

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 667 326**

51 Int. Cl.:

C12Q 1/68 (2008.01)

G01N 33/574 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.11.2013** **PCT/EP2013/074259**

87 Fecha y número de publicación internacional: **30.05.2014** **WO14079865**

96 Fecha de presentación y número de la solicitud europea: **20.11.2013** **E 13792689 (5)**

97 Fecha y número de publicación de la concesión europea: **03.01.2018** **EP 2922967**

54 Título: **Procedimiento para indicar la presencia o no presencia de cáncer de próstata agresivo**

30 Prioridad:

20.11.2012 SE 1251312

16.05.2013 SE 1350600

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
10.05.2018

73 Titular/es:

PHADIA AB (100.0%)

Box 6460

751 37 Uppsala, SE

72 Inventor/es:

GRÖNBERG, HENRIK y

EKLUND, MARTIN

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 667 326 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

Procedimiento para indicar la presencia o no presencia de cáncer de próstata agresivo

5 CAMPO DE LA INVENCION

La presente invención se refiere en general a la detección e identificación de diversas formas de marcadores genéticos y a diversas formas de proteínas que tienen posible utilidad como marcadores diagnósticos. En particular, la presente invención se refiere al uso simultáneo de múltiples marcadores diagnósticos para una mejor detección de formas agresivas de cáncer de próstata. Más en particular, la presente invención se refiere al uso simultáneo de múltiples marcadores diagnósticos para una mejor detección del cáncer de próstata agresivo en varones con un índice de masa corporal (IMC) por encima de 25.

ANTECEDENTES DE LA INVENCION

La medición del antígeno prostático específico (PSA) en suero es ampliamente utilizada para el cribado y detección precoz del cáncer de próstata (CaP). Como se discute en el artículo «Polygenic Risk Score Improves Prostate Cancer Risk Prediction: Results from the Stockholm-1 Cohort Study» de Markus Aly y coautores publicado en EUROPEAN UROLOGY 60 (2011) 21-28, el PSA sérico que se puede medir con los inmunoensayos clínicos actuales está principalmente en su forma libre «no formando complejos» (PSA libre) o formando un complejo con alantiquimotripsina (ACT). Se ha demostrado que la relación del PSA libre al total en suero mejora significativamente la detección del CaP. Otros factores, como la edad y los antecedentes familiares documentados, pueden mejorar adicionalmente la detección del CaP. La medición de los marcadores genéticos relacionados con el CaP, en particular los polimorfismos de un solo nucleótido (SNP, por sus siglas en inglés), es una modalidad emergente para el cribado y detección precoz del cáncer de próstata. El análisis de múltiples SNP relacionados con el CaP puede, en combinación con biomarcadores como PSA y con información general sobre el paciente, mejorar la evaluación del riesgo a través de una combinación de varios SNP en una puntuación genética.

El cribado y detección precoz del cáncer de próstata es una tarea complicada y, hasta la fecha, no se ha demostrado que ningún biomarcador individual sea suficientemente bueno para el mapeo específico y sensible de la población masculina. Por tanto, se han hecho intentos de combinar niveles de biomarcadores para obtener una fórmula que funcione mejor en el cribado y detección precoz del CaP. El ejemplo más frecuente es la prueba de PSA regular, que de hecho es una evaluación del PSA «libre» y el PSA «total». Existe una forma de PSA que no forma complejos y una forma en la que el PSA forma complejo con alfa-lantiquimotripsina. Otro de estos ejemplos es el uso de combinaciones de concentraciones de PSA libre, PSA total y una o más formas proenzima de PSA para los fines de diagnóstico, como se describe en el documento WO03100079 (METHOD OF ANALYZING PROENZYME FORMS OF PROSTATE SPECIFIC ANTIGEN IN SERUM TO IMPROVE PROSTATE CANCER DETECTION). La combinación posible de concentraciones de PSA y concentraciones de proenzima que puede resultar en un mejor rendimiento del cribado y la detección precoz del CaP es el índice de salud prostática (PHI, por sus siglas en inglés). El PHI se desarrolló como una combinación de PSA, PSA libre y una forma precursora de PSA [-2] proPSA para detectar mejor el CaP para hombres con un resultado en la prueba de PSA dudoso (p, ej., PSA 2-10ng/mL) y tacto rectal no sospechoso, como se describe en el artículo «Cost-effectiveness of Prostate Health Index for prostate cancer detection» de Nichol MB y coautores publicado en BJU Int. 2011 Nov 11. doi: 10.1111/j.1464-410X.2011.10751.x. Otro ejemplo es la combinación de psp94 y PSA, como se describe en el documento US2012021925 (DIAGNOSTIC ASSAYS FOR PROSTATE CANCER USING PSP94 AND PSA BIOMARKERS).

Existen otros biomarcadores con posible valor diagnóstico o pronóstico para valorar si un paciente sufre de CaP, como MIC-1 que se describe en el artículo «Macrophage Inhibitory Cytokine 1: A New Prognostic Marker in Prostate Cancer» de David A. Brown y coautores publicado en Clin Cancer Res 2009;15(21):OF1-7.

En el pasado se han descrito intentos de combinar información procedente de múltiples fuentes en un único modelo de algoritmo para la predicción del riesgo de CaP. En el artículo «Blood Biomarker Levels to Aid Discovery of Cancer-Related Single-Nucleotide Polymorphisms: Kallikreins and Prostate Cancer» de Robert Klein y coautores publicado en Cancer Prev Res (2010), 3(5):611-619, los autores debaten cómo los biomarcadores en sangre pueden ayudar al descubrimiento de nuevos SNP, pero también sugieren que existe un papel potencial para incorporar tanto el genotipo como los niveles de biomarcadores en los modelos predictivos. Asimismo, en este artículo se proporcionan evidencias de que la combinación no aditiva de marcadores genéticos y biomarcadores de manera conjunta pueden tener valor predictivo para la estimación del riesgo de CaP. Later, Xu y coinventores describieron un procedimiento para correlacionar marcadores genéticos con el cáncer de próstata de alto grado, principalmente con el fin de identificar sujetos idóneos para quimioterapia preventiva usando un medicamento inhibidor de la 5-alfa

reductasa (p. ej., dutasterida o finasterida) en la solicitud de patente WO2012031207. En conjunto, estas dos memorias descriptivas públicas resumen la técnica previa de la combinación de información genética y concentración de biomarcadores para los fines de estimación del riesgo de CaP, también para cánceres de alto grado.

- 5 Asimismo, Mattias Johansson *et al.* («Combining 33 genetic variants with prostate-specific antigen for prediction of prostate cancer», International J. Cancer 130,129-137, 2012) describen un estudio que investiga si una puntuación del riesgo genético que incluye 33 variantes genéticas frecuentes mejora la predicción del cáncer de próstata cuando se añade a las medidas del PSA. Se halló que la adición de una puntuación del riesgo genético al PSA daba lugar a una mejora marginal en la predicción del cáncer de próstata. I.M. Thompson *et al.* («Assessing prostate cancer risk: results from the prostate cancer prevention trial», JNCI, Journal of the National Cancer Institute, vol. 98, n.º 8, 18 de abril de 2006, p.529-534) utilizaron los datos de biopsias de próstata procedentes de hombres que participaron en el ensayo clínico de prevención del cáncer de próstata (PCPT, por sus siglas en inglés) para desarrollar un modelo predictivo del cáncer de próstata basado en variables que predecían el cáncer de próstata. En este modelo no se utilizó una puntuación genética.

El rendimiento actual del cribado y detección precoz con PSA es aproximadamente una sensibilidad del 80 % y especificidad del 30 %. Se estima que aproximadamente el 65 % se someterá innecesariamente a una biopsia de próstata y que el 15-20 % de los cánceres de próstata clínicamente importantes no se detectan con el cribado actual.

- 20 Solo en Estados Unidos se realizan aproximadamente un millón de biopsias al año, cuyo resultado es alrededor de 192.000 nuevos casos diagnosticados. Por lo tanto, también una pequeña mejora en el rendimiento diagnóstico dará como resultado ahorros importantes en los gastos de atención sanitaria debido a la menor cantidad de biopsias y al menor sufrimiento humano derivado de procedimientos diagnósticos invasivos.
- 25 La práctica clínica actual (en Suecia) es usar PSA total como biomarcador para la detección de cáncer de próstata asintomático e incipiente. El valor de corte general para la evaluación adicional con una biopsia de próstata es de 3 ng/mL. Sin embargo, debido a las consecuencias negativas del cribado con PSA, no existe un cribado con PSA organizado recomendado actualmente en Europa y Norteamérica.
- 30 Es especialmente importante identificar de forma precisa el cáncer de próstata agresivo (CaPa) en los individuos porque cuanto antes se le proporcione tratamiento a un individuo, mayor es la probabilidad de que el cáncer se cure. No obstante, la identificación del CaPa es difícil, en parte porque se requieren cohortes más grandes para proporcionar un número suficiente de casos y controles en el desarrollo de modelos estadísticos. Por tanto, la disponibilidad de modelos predictivos para el CaPa es baja. Sin embargo, esta invención proporciona modelos predictivos para la identificación del CaPa a través del análisis de biomarcadores y del perfil genético de un individuo.

RESUMEN DE LA INVENCION

- 40 La presente invención se basa en el descubrimiento de que la combinación de marcadores diagnósticos de diferente origen puede mejorar la capacidad de detectar el CaPa en una población general. En particular, la presente invención mejora la capacidad de detectar CaPa en individuos con un índice de masa corporal (IMC) elevado. Esto puede dar lugar a ahorros importantes para la sociedad, puesto que los cánceres agresivos que se identifican precozmente son más fáciles de tratar.

- 45 En consecuencia, según los descubrimientos de la presente invención, un aspecto de la presente invención proporciona un procedimiento basado en una combinación de datos diseñada de forma redundante para indicar la presencia o no presencia de cáncer de próstata (CaP) agresivo en un individuo, que comprende los pasos de:

- 50 1. Proporcionar al menos una muestra biológica de dicho individuo;

2. Analizar en dicha muestra biológica:

- a. una categoría de biomarcadores de CaP, midiendo la presencia o la concentración de al menos tres biomarcadores relacionados con CaP de tipo caliceína de dicha categoría de biomarcadores de CaP;
- 55 b. una categoría de SNP relacionados con el CaP (SNPcp), midiendo la presencia o ausencia de uno o dos alelo(s) de riesgo de cada uno de los diversos SNPcp de dicha categoría de SNPcp;

3. Combinar los datos relativos a al menos tres biomarcadores de CaP de tipo caliceína para formar un valor compuesto de biomarcador que represente el riesgo de desarrollar CaP relacionado con biomarcadores de CaP, donde el procedimiento permite ignorar datos de un subconjunto de al menos uno de dichos biomarcadores de CaP,
- 60

como por ejemplo uno o dos de dichos biomarcadores de CaP, al formar dicho valor compuesto de biomarcadores de al menos tres biomarcadores de CaP;

4. Combinar los datos relativos a dicha categoría de SNPcp para formar un valor compuesto de SNPcp que represente el riesgo de desarrollar CaP relacionado con SNPcp, donde el procedimiento permita ignorar datos de un subconjunto de aproximadamente el 10 % al 30 %, como por ejemplo el 15 %, 20 % o 30 % de los SNPcp de la categoría de SNPcp, al formar el valor compuesto de SNPcp, donde dicho valor compuesto de SNPcp se forma a partir de al menos 70 SNPcp seleccionados de entre cualquiera de los SNPcp de la tabla I o la tabla II;
5. Combinar el valor compuesto de biomarcadores y el valor compuesto de SNPcp para formar un valor compuesto global;
6. Correlacionar dicho valor compuesto global con la presencia o no presencia de CaP agresivo en dicho individuo comparando dicho valor compuesto global con un valor de corte predeterminado establecido con muestras de control de presencia conocida de CaP agresivo y de no presencia conocida de CaP agresivo, donde al menos uno, como por ejemplo dos, de dichos biomarcadores de CaP de tipo calicreína se selecciona de entre el grupo compuesto por (i) PSA, (ii) PSA total (tPSA), (iii) PSA intacto (iPSA), (iv) PSA libre (fPSA) y (v) hK2.

Según un aspecto de la invención, una o más de las etapas del procedimiento, normalmente las etapas 3, 4, 5 y/o 6, se proporcionan por medio de un programa informático ejecutado en un ordenador que comprende un procesador y memoria.

Habitualmente, la etapa 3 del procedimiento descrito anteriormente se realiza con un ordenador programado para formar o calcular un valor compuesto de biomarcador a partir de los datos de la etapa 2a; la etapa 4 del procedimiento descrito anteriormente se realiza con un ordenador programado para formar o calcular un valor compuesto de SNPcp a partir de los datos de la etapa 2b; la etapa 5 se realiza con un ordenador programado para formar o calcular un valor compuesto global a partir de los datos de las etapas 3 y 4; y/o la etapa 6 del procedimiento descrito anteriormente se realiza con un ordenador programado para correlacionar el valor compuesto global con la presencia o no presencia de CaP agresivo en dicho individuo comparando dicho valor compuesto global con un valor de corte predeterminado establecido con muestras de control de CaP agresivo conocido y diagnóstico de enfermedad benigna conocido. Asimismo, la presente invención se refiere a un medio de almacenamiento legible en ordenador, no transitorio y tangible con instrucciones ejecutables para realizar dichos cálculos o formar dichos valores compuestos y/o realizar la etapa de correlación como se describe anteriormente.

La elección del valor de corte (o nivel de corte) depende de muchos factores, que incluyen, pero sin limitaciones, el riesgo de la enfermedad como tal y el riesgo asociado con diagnosticar de forma inexacta como positivo a un individuo que no tiene la enfermedad (falso positivo). La elección del valor de corte se describe más en detalle a continuación.

La etapa 2(a) del procedimiento anterior comprende medir la presencia o la concentración de biomarcadores de CaP al menos parcialmente redundantes y, donde al menos uno, como por ejemplo dos, de los biomarcadores de CaP se selecciona de entre el grupo compuesto por (i) PSA, (ii) PSA total (tPSA), (iii) PSA intacto (iPSA), (iv) PSA libre (fPSA) y (v) hK2.

Más en particular, el procedimiento permite ignorar un subconjunto de al menos uno de dichos biomarcadores de CaP (i)-(v) de la categoría de biomarcadores de CaP al formar dicho valor compuesto de biomarcadores, como por ejemplo un subconjunto de uno, dos, tres o cuatro de dichos biomarcadores de CaP (i)-(v).

Además, en una realización, el procedimiento anterior permite ignorar al menos el 10 %, como por ejemplo el 15 %, como por ejemplo el 20 %, como por ejemplo el 30 % del SNPcp de la categoría de SNPcp al formar el valor compuestos de SNPcp.

Preferiblemente, los datos relativos a la categoría de biomarcadores de CaP se combinan según una ecuación predeterminada para formar dicho valor compuesto de biomarcadores, y/o los datos relativos a la categoría de SNPcp se combinan según una ecuación predeterminada para formar dicho valor compuesto de SNPcp. Además, dicho valor compuesto de biomarcadores y dicho valor compuesto de SNPcp se combinan preferiblemente según una ecuación predeterminada para formar dicho valor compuesto global.

En una realización, el procedimiento anterior además comprende una etapa de recomendación al individuo de una biopsia si el valor compuesto global es superior al valor de corte.

Aún en una realización, el procedimiento anterior además comprende una etapa de recomendar al individuo cambiar los hábitos alimentarios, perder peso, lograr un valor de IMC por debajo de 30, hacer ejercicio con regularidad y/o dejar de fumar, si el valor compuesto global es superior al valor de corte.

5

En una realización de la presente invención, los SNP relacionados con CaP (SNPcp) incluyen al menos uno de rs1 1672691, rs11704416, rs3863641, rs12130132, rs4245739, rs3771570, rs7611694, rs1894292, rs6869841, rs2018334, rs16896742, rs2273669, rs1933488, rs11135910, rs3850699, rs11568818, rs1270884, rs8008270, rs4643253, rs684232, rs11650494, rs7241993, rs6062509, rs1041449, rs2405942, rs12621278, rs9364554, 10 rs10486567, rs6465657, rs2928679, rs6983561, rs16901979, rs16902094, rs12418451, rs4430796, rs11649743, rs2735839, rs9623117 y rs138213197.

En una realización, el procedimiento además comprende analizar la categoría de SNP relacionado con una concentración de biomarcador de CaP (SNPbm), midiendo la presencia o ausencia de al menos un SNPbm; 15 combinar los datos relativos a dicho SNPbm para formar un valor compuesto de SNPbm e incluir dicho valor compuesto de SNPbm en dicho valor compuesto global.

