

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 732 371**

51 Int. Cl.:

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **01.12.2014 PCT/EP2014/076145**

87 Fecha y número de publicación internacional: **11.06.2015 WO15082418**

96 Fecha de presentación y número de la solicitud europea: **01.12.2014 E 14835550 (6)**

97 Fecha y número de publicación de la concesión europea: **27.03.2019 EP 3077537**

54 Título: **Nuevos biomarcadores de ARN para el diagnóstico del cáncer de próstata**

30 Prioridad:

02.12.2013 EP 13195377

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

22.11.2019

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27
80686 München, DE**

72 Inventor/es:

**HORN, FRIEDEMANN;
HACKERMÜLLER, JÖRG;
CHRIST, SABINA;
REICHE, KRISTIN;
WIRTH, MANFRED;
FRÖHNER, MICHAEL y
FÜSSEL, SUSANNE**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 732 371 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

Nuevos biomarcadores de ARN para el diagnóstico del cáncer de próstata

Campo de la invención

5 La presente invención se inscribe en el campo de la biología y la química. En particular, la invención se inscribe en el campo de la biología molecular. Más particularmente, la invención refiere al análisis de transcritos de ARN. Más precisamente, la invención se inscribe en el campo del diagnóstico del cáncer de próstata.

Antecedentes

10 El cáncer de próstata es el tipo de cáncer más frecuentemente diagnosticado en hombres. En 2012, se informaron aproximadamente 240 000 nuevos casos de diagnóstico de cáncer de próstata en Estados Unidos y aproximadamente 360 000 en la Unión Europea, 68 000 de ellos en Alemania. En Estados Unidos, el riesgo de ser diagnosticado cáncer de próstata en algún momento de la vida y el riesgo de morir de cáncer de próstata son de 15,9 % y 2,8 %, respectivamente. A pesar de la adopción generalizada de los estudios de detección del cáncer de próstata y de grandes avances en el tratamiento de la enfermedad metastásica, el cáncer de próstata todavía es el segundo tipo de cáncer más mortal en hombres, con más de 250 000 muertes al año en el mundo occidental.

15 En la actualidad, los análisis de niveles séricos de antígeno prostático específico (PSA) y los exámenes de tacto rectal son los dos principales métodos de detección. A los pacientes que tienen resultados anómalos en estos estudios generalmente se les recomienda realizarse una biopsia de próstata. Sin embargo, esto tiene consecuencias considerables. La falta de especificidad de la detección mediante análisis de PSA produce grandes cantidades de resultados falsos positivos, lo que da lugar a la realización de biopsias de próstata innecesarias a millones de hombres cada año (sobrediagnóstico). Además, la realización de biopsias conlleva un riesgo sustancial de complicaciones infecciosas. Así, existe una necesidad urgente de contar con análisis más sensibles y específicos para el diagnóstico temprano del cáncer de próstata con el fin de mejorar su detección y evitar la gran cantidad de biopsias de próstata innecesarias. La presente invención aborda este problema mediante un conjunto de biomarcadores para la detección y el diagnóstico del cáncer de próstata.

Resumen de la invención

25 Se identificaron transcritos diferencialmente expresos en muestras de tumor y de control mediante secuenciación de última generación de 64 muestras de pacientes con cáncer de próstata y controles, que luego se validaron mediante análisis con microarreglos y qRT-PCR de 256 y 56 muestras, respectivamente. La invención describe biomarcadores de ARN que no se habían identificado previamente como adecuados para el diagnóstico del cáncer de próstata.

La invención refiere a un método para el diagnóstico del cáncer de próstata que comprende los pasos de analizar el nivel de expresión de un ácido nucleico seleccionado de entre el grupo de las SEQ ID NO 1 a 42, donde, si al menos uno de dichos ácidos nucleicos está presente y/o el nivel de expresión de al menos uno de dichos ácidos nucleicos es superior a un valor umbral, la muestra se identifica como un resultado de cáncer de próstata positivo.

35 En una realización preferente, la invención refiere a un método para el diagnóstico del cáncer de próstata que comprende los pasos de analizar en una muestra de un paciente el nivel de expresión de un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11, donde, si el nivel de expresión de dicho ácido nucleico es superior a un valor umbral, la muestra se identifica como un resultado de cáncer de próstata positivo.

40 En una realización, la invención refiere a un cebador (*primer*) o una sonda (*probe*) que hibrida en condiciones estrictas a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11, o cualquier parte de estos.

45 La invención también refiere al uso de un cebador o una sonda que hibrida en condiciones estrictas a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11 para el diagnóstico del cáncer de próstata.

La invención refiere a una sonda o un cebador, donde la sonda o el cebador son específicos para una secuencia de entre el grupo de las SEQ ID NO 1 a 42, preferentemente para un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11.

50 La invención refiere a un ácido nucleico con una secuencia de entre el grupo de las SEQ ID NO 1 a 42 —o bien su complemento inverso— o a un ácido nucleico que tiene, preferentemente, al menos 85 %, 90 %, 95 % o 99 % de identidad a nivel de secuencia con un ácido nucleico de acuerdo con uno cualquiera de los ácidos nucleicos de las SEQ ID NO 1 a 42.

En una realización preferente, la invención refiere a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11 —o bien su complemento

inverso— o a un ácido nucleico que tiene, preferentemente, al menos 85 %, 90 %, 95 % o 99 % de identidad a nivel de secuencia con el ácido nucleico seleccionado.

La invención refiere al uso de un ácido nucleico con una secuencia de entre el grupo de las SEQ ID NO 1 a 42 para el diagnóstico del cáncer de próstata.

- 5 En una realización preferente, la invención refiere al uso de un transcripto del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11 para el diagnóstico del cáncer.

10 La invención también refiere a un kit para el diagnóstico del cáncer de próstata que comprende un cebador o una sonda que hibrida en condiciones estrictas a un transcripto del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11, y reactivos para la amplificación, cuantificación y/o detección de ácidos nucleicos.

Definiciones

Se proveen las siguientes definiciones de ciertos términos que se utilizan en la solicitud.

15 En la presente, "ácido(s) nucleico(s)" o "molécula de ácidos nucleicos" hacen referencia, en general, a cualquier ácido ribonucleico o ácido desoxirribonucleico, que puede ser modificado o no modificado. Los "ácidos nucleicos" incluyen, sin limitación, los ácidos nucleicos de hebra simple y de doble hebra. En la presente, el término "ácido(s) nucleico(s)" también incluye los ácidos nucleicos de acuerdo con la definición anterior que contienen una o más bases modificadas. Así, un ácido nucleico cuya cadena principal tiene una o más modificaciones para mejorar la estabilidad o para otros fines se considera de todas formas un "ácido nucleico". En la presente, el término "ácidos nucleicos" abarca tales formas modificadas química, enzimática o metabólicamente de los ácidos nucleicos, así como las formas químicas de los ácidos nucleicos características de los virus y las células, incluidas, por ejemplo, las células simples y complejas.

20 En el contexto de la presente, los términos "nivel" o "nivel de expresión" hacen referencia al nivel en que un biomarcador está presente en una muestra de un paciente. Para medir el nivel de expresión de un biomarcador, generalmente se compara su nivel de expresión con el nivel de uno o más genes constitutivos (*housekeeping*) en una muestra como forma de normalización. La muestra del paciente se identifica como un resultado de cáncer de próstata positivo si el nivel de expresión del biomarcador supera el nivel de expresión del mismo biomarcador en un control adecuado (por ejemplo, una muestra de tejido saludable) en un valor superior a un umbral.

30 En la presente, los términos "analizar una muestra para detectar la presencia y/o el nivel de ácidos nucleicos" o "estimar específicamente los niveles de ácidos nucleicos" refieren a los medios y métodos empleados para evaluar y cuantificar los niveles de ácidos nucleicos. Un método útil para tal fin es, por ejemplo, la PCR cuantitativa con transcripción inversa. De manera similar, el nivel de ARN también puede analizarse mediante Northern blot, secuenciación de última generación o, después de la amplificación, mediante técnicas espectrométricas que incluyan medir la absorbancia a 260 y 280 nm.

35 El término "transcripto" hace referencia a un ácido nucleico producido como una copia de un ácido nucleico molde (*template*). Esta definición incluye el ácido nucleico producido por la transcripción por parte de una polimerasa de un ácido nucleico molde (por ejemplo, un locus), para producir un nuevo ácido nucleico con secuencia complementaria a la del molde. El término "transcripto" también incluye todo ácido nucleico sujeto a procesamiento ulterior luego de la transcripción. Tal procesamiento ulterior incluye, sin carácter taxativo, el empalme (*splicing*), la poliadenilación, la edición, la digestión parcial, la ligación y el marcado. El término también comprende los ácidos nucleicos derivados del transcripto mediante, por ejemplo, transcripción, transcripción inversa, amplificación o elongación. El transcripto puede corresponder a cualquier porción del ácido nucleico molde, preferentemente a un mínimo de 20 nucleótidos. El transcripto también tiene una identidad de secuencia con el ácido nucleico molde o parte de este —o bien su complemento inverso— de al menos 90 %, preferentemente al menos 95 % y más preferentemente al menos 99%.

45 En la presente, el término "amplificada", en relación con una secuencia de ácidos nucleicos, hace referencia a un proceso por el cual se generan una o más copias de una secuencia particular de ácidos nucleicos a partir de una secuencia de ácidos nucleicos molde, preferentemente mediante el método de la reacción en cadena de la polimerasa. Otros métodos de amplificación incluyen, sin carácter taxativo, la reacción en cadena de la ligasa (LCR), la amplificación de polinucleótidos basada en la secuencia (NSBA) o cualquier otro método conocido en la técnica.

50 En la presente, el término "correlacionar", en relación con el uso de uno o más marcadores de diagnóstico o pronóstico, hace referencia a comparar la presencia o la cantidad del/de los marcador(es) en una muestra de un paciente con su presencia o nivel de expresión en una muestra de una persona de la cual se sabe que padece determinada afección o tiene riesgo de padecerla. El nivel de un marcador en una muestra de un paciente se puede comparar con un nivel asociado con un diagnóstico determinado.

55 En la presente, el término "diagnóstico" hace referencia a la identificación de la enfermedad —en este caso el cáncer de próstata— en cualquier etapa de su desarrollo, y también incluye la determinación de la predisposición de un

sujeto a desarrollar esta enfermedad.

El término "locus (hg19): Chr2: 1,550,437-1,629,191 (-)" hace referencia a un intervalo de secuencia en el cromosoma 2 del ensamblaje hg19 del genoma humano.

5 En la presente, el término "colorante fluorescente" hace referencia a cualquier producto químico que absorba energía lumínica a una longitud de onda específica y reemita luz a otra longitud de onda. Los colorantes fluorescentes adecuados para marcar ácidos nucleicos incluyen, por ejemplo: FAM (5-carboxifluoresceína o 6-carboxifluoresceína), VIC, NED, fluoresceína, FITC, IRD-700/800, CY3, CY5, CY3.5, CY5.5, HEX, TET, TAMRA, JOB, ROX, BODIPY TMR, Oregon Green, Rhodamine Green, Rhodamine Red, Texas Red, Yakima Yellow, Alexa Fluor, PET y otros similares.

