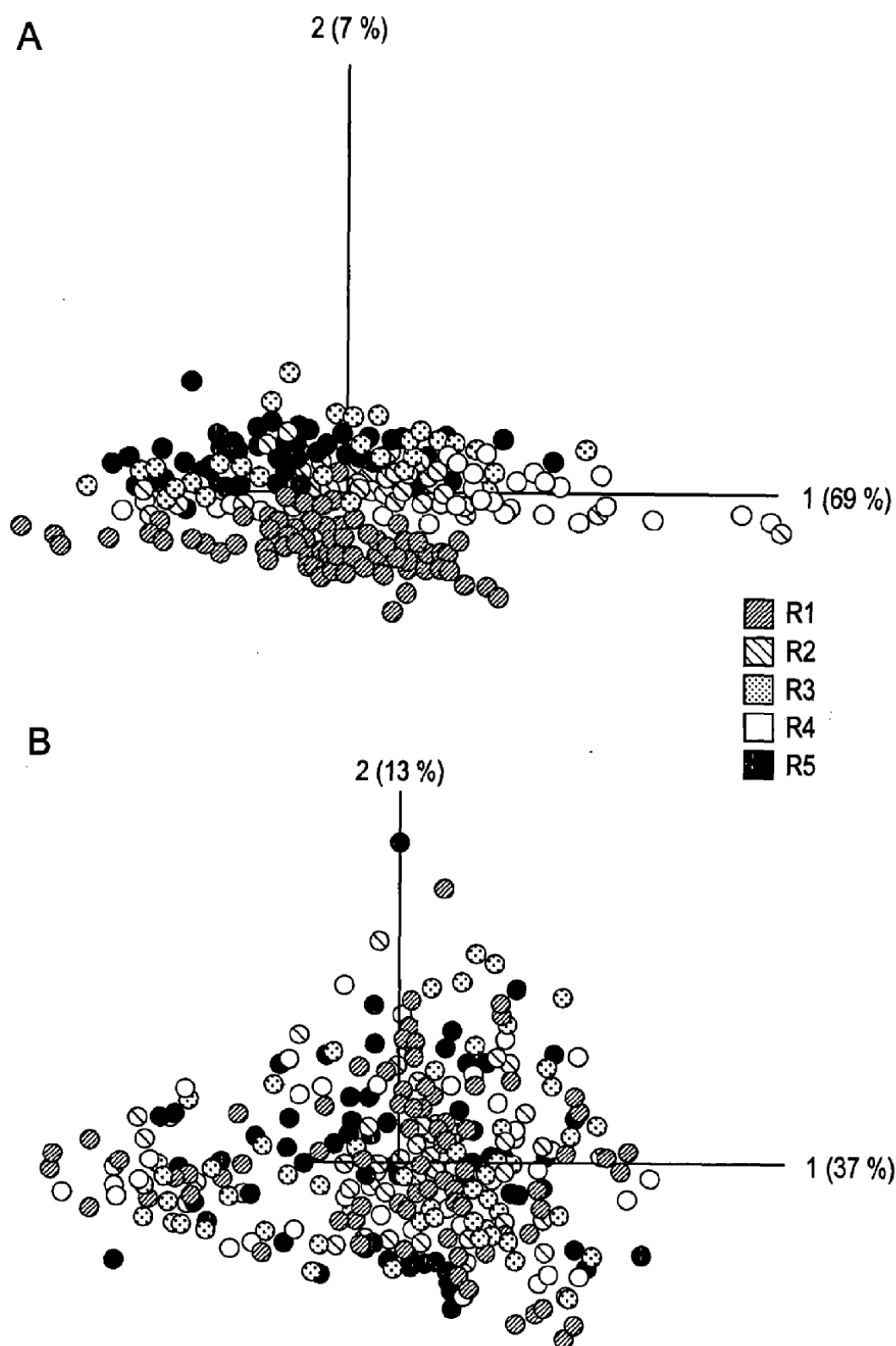


FIG. 5



FIG. 6