En una realización, el al menos un SNPbm incluye al menos uno de rs3213764, rs1354774, rs1227732, rs2736098, rs401681, rs10788160, rs11067228, rs1363120, rs888663 y rs1054564.

20

En una realización de la invención, el procedimiento además comprende analizar la categoría de SNP relacionado con el Índice de Masa Corporal de dicho individuo (SNPimc), midiendo la presencia o ausencia de al menos un SNPimc; los datos combinados relativos a dicho SNPimc para formar un valor compuesto de SNPimc e incluye dicho valor compuesto de SNPimc en dicho valor compuesto global.

25

En una realización, el al menos un SNPimc incluye al menos uno de rs3817334, rs10767664, rs2241423, rs7359397, rs7190603, rs571312, rs29941, rs2287019, rs2815752, rs713586, rs2867125, rs9816226, rs10938397, y rs1558902.

En otra realización de la invención, el procedimiento además comprende la obtención de los antecedentes familiares relativos al CaP, antecedentes de tratamiento y datos de la exploración física de dicho individuo y donde dichos antecedentes familiares, antecedentes de tratamiento y/o datos de la exploración se incluyen en los datos combinados que forman dicho valor compuesto global.

30

Aún en otra realización de la invención, el procedimiento además comprende analizar una categoría adicional de biomarcadores de CaP, midiendo la presencia o la concentración de uno o cada uno de una diversidad de biomarcadores de CaP de dicha categoría adicional de biomarcadores; combinar los datos relativos a dicha categoría adicional de biomarcadores de CaP para formar un valor compuesto adicional de biomarcadores para dicha categoría adicional de biomarcadores de CaP e incluir dicho valor compuesto adicional de biomarcadores en el valor compuesto global; donde la combinación de datos para formar el valor compuesto adicional de biomarcadores 40 está diseñada de forma redundante cuando la categoría adicional de biomarcadores de CaP comprende más de un biomarcador de CaP.

En una realización preferida, la categoría de biomarcadores de CaP adicional comprende el biomarcador MIC-1 y, opcionalmente, otros biomarcadores relacionados con MIC-1 o el biomarcador MSMB y, opcionalmente, otros 45 biomarcadores relacionados con MSMB.

En otra realización, el procedimiento comprende analizar cada uno de una diversidad de categorías adicionales de biomarcadores de CaP y formar un valor compuesto adicional de biomarcadores para cada una de las categorías de biomarcadores de CaP, según el procedimiento descrito anteriormente. Preferiblemente, se analizan al menos dos 50 categorías adicionales de biomarcadores de CaP, donde una categoría adicional de biomarcadores de CaP comprende el biomarcador MIC-1 y, opcionalmente, otros biomarcadores relacionados con MIC-1, y otra categoría adicional comprende el biomarcador MSMB y, opcionalmente, otros biomarcadores relacionados con MSMB.

En una realización, la muestra biológica es una muestra de sangre.

55

En una realización de la invención, el valor compuesto global se calcula usando un procedimiento en el que se utiliza el efecto no aditivo de un SNP relacionado con la concentración de un biomarcador de CaP (SNPbm) y se utiliza la concentración del correspondiente biomarcador de CaP.

60 En una realización preferida del procedimiento según la presente invención, el individuo tiene un valor de IMC

superior a 25, como por ejemplo superior a 30.

En una realización del procedimiento, la medición de la presencia o ausencia de SNP se realiza mediante el uso de espectrometría de masas MALDI.

5

En una realización del procedimiento, la medición de la presencia o la concentración de biomarcadores de CaP se realiza mediante el uso de tecnología de micromatrices.

10 En una realización preferida del procedimiento, la medición de la presencia o ausencia de un SNP (perteneciente a cualquier categoría de SNP) comprende la medición del número de alelos de dicho SNP. En una realización, uno o dos alelos corresponden a la presencia de dicho SNP y cero alelos corresponden a la ausencia de dicho SNP en dicho individuo; donde cero alelos corresponden a homocigoto negativo para dicho SNP, un alelo corresponde a heterocigoto positivo y dos alelos corresponden a homocigoto positivo.

15 En una realización, el procedimiento descrito anteriormente comprende el uso de un dispositivo de ensayo de tipo ELISA, un dispositivo de ensayo en micromatriz, un dispositivo de ensayo de inmunoprecipitación, un dispositivo de ensayo de inmunofluorescencia, un dispositivo de radioinmunoensayo o un dispositivo de espectrometría de masas usando desorción/ionización mediante láser asistida por matriz (MALDI), para la medición de la presencia o la concentración de un biomarcador de CaP.

20

En una realización, que puede combinarse con la realización mencionada anteriormente, el procedimiento descrito anteriormente comprende el uso de un dispositivo de espectrometría de masas usando desorción/ionización mediante láser asistida por matriz (MALDI), para la medición de la presencia o ausencia de un SNP.

25 Otro aspecto de la presente invención proporciona un dispositivo de ensayo para realizar la etapa 2a (es decir, medir una presencia o concentración de al menos un biomarcador de CaP) y la etapa 2b (es decir, medir una presencia o ausencia de al menos un SNPcp) del procedimiento descrito anteriormente para indicar la presencia o no presencia de cáncer de próstata agresivo en un individuo, comprendiendo dicho dispositivo de ensayo una fase sólida sobre la que hay inmovilizadas al menos dos categorías diferentes de ligando, donde:

30

- la primera categoría de dichos ligandos se une específicamente a al menos tres biomarcadores de CaP de tipo calicreína e incluye una diversidad de ligandos diferentes que se unen específicamente a cada uno de una diversidad de diferentes biomarcadores de CaP de tipo calicreína, que incluyen al menos uno de PSA, tPSA, fPSA, hK2 y, opcionalmente, MSMB y/o MIC-1; y

35

- la segunda categoría de dichos ligandos se une específicamente a una diversidad de SNPcp e incluye una diversidad de ligandos diferentes que se unen específicamente a cada uno de una diversidad de dichos SNPcp y donde la diversidad de SNPcp comprende al menos 70 SNPcp seleccionados de entre la lista de SNP de la tabla I o la tabla II.

40

En una realización, el dispositivo de ensayo está además adaptado para medir la presencia o ausencia de un SNPbm, en cuyo caso, la fase sólida del dispositivo de ensayo además tiene una tercera categoría de ligandos inmovilizados que se unen específicamente a un SNPbm e incluye uno o una diversidad de ligandos diferentes que se unen específicamente a uno o a cada uno de una diversidad de SNPbm diferentes, como por ejemplo uno de
45 rs1227732, rs3213764, rs1354774, rs2736098, rs401681, rs10788160, rs11067228, rs1363120, rs888663 y rs1054564.

En una realización, el dispositivo de ensayo está adaptado para medir la presencia o ausencia de un SNPimc, en cuyo caso, la fase sólida además tiene una cuarta categoría de ligandos inmovilizados que se une específicamente a un SNPimc, e incluye uno o una diversidad de ligandos diferentes que se unen específicamente a uno o a una diversidad de SNPimc diferentes, como por ejemplo al menos uno de rs3817334, rs10767664, rs2241423, rs7359397, rs7190603, rs571312, rs29941, rs2287019, rs2815752, rs713586, rs2867125, rs9816226, rs10938397 y rs1558902.

50

55 En una realización, el dispositivo de ensayo descrito anteriormente comprende un dispositivo de ensayo de tipo ELISA, un dispositivo de ensayo en micromatriz, un dispositivo de ensayo de inmunoprecipitación, un dispositivo de ensayo de inmunofluorescencia, un dispositivo de radioinmunoensayo o un dispositivo de espectrometría de masas usando desorción/ionización mediante láser asistida por matriz (MALDI), para la medición de la presencia o la concentración de un biomarcador de CaP.

60

En una realización, que puede combinarse con la realización mencionada anteriormente, el dispositivo de ensayo descrito anteriormente comprende un dispositivo de espectrometría de masas usando desorción/ionización mediante láser asistida por matriz (MALDI), para la medición de la presencia o ausencia de un SNP.

- 5 Según un aspecto adicional de la invención, se proporciona un kit de ensayo para realizar la etapa 2a (es decir, medir una presencia o concentración de al menos un biomarcador de CaP) y la etapa 2b (es decir, medir una presencia o ausencia de al menos un SNPcp) del procedimiento descrito anteriormente para indicar la presencia o no presencia de cáncer de próstata agresivo en un individuo, comprendiendo el correspondiente dispositivo de ensayo como se describe anteriormente y al menos dos categorías de moléculas de detección, donde:
 - 10 - la primera categoría de dichas moléculas de detección es capaz de detectar al menos tres biomarcadores de CaP de tipo caliceína, incluyendo al menos uno de PSA, iPSA, tPSA, fPSA y hK2 y, opcionalmente, MSMB y/o MIC-1; y
 - la segunda categoría de dichas moléculas de detección es capaz de detectar un SNPcp y donde al menos 70 SNP se seleccionan de entre las listas de SNP de la tabla I o II.
- 15 En una realización, el kit de ensayo comprende un dispositivo de ensayo que está adaptado además para medir la presencia o ausencia de al menos un SNPbm y una tercera categoría de moléculas de detección que es capaz de detectar un SNPimc, como por ejemplo al menos uno de rs1227732, rs3213764, rs1354774, rs2736098, rs401681, rs10788160, rs11067228, rs1363120, rs888663 y rs1054564.
- 20 En una realización, el kit de ensayo comprende un dispositivo de ensayo que está adaptado para medir la presencia o ausencia de al menos un SNPimc, y una cuarta categoría de moléculas de detección que es capaz de detectar un SNPimc, como por ejemplo al menos uno de rs3817334, rs10767664, rs2241423, rs7359397, rs7190603, rs571312, rs29941, rs2287019, rs2815752, rs713586, rs2867125, rs9816226, rs10938397 y rs1558902.
- 25 Aún otro aspecto de la presente invención proporciona un dispositivo de ensayo que comprende una fase sólida sobre la que hay inmovilizadas al menos dos categorías diferentes de ligando, donde:
 - la primera categoría de dichos ligandos se une específicamente a al menos tres biomarcadores de CaP de tipo caliceína e incluye una diversidad de ligandos diferentes que se unen específicamente a cada uno de una diversidad de diferentes biomarcadores de CaP de tipo caliceína, que incluyen al menos uno de PSA, iPSA, tPSA, fPSA y hK2 y, opcionalmente, MSMB y/o MIC-1; y
 - la segunda categoría de dichos ligandos se une específicamente a una diversidad de SNPcp e incluye una
- 30 diversidad de ligandos diferentes que se unen específicamente a cada uno de una diversidad de SNPcp diferentes y donde la diversidad de SNPcp comprende al menos 70 SNPcp seleccionados de entre las listas de SNP de la tabla I o la tabla II.
- En una realización del dispositivo de ensayo, la fase sólida además tiene una tercera categoría de ligandos inmovilizados que se unen específicamente a un SNPbm y que incluyen uno o una diversidad de ligandos diferentes que se unen específicamente a uno o a cada uno de una diversidad de SNPbm diferentes, seleccionados de entre al menos uno de rs1227732, rs3213764, rs1354774, rs2736098, rs401681, rs10788160, rs11067228, rs1363120, rs888663 y rs1054564.
- 40 En una realización adicional del dispositivo de ensayo, la fase sólida además tiene una cuarta categoría de ligandos inmovilizados que se unen específicamente a un SNPimc, y que incluyen uno o una diversidad de ligandos diferentes que se unen específicamente a uno o a cada uno de una diversidad de SNPimc diferentes, seleccionados de entre al menos uno de rs3817334, rs10767664, rs2241423, rs7359397, rs7190603, rs571312, rs29941, rs2287019, rs2815752, rs713586, rs2867125, rs9816226, rs10938397 y rs1558902.
- 45 Aún otro aspecto de la invención proporciona un programa informático que se puede cargar directamente en la memoria interna de un ordenador digital, donde el programa informático comprende medios de código de software para realizar al menos la etapa 3 (es decir, combinar los datos relativos a dicha categoría de biomarcadores de CaP para formar un valor compuesto de biomarcadores), la etapa 4 (es decir, combinar los datos relativos a dicha
- 50 categoría de SNPcp para formar un valor compuesto de SNPcp), la etapa 5 (es decir, combinar el valor compuesto de biomarcadores y el valor compuesto de SNPcp para formar un valor compuesto global) y/o la etapa 6 (correlacionar dicho valor compuesto global con la presencia o no presencia de CaP agresivo en dicho individuo comparando el valor compuesto global con un valor de corte predeterminado establecido con muestras de control de CaP agresivo conocido y diagnóstico de enfermedad benigna) del procedimiento descrito anteriormente para indicar
- 60 la presencia o no presencia de cáncer de próstata agresivo en un individuo; como por ejemplo la etapa 1 (es decir,

proporcionar al menos una muestra biológica de dicho individuo), etapa 2 (en la muestra biológica, analizar una categoría de biomarcadores de CaP midiendo la presencia o la concentración de cada uno de una diversidad de biomarcadores de CaP y analizar una categoría de SNPcp midiendo la presencia o ausencia de cada uno de una diversidad de SNPcp), etapas 3, 4, 5 y 6 de dicho procedimiento.

5

En una realización, el programa informático además comprende medios de código de software para analizar una categoría de SNPbm midiendo la presencia o ausencia de al menos un SNPbm.

En otra realización, el programa informático además comprende medios de código de software para analizar una categoría de SNPimc midiendo la presencia o ausencia de al menos un SNPimc.

10

Un aspecto adicional de la invención proporciona un aparato que comprende un dispositivo de ensayo como se describe anteriormente y el correspondiente programa informático como se describe anteriormente.

15 BREVE DESCRIPCIÓN DE LOS DIBUJOS

En la figura 1 se muestran las curvas ROC para el modelo lineal del ejemplo 1 que ilustra la diferencia en el rendimiento entre PSA (101) y un modelo multiparamétrico (102) en la predicción del CaPa.

En la figura 2 se muestra un ejemplo de un árbol de decisión para predecir si un sujeto debe ser remitido para una biopsia.

20

En la figura 3 se muestran las curvas ROC para el modelo lineal del ejemplo 1 que ilustra la diferencia en el rendimiento entre PSA (301) y un modelo multiparamétrico (302) en la predicción del CaPa en individuos con un valor de IMC superior a 25.

25 DESCRIPCIÓN DETALLADA DE LA INVENCION

Para los fines de esta solicitud y por claridad, se hacen las siguientes definiciones:

El término «PSA» se refiere al antígeno prostático específico sérico en general. PSA existe en diferentes formas, donde el término «PSA libre» se refiere a PSA que no está unido o no se une a otra molécula, el término «PSA unido» se refiere a PSA que se une o forma complejo con otra molécula y finalmente el término «PSA total» se refiere a la suma del PSA libre y el PSA unido (en complejo). El término «PSA F/T» es la relación entre PSA no unido y PSA total. Existen también derivados moleculares de PSA, donde el término «proPSA» se refiere a una forma precursora inactiva de PSA y «PSA intacto» se refiere a una forma adicional de proPSA que se encuentra intacta o inactiva.

35

El término «ensayo diagnóstico» se refiere a la detección de la presencia o naturaleza de una condición patológica. Se puede usar de forma intercambiable con el término «procedimiento diagnóstico». Los ensayos diagnósticos difieren en cuanto a sensibilidad y especificidad.

40

Una medida de la utilidad de una herramienta diagnóstica es el «área bajo la curva de características operativas del receptor», que se conoce habitualmente como estadística ROC-AUC. Esta medida ampliamente aceptada tiene en cuenta tanto la sensibilidad como la especificidad de la herramienta. La medida ROC-AUC oscila normalmente entre 0,5 y 1,0, donde un valor de 0,5 indica que la herramienta no tiene valor diagnóstico y un valor de 1,0 indica que la herramienta tiene un 100 % de sensibilidad y 100 % de especificidad.

45

El término «sensibilidad» se refiere a la proporción de todos los sujetos con CaP que son identificados correctamente como tales (lo que equivale al número de verdaderos positivos dividido entre la suma del número de verdaderos positivos y falsos positivos).

50

El término «especificidad» se refiere a la proporción de todos los sujetos sanos con respecto al CaP (es decir, que no padecen CaP) que son identificados correctamente como tales (lo que equivale al número de verdaderos negativos dividido entre la suma del número de verdaderos negativos y falsos negativos).