10 En la presente, el término "aislado", en referencia a un ácido nucleico, significa que una secuencia que ocurre naturalmente se ha extraído de su entorno celular (p. ej., cromosómico) normal o se sintetiza en un ambiente no natural (p. ej., se sintetiza artificialmente). Así, una secuencia "aislada" puede encontrarse en una solución sin células o colocarse en otro entrono celular.

15 En el contexto de la presente, un "kit" es una combinación de elementos en un paquete que, opcionalmente, incluye instrucciones para el uso de la combinación y/u otras reacciones o componentes para tal uso. Si el kit contiene ácidos nucleicos, el kit también puede comprender variantes sintéticas o no naturales de dichos ácidos nucleicos. Se entiende por ácido nucleico sintético o no natural aquel que comprende cualquier modificación química, bioquímica o biológica tal que el ácido nucleico no se encuentre en la naturaleza en esa forma. Tales modificaciones incluyen, sin carácter taxativo, el marcado con un colorante fluorescente, una fracción inhibidora de fluorescencia (*quencher*) o una marca de biotina, o bien una modificación en la cadena principal de un ácido nucleico o cualquier otra modificación que distinga al ácido nucleico de su contraparte natural. Lo mismo vale para otros compuestos naturales, como las proteínas, los lípidos, etc.

20

25 En la presente, el término "paciente" hace referencia a un organismo vivo humano o no humano que recibe atención médica o que debería recibir atención médica debido a una enfermedad, o del que se sospecha que tiene una enfermedad. Esto incluye a las personas que no tienen una enfermedad definida y que están siendo investigadas por presentar signos patológicos. Así, los métodos y análisis que se describen en la presente se aplican tanto a enfermedades humanas como a enfermedades veterinarias.

30 En la presente, el término "cebador" hace referencia a un ácido nucleico, sea que ocurra naturalmente (por ejemplo, en el producto purificado de una digestión enzimática) o se produzca sintéticamente, que sea capaz de actuar como punto de inicio de síntesis al colocarse en condiciones en las que se induce la síntesis de un producto de extensión del cebador, que es complementario a una hebra de ácido nucleico (p. ej., en presencia de nucleótidos y de un agente inductor, como una ADN polimerasa, y a una temperatura y un pH adecuados). El cebador puede ser de hebra simple o de doble hebra, y debe ser suficientemente extenso para iniciar la síntesis del producto de extensión deseado en presencia del agente inductor. La longitud exacta del cebador dependerá de muchos factores, incluida la temperatura, el origen del cebador y los métodos empleados. Preferentemente, los cebadores tienen una longitud de aproximadamente 15 a 100 bases, más preferentemente aproximadamente 20 a 50 bases, y con máxima preferencia aproximadamente 20 a 40 bases. Los factores involucrados en la determinación de la longitud adecuada el cebador son conocidos para la persona versada en el arte. Opcionalmente, el cebador puede ser un elemento sintético, en el sentido de comprender una modificación química, bioquímica o biológica. Tales modificaciones incluyen, sin carácter taxativo, el marcado con un colorante fluorescente o una fracción inhibidora de fluorescencia (*quencher*), o bien una modificación en la cadena principal de un ácido nucleico o cualquier otra modificación que distinga al cebador de su ácido nucleico contraparte natural.

35

40

45 El término "sonda" hace referencia a cualquier elemento que pueda usarse para detectar específicamente una entidad biológica, como un ácido nucleico, una proteína o un lípido. Además de la parte que le permite unirse específicamente a la entidad biológica en cuestión, la sonda también comprende al menos una modificación que permita su detección en un análisis. Tales modificaciones incluyen, sin carácter taxativo, las marcas como los colorantes fluorescentes, un elemento radioactivo introducido específicamente o una marca de biotina. La sonda también puede comprender una modificación en su estructura, como un ácido nucleico bloqueado.

50 En el contexto de la presente, el término "muestra" hace referencia a una muestra de un fluido o tejido corporal obtenida para fines de diagnóstico, pronóstico o evaluación de un sujeto de interés, como un paciente. Las muestras de análisis preferentes incluyen las muestras de sangre, suero, plasma, líquido cefalorraquídeo, orina, saliva, esputo y derrames pleurales. Además, una persona versada en la técnica notará que ciertas muestras se podrían analizar más adecuadamente con ayuda de un procedimiento previo de fraccionamiento o purificación (por ejemplo, la separación de la sangre en suero y plasma).

55 Así, en una realización preferente de la invención, la muestra se selecciona de entre el grupo que comprende una muestra de sangre, una muestra de suero, una muestra de plasma, una muestra de líquido cefalorraquídeo, una muestra de saliva y una muestra de orina o cualquier extracto de cualquiera de las muestras anteriores, así como células tumorales circulantes en la sangre o la linfa, cualquier tejido del que se sospecha que contiene metástasis y

5 cualquier fuente que pueda contener células tumorales prostáticas o partes de ellas, incluidas las vesículas como los exosomas, las microvesículas y otras, así como las moléculas de ARN libres o unidas a proteínas derivadas de células tumorales prostáticas. Preferentemente, la muestra es una muestra de sangre, más preferentemente una muestra de suero o una muestra de plasma. Cabe señalar que la orina (particularmente después de un examen de tacto rectal) y el eyaculado se encuentran entre las muestras de máxima preferencia. Las muestras de tejido también pueden ser biopsias o muestras de tejido obtenidas durante una cirugía.

10 En la presente, el término "área bajo la curva (AUC)" describe el área bajo una curva de característica operativa del receptor (ROC) o curva ROC. El AUC hace referencia a la especificidad y sensibilidad de un biomarcador. Un marcador perfecto (con un AUC de 1,0) correspondería a un punto en la esquina superior izquierda o las coordenadas (0,1) del espacio ROC, correspondiente a una sensibilidad de 100 % (ausencia de falsos negativos) y una especificidad de 100 % (ausencia de falsos positivos).

15 El término "valor p " hace referencia a la probabilidad de obtener los resultados observados en la muestra (o un valor más extremo) si la hipótesis nula es, en efecto, verdadera (p. ej., cuando no hay diferencias entre las medias de dos grupos). Cuanto menor es el valor p , mayor es la probabilidad de que la hipótesis alternativa explique los resultados observados mejor que la hipótesis nula.

El término "valor p ajustado" hace referencia a los valores p ajustados para compensar la realización de múltiples comparaciones (cantidad de genes o sondas analizados). El método aplicado se describe en detalle en la sección experimental.

Descripción detallada de la invención

20 La invención describe un método para el diagnóstico del cáncer de próstata. Este método comprende analizar una muestra tomada de un paciente y determinar, de forma específica, el nivel de un biomarcador o una combinación de biomarcadores en dicha muestra del paciente. Luego, el resultado se correlaciona con un valor umbral y, si se encuentra por encima de dicho valor umbral, dicha muestra del paciente se identifica como un resultado de cáncer de próstata positivo.

25 La invención refiere a un grupo de secuencias que comprenden las SEQ ID NO 1 a 42. Las secuencias se enumeran a continuación. Por cuestiones de espacio, solo se enumeran los primeros 100 nucleótidos. El resto de las secuencias se pueden consultar en el protocolo de secuencias.

30 Hay dos tipos de secuencias. En primer lugar, algunos transcritos son secuencias conocidas que ya están anotadas en las bases de datos relevantes. Estas se identifican con sus correspondientes anotaciones. En segundo lugar, se identificaron transcritos nuevos que aún no están anotados. Estos se identifican con la leyenda "XLOC_" seguida por un número. Estas identificaciones brindan información sobre el origen genómico de los

35 transcritos, pero podrían no representar necesariamente la secuencia completa de un transcripto dado. En algunos casos, las secuencias efectivamente detectadas podrían ser más largas o más cortas. En el caso de los transcritos XLOC, si se detectan fragmentos, estos pueden ser de apenas 1000, 500, 400, 300, 200, 150, 100, 50, 40, 30, 20, 10, 9, 8, 7, 6 o 5 nucleótidos.

Tabla 1: Lista de SEQ ID NOs. Se muestran las SEQ ID NO 1 a 42, junto con las anotaciones de transcritos y genes correspondientes. Se muestran los primeros 100 nucleótidos de cada SEQ ID NO.

ID DE SEC.	Transcripto	Anotación del gen/transcripto	Secuencia
1	1	Retro-RPL7 ID del gen en Ensemble: ENSG00000242899.1 Locus (hg19): Chr3: 131,962,301-131,963,125(-) ID del transcripto en Ensemble: ENST00000479738.1	Ttttcggctggaacca tggagggtgtgaagag aagaagaaggttctgc tgtccagaaaccctta agaaaaagtgaaggaat ttcacagagctgaag
2	2	XLOC_133897 ID del gen en Ensemble: ninguno Locus (hg19): chr20: 45,377,600- 45,380,719(-) ID del/de los transcripto(s) en Ensemble: ninguno Incluye las siguientes entradas de GenBank: AK128800.1, BC065739.1	gcccgcttctgtgactc cacccttacggaagt ctatgggactctctgaa atgtatgagtgactg ttagaaagcggaagaa aatgaaaagaaaacg

(continuación)

ID DE SEC.	Transcripto	Anotación del gen/transcripto	Secuencia
3	3	AC144450.2 ID del gen en Ensemble: ENSG00000203635.2 Locus (hg19): Chr2: 1,624,282-1,629,191(-) ID del transcripto(s) en Ensemble: ENST00000366424	attgccacagccggat ccacgggtgactaatc cgggaaggcgtccagcg tgagccgtgaggcctgc acctgcgccgactca ccaactcaccaggagt
4	4	RP11-279F6.1 ID del gen en Ensemble: ENSG00000245750.3 Locus (hg19): Chr15: 69,755,365-69,863,775(+) ID del transcripto en Ensemble: ENST00000558633	caggaatgggctggggc gcgttgtagtgggaa tcctgagcccgggctgt tgctggaggactcggg agcagcagtgatttcg gcgtaccaggagag
5	4	ENST00000558309	ttcggcgtaccaggag agctatgtataggaatg ccgctatggaagacat ccaggacacctgttaa gtgaaaaagacatgcc accattagggtca
6	4	ENST00000560882	gaggcccgacattgtgc tgggaaggagctccag aaagggccatccttct gtttgggtcagatct gaacactttgctaag gtctctggaagctc
7	4	ENST00000559029	Gactggagaggccagca cgcacagtgacttaac caagaagatggaataaA aaggcctacctcattgg gctcgtgtgggtgagga gaactgaagagtctg
8	4	ENST00000558781	Ctgggcttcagctcc aagccttctacctgtg aatgcttggccaatgT ctgggaccacctctt actcacaactcctccag atctgcagagtggcc
9	4	ENST00000498938	Ggagctggtccaggaa agaagggcacatgagca aacatgatggcccttt atgagagtaattfact gaaatgcacagcgatta cctgctcaccagcc
10	4	ENST00000559477	aggaactggaataact tgcagtgcttcagta ttgtgaaaccagcaacT tgtcacaattctctg aatttctgggaaattt gaagtggagtacctg
11	5	AC144450.1 ID del gen en Ensemble: ENSG00000228613.1 Locus (hg19): Chr2: 1,550,437-1,623,885(-) ID del transcripto(s) en Ensemble: ENST00000438247.1	cagtttcacagggcctg tgtgccgagagtgtcc ttaccatttttcaata ttattctgctaaggagg atttttagacattatgt tctagtaaacgct
12	6	AC012531.25 ID del gen en Ensemble: ENSG00000260597.1 Locus (hg19): Chr12: 54,413,694-54,416,373(+) ID del/de los transcripto(s) en Ensemble: ENST00000562848.1	caagacagaggcaagca gagaaggcatagcagca gcgaccggcgtctgtt ttcattttcactctgg ccaggggataaactgga ccccagtggactcca
13	7	XLOC_068574 _ ID del gen en Ensemble: ninguno Locus (hg19): chr14: 62,653,302-62,655,723(+) ID del/de los transcripto(s) en Ensemble: ninguno	ggtaacatgaaataat ggatgagcagttcaact atattaaaaataaacgt ggtaagagtgtcacc ttaagtgtaggattga aagtgtaggctctaa
14	8	RP1-207H1.3 ID del gen en Ensemble: ENSG00000231150.1 Locus (hg19): chr6: 38,890,805-38,920,875(-) ID del transcripto en Ensemble: ENST00000416948.1	Tgaagccatgagccac tagaagccacatgttct gccatgtggagaagaat gagagagtacatctca aattgaggtgtggcatg atgatttggctgcc
15	8	ENST00000453417.1	ctttcaagggcctgtgc ctgtgtaactgtctat gagccaggtatattctga agcatatttgacaacag aaaaagttaatgtaatt tcaaaggaaaaacg
16	8	ENST00000418399.1	atatctgaagcatattt gacaacagaaaaagta atgtaattttcaaagga aaaacgccaacttttt caaaaaggaaacagcaa ctggagagcagattt