El término «biomarcador» se refiere a una proteína, una parte de una proteína, un péptido o un polipéptido, que se puede usar como un marcador biológico, por ejemplo, para fines diagnósticos.

55

El término «biomarcador de tipo calicreína» se refiere a biomarcadores proteicos que pertenecen o están relacionados con la familia de proteínas de la calicreína, que incluye, pero sin limitaciones, el antígeno prostático específico (PSA) en forma libre o en forma de complejo, pro PSA (un conjunto de isoformas de PSA) y, en particular,

60

la forma truncada (-2) pro PSA, PSA intacto, fosfatasa ácida prostática (PAP) humana y calicreína 2 humana (abreviado hK2, HK2 o hk2 en la presente solicitud).

El término «polimorfismo de un solo nucleótido» (SNP) se refiere a las propiedades genéticas de un locus definido en el código genético de un individuo. Un SNP puede estar relacionado con un mayor riesgo de CaP por lo que se puede usar para evaluaciones diagnósticas y pronósticas de un individuo. La base de datos Single Nucleotide Polymorphism Database (dbSNP) es un archivo de la variación genética dentro y a lo largo de diferentes especies desarrollado y gestionado por el National Center for Biotechnology Information (NCBI) en colaboración con el National Human Genome Research Institute (NHGRI), ambos ubicados en EE. UU. Aunque el nombre de la base de datos implica únicamente una recopilación de una clase de polimorfismos (es decir, polimorfismos de un solo nucleótido [SNP]), contiene de hecho una serie de variaciones moleculares. Cada registro exclusivo de SNP enviado recibe un número de ID de SNP de referencia («rs#»; «grupo de refSNP»). En esta solicitud, los SNP se identifican principalmente usando números rs#. En consecuencia, dentro de la presente solicitud, SNP se usa para hacer referencia a la serie de variaciones moleculares incluidas en la dbSNP, y no solo a los polimorfismos de un solo nucleótido. Para los fines de la presente solicitud, el término «SNP» y los «SNP» se pueden usar de manera intercambiable para describir el singular y/o el plural del «polimorfismo de un solo nucleótido».

El término «índice de masa corporal» (IMC) se refiere a un indicador heurístico de la grasa corporal humana en relación con el peso y estatura de un individuo, según la fórmula $IMC = \text{peso} / (\text{estatura} * \text{estatura})$, donde «peso» es el peso de un individuo expresado en kilogramos y «estatura» es la estatura de un individuo expresado en metros. Un valor de IMC normal saludable se considera que está típicamente dentro del intervalo de 18,5 a 25 y se considera normalmente que los individuos que tienen un $IMC > 30$ son obesos.

El término «cáncer de próstata agresivo» (CaPa) se refiere a una afección más grave que la enfermedad promedio de cáncer de próstata. El CaPa se puede definir de diferentes formas, que incluyen, pero sin limitaciones, (a) cáncer de próstata con una puntuación de Gleason de 7 o superior, (b) cáncer de próstata en estadio tumoral de tres o superior, (c) cáncer de próstata en un individuo que tiene un valor de PSA superior a 10 ng/mL, (d) un individuo que tiene un valor de PSA aumentado (tiempo de duplicación inferior a un año) y (e) análisis de imágenes asistido por ordenador (p. ej., tomografía por emisión de positrones (PET) o tomografía computarizada por emisión de fotón único (SPECT) o tomografía computarizada por rayos X (TC) o resonancia magnética (RM) o ecografía, o cualquier otro tipo de análisis de imágenes asistido por ordenador) indicativo de un tamaño de tumor en el cuartil superior de la población de pacientes.

El término «antecedentes patológicos» se refiere a la información relacionada con exploraciones, diagnósticos y/o tratamientos para cualquier cáncer históricos. Un ejemplo no limitante de antecedentes patológicos es si se ha examinado previamente a un sujeto por la presencia de CaP por medio de una biopsia de próstata.

El término «categoría de parámetros» se refiere a un grupo o una familia de parámetros relacionados, como por ejemplo biomarcadores relacionados o SNP relacionados, que son parcial o completamente redundantes en términos de rendimiento predictivo. Un ejemplo de categoría de parámetros son los «biomarcadores de tipo calicreína», una categoría que incluye por ejemplo PSA, PSA total (tPSA), PSA intacto (iPSA), PSA libre (fPSA) y hK2. Otro ejemplo de categoría de parámetros es «SNP relacionado con IMC», una categoría que incluye los SNP que están relacionados con el IMC de un individuo. En los modelos de predicción de la presente invención, puede ser suficiente tener resultados de mediciones (datos) de un subconjunto de los miembros de cada categoría, de modo que esté representada cada categoría en el modelo de predicción, aunque usando solo un subconjunto de los miembros de las respectivas categorías. El término «categoría de parámetros» a veces se denomina únicamente «categoría» en la presente solicitud.

El término «valor compuesto» se refiere a la combinación de datos relacionados con una categoría de parámetros dentro de un valor representativo de dicha categoría de parámetros. La combinación de datos se puede realizar normalmente según una o más ecuaciones predeterminadas. Un valor compuesto es el resultado de la combinación de datos según una o más ecuaciones predeterminadas. Las ecuaciones diferentes son aplicables a diferentes resultados de mediciones (es decir, datos), dependiendo de para qué subconjuntos de los miembros de la categoría de parámetros están disponibles esos datos. Un ejemplo no limitante de un procedimiento para formar un valor compuesto de una categoría de parámetros en particular es usar el promedio de los resultados disponibles para los miembros de dicha categoría. El término «valor compuesto» a veces se denomina «puntuación» en la presente solicitud. Un ejemplo no limitante de un valor compuesto es «valor compuesto del biomarcador». Otro ejemplo no limitante de un valor compuesto es «valor compuesto genético» (o «puntuación genética») y más específicamente, «valor compuesto de SNP».

El término «combinación de datos diseñada de forma redundante» se refiere a una combinación de datos obtenida mediante una diversidad de mediciones para formar un valor compuesto de una o más categorías de parámetros o subconjunto de las mismas, donde la combinación de datos se realiza de forma que un valor compuesto que representa una categoría de parámetros se puede generar en función de un subconjunto de datos de dicha categoría, por ejemplo, donde algunos datos se omiten o son erróneos o en función del conjunto completo de datos de dicha categoría.

El término «una diversidad» como se usa en la presente solicitud significa «dos o más».

- 10 La presente invención proporciona procedimientos diagnósticos para ayudar en la indicación, estimación, detección y/o determinación de la presencia o no presencia de cáncer de próstata agresivo en un sujeto. La presente invención puede, si se desea, ajustarse a subpoblaciones definidas para aumentar el rendimiento y la utilidad de la invención en dicha subpoblación. Incluso aunque la presente invención se puede aplicar a la población general de individuos masculinos, es posible generar procedimientos diagnósticos para la detección de CaPa con rendimiento mejorado
- 15 para subpoblaciones definidas. Un ejemplo no limitante de una subpoblación definida son individuos que tienen un índice de masa corporal (IMC) alto, por ejemplo, $IMC > 25$ o $IMC > 30$ o $IMC > 35$. Otro ejemplo no limitante de una subpoblación definida son individuos que tienen un valor de PSA bajo, por ejemplo, $PSA < 4$ ng/mL o $PSA < 3$ ng/mL o $PSA < 2$ ng/mL o $PSA < 1$ ng/mL.
- 20 El principio básico de la invención es el uso de combinaciones de biomarcadores e información genética de tal manera que el uso combinatorio de la información evaluada sobre el individuo mejora la calidad del diagnóstico.
- Obtención de los antecedentes familiares relativos al CaP de dicho paciente (categoría ANT).
 - Obtención de datos de la exploración física del paciente, como peso, IMC, edad y similares (categoría DEF).
- 25 • Obtención de diversas muestras biológicas de dicho paciente.
- Medir o cuantificar en dichas muestras biológicas la presencia o la concentración de una diversidad de biomarcadores definidos (categoría biomarcadores), seguido por la combinación de datos relativos a dichos biomarcadores para formar un valor compuesto de biomarcadores.
 - Medir o cuantificar en dichas muestras biológicas el estado genético de dichos pacientes con respecto a una
- 30 diversidad de SNP definidos relacionados con el CaP (categoría SNPcp), midiendo o cuantificando la presencia o ausencia de una diversidad de SNP definidos relacionados con el CaP (SNPcp) y seguido por la combinación de los datos obtenidos relativos a los SNP relacionados con el CaP para formar un valor compuesto de SNPcp.
- Medir o cuantificar en dichas muestras biológicas el estado genético de dichos pacientes con respecto a una diversidad de SNP definidos relacionados con el nivel de expresión de biomarcadores o la concentración de
- 35 biomarcadores (categoría SNPbm), midiendo o cuantificando la presencia o ausencia de una diversidad de SNP definidos relacionados con el nivel de expresión de biomarcadores o la concentración de biomarcadores (SNPbm) para formar un valor compuesto de SNPbm.
- Medir o cuantificar en dichas muestras biológicas el estado genético de dichos pacientes con respecto a una diversidad de SNP definidos relacionados con el Índice de Masa Corporal (IMC) de dicho individuo (categoría
- 40 SNPimc), midiendo o cuantificando la presencia o ausencia de una diversidad de SNP definidos relacionados con el IMC (SNPimc) para formar un valor compuesto de SNPimc.
- Combinar los datos de al menos dos de las categorías definidas anteriormente para formar un valor compuesto global para su uso en la detección de cáncer de próstata agresivo incipiente.
 - Determinar mediante el uso de dicho valor compuesto global, solo o en combinación con datos adicionales, si es
- 45 probable que el paciente sufra CaPa.

Más en detalle, la etapa que comprende la obtención de los antecedentes familiares incluye, pero sin limitaciones, la identificación de si algún familiar allegado varón (como el padre, hermano o hijo del paciente) sufre o ha sufrido CaP.

- 50 La información de la exploración física del paciente se obtiene normalmente a través de una exploración física regular en la que se recaba la edad, peso, estatura, IMC y datos físicos similares.

La obtención de muestras biológicas de un paciente incluye, pero sin limitaciones, plasma, suero, ADN de leucocitos periféricos y orina.

- 55 La cuantificación de la presencia o la concentración de biomarcadores en una muestra biológica se puede hacer de muchas formas. Un procedimiento frecuente es el uso de ensayos de inmunoabsorción ligados a enzima (ELISA) que utilizan anticuerpos y una curva de calibración para evaluar la presencia y (si es posible) la concentración de un biomarcador seleccionado. Los ensayos de tipo ELISA son comunes y conocidos en la técnica, como resulta
- 60 evidente a partir de la publicación «Association between saliva PSA and serum PSA in conditions with prostate

- adenocarcinoma» de Shiiki N y coautores, publicado en *Biomarkers*, 2011 Sep;16(6):498-503. Otro procedimiento frecuente es el uso de un ensayo en micromatriz para la cuantificación de la presencia o la concentración de biomarcadores en una muestra biológica. Un ensayo en micromatriz típico comprende un portaobjetos de vidrio plano sobre el cual se adhiere en áreas no solapantes de una cara del portaobjetos una diversidad de reactivos de
- 5 captura diferentes (normalmente un anticuerpo) cada uno diseñado para capturar específicamente un tipo de biomarcador. Se deja que la muestra biológica entre en contacto, durante un periodo de tiempo definido, con el área en donde se localizan dichos reactivos de captura, seguido por el lavado del área de reactivos de captura. En este punto, en caso de que el biomarcador buscado estuviera presente en la muestra biológica, el correspondiente reactivo de captura habrá capturado una fracción del biomarcador buscado y lo mantendrá adherido también al
- 10 portaobjeto de vidrio después del lavado. A continuación, se añade un juego de reactivos de detección al área de reactivos de captura (que ahora mantienen unidos potencialmente los biomarcadores), siendo capaces dichos reactivos de detección de (i) unirse al biomarcador cuando está presente en el portaobjetos de vidrio y (ii) generar una señal detectable (normalmente a través de la conjugación de un colorante fluorescente). Normalmente se requiere añadir al portaobjetos de vidrio un reactivo de detección por biomarcador. Existen muchos otros
- 15 procedimientos capaces de cuantificar la presencia o la concentración de un biomarcador, lo que incluye, pero sin limitaciones, ensayos de inmunoprecipitación, ensayos de inmunofluorescencia, radioinmunoensayos y espectrometría de masas usando desorción/ionización mediante láser asistida por matriz (MALDI), por mencionar algunos ejemplos.
- 20 La cuantificación de la presencia de SNP a través del análisis de una muestra biológica normalmente implica el análisis mediante espectrometría de masas MALDI basado en extensiones de cebador específico de alelo, incluso aunque se pueden aplicar igualmente otros procedimientos. Esto se aplica a cualquier tipo de SNP, es decir, tanto SNP relacionados con el CaP (SNPcp), SNP relacionados con el IMC (SNPimc), como SNP relacionados con la expresión/concentración de biomarcadores (SNPbm).
- 25 La combinación de datos puede ser cualquier clase de combinación algorítmica de resultados, como por ejemplo, una combinación lineal de datos donde la combinación lineal mejora el rendimiento diagnóstico (por ejemplo, medido usando ROC-AUC). Otra posible combinación incluye una relación polinomial no lineal.
- 30 Entre los biomarcadores adecuados para el diagnóstico del CaPa se incluyen, pero sin limitaciones, antígeno prostático específico (PSA) tanto en forma libre como en forma de complejo, pro PSA (un conjunto de isoformas de PSA) y, en particular, la forma truncada (-2) pro PSA, PSA intacto, fosfatasa ácida prostática (PAP) humana, calicreína 2 humana (hK2), antígeno temprano del cáncer de próstata (EPCA), proteína secretora prostática (PSP94, también conocida como microseminoproteína-beta y MSMB), glutatión S-transferasa π (GSTP1) y α -metilacil
- 35 coenzima A racemasa (AMACR). Entre los biomarcadores relacionados que pueden ser útiles para mejorar la exactitud diagnóstica del procedimiento se incluye la citoquina inhibidora de macrófagos 1 (MIC-1, también conocida como GDF-15).
- Entre los SNP adecuados relacionados con el CaP se incluyen, pero sin limitaciones, rs12621278 (cromosoma 2, locus 2q31.1), rs9364554 (cromosoma 6, locus 6q25.3), rs10486567 (cromosoma 7, locus 7p15.2), rs6465657 (cromosoma 7, locus 7q21.3), rs2928679 (cromosoma 8, locus 8p21), rs6983561 (cromosoma 8, locus 8q24.21), rs16901979 (cromosoma 8, locus 8q24.21), rs16902094 (cromosoma 8, locus 8q24.21), rs12418451 (cromosoma 11, locus 11q13.2), rs4430796 (cromosoma 17, locus 17q12), rs11649743 (cromosoma 17, locus 17q12), rs2735839 (cromosoma 19, locus 19q13.33), rs9623117 (cromosoma 22, locus 22q13.1) y rs138213197 (cromosoma 17, locus
- 40 17q21).
- Entre los SNP adecuados relacionados con el CaP además se incluyen, pero sin limitaciones, rs11672691, rs11704416, rs3863641, rs12130132, rs4245739, rs3771570, rs7611694, rs1894292, rs6869841, rs2018334, rs16896742, rs2273669, rs1933488, rs11135910, rs3850699, rs11568818, rs1270884, rs8008270, rs4643253,
- 50 rs684232, rs11650494, rs7241993, rs6062509, rs1041449 y rs2405942.
- Entre los SNP adecuados relacionados con el CaP además se incluyen, pero sin limitaciones, rs138213197 como se describe en el artículo «Germline mutations in HOXB13 and prostate-cancer risk» de Ewing CM y coautores publicado en *N Engl J Med.*, 12 de enero de 2012;366(2):141-9, 1100delC (22q12.1) e I157T (22q12.1) como se
- 55 describe en el artículo «A novel founder CHEK2 mutation is associated with increased prostate cancer risk» de Cybulski C y coautores publicado en *Cancer Res.*, 15 de abril de 2004; 64(8):2677-9 y 657 del 5 (8q21) como se describe en el artículo «NBS1 is a prostate cancer susceptibility gene» de Cybulski C y coautores publicado en *Cancer Res.*, 15 de febrero de 2004; 64(4):1215-9.
- 60 Es posible definir una categoría de parámetros como «SNP relacionados con CaP» que incluya SNP relacionados

con CaP. Entre los miembros adecuados se incluyen (pero sin limitaciones) los SNP enumerados anteriormente. Un subconjunto de los miembros de esta categoría sería suficiente para representar la categoría como tal en un modelo predictivo.