(continuación)

ID DE SEC.	Transcripto	Anotación del gen/transcripto	Secuencia
17	9	XLOC_016724 ID del gen en Ensemble: ninguno Locus (hg19): chr1: 177,827,793-177,841,757(-)	atcccctctgagaattt atcagaaaaacaagcaa taagtgagaccaacgtt gtgaggattactcgg aacgctcatctatcctt gtggagaaaaacccg
18	10	RP11-314013.1 ID del gen en Ensemble: ENSG00000260896 Locus (hg19): Chr16: 80,862,632-80,926,492(-) ID del transcripto en Ensemble: ENST00000562231	ttctttgtttgctgc cttccgtagaagatgtg gcttgctcatgctgac ttctgcatggtgtga ggcctcccagccatgt ggaactgtttcagg
19	10	ENST00000569356	Aggggtttccgctttg cttctcctcattttct ctgtctgctgcaattT cgctcccgcctgatt ctgaggcctcccagct attggaactgtaag
20	10	ENST00000561519	Aaaagactatctctcc cattgaattaaattgga actttggaatcttaata gaaaccaactgacttg gcttggtttcagggtc tgggtccatggctct
21	10	ENST00000563626	Cttgctcatgcttact tctgcatggtgtgag gcctcccagccatgtG gaactgtttcagggtc tgggtccatggctctc ctgagccgaaaataa
22	11	XLOC_167596 ID del gen en Ensemble: ninguno Locus (hg19): chr4: 67,964,836- 67,975,652(-)	Ctcttctctcctctc cctcctcctccctcc ctcccctctctcctc tttcttcttcttctc tcttcttcttctt ctttcttcttctt
23	12	XLOC_167595 ID del gen en Ensemble: Locus (hg19): chr4: 67,946,236-67,964,614(-)	aaacatacgtgtgatg tgtctttatggcagcat gatttataatcctttgg ggataactcagtaatg ggatggctgggtcaaat ggtatttctagtct
24	13	XLOC_156132 ID del gen en Ensemble: ninguno Locus (hg19): chr3: 193,632,725- 193,636,178(-)	agtatgtgcattgtac cttgctttgtttctc aactttgtcttgttC tgtaattccctcattca ttcctacctctgcatgc ttgaaagtctttgt
25	14	XLOC_156120 ID del gen en Ensemble: ninguno Locus (hg19): chr3: 193,580,748-193,608,459(-)	accaaggacatgcgaa aactttgggtgtgatg gatatagtcataatctt tattgtgggtactgttt cacacatgtgtacatat atcacaactcatcaa
26	15	RP11-627G23.1 ID del gen en Ensemble: ENSG00000255545.3 Locus (hg19): Chr11: 134,306,367-134,375,555 (+) ID del transcripto en Ensemble: ENST00000533390	cttctcgggggttctc tccaggcctgactttta ctccccttctaagtgt gcagatgggatgtgctt ctccacaggaggccca cggttccccacccc
27	15	ENST00000531319	ctgtctcaagcctcaa tcaacagatcagacagc ttgtactcacaggcaa ggacacgtggaagagg ctcaattttctagatgg gtggcaacagccatg
28	15	ENST00000528482	gaggcagccatgactgg ccacttcatgtgctct ggagaagggctgcacc agccgtttcaggaag tcaagcagctgtgact cctgagctgggtga
29	15	ENST00000532886	caaatgcctggcagcgt cctcgtgctcacctg ccatagccgacagtggc tgacctccatgcctgt tgcttttcttctgtt ggatcagggatacac
30	16	XLOC_047797 ID del gen en Ensemble: ninguno Locus (hg19): chr12: 75,378,181-75,383,176(+)	aagatgggacaatttt tttctctgtgttctt tataattattgacccc tttctggaataatctt ttcatctgttcatctg tcaatgctgtctgt
31	17	ANKRD34B ID del gen en Ensemble: ENSG00000189127.3 Locus (hg19): Chr5: 79,852,574-79,866,307(-) ID del transcripto en Ensemble: ENST00000338682	agctgctggccccctg ggtccagaggagccttg ccgccctcacctgcga gagcctggagccgacgc gtcacccccagcggaaag cgctcgtgccccg
32	17	ENST00000508916	agctcagctcagacggc gccctaggccgcacag agggctgggcagtgccg gagagaggttgaaagc gcccgcgccaactcgac agcgcgtcccaggaa

(continuación)

ID DE SEC.	Transcripto	Anotación del gen/transcripto	Secuencia
33	18	XLOC_243739 ID del gen en Ensemble: ninguno Locus (hg19): chr9: 79,530,077-79,542,427(-)	aaacaggaaaagaatt gggatTTTTatgaaaa tgtaaaggctagctct gttaggattccccatga cattgcagtggtgacat ggctggtgatgtgcc
34	19	XLOC_198292 ID del gen en Ensemble: ninguno Locus (hg19): chr6: 148,396,831-148,428,362(+)	tccctccctccttctcct tccctcctccttctt cccttcagtttctctc ctctaatgccccctgt ccttaaaaatgtctcca ttcaggcactatgca
35	20	XLOC_068639 ID del gen en Ensemble: ninguno Locus (hg19): chr14: 62,931,844-62,933,233(+)	ccaagatttctcatcca tggttcaactaagaat atTTattctctccagt gaaatTTTTacaatta ggattgcaaaactacat acattcaggtagatc
36	21	XLOC_172083 ID del gen en Ensemble: ninguno Locus (hg19): chr4: 169,961,616-169,999,957(-)	cactgcagtctcctcct ccctggttcaagcaatt ctctgcctcagctccc tgagtagctgggaccac aggcgctcaccaccacg catggctcattttg
37	22	XLOC_172082 ID del gen en Ensemble: ninguno Locus (hg19): chr4: 169,947,628-169,961,481(-)	agtgatccgcccgcctc cgcctcccaagtgctg ggattacaggtgtgagc cactgcgcctggccgct gctcttatactattttg aatgtaggccggccg
38	23	XLOC_112832 ID del gen en Ensemble: ninguno Locus (hg19): chr2: 123,297,707-123,644,538(+)	agcagatggcatttgag caaacactgcaaaaagg tgaggaagatagccatc atagctgatggaacaag caaaacaaaagtataa ggaagaattgtactc
39	24	XLOC_243747 ID del gen en Ensemble: ninguno Locus (hg19): chr9: 79,622,778-79,633,361(-)	cccgcagctgcgcccc cccgggccaccaagcac ggtggagggggaacagg acactgcctcttgctt ctcttctctggcatc tcctctccgcccc
40	25	XLOC_243744 ID del gen en Ensemble: ninguno Locus (hg19): chr9: 79,601,892-79,606,132(-)	atgtgccaccacacctg gctgatttttgtatt ttagtagagatgggata tcaccatattaaccaag atggctcgtattacctg acctcgtgatccgcc
41	26	XLOC_126289 ID del gen en Ensemble: ninguno Locus (hg19): chr2: 180,988,687-180,989,287(-)	cctgtgatctaattta gtggggggcagacctgt ttacaagccaaaataa caggctgcaataactga ggattttatatacc tgaccaaaagaagttt
42	27	XLOC_172084 ID del gen en Ensemble: ninguno Locus (hg19): chr4: 169,983,995-169,984,246(-)	attgtggaactgctct tctccctgcgattcaga ggggaaaagataaagcc acacagccctggggcct cttgcttaagaacacat ctcagtttaaccacc

El biomarcador PCA3 se usa rutinariamente para el diagnóstico del carcinoma de próstata (CaP). Así, como era de esperar, los niveles de expresión de PCA3 fueron indicativos del CaP en los sujetos analizados mediante secuenciación de última generación por los inventores (Fig. 2). Sin embargo, se observó que el biomarcador tenía su máximo nivel de expresión en tumores de muy bajo riesgo (V) y disminuía a medida que aumentaba el factor de riesgo de los tumores. Este hallazgo convierte al PCA3 en un marcador poco confiable en los tumores de riesgo medio y alto e indica la necesidad de contar con biomarcadores superiores para el cáncer de próstata.

Muchos de los nuevos biomarcadores descubiertos por los inventores son considerablemente superiores en términos de especificidad y sensibilidad que el PCA3. Por ejemplo, el biomarcador Retro-RPL7 (SEQ ID NO 1) arrojó un área bajo la curva ROC (AUC) de 0,935, en comparación con un valor de 0,851 en el caso del PCA3 (Fig. 3).

Dos nuevos biomarcadores corresponden a transcritos con EC 3 y 11 derivados del locus (hg19): Chr2: 1,550,437-1,629,191 (-), donde se encuentran cerca uno del otro. Estos dos transcritos también mostraron una marcada expresión diferencial entre los pacientes con tumores y los pacientes de control. Nuestros datos de secuenciación de última generación de ARN demuestran que ambos transcritos se expresan considerablemente en el tejido con cáncer de próstata, mientras que solo se expresan de manera tenue en el tejido de control o no tumoral, por lo que parecen estar corregulados, como se muestra en las Figuras 4 y 5, respectivamente. En el experimento de secuenciación, los valores de área bajo la curva ROC (AUC) son 0,991 en el caso de la SEQ ID NO 3 (Fig. 4) y 0,994 en el caso de la SEQ ID NO 11 (Fig. 5). Además, los datos de secuenciación demuestran claramente una región transcripta continuamente que se extiende desde el último exón de AC144450.2 (SEQ ID NO 3) hasta el primer exón de AC144450.1 (SEQ ID NO 11), incluida la región interadyacente entre ambos transcritos (Fig. 6). Así, los datos de secuenciación sugieren fuertemente que ambos transcritos derivan de un precursor común (transcripto primario). La expresión diferencial de las SEQ ID NO 3 y 11 pudo validarse mediante análisis con microarreglos hechos a medida de 256 muestras de tejido (Fig. 7 y 8, respectivamente).

Los transcritos correspondientes a las SEQ ID NO 3 y 11 se expresan a partir de las regiones genómicas contiguas (hg19): Chr2: 1,624,282-1,629,191 (-) y (hg19): Chr2: 1,550,437-1,623,885 (-), respectivamente. Juntas, estas

regiones genómicas forman el locus (hg19): Chr2: 1,550,437-1,629,191 (-). Otras denominaciones equivalentes de la SEQ ID NO 3 en otros sistemas de anotación son, por ejemplo, AC144450.2 o ENSG00000203635.2. En el caso de la SEQ ID NO 11, otras denominaciones equivalentes son, por ejemplo, AC144450.1 o ENSG00000228613.1.

5 Las secuencias de ácidos nucleicos de las SEQ ID NO 3 y 11 corresponden a transcritos empalmados del locus (hg19): Chr2: 1,550,437-1,629,191 (-).

La invención refiere a un método para el diagnóstico del cáncer de próstata que comprende los pasos de analizar el nivel de expresión del ácido nucleico de acuerdo con las SEQ ID NO 1 a 42, donde, si al menos uno de dichos ácidos nucleicos está presente y/o el nivel de expresión de al menos uno de dichos ácidos nucleicos es superior a un valor umbral, la muestra se identifica como un resultado de cáncer de próstata positivo.

10 En una realización preferente, la invención refiere a un método para el diagnóstico del cáncer de próstata que comprende los pasos de analizar en una muestra de un paciente el nivel de expresión de un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11, donde, si el nivel de expresión de dicho es superior a un valor umbral, la muestra se identifica como un resultado de cáncer de próstata positivo.

15 En una realización más preferente, la invención refiere a un método para el diagnóstico del cáncer de próstata que comprende los pasos de analizar en una muestra de un paciente el nivel de expresión de un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que consiste en la SEQ ID NO 3 y la SEQ ID NO 11, donde, si el nivel de expresión de dicho ácido nucleico es superior a un valor umbral, la muestra se identifica como un resultado de cáncer de próstata positivo.

20 En una realización alternativa, analizar el nivel de expresión de un ácido nucleico significa analizar el complemento inverso o el ADNc del ácido nucleico.

25 En una realización preferente, la muestra se selecciona de entre el grupo que comprende tejido prostático, una biopsia, nodos linfáticos, orina, eyaculado, sangre, suero sanguíneo, plasma sanguíneo, células tumorales circulantes en la sangre o la linfa, y cualquier tejido del que se sospeche que contiene metástasis, así como cualquier otra fuente que pueda contener células prostáticas tumorales o partes de ellas, incluidas las vesículas como los exosomas, las microvesículas y otras, así como las moléculas de ARN libres o unidas a proteínas derivadas de células tumorales prostáticas o partes de ellas. Más preferentemente, la muestra es orina y, con máxima preferencia, la muestra es orina obtenida de un paciente tras un examen de tacto rectal.

30 Los resultados experimentales demuestran que los nuevos biomarcadores tienen alta especificidad y sensibilidad para la detección del CaP.

35 Idealmente, el nivel de expresión de un transcrito de los ácidos nucleicos de acuerdo con las SEQ ID NO 1 a 42, más preferente un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11, se compara con el nivel de expresión de uno o varios otros transcritos de genes en la muestra, por ejemplo, de genes constitutivos (*housekeeping*). En la Tabla 2 se muestran algunos ejemplos de genes constitutivos adecuados:

Tabla 2: Ejemplos de genes constitutivos (*housekeeping*) adecuados.

<p>Nombre del gen constitutivo (<i>housekeeping</i>) GAPDH: gliceraldehído-3-fosfato deshidrogenasa HPRT1: hipoxantina fosforibosiltransferasa 1 HMBS: hidroximetilbilano sintasa TBP: proteína de unión a TATA</p>
--

El umbral es la mínima diferencia de expresión entre la muestra bajo análisis y la muestra de control para la cual la muestra se identifica como un resultado de cáncer positivo.

40 Idealmente, el valor umbral para la diferencia en el nivel de expresión del biomarcador entre la muestra bajo análisis y la muestra de control es de 1,5 veces ($\pm 20\%$), 2 veces ($\pm 20\%$), 3 veces ($\pm 20\%$), 4 veces ($\pm 20\%$) y, con máxima preferencia, 5 veces ($\pm 20\%$) o más. El valor p (prueba T) es menor a 2×10^{-5} . La tasa de falsos descubrimientos (FDR) es, preferentemente, menor a 5×10^{-4} .

45 En el caso de un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11, el umbral para asignar a una muestra un resultado positivo de cáncer de próstata es, preferentemente, un aumento de 2 veces en el nivel de expresión entre la muestra bajo análisis y la muestra de control.

La invención refiere a la cuantificación del nivel de expresión de biomarcadores de ARN. Luego de la amplificación, la cuantificación es sencilla y puede lograrse mediante diversos métodos. En el caso de que se utilicen cebadores al

menos uno de los cuales tiene un colorante fluorescente unido, se puede realizar una cuantificación a partir de la señal fluorescente del colorante. Existen distintos tipos de sistemas de cebadores y de colorantes, como SYBR Green, las sondas múltiplex, las sondas TaqMan, las sondas moleculares y los cebadores Scorpion. Estos son adecuados, por ejemplo, para realizar métodos basados en PCR, como la PCR cuantitativa con transcripción inversa (qRT-PCR). Otros medios posibles de cuantificación son los Northern blots, la secuenciación de última generación y las mediciones de absorbancia a 260 y 280 nm.

Cualquier método adecuado para la cuantificación de ácidos nucleicos puede usarse para analizar los niveles de expresión de los ácidos nucleicos. En una realización de la invención, el análisis del método se realiza mediante una determinación basada en fluorescencia. En una realización preferente, el análisis se realiza midiendo la fluorescencia de un cebador marcado, una sonda marcada o un agente de detección fluorescente (como SYBR Green). Más preferentemente, este análisis del nivel de expresión se realiza mediante qRT-PCR. En este método, luego de la transcripción inversa, la muestra se mezcla con un cebador directo (*forward*) o inverso (*reverse*) que es específico para al menos un ácido nucleico seleccionado de entre el grupo de las SEQ ID NO 1 a 42, preferentemente un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11, y, luego, se realiza una amplificación. Las sondas o los cebadores están diseñados de tal manera que hibridan en condiciones estrictas con tal secuencia blanco.

En una realización, el análisis del nivel de expresión se realiza mediante secuenciación de última generación.

En una realización alternativa, se analiza y/o cuantifica el producto proteico de una de las SEQ ID NO 1 a 42, preferentemente las SEQ ID NO 3 y 11.

La invención también refiere a un cebador o una sonda que hibrida en condiciones estrictas con uno de los ácidos nucleicos con las SEQ ID NO 1 a 42.

En una realización preferente, la invención refiere a un cebador (*primer*) o una sonda (*probe*) que hibrida en condiciones estrictas a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11, o cualquier parte de estos, donde dicho cebador o sonda es, preferentemente, una sonda marcada.

En una realización preferente de la invención, el cebador o la sonda que hibrida en condiciones estrictas a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11 tiene aproximadamente entre 5 y 500 nt de longitud, más preferentemente entre 10 y 200 nt, incluso más preferentemente entre 10 y 100 nt. En la realización de máxima preferencia, tal ácido nucleico tiene 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30 nt de longitud.

En una realización de la invención, el cebador o la sonda que hibrida en condiciones estrictas a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11 comprende una marca detectable. En una realización incluso más preferente, el cebador o la sonda que hibrida en condiciones estrictas a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11 comprende, además, una fracción inhibidora de fluorescencia (*quencher*).

La invención también refiere al uso de un cebador o una sonda que hibrida en condiciones estrictas a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11 para el diagnóstico del cáncer de próstata.

En una realización preferente, la invención refiere a un cebador (*primer*) o una sonda (*probe*) que hibrida en condiciones estrictas a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que consiste en la SEQ ID NO 3 y la SEQ ID NO 11, o cualquier parte de estos, donde dicho cebador o sonda es, preferentemente, una sonda marcada.

La invención refiere, además, a un ácido nucleico con una secuencia de entre el grupo de las SEQ ID NO 1 a 42 —o bien su complemento inverso— o a un ácido nucleico que tiene, preferentemente, al menos 85 %, 90 %, 95 % o 99 % de identidad a nivel de secuencia con un ácido nucleico de acuerdo con uno cualquiera de los ácidos nucleicos de las SEQ ID NO 1 a 42.

En una realización preferente, la invención refiere a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11 —o bien su complemento inverso— o a un ácido nucleico que tiene, preferentemente, al menos 85 %, 90 %, 95 % o 99 % de identidad a nivel de secuencia con el ácido nucleico seleccionado.

En una realización más preferente, la invención refiere a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que consiste en la SEQ ID NO 3 y la SEQ ID NO 11 —o bien su complemento inverso— o a un ácido nucleico que tiene, preferentemente, al menos 85 %, 90 %, 95 % o 99 % de identidad a nivel de secuencia con el ácido nucleico seleccionado.

La invención también refiere al uso de un ácido nucleico con una secuencia de entre el grupo de las SEQ ID NO 1 a 42 para el diagnóstico del cáncer de próstata.

5 En una realización preferente, la invención refiere al uso de un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11 —o bien su complemento inverso— para el diagnóstico del cáncer.

En una realización más preferente, la invención refiere al uso de un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que consiste en la SEQ ID NO 3 y la SEQ ID NO 11 —o bien su complemento inverso— para el diagnóstico del cáncer.

10 La invención también refiere a un kit para la detección y/o el diagnóstico del cáncer de próstata que comprende una sonda o un cebador que hibrida en condiciones estrictas a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11. El kit puede contener más de un ácido nucleico. En una realización preferente, el kit comprende, además, reactivos

para la amplificación, cuantificación y/o detección de ácidos nucleicos. En otra realización, el kit comprende muestras de control.

15 En una realización preferente, la invención refiere a un kit para la detección y/o el diagnóstico del cáncer de próstata que comprende una sonda o un cebador que hibrida en condiciones estrictas a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que consiste en la SEQ ID NO 3 y la SEQ ID NO 11. El kit puede contener más de un ácido nucleico. En una realización preferente, el kit comprende, además, reactivos para la amplificación, cuantificación y/o detección de ácidos nucleicos. En otra realización, el kit comprende muestras de control.

20 En una realización alternativa, la invención refiere a un método para el tratamiento y el diagnóstico del cáncer de próstata que comprende los pasos de analizar en una muestra de un paciente el nivel de expresión de un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11, donde, si el nivel de expresión de dicho ácido nucleico es superior a un valor umbral, la muestra se identifica como un resultado de cáncer de próstata positivo, y administrar al paciente uno o más agentes terapéuticos para el tratamiento del cáncer de próstata.