- 5 Entre los SNP adecuados relacionados con otros procesos diferentes al CaP se incluyen, pero sin limitaciones, rs3213764, rs1354774, rs2736098, rs401681, rs10788160 y rs11067228, estando todos ellos relacionados con el nivel de expresión de PSA. Es posible definir una categoría de parámetros como «SNP relacionados con la concentración de PSA» o «SNP relacionados con el nivel de expresión de PSA», que incluye SNP relacionados con la concentración o el nivel de expresión de PSA. Un subconjunto de los miembros de esta categoría sería suficiente para representar la categoría como tal en un modelo predictivo. Los SNP rs3213764 y rs1354774 están especialmente relacionados con el nivel de expresión de PSA libre.

- 15 Entre los SNP adecuados relacionados con otros procesos diferentes al CaP se incluyen además, pero sin limitaciones, rs1363120, rs888663, rs1227732 y rs1054564, estando todos ellos relacionados con el nivel de expresión del biomarcador de citoquina de inflamación MIC-1. Es posible definir una categoría de parámetros como «SNP relacionados con la concentración de MIC-1» o «SNP relacionados con el nivel de expresión de MIC-1», que incluye SNP relacionados con la concentración o el nivel de expresión de MIC-1. Un subconjunto de los miembros de esta categoría sería suficiente para representar la categoría como tal en un modelo predictivo.

- 20 Es posible definir una categoría de parámetros como «SNP relacionados con la concentración de biomarcador de CaP» o «SNP relacionados con el nivel de expresión de biomarcador de CaP» que incluye SNP relacionados con la concentración o el nivel de expresión de biomarcadores relevantes como por ejemplo antígeno prostático específico (PSA) en forma libre o en forma de complejo, pro PSA (un conjunto de isoformas de PSA) y, en particular, la forma truncada (-2) pro PSA, PSA intacto, fosfatasa ácida prostática (PAP) humana, calicreína 2 humana (hK2), antígeno temprano de cáncer de próstata (EPCA), proteína secretora prostática (PSP94, también conocida como microseminoproteína-beta y MSMB), glutatión S-transferasa π (GSTP1), α -metilcil coenzima A racemasa (AMACR) y citoquina inhibidora de macrófagos 1 (MIC-1; también conocida como GDF-15). Un subconjunto de los miembros de esta categoría sería suficiente para representar la categoría como tal en un modelo predictivo.

- 30 Entre los SNP adecuados relacionados con otros procesos diferentes al CaP además se incluyen, pero sin limitaciones, rs3817334, rs10767664, rs2241423, rs7359397, rs7190603, rs571312, rs29941, rs2287019, rs2815752, rs713586, rs2867125, rs9816226, rs10938397 y rs1558902, estando todos ellos relacionados con el IMC de un individuo. Otro SNP adecuado relacionado con el IMC se describe en el artículo «Contribution of 32 GWAS-identified common variants to severe obesity in European adults referred for bariatric surgery» de Mägi y coautores publicado en PLoS One. 7 de agosto de 2013; 8(8):e70735. Es posible definir una categoría de parámetros como «SNP relacionados con el nivel de expresión del IMC» que incluya SNP relacionados con el IMC del individuo. Un subconjunto de los miembros de esta categoría sería suficiente para representar la categoría como tal en un modelo predictivo.

- 40 Un conjunto preferido de SNP usado en la evaluación de la presencia o no presencia de cáncer de próstata agresivo en un sujeto es rs582598, rs439378, rs2207790, rs1046011, rs10458360, rs7525167, rs10489871, rs7529518, rs4245739, rs4512641, rs10178804, rs11900952, rs1873555, rs10191478, rs6755901, rs6545962, rs721048, rs2710647, rs12612891, rs2028900, rs1009, rs12233245, rs6760417, rs10496470, rs10199796, rs12475433, rs16860513, rs12151618, rs3765065, rs13017302, rs12988652, rs871688, rs749264, rs3771570, rs4346531, rs6770955, rs12637074, rs2660753, rs13319878, rs6437715, rs2162185, rs1515542, rs2270785, rs9830294, rs1439024, rs6762443, rs888507, rs6794467, rs12490248, rs1477886, rs4833103, rs3796547, rs17779822, rs2366711, rs16849146, rs1894292, rs12640320, rs3805284, rs12500426, rs4699312, rs17021918, rs7679673, rs2047408, rs2647262, rs12506850, rs7658048, rs2078277, rs12505546, rs13113975, rs4246742, rs2736098, rs401681, rs11134144, rs10060513, rs40485, rs2087724, rs1482679, rs16901841, rs1295683, rs2070874, rs7752029, rs2018334, rs9358913, rs1140809, rs409558, rs3096702, rs9267911, rs2025645, rs9359428, rs6569371, rs2813532, rs1933488, rs712242, rs6934898, rs9456490, rs651164, rs3120137, rs9364554, rs9457937, rs10486562, rs10807843, rs7801918, rs6962297, rs2465796, rs6957416, rs7777631, rs2272316, rs6961773, rs2132276, rs13265330, rs16887736, rs2911756, rs2272668, rs2339654, rs1380862, rs9297746, rs12543663, rs10086908, rs16901922, rs1016343, rs17832285, rs16901979, rs4871779, rs10107982, rs16902094, rs620861, rs17467139, rs6983267, rs9297756, rs10094059, rs7818556, rs1992833, rs986472, rs12552397, rs4273907, rs4237185, rs753032, rs11253002, rs2386841, rs10795841, rs10508422, rs7075945, rs10508678, rs539357, rs10826398, rs3818714, rs7090755, rs10993994, rs4382847, rs1891158, rs10887926, rs10788160, rs6579002, rs10832514, rs7358335, rs1944047, rs3019779, rs10896437, rs12793759, rs7106762, rs7102758, rs2449600, rs585197, rs2509867, rs11568818, rs7125415, rs11601037, rs11222496, rs4570588, rs6489721, rs3213764, rs17395631, rs4423250, rs11168936, rs10875943, rs3759129, rs902774, rs1827611, rs4760442, rs11610799,

rs6539333, rs11067228, rs7485441, rs6489794, rs4119478, rs17070292, rs2293710, rs17256058, rs1950198, rs2331780, rs7141529, rs12880777, rs17123359, rs785437, rs524908, rs12903579, rs7178085, rs7164364, rs896615, rs11634741, rs9972541, rs12594014, rs11631109, rs1558902, rs8044335, rs2738571, rs885479, rs385894, rs684232, rs4925094, rs17138478, rs11649743, rs2107131, rs7213769, rs12946864, rs306801, rs138213197, rs1863610, rs17224342, rs9911515, rs12947919, rs966304, rs17744022, rs7234917, rs1943821, rs2227270, rs1363120, rs888663, rs1227732, rs1054564, rs4806120, rs11672691, rs758643, rs3745233, rs6509345, rs2659051, rs2735839, rs1354774, rs2691274, rs6090461, rs2297434, rs6062509, rs2315654, rs2823118, rs2838053, rs398146, rs16988279, rs2269640, rs4822763, rs132774, rs747745, rs5978944, rs6530238, rs5934705, rs5935063, rs4830488, rs17318620, rs5945619, rs5945637, rs11091768, rs2473057, rs5918762, rs4844228, rs6625760 y rs17324573 (tabla I). Incluso aunque es preferible el uso de la lista completa, cualquier subconjunto de esta lista es adecuado para su uso en la evaluación de la presencia o no presencia de cáncer de próstata agresivo en un sujeto. Los SNP de la lista (todos o un subconjunto que comprende aproximadamente el 95 %, o el 90 %, o el 85 %, o el 80 %, o el 75 %, o el 70 % de los SNP de esta lista) se pueden colocar en el mismo soporte sólido, por ejemplo, el mismo portaobjetos de vidrio, para la detección simultánea en un instrumento analítico adecuado.

Como se ha discutido previamente, la evaluación del rendimiento de la eficiencia del cribado del CaP es difícil. Aunque las características de ROC-AUC proporcionan cierta información relativa al rendimiento, son deseables procedimientos adicionales. Un método alternativo para evaluar el rendimiento del cribado del CaP es calcular el porcentaje de biopsias positivas a un nivel dado de sensibilidad y comparar el rendimiento del cribado usando solo PSA con cualquier nuevo procedimiento de cribado. Esto sin embargo requiere que el rendimiento del PSA se defina de forma exacta.

Un ejemplo de evaluación del rendimiento del cribado de PSA ha sido descrito por IM Thompson y coautores en el artículo «Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial» publicado en J Natl Cancer Inst., 19 de abril de 2006; 98(8):529-34. En este artículo, los datos de biopsias de próstata de los hombres que participaron en el ensayo clínico de prevención del cáncer de próstata (PCPT) se utilizaron para determinar la sensibilidad del PSA. En total se incluyeron 5519 varones del grupo de placebo del PCPT que se sometieron a biopsia de próstata tenían al menos una medición del PSA y se les había realizado un tacto rectal durante el año previo a la biopsia, y contaban con al menos dos mediciones de PSA realizadas durante los 3 años previos a la biopsia de próstata. En este artículo se describe que si se usa un valor de PSA de 3 ng/mL como valor de corte, se omitirá aproximadamente el 41 % de los cánceres de alto grado (es decir, cánceres con una puntuación de Gleason de 7 o más).

Se ha descrito un segundo análisis usando la misma población de estudio por IM Thompson y coautores en el artículo «Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/mL or lower» publicado en JAMA, 6 de julio de 2005; 294(1):66-70.

En este artículo, los autores presentan una estimación de la sensibilidad y la especificidad de PSA para todos los cánceres de próstata, Gleason 7+ y Gleason 8+. Cuando se utilizó 3,1 ng/mL como valor de corte de PSA para biopsias, se estimaron una sensibilidad del 56,7 % y una especificidad del 82,3 % para tumores Gleason 7+. En este artículo, los autores concluyeron que no existe punto de corte de PSA con una sensibilidad alta y una especificidad alta simultáneamente para controlar el cáncer de próstata en varones sanos, sino más bien un continuo de riesgo de cáncer de próstata en todos los valores de PSA. Esto ilustra la complicación que surge con el PSA como prueba de cribado a la vez que se sigue reconociendo la conexión del PSA con el cáncer de próstata.

Una consecuencia inevitable de las dificultades en la obtención de estimaciones exactas y comparables del rendimiento predictivo de cualquier modelo diagnóstico o pronóstico dado en el cribado del CaP es que cuando se calcula la mejora relativa de un nuevo procedimiento en comparación con el uso solo de PSA, la mejora relativa calculada variará dependiendo de muchos factores. Un factor importante que influye en la mejora relativa calculada es la forma en la que se obtiene el grupo control (es decir, negativos conocidos). Puesto que no es ético realizar biopsias a sujetos para los que no existen indicaciones de CaP, el grupo control será seleccionado con sesgo. Por tanto, la mejora relativa de un procedimiento nuevo dependerá de cómo se seleccionó el grupo control y existen muchos procedimientos bien conocidos para seleccionar grupos control. Cualquier mejora estimada documentada debe considerarse, por tanto, a la luz de tal variación. Según nuestra experiencia, estimamos que si se documenta que la mejora relativa de un procedimiento nuevo es del 15 % en comparación con el valor de PSA solo usando un procedimiento bien conocido para seleccionar el grupo control, dicho nuevo procedimiento sería al menos un 10 % mejor que el valor de PSA solo usando cualquier otro procedimiento bien conocido para seleccionar el grupo control.

Para su uso generalizado en la sociedad, el rendimiento de un cribado debe cumplir con ventajas económico-

- sanitarias razonables. Una estimación aproximada es que un procedimiento de cribado con un rendimiento de aproximadamente el 15 % mejor que el PSA (es decir, que evita el 15 % de las biopsias innecesarias) al mismo nivel de sensibilidad (es decir, que detecta el mismo número de cánceres de próstata en la población) tendría la posibilidad de ser utilizado de forma generalizada al nivel de coste actual de los sistemas públicos de sanidad. Sin embargo, para subpoblaciones definidas de individuos, un nuevo procedimiento de cribado puede tener también ventajas económicas para mejoras más pequeñas en comparación con el rendimiento del valor de PSA. Es de destacar que, aunque se han hecho esfuerzos importantes por encontrar un modelo combinado para la estimación del riesgo de CaP (como se ilustra en varios de los documentos citados en esta solicitud de patente), ninguno de dichos procedimientos combinados se usa actualmente de forma rutinaria en Europa. Por tanto, los procedimientos multiparamétricos previos conocidos no cumplen los estándares socioeconómicos para ser útiles en la atención sanitaria moderna. El procedimiento de la presente invención tiene mejor rendimiento que los procedimientos combinados presentados previamente y cumplen los requisitos de rendimiento socioeconómico para ser considerado del todo por un sistema sanitario.
- 15 Un posible procedimiento para obtener un procedimiento de cribado para el CaPa que cumpla los requisitos para su uso generalizado es combinar la información procedente de múltiples fuentes. Desde un punto de vista general, esto comprende combinar los valores obtenidos del análisis de biomarcadores (p. ej., valores de PSA), los perfiles genéticos (p. ej., el perfil de SNP), antecedentes familiares y otras fuentes. La combinación como tal tiene la posibilidad de producir una declaración de diagnóstico mejor que cualquiera de los factores incluidos por sí solos. Se han descrito en el pasado intentos por combinar los valores en un modelo multiparamétrico para producir mejores declaraciones de diagnóstico, como se describe en otra parte de la presente solicitud.

- La combinación de datos puede ser cualquier clase de combinación algorítmica de resultados, como por ejemplo, una combinación lineal de datos donde la combinación lineal mejora el rendimiento diagnóstico (por ejemplo, medido usando ROC-AUC). Otros métodos posibles de combinación en un modelo capaz de producir una estimación diagnóstica incluyen (pero sin limitaciones) polinomiales no lineales, máquinas de vectores de soporte, clasificadores basados en redes neuronales, análisis discriminatorio, bosque aleatorio, potenciación de gradiente, mínimos cuadrados parciales, regresión contraída, lasso, redes elásticas, k vecinos más próximos. Asimismo, en el libro «The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition» de T Hastie, R Tibshirani y J Friedman publicado por Springer Series in Statistics, ISBN 978-0387848570 se describen muchos métodos adecuados de combinación de datos para predecir o clasificar un resultado en particular.

- El algoritmo que transforma los datos de las diferentes categorías en un único valor indicativo de si el paciente tiene probabilidad de sufrir un CaPa es preferiblemente una función lineal, donde la dependencia de las diferentes categorías se emplea para aumentar adicionalmente el rendimiento diagnóstico del procedimiento. Por ejemplo, una dependencia importante es el nivel medido de un biomarcador seleccionado combinado con cualquier marcador genético asociado relacionado con el nivel de expresión esperado de dicho biomarcador. En casos en los que se encuentre una concentración elevada del biomarcador en una muestra de un paciente y al mismo tiempo dicho paciente esté genéticamente predispuesto a tener niveles bajos de dichos biomarcadores, la importancia del nivel elevado de biomarcador se incrementa. Del mismo modo, si el nivel de un biomarcador es claramente inferior al normal en un paciente que está genéticamente predispuesto a tener niveles altos de dichos biomarcadores, el hallazgo contradictorio incrementa la importancia de la interpretación del nivel de biomarcadores. El algoritmo utilizado para predecir el riesgo de CaP agresivo puede beneficiarse del uso de variables transformadas, por ejemplo, mediante el uso del valor de $\log_{10}(\text{PSA})$. La transformación es especialmente beneficiosa para variables con una distribución que se desvía claramente de la distribución normal. Las posibles transformaciones de variables incluyen, pero sin limitaciones, logaritmo, inversa, cuadrado y raíz cuadrada. Es además frecuente centrar cada variable a media cero y varianza uno.

- Aunque la combinación de datos puede realizarse de diferentes formas, un procedimiento típico de acuerdo con la presente invención puede ilustrarse de la siguiente forma no limitante.

- En un caso típico, los datos relativos a biomarcadores pertenecientes a una categoría de parámetros se combinarán según una ecuación predeterminada para formar un valor compuesto que está relacionado con el riesgo relacionado con la categoría de parámetros como tal. Un ejemplo no limitante es calcular el valor promedio de todos los valores de mediciones disponibles (datos) de los miembros de una categoría de biomarcadores, y usar dicho valor promedio como valor compuesto que representa dicha categoría de biomarcadores. Este procedimiento puede aplicarse claramente con independencia de cuántos biomarcadores sean miembros que pertenecen a la categoría. Si solo se dispone de los datos de uno de los biomarcadores incluidos en una categoría, se puede usar por sí mismo para representar a la categoría de biomarcadores. En el caso de biomarcadores, el valor medido usado frecuentemente en la etapa de combinación de datos es la concentración de dicho biomarcador encontrado en la muestra biológica.

Por ejemplo, para los biomarcadores PSA y HK2, esta es con más frecuencia la concentración de biomarcador en una muestra de sangre expresada en ng/mL.

El cálculo de la puntuación genética (es decir, el valor compuesto genético, o más específicamente, el valor compuesto de SNP) se basa normalmente en un cociente de probabilidades predeterminado para cada SNP individual incluido en una categoría de parámetros. Para cada SNP, se determina por anticipado el cociente de probabilidades, es decir, la probabilidad de que un individuo portador de un SNP (es decir, que tiene el alelo de riesgo definido por el SNP) tenga la enfermedad o afección en estudio. La determinación del cociente de probabilidades para un SNP se realiza habitualmente en estudios prospectivos grandes que implican a miles de sujetos con afecciones o enfermedades conocidas.

La puntuación genética de un individuo puede calcularse, como ejemplo no limitante, según el siguiente algoritmo: para el individuo en análisis, cada SNP se procesa de la siguiente manera. Para cada SNP, el individuo puede portar dos alelos de riesgo SNP (homocigoto positivo para dicho SNP), o un alelo de riesgo (heterocigoto positivo para dicho SNP) o cero alelos de riesgo (homocigoto negativo para dicho SNP). El número de alelos de un SNP se multiplica por el logaritmo natural del cociente de probabilidades para dicho SNP para formar un valor de evaluación del riesgo para este SNP en particular. Esto significa que un individuo que es negativo para un SNP en particular (es decir, tiene cero alelos de riesgo de SNP) no tendrá ninguna contribución al riesgo procedente de dicho SNP en particular. Este procedimiento se repite para todos los SNP de los que se dispone de los datos de mediciones. Cuando todos los valores de evaluación del riesgo se han calculado, se calcula el promedio de la contribución al riesgo de los SNP de los que se dispone de los datos de medición y se usa como la puntuación genética de dicho individuo, es decir, el valor compuesto genético con respecto a una determinada categoría de SNP. Este procedimiento puede aplicarse claramente con independencia de cuántos SNP sean miembros que pertenecen a la categoría de SNP.

Para ilustrar adicionalmente un procedimiento típico de acuerdo con la presente invención, cuando se aplica a un individuo, se hacen los siguientes supuestos. Se definen dos categorías de parámetros, en primer lugar, una categoría de biomarcadores proteicos (o categoría de biomarcadores) con los miembros Prot1 y Prot2, y en segundo lugar, una categoría genética (o más específicamente, una categoría de SNP) con los miembros Snp1, Snp2 y Snp3. En un experimento en 100 individuos con la enfermedad E conocida y 100 individuos que se sabe no tienen la enfermedad E, se establece la relación de Prot1, Prot2, Snp1, Snp2 y Snp3 con la enfermedad E y se formula como un valor compuesto de biomarcadores proteicos para Prot1 y Prot2, y un valor compuesto genético para Snp1, Snp2 y Snp3, y además un valor compuesto global que a su vez se relaciona con el riesgo de tener la enfermedad E. El valor compuesto para la categoría de biomarcadores proteicos se calcula usando las siguientes ecuaciones predeterminadas:

$$P = (\text{Prot1} + \text{Prot2}) / 3 \text{ [si se dispone de los datos relativos a ambos Prot 1 y Prot 2 (es decir, ambos valores de los Prot1 y Prot2)]}$$

$$P' = \text{Prot1} \text{ [en caso de que solo se disponga de los datos relativos a Prot1 (es decir, el valor de Prot1)]}$$

$$P'' = \text{Prot2} \text{ [en caso de que solo se disponga de los datos relativos a Prot2 (es decir, el valor de Prot2)]}$$

Por tanto, en este caso hipotético se encontró en el experimento que (a) Prot1 y Prot2 tienen la misma escala y (b) el valor de Prot2 tiene el doble de importancia para la evaluación si un individuo tiene la enfermedad E que Prot1. Si solo se dispone de los datos de uno de los biomarcadores proteicos, se puede usar por sí mismo para representar a la categoría de biomarcadores proteicos.

Los cocientes de probabilidades para los miembros de la categoría genética se han determinado de antemano y fueron los siguientes: Snp1 = 1,1; Snp2 = 1,2 y Snp3 = 1,3. El valor compuesto para la categoría genética se calcula como la puntuación genética descrita anteriormente.

El valor compuesto de biomarcadores proteicos y la puntuación genética (que en este caso es equivalente al valor compuesto de la categoría genética, o al valor compuesto de SNP) se combinan después en un valor compuesto global según la siguiente ecuación predeterminada:

$$Y = P + 10 \times \text{puntuación}$$

donde Y está relacionado con el riesgo de tener la enfermedad E, P es el valor compuesto de biomarcadores proteicos (y P puede estar sustituido por P' o P'' como se define anteriormente) y «puntuación» es la puntuación genética. Todas las ecuaciones deben desarrollarse en función de un grupo grande de individuos (en este caso

hipotético, los 100 + 100 individuos) en el que se deriva la relación entre Y y la enfermedad o afección en estudio. En este caso hipotético se asume que si $Y > 5$, el riesgo de que el individuo tenga la enfermedad E es elevado y si $Y > 10$, el riesgo es muy alto.

- 5 Ahora se asume que se está analizando Prot1, Prot2, Snp1, Snp2 y Snp3 en un primer individuo A. En este caso en particular, se hicieron con éxito todas las mediciones y produjeron los siguientes resultados:

Prot1 = 3 ng/mL

Prot2 = 6 ng/mL

- 10 Snp1 = homocigoto negativo, es decir, sin alelos de riesgo = 0

Snp2 = heterocigoto positivo, es decir, un alelo de riesgo = 1

Snp3 = homocigoto positivo, es decir, dos alelos de riesgo = 2

- El valor compuesto para la categoría de biomarcadores proteicos será en este caso $P = (3 + 2 \cdot 6)/3 = 5$. El valor
15 compuesto para la categoría genética, también conocida como puntuación genética, es puntuación = $(0 \cdot \log(1,1) + 1 \cdot \log(1,2) + 2 \cdot \log(1,3))/3 = 0,2357$. El valor compuesto global es $Y = 5 + 10 \cdot 0,2357 = 7,357$. Por tanto, se estima que el riesgo de que el individuo A tenga la enfermedad E es elevado pero no muy alto.

- Ahora se asume además que se está analizando Prot1, Prot2, Snp1, Snp2 y Snp3 en un segundo individuo B. En
20 este caso en particular, se hicieron tres mediciones con éxito y produjeron los siguientes resultados:

Prot1 = 2 ng/mL

Prot2 = DATOS NO DISPONIBLES

Snp1 = homocigoto positivo, es decir, dos alelos de riesgo = 2

- 25 Snp2 = DATOS NO DISPONIBLES

Snp3 = heterocigoto positivo, es decir, un alelo de riesgo = 1

- El valor compuesto para la categoría de biomarcadores proteicos será en este caso $P' = 2$, porque solo se dispone de los resultados de Prot1. El valor compuesto para la categoría genética, también conocida como puntuación
30 genética, es puntuación = $(2 \cdot \log(1,1) + 1 \cdot \log(1,3))/2 = 0,2264$. El valor compuesto global es $Y = 2 + 10 \cdot 0,2264 = 4,264$. Por tanto, se estima que el riesgo de que el individuo B tenga la enfermedad E es bajo.

- Por lo general, en modelos que predicen el riesgo de desarrollar CaPa, existe a menudo uno o dos valores de corte
35 definidos. La elección del valor de corte (o nivel de corte) depende de muchos factores, que incluyen, pero sin limitaciones, el riesgo de la enfermedad como tal y el riesgo asociado con diagnosticar de forma inexacta como positivo a un individuo que no tiene la enfermedad (falso positivo). En el caso general, un modelo predictivo habitualmente es una función monotónica $Y = f(x_1, x_2, \dots, x_N)$ donde el riesgo estimado de tener la enfermedad se correlaciona con el valor creciente de Y. Esto significa que si el valor de corte se establece a un nivel bajo, la prueba producirá un gran número de resultados falsos positivos, pero por otro lado, detectará a la mayoría de los individuos
40 que tienen realmente la enfermedad. Si el nivel de corte se establece a un valor alto ocurre lo contrario, donde los individuos que tienen un valor de Y por encima del nivel de corte tendrán la enfermedad con una probabilidad muy alta, pero un gran número de individuos con la enfermedad recibirán un resultado negativo de la prueba (es decir, un gran número de resultados falsos negativos). La elección del nivel de corte depende de muchos factores, incluido el resultado socio-económico de equilibrar (a) los individuos en los que no se detecta la enfermedad y (b) los individuos
45 tratados que no tienen la enfermedad.

- Cuando se aplica en la práctica, ocurrirá ocasionalmente que una o algunas mediciones fallarán debido a, por ejemplo, problemas técnicos imprevistos, un error humano o cualquier otro motivo imprevisto o poco frecuente. En tales casos, el conjunto de datos obtenidos de un individuo estará incompleto. Normalmente, dicho conjunto de
50 datos incompleto sería difícil o incluso imposible de evaluar. No obstante, la presente invención se basa en mediciones de un gran número de características de las cuales muchas son parcialmente redundantes. Esto significa que también en el caso de individuos cuyo conjunto de datos esté incompleto, será posible en muchos casos realizar una evaluación de alta calidad de acuerdo con la invención. Esto es particularmente cierto dentro de categorías, donde por ejemplo los biomarcadores de tipo calicreína están correlacionados y son parcialmente
55 redundantes. Técnicamente es posible, por tanto, aplicar una estrategia algorítmica en dos etapas, donde la contribución de biomarcadores de tipo calicreína se resume en una puntuación de calicreína (o valor de calicreína). Esta puntuación de calicreína se combina en una segunda etapa con otros datos (tales como la puntuación genética, la edad y los antecedentes familiares, por mencionar algunos ejemplos no limitantes) para producir una declaración diagnóstica o pronóstica de CaP. Se pueden implementar procedimientos similares en dos etapas para otras clases
60 de marcadores, tales como marcadores genéticos relacionados con el IMC o biomarcadores proteicos relacionados

con la superfamilia del factor de crecimiento transformante beta (una gran familia de proteínas reguladoras celulares relacionadas estructuralmente en la que se incluye MIC-1), por mencionar dos ejemplos no limitantes.

El aspecto de la redundancia se puede expresar de muchas maneras diferentes. Una posible forma de implementar el aspecto de la redundancia es definir un conjunto de biomarcadores que representen los biomarcadores relacionados con una familia o campo común. Un ejemplo no limitante de dicha familia o campo son los biomarcadores de tipo calicreína. Se puede determinar más de un conjunto (o categoría) definido de biomarcadores y, además, se pueden aplicar incluso otros biomarcadores fuera de dicho conjunto. Normalmente, las categorías no se solapan, es decir, cualquier biomarcador definido es miembro exclusivo de una categoría definida o se utiliza de forma individual. A continuación, se hace un intento de determinar la presencia o la concentración de todos los biomarcadores. En la mayoría de los casos, la determinación de todos los biomarcadores tendrá éxito, pero en ocasiones no se dispondrá de uno o algunos valores. Para inducir la solidez del modelo a valores no disponibles, es posible definir un valor compuesto de la categoría de biomarcadores que se pueda determinar usando todos o un subconjunto de los miembros de la categoría definida. Para trabajar en la práctica, se requiere que los miembros de la categoría definida de biomarcadores sean al menos parcialmente redundantes. En la etapa siguiente, el valor compuesto de la categoría de biomarcadores se combina con otros valores de biomarcadores, otros valores compuestos de categorías de biomarcadores (si se definen dos o más categorías de biomarcadores), la puntuación genética relacionada con el riesgo de CaP, la puntuación genética relacionada con otras características (tales como IMC o concentración de biomarcadores, por mencionar dos ejemplos no limitantes), los antecedentes familiares, la edad y otros portadores de información relacionada con el riesgo de CaP en un valor compuesto global. El valor compuesto global se usa finalmente para la estimación del riesgo de CaP.

El objetivo del valor compuesto de la categoría de biomarcadores es, por tanto, servir como valor intermedio que se puede estimar usando datos incompletos. Supongamos que una categoría definida de biomarcadores comprende N biomarcadores diferentes denominados B1, B2, B3, ... BN, todos relacionados con la familia B de biomarcadores. En este caso, podría haber N modelos diferentes disponibles para calcular el valor compuesto C de biomarcadores de la familia B:

$$C = f1(B1, B2, B3, \dots BN)$$

$$C = f2(B2, B3, \dots BN)$$

$$C = f3(B1, B3, \dots BN)$$

...

$$C = fN(B1, B2, B3, \dots BN-1)$$

Donde f1(), f2() ... fN() son funciones matemáticas que usan los valores de los biomarcadores B1, ... BN como valor de entrada y, en cierta manera, producen un único valor de salida C que representa el valor compuesto de biomarcadores de la familia B. Un ejemplo no limitante de las funciones f1(), ... fN() incluye combinaciones lineales de los presentes argumentos. Con dicho conjunto de funciones múltiples capaces de calcular C para todos los casos de un único valor de biomarcador no disponible, el cálculo del valor compuesto global es menos sensible a los datos que faltan. Se entiende que la estimación de C podría ser de menos calidad cuando no todos los datos están presentes, pero aún pueden ser lo suficientemente buena para utilizarla en la evaluación del riesgo de CaP. De este modo, usando dicha estrategia, solo las determinaciones del biomarcador N-1 tienen que tener éxito para producir una estimación de C. Es posible además desarrollar estimaciones para cualquier número de datos perdidos, es decir, si las determinaciones del biomarcador N-2 tienen que tener éxito, se podría desarrollar otro conjunto de funciones f() y aplicarse a la estimación de C.

Por tanto, con respecto a los biomarcadores de CaP, la presente invención se refiere a un procedimiento que se basa en una combinación de datos diseñada de forma redundante, como se define en otra parte de la presente solicitud. Más específicamente, el procedimiento comprende medir la presencia o la concentración de biomarcadores de CaP al menos parcialmente redundantes, y donde al menos uno, como por ejemplo dos, de los biomarcadores de CaP se selecciona de entre el grupo compuesto por (i) PSA, (ii) PSA total (tPSA), (iii) PSA intacto (iPSA), (iv) PSA libre (fPSA) y (v) hK2. El procedimiento permite ignorar un subconjunto de al menos uno de los biomarcadores de CaP (i)-(v) al formar el valor compuesto de biomarcadores. En otras palabras, el procedimiento permite que el valor compuesto de biomarcadores se forme a partir de los datos relativos a menos de todos los biomarcadores de CaP de la categoría de biomarcadores, más específicamente, los datos relativos a un subconjunto de a lo sumo cuatro de dichos biomarcadores de CaP. Como el experto apreciará, esto será equivalente a un procedimiento en el que se requieren los datos relativos a un subconjunto de a lo sumo cuatro de dichos biomarcadores de CaP para formar

dicho valor compuesto de biomarcadores. Una ventaja del procedimiento de acuerdo con la presente invención es que la omisión, falta o pérdida de los datos relativos a un subconjunto de dichos biomarcadores de CaP es aceptable al formar el valor compuesto de biomarcadores.

- 5 Como el experto apreciará, la presente invención incluye que el procedimiento comprenda la formación del valor compuesto de biomarcadores a partir de los datos relativos a todos los biomarcadores de la categoría de biomarcadores, siempre que los datos relativos a todos los biomarcadores estén disponibles.

En una realización, el procedimiento permite ignorar un subconjunto de uno, dos, tres o cuatro de los biomarcadores de CaP (i) PSA, (ii) PSA total (tPSA), (iii) PSA intacto (iPSA), (iv) PSA libre (fPSA) y (v) hK2. En otras palabras, el procedimiento permite que dicho valor compuesto de biomarcadores se forme a partir de datos relativos a un subconjunto de cuatro, tres, dos o uno de los biomarcadores de CaP (i)- (v), respectivamente.