25 En una realización preferente, la invención refiere a un método para el tratamiento y el diagnóstico del cáncer de próstata que comprende los pasos de analizar en una muestra de un paciente el nivel de expresión de un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que consiste en la SEQ ID NO 3 y la SEQ ID NO 11, donde, si el nivel de expresión de dicho ácido nucleico es superior a un valor umbral, la muestra se identifica como un resultado de cáncer de próstata positivo, y administrar al paciente uno o más agentes terapéuticos para el tratamiento del cáncer de próstata.

30 En una realización, los agentes terapéuticos para el tratamiento del cáncer de próstata comprenden: docetaxel (Taxotere®); cabazitaxel (Jevtana®); mitoxantrona (Novantrone®); estramustina (Emcyt®); doxorubicina (Adriamycin®); etopósido (VP-16); vinblastina (Velban®); paclitaxel (Taxol®); carboplatino (Paraplatin®); acetato de abiraterona, bicalutamida (Casodex®), degarelix, enzalutamida, acetato de goserelina, acetato de leuprolida, prednisona, sipuleucel-T, dicloruro de radio 223 y/o vinorelbina (Navelbine®).

Como se apreciará en los ejemplos siguientes, la invención divulga biomarcadores para el cáncer de próstata que permiten un diagnóstico de la enfermedad más preciso y sensible que los biomarcadores actuales.

40 Ejemplos

Materiales y métodos

Cohorte clínica

45 Pacientes con carcinoma de próstata (CaP) sometidos a prostatectomía radical (PR) o cirugía para extirpar una hiperplasia prostática benigna (HPB) en el Hospital Universitario de Dresde se incluyeron en una cohorte clínica retrospectiva con el objetivo de identificar nuevos biomarcadores para el CaP. Se obtuvo la aprobación del comité de ética local, así como el consentimiento informado de los pacientes, de conformidad con las normas legales vigentes. Se recogieron datos sobre el seguimiento clínico de los pacientes con CaP durante al menos cinco años.

50 Se usó secuenciación de ARN a nivel de todo el genoma sobre muestras de tejido prostático de una cohorte de 40 pacientes con CaP y 8 pacientes con HPB para la identificación de biomarcadores relevantes para el diagnóstico. Se definieron cuatro grupos de CaP de acuerdo con las etapas definidas por Gleason (The Veteran's Administration Cooperative Urologic Research Group: histologic grading and clinical staging of prostatic carcinoma; en Tannenbaum, M. Urologic Pathology: The Prostate, Philadelphia: Lea and Febiger. Pp. 171-198), así como la presencia de metástasis en los nodos linfáticos adyacentes al momento de la PR (véase la Tabla 3).

Tabla 3: Cohorte de CaP para detección mediante secuenciación de ARN a nivel de todo el genoma. El grupo de control (C) consiste en muestras de HPB. Los grupos de muy bajo riesgo (V) y de bajo riesgo (L) comprenden muestras de pacientes con un puntaje de Gleason (GS) menor a 7 o igual a 7, respectivamente, todos sin metástasis en nodos linfáticos (pN0). El grupo de riesgo medio (M) comprende los casos con un GS menor o igual a 7 que exhiben metástasis en nodos linfáticos (pN+), mientras que el grupo de alto riesgo (H) consiste en los tejidos con un GS mayor a 7. En este último, se analizaron pares de muestras con y sin tumor obtenidas del mismo paciente.

Grupo	C	V	L	M	H			
Puntaje de Gleason (GS)	HPB	GS<7	GS=7	GS<=7	GS>7			
Metástasis en nodos linfáticos	-	pN0	pN0	pN+	pN0		pN+	
Tejido	control	tumor	tumor	tumor	tumor	Sin tumor	tumor	sin tumor
cantidad de muestras	8	8	8	8	8	8	8	8

Los biomarcadores candidatos seleccionados se validaron posteriormente mediante el uso de microarreglos hechos a medida y de PCR cuantitativa en tiempo real con transcripción inversa (qRT-PCR) en cohortes de 256 pacientes (40 muestras de HPB como control y 216 muestras de tumor) y 56 pacientes (16 muestras de HPB como control y 40 muestras de tumor), respectivamente.

Muestras de tejido prostático

Se obtuvieron muestras de tejido prostático de cirugías realizadas en el Departamento de Urología del Hospital Universitario de Dresde y se las almacenó en nitrógeno líquido en el Centro Integral del Cáncer de la Universidad de Dresde. Las muestras de tejido prostático obtenidas de prostatectomías radicales (PR) de pacientes con carcinoma de próstata (CaP) se dividieron en muestras con tumor y muestras sin tumor. Se usaron muestras de tejido prostático de pacientes con hiperplasia prostática benigna (HPB) como controles. En todos los casos, se obtuvo el consentimiento de los pacientes.

Para verificar el estado de las muestras y su contenido de células tumorales, todas las muestras se dividieron en una serie de criosecciones. Para ello, muestras de tejido congeladas se sumergieron en compuesto Tissue-Tek OCT (Sakura Finetek GmbH) y se fijaron sobre buriles metálicos por congelación. Las criosecciones se prepararon con un criomicrotomo (Leica) equipado con una cuchilla de microtomo C35 (FEATHER) enfriada a -28 °C. Cada muestra se cortó en un total de 208 criosecciones, 4 de las cuales se tiñeron con HE y fueron evaluadas por un patólogo en relación con su contenido de células tumorales (Fig. 1). De este modo, se obtuvieron 3 pilas de criosecciones consecutivas, cada una con secciones teñidas con HE a cada lado. Únicamente se usaron como muestras de tumor y muestras sin tumor pilas que tuvieran a cada lado secciones con al menos 60 % o a lo sumo 5 % de células tumorales, respectivamente. Luego, 50 criosecciones de las pilas seleccionadas se sometieron a una preparación de ARN.

Aislamiento de ARN

Se aisló ARN total del tejido criopreservado con Qiazol y el kit miRNeasy Mini Kit en un instrumento QIAcube (todos ellos de Qiagen), seguido por una digestión manual con ADNasa I. La concentración de ARN se determinó con un instrumento Nanodrop 1000 (Peqlab). La integridad del ARN se verificó en un instrumento Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, California), y únicamente se continuaron procesando las muestras con un número de integridad del ARN (RIN) mayor o igual a 6.

Secuenciación de última generación de ARN largo a nivel de todo el genoma

Se realizó secuenciación de última generación de ARN largo a nivel de todo el genoma sobre un subconjunto de la cohorte retrospectiva de CaP que comprendía 8 muestras de tejido prostático de hiperplasia prostática benigna (HPB) como control y 56 muestras de pacientes con cáncer de próstata (incluidos pares de tejido tumoral y sin tumor de muestras con un puntaje de Gleason mayor a 7). Se eliminó el ARN ribosómico de 1 µg de ARN total con el kit Ribo-Zero rRNA Removal Kit (Epicentre). Las bibliotecas de secuenciación se prepararon a partir de 50 ng de ARN sin ARNr con el kit ScriptSeq v2 RNA-Seq Library Preparation Kit (Epicentre). El ADNc doblemente marcado se purificó con el kit Agencourt AMPure XP System Kit (Beckman Coulter). Se realizaron 10 ciclos de PCR para incorporar los códigos de barra indizados para la multiplexación de muestras y amplificar las bibliotecas de ADNc. La calidad y concentración de las bibliotecas amplificadas se determinó con un kit DNA High Sensitivity Kit en un instrumento Agilent Bioanalyzer (Agilent Technologies). 4 ng de cada una de las 8 muestras se combinaron y se sometieron a una selección de tamaño por electroforesis en geles de agarosa al 2 %. El rango entre 150 pb y 600 pb se extrajo del gel y se purificó con el kit MinElute Gel Extraction Kit (Qiagen) de acuerdo con las instrucciones del fabricante. Las bibliotecas purificadas se cuantificaron en un instrumento Agilent Bioanalyzer con un chip DNA High Sensitivity Chip (Agilent Technologies). A continuación, cada biblioteca combinada, sometida a selección de tamaño y purificada se cargó en una celda de flujo (*flow cell*) HiSeq2000, distribuyéndola entre todas las hileras (*lanes*). La generación de

clústers se realizó con kits TruSeq PE Cluster Kit v3 (Illumina Inc.) en un instrumento Illumina cBOT de acuerdo con el protocolo del fabricante. La secuenciación se realizó en un secuenciador Illumina HiSeq2000 (Illumina, Inc.). Los detalles de los experimentos de secuenciación fueron los siguientes: estrategia de secuenciación con lecturas apareadas (*paired-end*); 101 ciclos para la lectura 1, 7 ciclos para las secuencias de índice, y 101 ciclos para la lectura 2.

Análisis de los datos de secuenciación: Preparación de las lecturas crudas

Los datos de secuenciación brutos, que comprenden los archivos de identificación de bases (archivos BCL), se procesaron con el software CASAVA v1.8.1 (Illumina) para obtener archivos FASTQ. Estos contienen, por cada muestra clínica, todos los fragmentos de ARN secuenciados (en lo sucesivo, "lecturas"). Las secuencias de los adaptadores específicos de la plataforma se eliminaron con el software cutadapt (<http://code.google.com/p/cutadapt/>).

Análisis de los datos de secuenciación: Mapeo al genoma y ensamblaje de transcritos

Las lecturas se mapearon al genoma humano (ensamblaje hg19) con segemehl v0.1.4-382 y TopHat v2.0.9. Los transcritos nuevos (es decir, los transcritos no anotados en Gencode v17) se ensamblaron con Cufflinks v2.1.1 y Cuffmerge v2.1.1. Todos los transcritos nuevos y todos los transcritos conocidos en Gencode v17 se combinaron en un conjunto de anotación integral.

Análisis de los datos de secuenciación: Análisis estadístico

Se usó Htseq-count v0.5.4p1 (<http://www-huber.embl.de/users/anders/HTSeq/doc/count.html>) para calcular los recuentos de lecturas por transcripto y por gen contenidos en el conjunto de anotación integral de transcritos nuevos y conocidos. Los transcritos y genes diferencialmente expresados se identificaron con R y con la biblioteca edgeR de Bioconductor. Para compensar las diferencias en la composición de ARN de las muestras clínicas, se escaló el tamaño de la biblioteca de cada muestra (método TMM). Se ajustó un modelo logarítmico-lineal-binominal negativo al recuento de lecturas correspondientes a cada transcripto o gen y se usó una prueba de cocientes de verosimilitud para identificar coeficientes distintos de cero para el modelo. La tasa de falsos descubrimientos (FDR) se controló mediante el ajuste de Benjamini-Hochberg.

Validación mediante microarreglos hechos a medida

Sobre la base de los resultados de secuenciación, se diseñaron a medida microarreglos con 180 000 sondas (Agilent SurePrint G3 Custom Exon Array, 4x180K, ID de diseño 058029) que comprendían sondas para ARNm, ARN no codificante (gencode v15), nuevos transcritos y todos los transcritos expresados diferencialmente entre las muestras de tumor y de control de acuerdo con la secuenciación de ARN. El diseño de las sondas se realizó mediante la herramienta de diseño personalizado eArray de Agilent.