Como se menciona previamente en la presente solicitud, el procedimiento puede comprender además analizar uno o cada uno de una diversidad de categorías adicionales de biomarcadores de CaP, donde la combinación de datos para formar cada valor compuesto de biomarcadores adicional está diseñada de forma redundante donde la categoría adicional de biomarcadores de CaP comprende más de un biomarcador de CaP. El procedimiento permite ignorar un subconjunto de biomarcadores de CaP al formar el valor compuesto de biomarcadores. En otras palabras, el procedimiento permite que el valor compuesto de biomarcadores se forme a partir de los datos relativos a menos de todos los biomarcadores de CaP de la categoría adicional de biomarcadores, tales como los datos relativos a un subconjunto del 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 % o 90 % de los biomarcadores de CaP de la categoría adicional de biomarcadores de CaP. Como el experto apreciará, la presente invención incluye que el procedimiento comprenda la formación de cada valor compuesto adicional de biomarcadores a partir de los datos relativos a todos los biomarcadores de CaP de la categoría de biomarcadores de CaP, siempre que los datos relativos a todos los biomarcadores de CaP estén disponibles.

Las puntuaciones de riesgo genético (es decir, puntuaciones genéticas o valores compuestos genéticos, más en particular, valores compuestos de SNP) son también insensibles a pérdidas pequeñas de datos debido a, por ejemplo, problemas técnicos imprevistos, error humano o cualquier otro motivo inesperado o poco frecuente. La contribución de un SNP a la puntuación del riesgo normalmente no se correlaciona con ningún otro SNP. En el caso de SNP, el cambio en el riesgo debido a cada SNP es pequeño, pero usando varios SNP relacionados con una afección en concreto, el cambio en el riesgo para dicha afección se hace suficientemente grande como para afectar al rendimiento del modelo. El número preferido de SNP para formar una puntuación genética es de al menos 3 SNP, preferiblemente 10 SNP, más preferiblemente 25 SNP, aún más preferiblemente 50 SNP, más preferiblemente 60 SNP, aún más preferiblemente 70 SNP, incluso más preferiblemente 80 SNP, más preferiblemente 90 SNP, incluso más preferiblemente 100 SNP, aún más preferiblemente 150 SNP, incluso más preferiblemente 200 SNP, aún más preferiblemente 250 y aún incluso más preferiblemente 300 SNP. Esto significa que el impacto de cualquier SNP individual sobre el resultado total normalmente es pequeño, y la omisión de unos pocos SNP normalmente no alterará en gran medida la evaluación del riesgo por medio de la puntuación genética global, es decir, no alterará normalmente el valor compuesto de SNP de forma significativa. En el estado actual de la técnica, la pérdida de datos típica en las mediciones genéticas a gran escala es del orden del 1-2 %, lo que significa que si una puntuación está compuesta de 100 SNP diferentes, la caracterización genética típica de un individuo proporcionaría información sobre 98-99 de estos SNP. El modelo presente como tal, como se descubre en el trabajo de la presente invención, puede sin embargo soportar una pérdida o falta de datos más grande, como por ejemplo, 5-7 % de pérdida de información, o 7-15 %, o incluso 15-30 %. En este sentido, la combinación de datos relativos a SNPcp es al menos parcialmente redundante.

En consecuencia, también con respecto a los marcadores genéticos (SNP), la presente invención se refiere a un procedimiento que se basa en una combinación de datos diseñada de forma redundante, como se define en otra parte de la presente solicitud. El procedimiento permite ignorar al menos el 5 % de los SNPcp al formar el valor compuesto de SNP. En otras palabras, el procedimiento permite que el valor compuesto de SNPcp se forme a partir de los datos relativos a menos de todos los SNPcp de la categoría de SNPcp, más específicamente, los datos relativos a un subconjunto de a lo sumo el 95 % de dichos SNPcp. Como el experto apreciará, esto será equivalente a un procedimiento en el que se requieren los datos relativos a un subconjunto de a lo sumo el 95 % de dichos SNPcp para formar dicho valor compuesto de SNPcp. Una ventaja del procedimiento de acuerdo con la presente invención es que la omisión, falta o pérdida de los datos relativos a un subconjunto de dichos SNPcp es aceptable al formar el valor compuesto de SNPcp.

Como el experto apreciará, la presente invención incluye que el procedimiento comprenda la formación del valor compuesto de SNPcp a partir de los datos relativos a todos los SNPcp de la categoría de SNPcp, siempre que los

datos relativos a todos los SNPcp estén disponibles. De forma similar, la presente invención incluye que el procedimiento comprenda la formación del valor compuesto de SNPcp a partir de los datos relativos a un subconjunto del 99 %, 98 %, 97 % o 96 % de dichos SNPcp.

- 5 En una realización, el procedimiento permite ignorar el 6 %, 7 %, 8 %, 9 %, 10 %, 15 %, 20 %, 25 % o 30 % de los SNPcp al formar el valor compuesto de SNPcp. En otras palabras, el procedimiento permite que dicho valor compuesto de SNPcp se forme a partir de los datos relativos a un subconjunto del 94 %, 93 %, 92 %, 91 %, 90 %, 85 %, 80 %, 75 % o 70 % de los SNPcp, respectivamente.
- 10 Un ejemplo no limitante de dicha combinación de datos diseñada de forma redundante es un cálculo del promedio del riesgo relacionado con cada SNP para el que existen datos de mediciones. Otro ejemplo no limitante de dicha combinación de datos diseñada de forma redundante es proporcionar varias ecuaciones independientes para calcular el valor compuesto, una ecuación para cada subconjunto de datos que se puede usar para producir dicho valor compuesto.
- 15 Un procedimiento adecuado para asociar un SNP con una afección (por ejemplo CaP o IMC >25 o concentración elevada del biomarcador hk2 en sangre) se ha descrito en el artículo «Blood Biomarker Levels to Aid Discovery of Cancer-Related Single-Nucleotide Polymorphisms: Kallikreins and Prostate Cancer» de Robert Klein y coautores publicado en Cancer Prev Res 2010; 3:611-619. En este artículo, los autores describen cómo se podría asociar el
- 20 SNP rs2735839 a un valor elevado de (PSA libre)/(PSA total). Además, se podría asociar el SNP rs10993994 con el riesgo elevado de CaP, el valor elevado de PSA total, el valor elevado de PSA libre y el valor elevado de hk2, y finalmente, SNP rs198977 se asoció con el riesgo elevado de CaP, el valor elevado de (PSA libre)/(PSA total) y el valor elevado de hk2.
- 25 En la práctica, un procedimiento común para asociar un SNP con una afección se basa en el acceso a un ensayo clínico de casos y controles en el que se comparan dos grupos grandes de individuos, un grupo de controles sanos y un grupo de casos que padecen la afección en estudio. Se obtiene el genotipo para la mayoría de los SNP comunes conocidos de todos los individuos de cada grupo. Cuando se dispone de todos los datos de genotipado, se investiga si la frecuencia de los alelos está alterado de forma significativa entre el grupo de casos y el grupo de controles. En
- 30 dicha configuración, la unidad típica para la notificación de tamaños del efecto es el cociente de probabilidades. El cociente de probabilidades proporciona información sobre el cociente entre dos proporciones: la proporción de individuos del grupo de casos que tienen un alelo específico y la proporción de individuos del grupo de controles que tienen el mismo alelo. Si la frecuencia del alelo en el grupo de casos es significativamente superior a la frecuencia del alelo en el grupo de controles, el cociente de probabilidades será superior a 1. Si la frecuencia del alelo en el
- 35 grupo de casos es significativamente inferior a la frecuencia del alelo en el grupo de controles, el cociente de probabilidades será inferior a 1.

- Se ha descrito un procedimiento preferido para combinar información procedente de diversas fuentes en el artículo «Polygenic Risk Score Improves Prostate Cancer Risk Prediction: Results from the Stockholm-1 Cohort Study» de
- 40 Markus Aly y coautores publicado en EUROPEAN UROLOGY 60 (2011) 21-28. Las asociaciones entre cada SNP y el CaP en biopsia se evaluaron empleando una prueba de tendencia de Cochran-Armitage. Los cocientes de probabilidades (CP) alélicos con intervalos de confianza del 95 % se calcularon usando modelos de regresión logística. Para cada paciente se generó una puntuación de riesgo genético sumando el número de alelos de riesgo (0,1 o 2) en cada uno de los SNP multiplicado por el logaritmo del CP de ese SNP. Las asociaciones entre el
 - 45 diagnóstico de CaP y los factores de riesgo evaluados se estudiaron en un análisis de regresión logística. La porción del modelo relacionado con información no genética incluía el PSA total transformado logarítmicamente, la relación PSA libre/total transformada logarítmicamente, la edad en el momento de la biopsia y los antecedentes familiares de CaP (sí o no). Se utilizó una validación cruzada repetida diez veces para estimar las probabilidades predichas de CaP en la biopsia. Se generaron los intervalos de confianza del 95 % para los valores ROC-AUC usando una
 - 50 aproximación normal. Todos los valores de p notificados se basan en hipótesis bilaterales.

- Existen muchas razones justificadas para distinguir entre el cáncer de próstata en general y el cáncer de próstata agresivo. En la mayoría de los casos, el cáncer de próstata es una enfermedad que progresa lentamente. El hecho de que la mayoría de los hombres sean diagnosticados tarde en su vida significa que una gran parte de los hombres
- 55 diagnosticados de cáncer de próstata mueren por otras causas. Por tanto, la capacidad para estimar si un individuo está en riesgo elevado de padecer cáncer de próstata agresivo, antes de la biopsia, hace posible, por ejemplo, que el paciente se motive para cambiar de estilo de vida. Dejar de fumar, lograr un valor de IMC inferior a 30 y hacer ejercicio regularmente (aproximadamente 30 minutos 3 a 6 días a la semana) son todos factores que, en general, favorecen la supervivencia en condiciones de enfermedad grave, como el cáncer de próstata. Por tanto, si se
 - 60 encuentra que un individuo tiene un riesgo elevado de CaPa, es razonable sugerir al individuo que deje de fumar,

intente conseguir un IMC <30 y empiece a hacer ejercicio. Otro aspecto importante son los problemas alimentarios. A través de cambios en la dieta, el desarrollo del CaP se puede reducir o retrasar. Existen indicios que sugieren que una reducción del consumo de productos lácteos puede disminuir el riesgo de aparición de CaP, como describen Song y coautores en el artículo «Whole milk intake is associated with prostate cancer-specific mortality among U.S. male physicians» publicado en J Nutr. Feb 2013; 143(2): 189-96. Existen indicios similares de los efectos positivos del consumo de té verde y de productos derivados de la soja. Por tanto, si se encuentra que un individuo tiene un riesgo elevado de CaPa, es razonable sugerir a este individuo que disminuya el consumo de productos lácteos y/o que aumente el consumo de té verde y productos derivados de la soja.

10 Ejemplo 1

Para ilustrar la presente invención, se extrajo un conjunto de datos que comprenden 215 casos (sujetos que se sabe sufren CaPa con un grado de Gleason de 7 o superior) y 627 controles (sujetos que se sabe que no padecen CaPa) del conjunto de datos de STHLM2. El conjunto de datos de STHLM2 se ha discutido en el dominio público como es evidente en la página web <http://sthlm2.se/>. En resumen, durante 2010-2012 se incluyeron en el estudio STHLM2 aproximadamente 26.000 varones que se sometieron a una prueba de PSA en la zona de Estocolmo. Se caracterizó a 215 + 627 = 842 sujetos con respecto a los siguientes biomarcadores y SNP.

Biomarcadores:

- 20 Antígeno prostático específico total (tPSA) [ng/mL]
- Antígeno prostático específico intacto (iPSA) [ng/mL]
- Antígeno prostático específico libre (fPSA) [ng/mL]
- Calicreína 2 humana (hK2) [ng/mL]
- 25 Citoquina inhibidora de macrófagos 1 (MIC-1) [ng/mL]
- Microseminoproteína-beta (MSMB) [ng/mL]

SNP:

- 30 657del5, rs10086908, rs1016343, rs10187424, rs1041449, rs10486567, rs1054564, rs10875943, rs10896449, rs10934853, rs10993994, rs11067228, rs11135910, rs11228565, rs11568818, rs11649743, rs11650494, rs11672691, rs11704416, rs12130132, rs12409639, rs12418451, rs12500426, rs12543663, rs12621278, rs12653946, rs1270884, rs130067, rs13252298, rs13385191, rs1354774, rs1363120, rs137853007, rs138213197, rs1447295, rs1465618, rs1512268, rs1571801, rs16901979, rs16902094, rs17021918, rs17632542, rs17879961,
- 35 rs1859962, rs1894292, rs1933488, rs1983891, rs2018334, rs2121875, rs2242652, rs2273669, rs2292884, rs2405942, rs2660753, rs2735839, rs2736098, rs2928679, rs3213764, rs339331, rs3771570, rs3850699, rs3863641, rs401681, rs4245739, rs4430796, rs445114, rs4643253, rs4857841, rs4962416, rs5759167, rs5919432, rs5945619, rs6062509, rs620861, rs6465657, rs6763931, rs684232, rs6869841, rs6983267, rs6983561, rs7127900, rs7210100, rs721048, rs7241993, rs7611694, rs7679673, rs7931342, rs8008270, rs8102476, rs888663, rs902774,
- 40 rs9364554, rs9600079, rs9623117 (tabla II) Se recopiló la información básica de cada sujeto, incluido edad y antecedentes familiares (sí o no). La edad se expresó en años.

Con el fin de decidir qué sujetos deben ser remitidos a biopsia, se requiere predecir un valor para cada sujeto analizado que se correlacione con la probabilidad de que dicho sujeto tenga cáncer de próstata con un grado de Gleason de 7 o superior. Esto se puede realizar combinando los valores medidos de los biomarcadores en la

45 siguiente ecuación predeterminada:

$$y = -0,4366579 + 0,0577639 * \text{puntuación} - 0,1026622 * \text{HK2} - 0,0312050 * \text{fPSA} + 0,0640730 * \text{iPSA} + 0,0256631 * \text{MIC1} - 0,0069049 * \text{MSMB} + 0,0012231 * \text{tPSA} + 0,0069759 * \text{edad}$$

En esta ecuación, «puntuación» se refiere a la puntuación genética calculada como se describe en el artículo «Polygenic Risk Score Improves Prostate Cancer Risk Prediction: Results from the Stockholm-1 Cohort Study» de Markus Aly y coautores publicado en EUROPEAN UROLOGY 60 (2011) 21-28, que contiene los SNP validados de susceptibilidad al cáncer de próstata (estando dichos SNP relacionados con la susceptibilidad al cáncer de próstata o relacionados con los niveles plasmáticos de los biomarcadores PSA, PSA libre, MSMB y MIC-1) enumerados en el presente ejemplo. Los parámetros «HK2», «fPSA», «iPSA», «MIC-1», «MSMB» y «tPSA» se refieren a los respectivos valores medidos (no transformados) de estos biomarcadores y «edad» se refiere a la edad del sujeto. La

ecuación se obtuvo usando el estimador ordinario de mínimos cuadrados (también pueden utilizarse fácilmente otros estimadores lineales, p. ej., el estimador de regresión logística). En este modelo en particular se omitió la información relativa a los antecedentes familiares.

- 5 El valor resultante «y» estará fuertemente correlacionado con el riesgo de tener cáncer de próstata con un grado de Gleason de 7 o superior, como se ilustra en la figura 1. Las curvas ROC de la figura 1 representan PSA (101) solo y el modelo descrito en este ejemplo (102). Si «y» está por encima de un valor de corte, se debe recomendar al hombre que acuda a un urólogo para una exploración de la próstata mediante biopsia. El hecho de que este modelo pronostique el CaP agresivo de alto grado significa implícitamente que si el valor resultante «y» es bajo, existe aún un riesgo de que el paciente tenga CaP, aunque de una forma no agresiva. Un valor resultante «y» bajo también puede indicar que el paciente no tiene CaP.

- El valor del punto de corte depende del equilibrio entre la sensibilidad y la especificidad de la prueba. Si, por ejemplo, se usa el valor de corte de 0,166, esta prueba en particular dará como resultado una sensibilidad de 0,9 y una especificidad de 0,38. Esto se puede comparar con el uso del valor de PSA solo como prueba de cribado, que tiene como resultado una sensibilidad de 0,9 y una especificidad de 0,22. En la práctica, esto significa que este modelo en particular aplicado a la población de 827 sujetos tendría como resultado el mismo número de cánceres de alto riesgo (grado de Gleason de 7 y superior) detectados que la prueba de PSA, pero con 100 sujetos menos remitidos para biopsia, lo que corresponde a una mejora de aproximadamente el 15 % en comparación con la prueba de PSA solo. Si, como en un segundo ejemplo, se usa el valor de corte de 0,201, esta prueba en particular dará como resultado una sensibilidad de 0,8 y una especificidad de 0,52. Al nivel de sensibilidad de 0,8, se evitaría aproximadamente el 20 % de las biopsias que se pronostican con el PSA.

Ejemplo 2

- 25 Para ilustrar adicionalmente la presente invención, se aplicó un método informático alternativo para obtener una predicción. Las ecuaciones como la presentada en el ejemplo 1 no son la única forma en la que se pueden combinar los biomarcadores para predecir el CaPa. De hecho, el procedimiento para calcular «y» con el fin de predecir el CaPa puede ser intrincado e incluso imposible de anotar en una hoja de papel. Un ejemplo más complicado, aunque muy convincente, de cómo se pueden combinar los biomarcadores es el uso de un bosque de árboles de decisiones. En la figura 2 se representa un ejemplo de árbol de decisiones (200). Supongamos que a un sujeto de 81 años se le hicieron las pruebas de biomarcadores y SNP, con los resultados de HK2 = 0,2425 y PSA = 84,1. Si se aplica el árbol de decisiones (200) como se muestra en el ejemplo 2, el nodo superior (201) se refiere al valor de hk2. Puesto que el sujeto tiene un valor de HK2 que no satisface la condición del nodo, se sigue la rama izquierda de ese nodo. El segundo nodo (202) también está relacionado con el valor de hk2, y en este caso el sujeto tiene un valor de hk2 que no satisface la condición del nodo, por lo que se sigue la rama derecha de ese nodo. El nodo de tercer nivel (203) está relacionado con la edad. Puesto que la edad del sujeto no satisface la condición del nodo, se sigue la rama derecha de ese nodo. El nodo de cuarto nivel (204) está relacionado con el valor de PSA, y puesto que el valor de PSA del sujeto satisface la condición del nodo, se sigue la rama izquierda de ese nodo. En este punto, no hay más nodos, lo que significa que se ha llegado a una hoja del árbol de decisiones. Cada hoja tiene su correspondiente resultado; en este ejemplo en particular, un valor de hoja de «1» significa «remitir para biopsia» y «0» significa «no remitir para biopsia». El sujeto del ejemplo en este caso acabó en una hoja con un valor de «0», lo que significa que la predicción que proporciona este árbol de decisiones es «no remitir para biopsia».
- 45 Un problema al basarse en un solo árbol de decisiones para calcular «y» en la predicción del CaPa es que un único árbol de decisiones tiene una varianza muy alta (es decir, si los datos cambian ligeramente, también es probable que cambie el valor calculado de «y», lo que conlleva una varianza en la predicción del CaPa), aunque su sesgo es muy bajo. Un posible procedimiento para reducir la alta varianza es construir un bosque de árboles descorrelacionados usando el algoritmo de bosque aleatorio como se describe en el artículo «Random Forests» de Leo Breiman publicado en Machine Learning 45 (1): 5-32 (2001)). Crece un gran número de árboles y, antes del crecimiento de cada árbol, los datos se alteran aleatoriamente de tal modo que el valor esperado de su predicción no cambia. Para predecir el CaPa, todos los árboles emiten un voto para decidir si un sujeto debe ser remitido para biopsia. Esta predicción derivada de los votos conserva las propiedades no sesgadas de los árboles de decisiones, aunque reduce considerablemente la varianza (de forma similar a cómo la varianza de una media es menor que la varianza de las mediciones individuales usadas para calcular la media). Puesto que el algoritmo de bosque aleatorio depende de la generación aleatoria de números, anotar el algoritmo de predicción resultante de forma cerrada es un procedimiento complejo.

- Cuando se aplica al conjunto de datos descrito en el ejemplo 1, este modelo, a una sensibilidad de 0,9, puede evitar aproximadamente el 20 % del número de biopsias en comparación con PSA solo.

Ejemplo 3

Para ilustrar incluso más la presente invención, se extrajo un conjunto de datos que comprenden 51 casos (sujetos que se sabe sufren CaPa con un grado de Gleason de 7 o superior) y 195 controles (sujetos que se sabe que no sufren CaPa) del conjunto de datos de STHLM2. Todos estos casos y controles tenían un valor de IMC superior a 25. Se caracterizó a $51 + 195 = 246$ sujetos con respecto a los siguientes biomarcadores.

Biomarcadores:

- 10 Antígeno prostático específico total (tPSA) [ng/mL]
- Antígeno prostático específico intacto (iPSA) [ng/mL]
- Antígeno prostático específico libre (fPSA) [ng/mL]
- Calicreína 2 humana (hk2) [ng/mL]
- Citoquina inhibidora de macrófagos 1 (MIC-1) [ng/mL]
- 15 Microseminoproteína-beta (MSMB) [ng/mL]

En este ejemplo se aplicaron también los mismos SNP descritos en el ejemplo 1. Se recopiló la información básica de cada sujeto, incluyendo si el sujeto se había sometido a una biopsia previa (prevBiop) de próstata, edad y antecedentes familiares (sí o no). La edad se expresó en años, la estatura en metros y el peso en kilogramos.

20 Con el fin de decidir qué sujetos deben ser remitidos a biopsia, se requiere predecir un valor para cada sujeto analizado que se correlacione con la probabilidad de que dicho sujeto tenga cáncer de próstata con un grado de Gleason de 7 o superior. Esto se puede realizar combinando los valores medidos de los biomarcadores en un valor compuesto global usando la siguiente ecuación predeterminada:

$$\begin{aligned}
 25 \quad y = & 21,487704 + 0,548938 * \text{preBiopsia} + 0,014242 * \text{PuntuaciónGen} + 0,311481 * \text{hk2} - \\
 & 0,043471 * \text{fPSA} + 0,047176 * \text{iPSA} + 0,068407 * \text{mic1} - 0,008860 * \text{msmb} + 0,002693 \\
 & * \text{tPSA} + 0,006325 * \text{edad} - 0,121356 * \text{altura} + 0,119005 * \text{peso} - 0,388930 * \text{imc}
 \end{aligned}$$

En esta ecuación, «puntuación» es la variable puntuación genética calculada como se describe en el ejemplo 1 anterior. Los parámetros «HK2», «fPSA», «iPSA», «MIC-1», «MSMB» y «tPSA» se refieren a los respectivos valores medidos (no transformados) de estos biomarcadores y «edad», «estatura», «peso» e «IMC» se refiere a la edad, estatura, peso e IMC del sujeto. El parámetro «prevBiop» indica si el sujeto se ha sometido previamente a una biopsia de próstata, lo que refleja los antecedentes patológicos de dicho sujeto. La ecuación se obtuvo usando el estimador ordinario de mínimos cuadrados (también pueden utilizarse fácilmente otros estimadores lineales, p. ej., el estimador de regresión logística). En este modelo en particular se omitió la información relativa a los antecedentes familiares.

El valor resultante «y» estará fuertemente correlacionado con el riesgo de tener cáncer de próstata agresivo con un grado de Gleason de 7 o superior, como se ilustra en la figura 3. Las curvas ROC de la figura 3 representan PSA (301) solo y el modelo descrito en este ejemplo (302). Si «y» está por encima de un valor de corte, se debe recomendar al hombre que acuda un urólogo para una exploración de la próstata mediante biopsia.

El valor del punto de corte depende del equilibrio entre la sensibilidad y la especificidad de la prueba. Si, por ejemplo, se usa el valor de corte de 0,201, esta prueba en particular dará como resultado una sensibilidad de la prueba de 0,8 y la prueba evitará aproximadamente el 44 % de las biopsias en comparación con el uso de PSA solo.

Ejemplo 4

Para ilustrar incluso más los aspectos de las categorías de parámetros y la redundancia dentro de una categoría, se caracterizó el conjunto de datos del ejemplo 1 con respecto a lo siguiente:

Biomarcadores:

- Antígeno prostático específico total (tPSA) [ng/mL]
- Antígeno prostático específico intacto (iPSA) [ng/mL]
- 55 Antígeno prostático específico libre (fPSA) [ng/mL]
- Calicreína 2 humana (HK2) [ng/mL]

Citoquina inhibidora de macrófagos 1 (MIC-1) [ng/mL]
Microseminoproteína-beta (MSMB) [ng/mL]

SNP; pertenecientes a la categoría de SNP relacionados con el CaP (SNPcp):

- 5 657del5, rs10086908, rs1016343, rs10187424, rs1041449, rs10486567, rs1054564, rs10875943, rs10896449, rs10934853, rs10993994, rs11067228, rs11135910, rs11228565, rs11568818, rs11649743, rs11650494, rs11672691, rs11704416, rs12130132, rs12409639, rs12418451, rs12500426, rs12543663, rs12621278, rs12653946, rs1270884, rs130067, rs13252298, rs13385191, rs1354774, rs1363120, rs137853007, rs138213197, rs1447295, rs1465618, rs1512268, rs1571801, rs16901979, rs16902094, rs17021918, rs17632542, rs17879961, rs1859962, rs1894292, rs1933488, rs1983891, rs2018334, rs2121875, rs2242652, rs2273669, rs2292884, rs2405942, rs2660753, rs2735839, rs2736098, rs2928679, rs3213764, rs339331, rs3771570, rs3850699, rs3863641, rs401681, rs4245739, rs4430796, rs445114, rs4643253, rs4857841, rs4962416, rs5759167, rs5919432, rs5945619, rs6062509, rs620861, rs6465657, rs6763931, rs684232, rs6869841, rs6983267, rs6983561, rs7127900, rs7210100, rs721048, rs7241993, rs7611694, rs7679673, rs7931342, rs8008270, rs8102476, rs888663, rs902774, rs9364554, rs9600079, rs9623117

Se recopiló la información básica de cada sujeto, incluyendo la edad y si el sujeto se había sometido a una biopsia previa (sí o no). La edad se expresó en años.

- 20 La ecuación para el valor compuesto global, que se usa como modelo predictivo, se diseñó según la ecuación predeterminada:

$$Y = -0,632820 + 0,118107 * K + 0,139045 * \text{preBiopsia} + 0,051609 * \text{puntuación} + 0,048033 * \text{MIC1} - 0,001368 * \text{MSMB} + 0,008002 * \text{edad}$$

- 25 Donde «puntuación» es la puntuación genética, es decir, el valor compuesto obtenido a partir de los SNP relacionados con el CaP (es decir, valor compuesto de SNPcp), como se describe anteriormente, y K es el valor compuesto para la categoría de parámetros para biomarcadores de tipo calicreína, MIC-1 es la concentración de MIC-1, MSMB es la concentración de MSMB, «edad» es la edad del individuo y «PrevBiop» es 1 si el individuo se sometió previamente a una biopsia (y 0 si no lo hizo). Dependiendo de la disponibilidad de los datos de calicreína de un individuo en particular, el valor compuesto para biomarcadores de tipo calicreína K de la categoría se calculó de diferentes formas.

$$K = (0,6122516 + 0,0012714 * f\text{PSA} + 0,0001864 * \text{PSA} + 0,0200385 * i\text{PSA} - 0,0377976 * \text{HK2} - 1,3108243 f/t\text{PSA}) / 0,1559314$$

$$K' = (0,3961801 + 0,0001864 * \text{PSA} + 0,0200385 * i\text{PSA} - 0,0377976 * \text{HK2}) / 0,109478$$

$$K'' = (0,3961967 + 0,0012714 * f\text{PSA} + 0,0200385 * i\text{PSA} - 0,0377976 * \text{HK2}) / 0,1090876$$

$$K''' = (0,3987352 + 0,0200385 * i\text{PSA} - 0,0377976 * \text{HK2}) / 0,1033296$$

$$K'''' = (0,6548828 + 0,0012714 * f\text{PSA} + 0,0001864 * \text{PSA} - 1,3108243 f/t\text{PSA}) / 0,1068742$$

- 35 En esta ecuación, PSA es la concentración de PSA, fPSA en la concentración de PSA libre, iPSA es la concentración de PSA intacto, HK2 es la concentración de HK2 y f/tPSA es el cociente entre PSA libre y PSA total. K es el valor del parámetro adecuado para su uso cuando todos los datos de calicreína están disponibles. Los parámetros K', K'', K''' y K'''' son aproximaciones de K que son adecuadas para su uso en casos en los que uno o varios de los datos de calicreína no están disponibles.

40

Cuando se probó el modelo descrito anteriormente, se obtuvieron los siguientes resultados:

- Modelo completo, todos los datos incluidos: ROC-AUC = 0,77

- Usando todos los SNP y la aproximación K': ROC-AUC = 0,70
- Usando todos los SNP y la aproximación K'': ROC-AUC = 0,70
- Usando todos los SNP y la aproximación K''': ROC-AUC = 0,70
- Usando todos los SNP y la aproximación K''': ROC-AUC = 0,75
- 5 • Usando los datos de K'''' y dejando fuera aleatoriamente el 10 % de los datos de SNP: ROC-AUC = 0,74
- Usando los datos de K'''' y dejando fuera aleatoriamente el 30 % de los datos de SNP: ROC-AUC = 0,73

Como punto de referencia, cuando se usa solo PSA para predecir el riesgo de CaPa, ROC-AUC = 0,65. Por tanto, el modelo en el presente ejemplo es (a) mejor que el modelo de referencia cuando se usan todos los datos, pero

10 también (b) sólido ante la pérdida de datos de entrada, gracias a la redundancia en las categorías de parámetros. Es posible dejar fuera uno o más resultados de mediciones (es decir, datos) de biomarcadores de tipo calicreína, también en combinación con una pérdida del 10 % (o incluso el 30 %) de la información de SNP y generar aún resultados útiles que son mejores que el modelo de referencia en el que solo se usa el PSA. En un contexto práctico,

15 dicha operación sólida hace posible estimar el riesgo para un individuo que tiene CaPa incluso en casos en los que faltan algunos datos debido a un fallo de la tecnología, falta de material de la muestra, error humano o cualquier otro motivo. Esto tiene el potencial de reducir los costes para el profesional sanitario porque se reduciría el número de repeticiones de la prueba. Además haría la situación más cómoda para el individuo, con una respuesta más rápida y aliviando su necesidad de acudir al profesional sanitario para proporcionar una muestra adicional para la repetición de las pruebas.

20 Aunque la invención se ha descrito con respecto a sus realizaciones preferidas, que constituyen el mejor modo actualmente conocido por el inventor, debe entenderse que se pueden realizar diversos cambios y modificaciones que serían obvios para el experto en la materia sin apartarse del alcance de la invención como se establece en las reivindicaciones adjuntas.