Los estudios de detección con microarreglos se realizaron con la cohorte retrospectiva de CaP que comprendía 40 muestras de tejido prostático de pacientes con hiperplasia prostática benigna (HPB) como control, así como 164 y 52 muestras con y sin tumor, respectivamente, de pacientes con cáncer de próstata tras una prostatectomía radical. Se usó el kit Quick Amp Labeling Kit (Agilent) para sintetizar ARNc a partir de 200 ng de ARN total y se hibridaron 1650 ng de ARNc a los arreglos (Agilent Gene Expression Hybridization Kit).

Análisis de datos de microarreglos de ARN hechos a medida

Las sondas con expresión diferencial se identificaron con R y la biblioteca de Bioconductor "limma". Para realizar un control de calidad de los arreglos, se verificó la distribución de las sondas de la "esquina brillante" y la "esquina oscura", así como la concentración de los controles de calibración (*spike-ins*) frente a la señal normalizada. Para obtener un conjunto de sondas que mapearan a posiciones genómicas únicas en la referencia hg19, se usó BLAT configurado con el parámetro $-minIdentity = 93$. Se descartaron todas las sondas que mapearan a más de una región genómica distinta. La normalización entre los arreglos se realizó mediante normalización por cuantiles. Para reducir la cantidad de pruebas, se aplicó el siguiente filtrado no específico: La expresión de una sonda debe ser mayor a la expresión de fondo en 10 % de los arreglos. La expresión de fondo se define como la intensidad media más tres veces la desviación estándar de los puntos de control negativo (puntos 3xSLV de Agilent). Además, una sonda debe exhibir un cambio de expresión no específico con un rango intercuartílico (RIC) mayor a 0,5. Por último, se ajustó un modelo lineal con el paquete de R limma y se obtuvieron estimaciones confiables de la varianza mediante estadísticos moderados mediante el método empírico de Bayes. La tasa de falsos descubrimientos (FDR) se controló mediante el ajuste de Benjamini-Hochberg.

Validación mediante PCR cuantitativa en tiempo real

Para la validación de los resultados obtenidos con secuenciación de última generación y detección mediante microarreglos, se analizaron muestras de tejido (16 sin tumor y 40 de tumor) mediante PCR cuantitativa en tiempo real. Se sintetizó ADNc a partir de 100 ng de ARN total con el kit High-Capacity Reverse Transcription Kit (Applied Biosystems) y cebadores aleatorios, de acuerdo con las instrucciones del fabricante. Los análisis por PCR posteriores se realizaron con 4 μ l del ADNc diluido. Se realizó PCR cuantitativa en tiempo real mediante ensayos de

expresión génica prediseñados TaqMan (Applied Biosystems) dirigidos a los transcritos constitutivos y blanco en un sistema 7900HT Real-Time PCR System de Applied Biosystems.

Tabla 4: Identificadores de los ensayos de expresión génica TaqMan de Applied Biosystems usados para la validación por qRT-PCR en muestras de tejido prostático.

	Nombre del gen constitutivo/blanco	ID de ensayo TaqMan
Gen constitutivo	GAPDH	Hs02758991_gi
	HPRT1	Hs02800695_ml
	HMBS	Hs00609293_gi
Blanco	SEQ ID NO 1	AJ70L28
	SEQ ID NO 9	Hs01388451_m1
	SEQ ID NO 3	AJCSVRJ
	PCA3	Hs01371939_gi

5

Todas las muestras se midieron por triplicado y las medias de estas mediciones se usaron para los cálculos posteriores.

Análisis estadístico de los resultados de qRT-PCR

La normalización de los datos se realizó tomando como referencia los genes constitutivos no regulados GAPDH y HPRT1. Para la cuantificación relativa, se analizaron los cambios en la expresión génica de cada muestra en relación con la expresión mediana de las muestras de control. Todos los análisis estadísticos se realizaron con el software estadístico R.

Se compararon los niveles de expresión relativos de los biomarcadores transformados por el logaritmo en base 2 (\log_2) entre las muestras de tumor y de control con una prueba t de Student. Luego, se calcularon con el paquete pROC las curvas de característica operativa del receptor (ROC), cuyo área bajo la curva (AUC) es una medida del poder diagnóstico de cada marcador.

Validación en muestras de orina de ETR: recolección de las muestras y aislamiento del ARN

Se recolectaron muestras de orina tras un examen de tacto rectal (ETR) de la próstata (orina de ETR). Este método de examen de rutina permite obtener muestras de orina que contienen cierta cantidad de células prostáticas. Las muestras de orina de ETR se centrifugaron y lavaron dos veces con PBS. El precipitado de células resultante se suspendió en 700 μ l de Qiazol. Se aisló ARN total con el kit miRNeasy Mini Kit en un instrumento QIAcube (todos ellos de Qiagen), seguido por una digestión manual con ADNasa I. La concentración de ARN se determinó con un instrumento Nanodrop 1000 (Peqlab). La integridad del ARN se verificó en un Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, California).

Detección en muestras de orina de ETR por PCR cuantitativa en tiempo real

Se sintetizó ADNc a partir de 2x50 ng de ARN total con la enzima SuperScript III Reverse Transcriptase (Applied Biosystems) y cebadores aleatorios, de acuerdo con las instrucciones del fabricante. Los análisis por PCR posteriores se realizaron con 4 μ l de ADNc. Se realizó PCR cuantitativa en tiempo real mediante ensayos de expresión génica prediseñados TaqMan (Applied Biosystems) dirigidos a los transcritos constitutivos (PSA) y blanco en un sistema 7900HT Real-Time PCR System de Applied Biosystems. Todas las muestras se midieron por duplicado y las medias de estas mediciones se usaron para los cálculos posteriores.

Secuenciación de última generación de ARN largo a nivel de todo el genoma en muestras de orina de ETR

Para la secuenciación de última generación de ARN largo a nivel de todo el genoma, se precipitó ARN total de 7 muestras de orina de ETR con etanol para concentrar el ARN y se lo resuspendió en 10 μ l de agua sin RNasa. La eliminación del ARNr se realizó sobre 4 ng de ARN total con el kit Low Input Ribo-Zero rRNA Removal Kit (Epicentre, modificado por Clontech), obteniéndose 10 μ l de ARN sin ARNr. Las bibliotecas de secuenciación se prepararon a partir de 8 μ l de ARN sin ARNr con el kit SMARTER stranded RNAseq Kit (Clontech). El ADNc doblemente marcado se purificó con el kit Agencourt AMPure XP System Kit (Beckman Coulter). Se realizaron 18 ciclos de PCR para amplificar las bibliotecas de ADNc e incorporar los códigos de barra indizados para la multiplexación de muestras. La calidad y concentración de las bibliotecas amplificadas se determinó con un kit DNA High Sensitivity Kit en un instrumento Agilent Bioanalyzer (Agilent Technologies). Las muestras se combinaron y la generación de clústers se realizó con 15 pmol/l de la biblioteca combinada y el kit TruSeq PE Cluster Kit v4 (Illumina Inc.) en un instrumento Illumina cBOT de acuerdo con el protocolo del fabricante. La secuenciación se realizó con los reactivos de secuenciación HiSeq SBS v4 (250 ciclos) en un secuenciador Illumina HiSeq2500 (Illumina, Inc.). Los detalles del experimento de secuenciación fueron los siguientes: estrategia de secuenciación con lecturas apareadas (*paired-*

45

end); 126 ciclos para la lectura 1, 7 ciclos para las secuencias de índice, y 126 ciclos para la lectura 2.

Análisis estadístico de los resultados de qRT-PCR de muestras de orina de ETR

5 Para el análisis de los resultados de qRT-PCR de muestras de orina de ETR, se realizó una normalización de los datos tomando como referencia el antígeno prostático específico (PSA). Para la cuantificación relativa, se analizaron los cambios en la expresión génica de cada muestra en relación con la expresión mediana de las muestras de control. Todos los análisis estadísticos se realizaron con el software estadístico R.

Tabla 5: Identificadores de los ensayos de expresión génica TaqMan de Applied Biosystems usados para la validación por qRT-PCR en muestras de orina de ETR.

	Nombre	ID de ensayo TaqMan
Gen constitutivo	GAPDH	Hs02758991_gl
	HPRT1	Hs02800695_ml
Blanco	SEQ ID NO 1	AJ70L28
	PSA	Hs02576345_ml

10 **Resultados**

Se analizaron los transcriptomas de 40 muestras de tumor de CaP y 16 muestras sin tumor obtenidas tras un ETR y 8 muestras de tejido prostático de HPB como controles benignos no tumorales se analizaron mediante secuenciación de última generación (NGS) de ARN largo con lecturas apareadas (paired-end). Se prepararon aproximadamente 150 criosecciones por muestra en al menos tres segmentos en un esfuerzo por optimizar la calidad de los datos y mejorar la solidez del análisis. Tras una evaluación patológica, únicamente se retuvieron para el análisis posterior los segmentos con un porcentaje de células tumorales mínimo de 60 % o máximo de 5 % en las muestras tumorales y sin tumor, respectivamente. El enfoque de secuenciación transcriptómica (RNA-seq) buscó identificar y cuantificar de forma exhaustiva el ARN expresado en tejido normal y en tejido prostático canceroso. Se secuenciaron todas las clases de transcritos codificantes y no codificantes, independientemente del estado de poliadenilación. Se usaron grandes cantidades de ARN de entrada para garantizar que la biblioteca tuviera buena complejidad. Se secuenciaron, en promedio, 200 M de lecturas apareadas (paired-end) de 2 x 100 nt por biblioteca, permitiendo en ensamblaje de nuevos transcritos con baja expresión gracias a la alta cobertura. Este enfoque superó la mayoría de los estudios comparables publicados que analizaron grandes cantidades de muestras. En total, se ensamblaron aproximadamente 3000 transcritos nuevos que no mostraron solapamiento exónico con los transcritos anotados en Gencode v17. Con una tasa de falsos descubrimientos (FDR) de 0,01, se observaron 6442 genes diferencialmente expresados en todos los contrastes. En la Tabla 6 se indican las cantidades de genes diferencialmente expresados en contrastes específicos.

Contraste	Total	Codificantes de proteínas	ARNlinc	Antisentido	Intrónico directo	Pseudogénico	Transcripto nuevo	No codificante de proteínas
Tumor vs. control	5615	3882	116	96	13	456	847	1733
Tumor, puntaje Gleason > 7 vs. control	2677	1812	73	40	4	88	552	865
Tumor, riesgo alto y medio vs. tumor, bajo y muy bajo riesgo	138	51	3	2	0	7	72	87
Tumor, puntaje de Gleason = 7 vs. Tumor, puntaje de Gleason < 7	12	6	0	1	0	0	5	6
Tumor, puntaje de Gleason > 7 vs. Tumor, puntaje de Gleason = 7	14	7	0.	.0	0	1	6	7

Tabla 6: Cantidad de genes diferencialmente expresados en distintos contrastes y con distintos biotipos de Gencode.