25

REIVINDICACIONES

1. Un procedimiento basado en una combinación de datos diseñada de forma redundante para indicar la presencia o no presencia de cáncer de próstata (CaP) agresivo en un individuo, que comprende los pasos de:
 1. Proporcionar al menos una muestra biológica de dicho individuo;
 2. Analizar en dicha muestra biológica
- a. una categoría de biomarcadores de CaP, midiendo la presencia o la concentración de al menos tres biomarcadores de CaP de tipo calicreína; y
- b. una categoría de SNP relacionados con el CaP (SNPcp), midiendo la presencia o ausencia de uno o dos alelos de riesgo de cada uno de los diversos SNPcp de dicha categoría de SNPcp;
3. Combinar los datos relativos a al menos tres biomarcadores de CaP de tipo calicreína para formar un valor compuesto de biomarcador que represente el riesgo de desarrollar CaP relacionado con biomarcadores de CaP, donde el procedimiento permite ignorar datos de un subconjunto de al menos uno de dichos biomarcadores de CaP, como por ejemplo uno o dos de dichos biomarcadores de CaP, al formar dicho valor compuesto de biomarcadores de al menos tres biomarcadores de CaP;
4. Combinar los datos relativos a dicha categoría de SNPcp para formar un valor compuesto de SNPcp que represente el riesgo de desarrollar CaP relacionado con SNPcp, donde el procedimiento permita ignorar datos de un subconjunto de aproximadamente del 10 % al 30 %, como por ejemplo el 15 %, 20 % o 30 % de los SNPcp de la categoría de SNPcp, al formar el valor compuesto de SNPcp, donde dicho valor compuesto de SNPcp se forma a partir de al menos 70 SNPcp seleccionados de entre cualquiera de los SNPcp de la tabla I o la tabla II;
5. Combinar el valor compuesto de biomarcadores y el valor compuesto de SNPcp para formar un valor compuesto global;
6. Correlacionar dicho valor compuesto global con la presencia o no presencia de CaP agresivo en dicho individuo comparando dicho valor compuesto global con un valor de corte predeterminado establecido con muestras de control de presencia conocida de CaP agresivo y de no presencia conocida de CaP agresivo, donde al menos uno, como por ejemplo dos, de dichos biomarcadores de CaP de tipo calicreína se selecciona de entre el grupo compuesto por (i) PSA, (ii) PSA total (tPSA), (iii) PSA intacto (iPSA), (iv) PSA libre (fPSA) y (v) hK2.
2. El procedimiento de la reivindicación 1, donde dicho valor compuesto de SNPcp se forma a partir de al menos 80 SNPcp, 90 SNPcp, 100 SNPcp o 150 SNPcp o a partir de al menos el 70 % de los SNPcp de la tabla I o la tabla II.
3. El procedimiento de la reivindicación 1 o 2, donde el procedimiento permite ignorar un subconjunto de al menos uno de dichos biomarcadores de CaP (i)-(v) de la categoría de biomarcadores de CaP al formar dicho valor compuesto de biomarcadores, como por ejemplo un subconjunto de uno, dos, tres o cuatro de dichos biomarcadores de CaP (i)-(v).
4. El procedimiento de una cualquiera de las reivindicaciones precedentes, donde los datos relativos a dicha categoría de biomarcadores de CaP se combinan según una ecuación predeterminada para formar dicho valor compuesto de biomarcadores.
5. El procedimiento de una cualquiera de las reivindicaciones precedentes, donde los datos relativos a dicha categoría de SNPcp se combinan según una ecuación predeterminada para formar dicho valor compuesto de SNPcp.
6. El procedimiento de una cualquiera de las reivindicaciones precedentes, donde dicho valor compuesto de biomarcadores y dicho valor compuesto de SNPcp se combinan según una ecuación predeterminada para formar dicho valor compuesto global.
7. El procedimiento de una cualquiera de las reivindicaciones precedentes, que además comprende recomendar al individuo una biopsia si el valor compuesto global es superior al valor de corte.
8. El procedimiento de una cualquiera de las reivindicaciones precedentes, que además comprende

recomendar al individuo cambiar los hábitos alimentarios, perder peso, lograr un valor de IMC por debajo de 30, hacer ejercicio con regularidad y/o dejar de fumar, si el valor compuesto global es superior al valor de corte.

9. El procedimiento de una cualquiera de las reivindicaciones precedentes, donde los al menos 70 SNP se seleccionan de entre el grupo consistente en 657del5, rs10086908, rs1016343, rs10187424, rs1041449, rs10486567, rs1054564, rs10875943, rs10896449, rs10934853, rs10993994, rs11067228, rs11135910, rs11228565, rs11568818, rs11649743, rs11650494, rs11672691, rs11704416, rs12130132, rs12409639, rs12418451, rs12500426, rs12543663, rs12621278, rs12653946, rs1270884, rs130067, rs13252298, rs13385191, rs1354774, rs1363120, rs137853007, rs138213197, rs1447295, rs1465618, rs1512268, rs1571801, rs16901979, rs16902094, rs17021918, rs17632542, rs17879961, rs1859962, rs1894292, rs1933488, rs1983891, rs2018334, rs2121875, rs2242652, rs2273669, rs2292884, rs2405942, rs2660753, rs2735839, rs2736098, rs2928679, rs3213764, rs339331, rs3771570, rs3850699, rs3863641, rs401681, rs4245739, rs4430796, rs445114, rs4643253, rs4857841, rs4962416, rs5759167, rs5919432, rs5945619, rs6062509, rs620861, rs6465657, rs6763931, rs684232, rs6869841, rs6983267, rs6983561, rs7127900, rs7210100, rs721048, rs7241993, rs7611694, rs7679673, rs7931342, rs8008270, rs8102476, rs888663, rs902774, rs9364554, rs9600079 y rs9623117.

10. El procedimiento de una cualquiera de las reivindicaciones 1-8, donde los al menos 70 SNP se seleccionan de entre el grupo consistente en rs582598, rs439378, rs2207790, rs1046011, rs10458360, rs7525167, rs10489871, rs7529518, rs4245739, rs4512641, rs10178804, rs11900952, rs1873555, rs10191478, rs6755901, rs6545962, rs721048, rs2710647, rs12612891, rs2028900, rs1009, rs12233245, rs6760417, rs10496470, rs10199796, rs12475433, rs16860513, rs12151618, rs3765065, rs13017302, rs12988652, rs871688, rs749264, rs3771570, rs4346531, rs6770955, rs12637074, rs2660753, rs13319878, rs6437715, rs2162185, rs1515542, rs2270785, rs9830294, rs1439024, rs6762443, rs888507, rs6794467, rs12490248, rs1477886, rs4833103, rs3796547, rs17779822, rs2366711, rs16849146, rs1894292, rs12640320, rs3805284, rs12500426, rs4699312, rs17021918, rs7679673, rs2047408, rs2647262, rs12506850, rs7658048, rs2078277, rs12505546, rs13113975, rs4246742, rs2736098, rs401681, rs11134144, rs10060513, rs40485, rs2087724, rs1482679, rs16901841, rs1295683, rs2070874, rs7752029, rs2018334, rs9358913, rs1140809, rs409558, rs3096702, rs9267911, rs2025645, rs9359428, rs6569371, rs2813532, rs1933488, rs712242, rs6934898, rs9456490, rs651164, rs3120137, rs9364554, rs9457937, rs10486562, rs10807843, rs7801918, rs6962297, rs2465796, rs6957416, rs7777631, rs2272316, rs6961773, rs2132276, rs13265330, rs16887736, rs2911756, rs2272668, rs2339654, rs1380862, rs9297746, rs12543663, rs10086908, rs16901922, rs1016343, rs17832285, rs16901979, rs4871779, rs10107982, rs16902094, rs620861, rs17467139, rs6983267, rs9297756, rs10094059, rs7818556, rs1992833, rs986472, rs12552397, rs4273907, rs4237185, rs753032, rs11253002, rs2386841, rs10795841, rs10508422, rs7075945, rs10508678, rs539357, rs10826398, rs3818714, rs7090755, rs10993994, rs4382847, rs1891158, rs10887926, rs10788160, rs6579002, rs10832514, rs7358335, rs1944047, rs3019779, rs10896437, rs12793759, rs7106762, rs7102758, rs2449600, rs585197, rs2509867, rs11568818, rs7125415, rs11601037, rs11222496, rs4570588, rs6489721, rs3213764, rs17395631, rs4423250, rs11168936, rs10875943, rs3759129, rs902774, rs1827611, rs4760442, rs11610799, rs6539333, rs11067228, rs7485441, rs6489794, rs4119478, rs17070292, rs2293710, rs17256058, rs1950198, rs2331780, rs7141529, rs12880777, rs17123359, rs785437, rs524908, rs12903579, rs7178085, rs7164364, rs896615, rs11634741, rs9972541, rs12594014, rs11631109, rs1558902, rs8044335, rs2738571, rs885479, rs385894, rs684232, rs4925094, rs17138478, rs11649743, rs2107131, rs7213769, rs12946864, rs306801, rs138213197, rs1863610, rs17224342, rs9911515, rs12947919, rs966304, rs17744022, rs7234917, rs1943821, rs2227270, rs1363120, rs888663, rs1227732, rs1054564, rs4806120, rs11672691, rs758643, rs3745233, rs6509345, rs2659051, rs2735839, rs1354774, rs2691274, rs6090461, rs2297434, rs6062509, rs2315654, rs2823118, rs2838053, rs398146, rs16988279, rs2269640, rs4822763, rs132774, rs747745, rs5978944, rs6530238, rs5934705, rs5935063, rs4830488, rs17318620, rs5945619, rs5945637, rs11091768, rs2473057, rs5918762, rs4844228, rs6625760 y rs17324573.

11. El procedimiento de una cualquiera de las reivindicaciones precedentes, que además comprende la obtención de los antecedentes familiares relativos al CaP, antecedentes de tratamiento y datos de la exploración física de dicho individuo; y donde dichos antecedentes familiares, antecedentes de tratamiento y/o datos de la exploración se incluyen en los datos combinados que forman dicho valor compuesto global.

12. El procedimiento de una cualquiera de las reivindicaciones 2-11, que además comprende analizar una categoría adicional de biomarcadores de CaP, midiendo la presencia o la concentración de uno o cada uno de una diversidad de biomarcadores de CaP de dicha categoría adicional de biomarcadores; combinar los datos relativos a dicha categoría adicional de biomarcadores de CaP para formar un valor compuesto adicional de biomarcadores para dicha categoría adicional de biomarcadores de CaP e incluir dicho valor compuesto adicional de biomarcadores en el valor compuesto global; donde la combinación de datos para formar el valor compuesto adicional de biomarcadores está diseñada de forma redundante cuando la categoría adicional de biomarcadores de CaP

comprende más de un biomarcador de CaP, preferiblemente, donde la categoría adicional de biomarcadores de CaP comprende el biomarcador MIC-1 y, opcionalmente, otros biomarcadores relacionados con MIC-1, o el biomarcador MSMB y, opcionalmente, otros biomarcadores relacionados con MSMB.

- 5 13. El procedimiento de una cualquiera de las reivindicaciones precedentes, donde dicho individuo tiene un valor de IMC superior a 25.
14. El procedimiento de una cualquiera de las reivindicaciones precedentes, donde la medición de la presencia o la concentración de dichos biomarcadores de CaP se realiza mediante el uso de tecnología de micromatrices.
- 10 15. Un dispositivo de ensayo, que comprende una fase sólida sobre la que hay inmovilizada al menos dos categorías diferentes de ligandos, donde:
 - la primera categoría de dichos ligandos se une específicamente a al menos tres biomarcadores de CaP de tipo calicreína e incluye una diversidad de ligandos diferentes que se unen específicamente a cada uno de una diversidad de diferentes biomarcadores de CaP de tipo calicreína, que incluyen al menos uno de PSA, iPSA, tPSA, fPSA y hK2 y, opcionalmente, MSMB y/o MIC-1; y
 - la segunda categoría de dichos ligandos se une específicamente a una diversidad de SNPcp e incluye una diversidad de ligandos diferentes que se unen específicamente a cada uno de dichos SNP, y donde la diversidad de SNPcp comprende al menos 70 SNPcp seleccionados de entre las listas de SNP de la tabla I o la tabla II.
16. Un kit de análisis, que comprende un dispositivo de ensayo de acuerdo con la reivindicación 15 y al menos dos categorías de moléculas de detección, donde:
 - la primera categoría de dichas moléculas de detección es capaz de detectar al menos tres biomarcadores de CaP de tipo calicreína, incluyendo al menos uno de PSA, iPSA, tPSA, fPSA y hK2 y, opcionalmente, MSMB y/o MIC-1; y
 - la segunda categoría de dichas moléculas de detección es capaz de detectar un SNPcp y donde al menos 70 SNP se seleccionan de entre las listas de SNP de la tabla I o la tabla II.
17. El kit de análisis del dispositivo de ensayo o de acuerdo con la reivindicación 15 o 16, donde se detectan al menos 80 SNPcp, 90 SNPcp, 100 SNPcp o 150 SNPcp, o al menos el 70 % de los SNPcp de la tabla I o la tabla II.
18. Un programa informático que se puede cargar directamente en la memoria interna de un ordenador digital, **caracterizado porque** dicho programa informático comprende medios de código de software para realizar al menos las etapas 3, 4 y 5 de la reivindicación 1, como por ejemplo las etapas 1-6 de la reivindicación 1.

FIG. 1

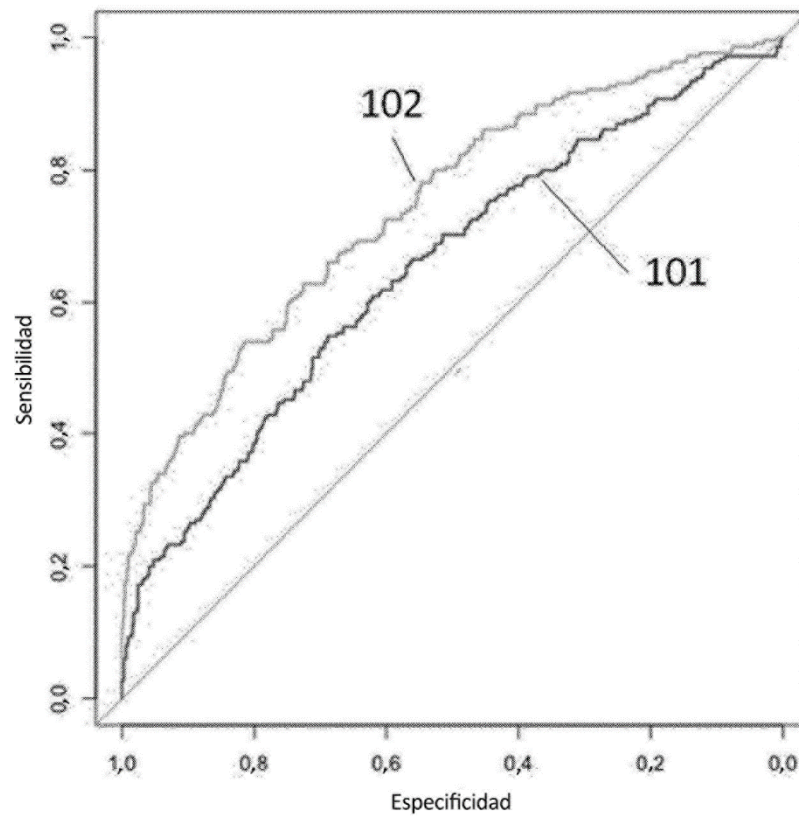


FIG. 2

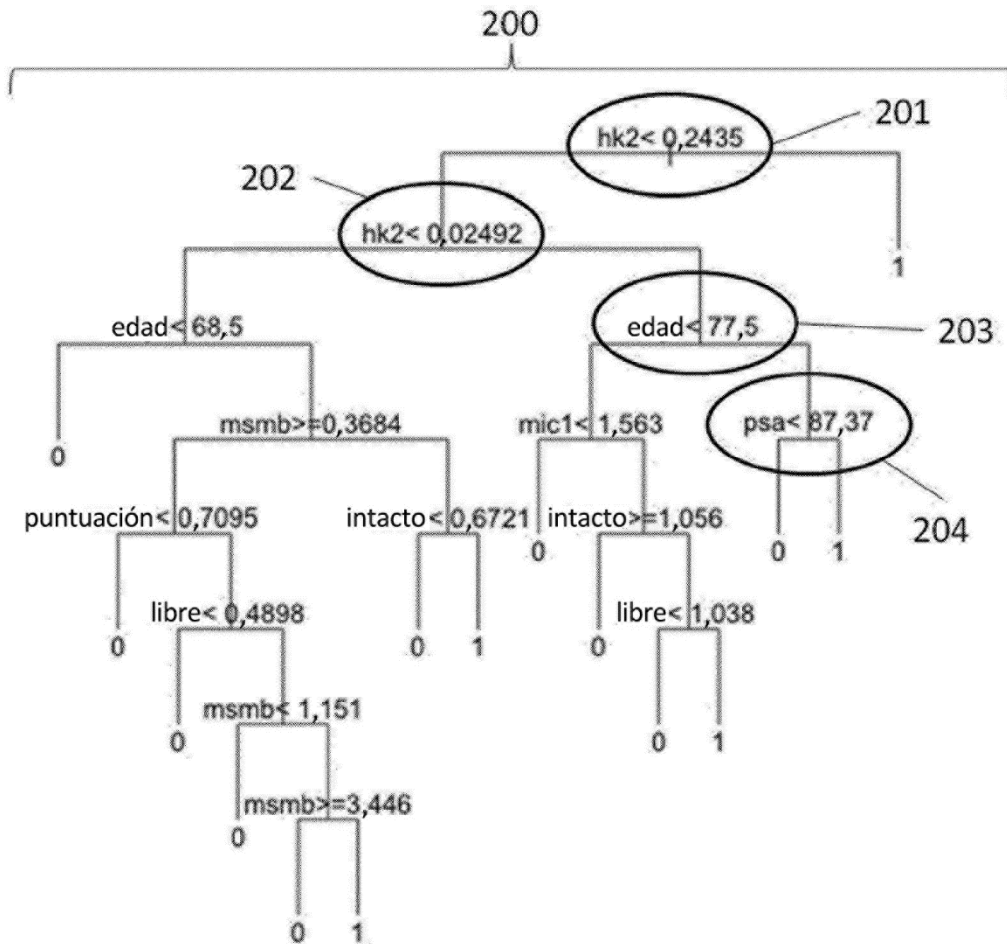


FIG. 3