30 Los resultados reprodujeron exitosamente la mayoría de los transcritos informados anteriormente como diferencialmente expresados entre el tumor de próstata y el tejido normal. Además, se identificó una serie de nuevos transcritos asociados con el CaP, que pueden usarse para desarrollar análisis para el diagnóstico del CaP. Los transcritos más prometedores se seleccionaron para su validación en una cohorte de prueba de muestras de tumor de CaP y muestras de control de HPB por qRT-PCR.

Varios de estos nuevos biomarcadores candidatos tienen una especificidad y una sensibilidad considerablemente mayores a las del biomarcador PCA3, que ya se utiliza para el diagnóstico del CaP. En la cohorte de secuenciación, el PCA3 mostró una asociación clara con el CaP, pero una tendencia marcada a disminuir en el grupo de alto riesgo (Fig. 2).

- 5 Los resultados experimentales demuestran que los nuevos biomarcadores tienen alta especificidad y sensibilidad para la detección del CaP. Por lo tanto, pueden crearse ensayos basados en la medición de estos nuevos biomarcadores solos o combinados (o incluso combinados con otros marcadores) en cualquier fuente que pueda contener células tumorales prostáticas o partes de ellas (incluidas las vesículas como los exosomas, las microvesículas y otras, así como las moléculas de ARN libre o unido a proteínas derivadas de células tumorales prostáticas) para usarse para el diagnóstico del CaP. Estas fuentes incluyen, sin carácter taxativo, el tejido prostático, las biopsias, los nodos linfáticos, la orina, el eyaculado, la sangre, el suero sanguíneo, el plasma sanguíneo y las células tumorales circulantes en sangre o linfa, así como cualquier tejido del que se sospeche que contiene metástasis de CaP. La medición de nuestros biomarcadores de ARN puede realizarse por medio de cualquier método apto para estimar específicamente niveles de ARN (p. ej., los métodos basados en PCR, como la qRT-PCR). Los análisis pueden aplicarse para el diagnóstico temprano del CaP (detección), para predecir la agresividad de los tumores (pronóstico) y/o para contribuir con la elección de la terapia.

En la Fig. 6 se muestran los resultados de la detección de una selección de biomarcadores en orina. Los niveles de expresión de todos los biomarcadores que se muestran en esta figura son mayores en la orina de los pacientes que padecen cáncer de próstata que en la de individuos saludables. Esto muestra que analizar el nivel de expresión de uno de estos biomarcadores en orina permite diagnosticar el cáncer de próstata. Este resultado es inesperado porque Fontenete et al, (Int. braz j urol. vol.37 no.6 Río de Janeiro Nov./Dic. 2011) mostraron que el ARNm del PSA no es un biomarcador adecuado para el cáncer de próstata en muestras de orina, dado que se descubrió que se sobreexpresaba más frecuentemente en pacientes saludables que en pacientes con CaP en estas muestras. Por lo tanto, no era evidente *a priori* que el análisis de los niveles de expresión de los biomarcadores en muestras de orina pudiera usarse para diagnosticar de forma confiable el cáncer de próstata.

Un análisis para diagnóstico basado en estos biomarcadores ofrece como ventajas una tasa de falsos positivos drásticamente menor que la de los análisis actuales y la posibilidad de medir los niveles de expresión en una muestra de orina sin necesidad de realizar biopsias de próstata invasivas.

Leyendas de las figuras

30 Fig.1: Verificación de la calidad de las muestras de tejido: para determinar la cantidad de células tumorales de las muestras de tejido, se prepararon criosecciones de las muestras congeladas como se muestra. HE: hematoxilina/eosina; IHC: inmunohistoquímica. Verificación de la calidad de las muestras de tejido: se prepararon criosecciones de 4 µm a partir de las muestras congeladas como se muestra en la figura para tinción con HE (para verificar el contenido de células tumorales de las muestras de tejido), para aislamiento de ARN y ADN, y para inmunohistoquímica (IHC). HE: hematoxilina/eosina; IHC: inmunohistoquímica.

Fig.2: Diagrama de caja de los datos de RNA-seq correspondientes al transcripto de PCA3. Resultados de secuenciación de ARN de la cohorte de CaP retrospectiva que comprende 8 muestras de tejido prostático de hiperplasia de próstata benigna como control (C), 8 muestras de tumor de CaP de cada uno de los grupos V (muy bajo riesgo; puntaje de Gleason < 7, pN0), L (bajo riesgo; puntaje de Gleason = 7, pN0), y M (riesgo medio; puntaje de Gleason ≤ 7, pN+), así como 16 pares de muestras de tejido tumoral y sin tumor del grupo H (alto riesgo; puntaje de Gleason > 7).

Fig.3: Curvas ROC correspondientes a Retro-RPL7 (SEQ ID NO 1) y PCA3 obtenidas mediante análisis por qRT-PCR de 56 muestras de tejido prostático.

45 Fig.4: Datos de secuenciación de última generación de ARN de la SEQ ID NO 3 de 64 muestras de tejido. Se obtuvieron 8 muestras de tejido de control de pacientes con hiperplasia de próstata benigna (HPB) y 56 muestras de tejido de pacientes con cáncer de próstata tras una prostatectomía radical (PR). Entre estas, 40 muestras representaban tejido tumoral (recuento de células tumorales de al menos 60 %), mientras que 16 muestras representaban tejido adyacente no tumoral (recuento de células tumorales no mayor a 5 %) derivadas del mismo paciente.

50 (A) Diagrama de caja que muestra los recuentos normalizados del ácido nucleico con la SEQ ID NO 3.
(B) Curva ROC para la comparación de los niveles de expresión del ácido nucleico con la SEQ ID NO 3 entre las muestras de tumor y de control: Área bajo la curva ROC (AUC): 0,9906.

55 Fig.5: Datos de secuenciación de última generación de ARN de la SEQ ID NO 11 de 64 muestras de tejido. Se obtuvieron 8 muestras de tejido de control de pacientes con hiperplasia de próstata benigna (HPB) y 56 muestras de tejido de pacientes con cáncer de próstata tras una prostatectomía radical (PR). Entre estas últimas, 40 representaban tejido tumoral con un recuento de células tumorales de al menos 60 %, mientras que 16 muestras representaban tejido adyacente no tumoral (recuento de células tumorales no mayor a 5 %) derivadas del mismo paciente.

(A) Diagrama de caja que muestra los recuentos normalizados del ácido nucleico con la SEQ ID NO 11.

(B) Curva ROC para la comparación de los niveles de expresión del ácido nucleico con la SEQ ID NO 11 entre las muestras de tumor y de control. Área bajo la curva ROC (AUC): 0,9938.

5 Fig.6: Datos de secuenciación de última generación de ARN del locus (hg19): Chr2: 1,550,437-1,629,191. La secuenciación de ARN largo a nivel de todo el genoma en muestras de tejido prostático se realizó en un secuenciador de última generación Illumina HiSeq2000. Las lecturas se mapearon al genoma humano y se ensamblaron. La visualización de las lecturas (cajas grises, verdes y rojas, cada una de 100 bases de longitud), así como de lecturas partidas que indican la presencia de uniones de empalme o *splice junctions* (líneas celestes horizontales), se realiza con el software Integrative Genomics Viewer (IGV, Broad Institute). Las anotaciones se toman del UCSC Genome Browser (genoma humano versión 19). La sección ampliada muestra la transcripción continua de la región interadyacente entre los transcritos anotados.

10 Fig.7: Datos de microarreglos hechos a medida correspondientes a la SEQ ID NO 3 de 256 muestras de tejido. Se obtuvieron 40 muestras de tejido de control de pacientes con hiperplasia de próstata benigna (HPB) y 216 muestras de tejido de pacientes con cáncer de próstata tras una prostatectomía radical (PR). Entre estas últimas, 164 representaban tejido tumoral mientras que 52 muestras representaban un tejido sin tumor adyacente de los mismos pacientes.

(A) Diagrama de caja que muestra los recuentos normalizados del ácido nucleico con la SEQ ID NO 3.

(B) Curva ROC para la comparación de los niveles de expresión del ácido nucleico con la SEQ ID NO 3 entre las muestras de tumor y de control: Área bajo la curva ROC (AUC): 0,9354.

20 Fig.8: Datos de microarreglos hechos a medida correspondientes a la SEQ ID NO 11 de 256 muestras de tejido. Se obtuvieron 40 muestras de tejido de control de pacientes con hiperplasia de próstata benigna (HPB) y 216 muestras de tejido de pacientes con cáncer de próstata tras una prostatectomía radical (PR). Entre estas últimas, 164 representaban tejido tumoral mientras que 52 muestras representaban un tejido sin tumor adyacente de los mismos pacientes.

25 (A) Diagrama de caja que muestra los recuentos normalizados del ácido nucleico con la SEQ ID NO 11.

(B) Curva ROC para la comparación de los niveles de expresión del ácido nucleico con el SEQ ID NO 11 entre las muestras de tumor y de control: Área bajo la curva ROC (AUC): 0,9345.

30 Fig.9: Muestras de orina de pacientes con carcinoma de próstata (tumor) o pacientes saludables (control) obtenidas tras un examen de tacto rectal realizado por un urólogo. El ARN aislado de estas muestras se sometió a secuenciación de ARN de todo el transcriptoma con un secuenciador de última generación Illumina HiSeq2500. Las lecturas se mapearon al genoma con algoritmos estándar. Las lecturas que mapeaban a los *loci* genómicos del transcrito de la SEQ ID NO. que se muestra se contaron y normalizaron respecto a las lecturas derivadas del locus génico del antígeno prostático específico como medida de la presencia de células epiteliales prostáticas en la orina a modo de normalización. La cantidad de lecturas (en millones) se muestra en una escala logarítmica en base 2 (\log_2).

35

REIVINDICACIONES

1. Un método para el diagnóstico del cáncer de próstata que comprende los pasos de
 - a) analizar en una muestra de un paciente el nivel de expresión de un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11,
 - b) donde, si el nivel de expresión de dicho ácido nucleico es superior a un valor umbral, la muestra se identifica como un resultado de cáncer de próstata positivo.
2. Un método de acuerdo con la reivindicación 1, donde la muestra se selecciona de entre el grupo que comprende tejido prostático, material de biopsia, nodos linfáticos, orina, eyaculado, sangre, suero sanguíneo, plasma sanguíneo, células tumorales circulantes en sangre o linfa, cualquier tejido del que se sospeche que contiene metástasis y cualquier otra fuente que pueda contener células prostáticas tumorales o partes de ellas, incluidas las vesículas como los exosomas, las microvesículas y otras, así como las moléculas de ARN libres o unidas a proteínas derivadas de células tumorales prostáticas.
3. Un método de acuerdo con la reivindicación 1, donde la muestra es una muestra de orina.
4. Un método de acuerdo con cualquiera de las reivindicaciones anteriores, donde el análisis del nivel de expresión se realiza midiendo la fluorescencia de un cebador marcado, una sonda marcada o un agente de detección fluorescente.
5. Un método de acuerdo con una cualquiera de las reivindicaciones anteriores, donde el análisis del nivel de expresión se realiza mediante qRT-PCR.
6. El uso de un cebador o una sonda que hibrida en condiciones estrictas a un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11 para el diagnóstico del cáncer de próstata.
7. El uso de un ácido nucleico de acuerdo con la reivindicación 6, donde el cebador o la sonda tiene aproximadamente entre 10 y 100 nucleótidos de longitud.
8. El uso de un cebador o una sonda de acuerdo con las reivindicaciones 6 o 7, donde el cebador o la sonda comprende una marca detectable.
9. El uso de un transcrito del locus (hg19): Chr2: 1,550,437-1,629,191 (-) seleccionado de entre el grupo que comprende la SEQ ID NO 3 y la SEQ ID NO 11 —o bien su complemento inverso— o de un ácido nucleico que tiene, preferentemente, al menos 85 %, 90 %, 95 % o 99 % de identidad a nivel de secuencia con el ácido nucleico seleccionado para el diagnóstico del cáncer de próstata.

30

FIG. 1:

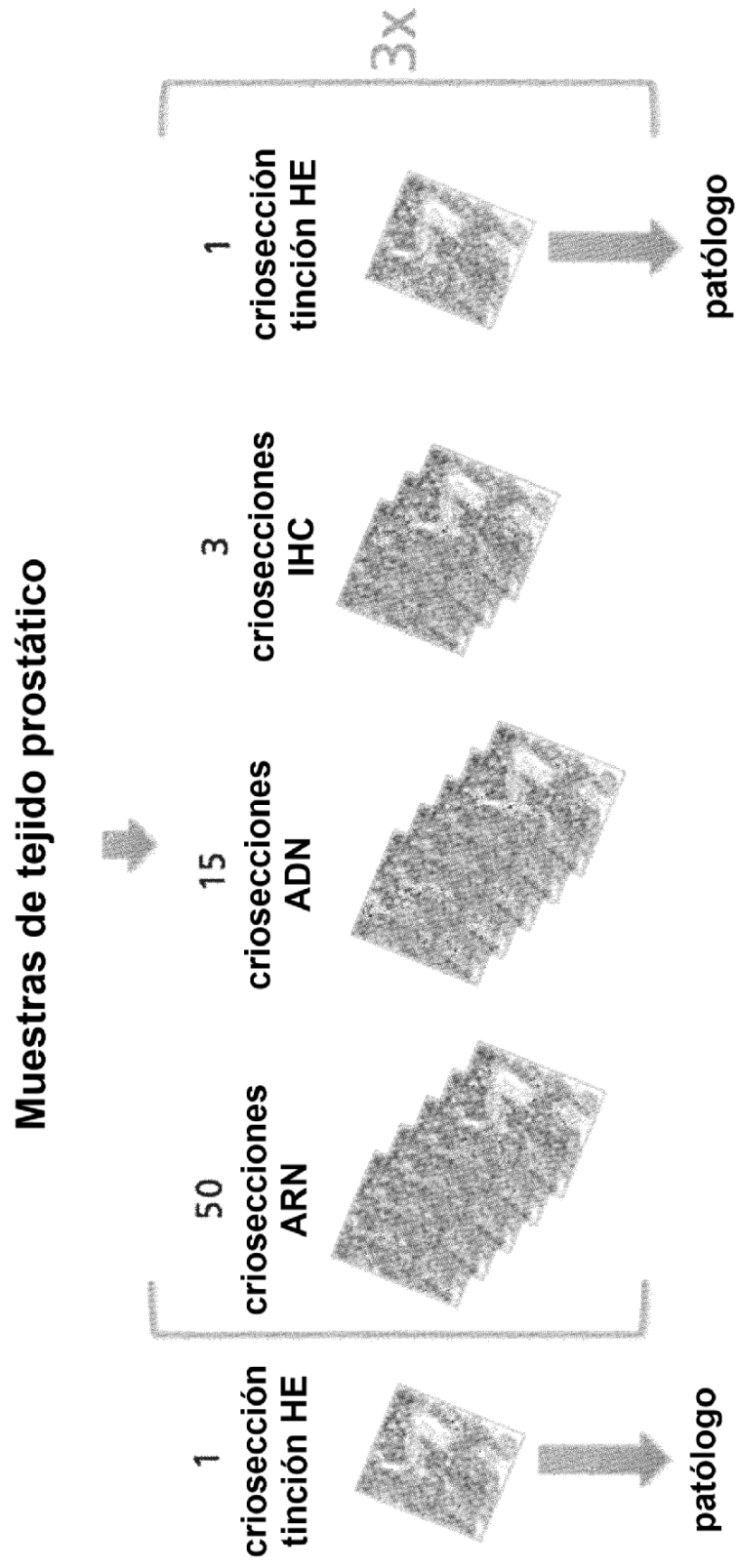


FIG. 2:

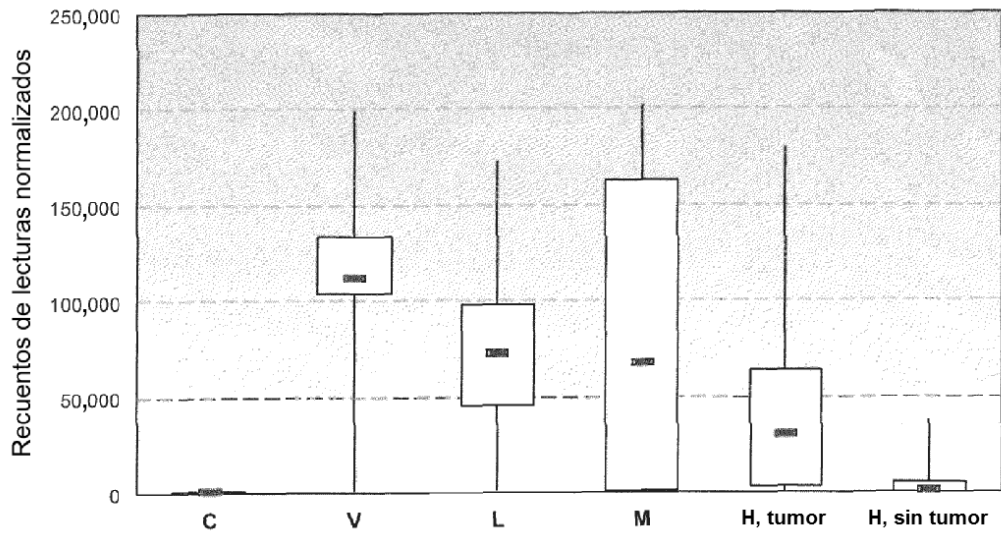


FIG. 3:

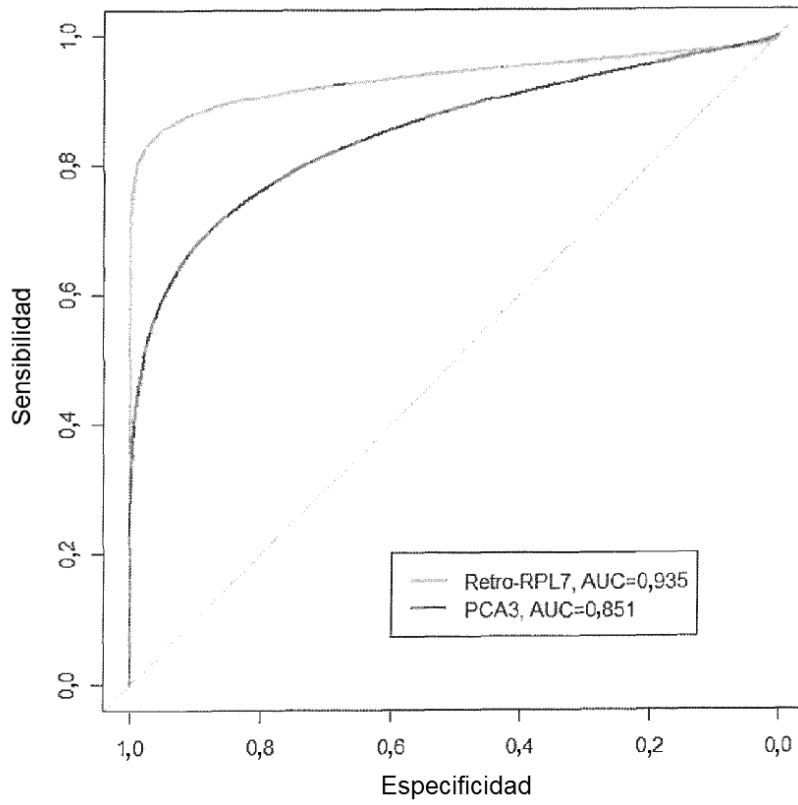
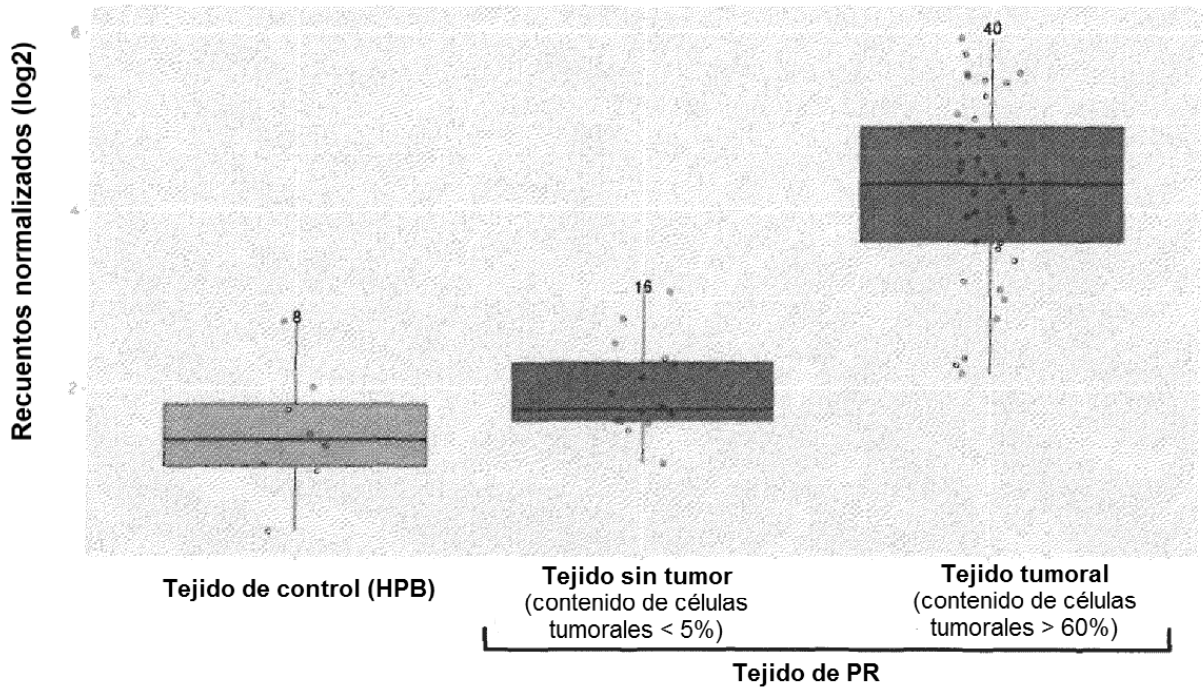


FIG. 4

Secuencias de ARN de la SEQ ID. NO 3:
A) Diagramas de caja



B) Curva ROC

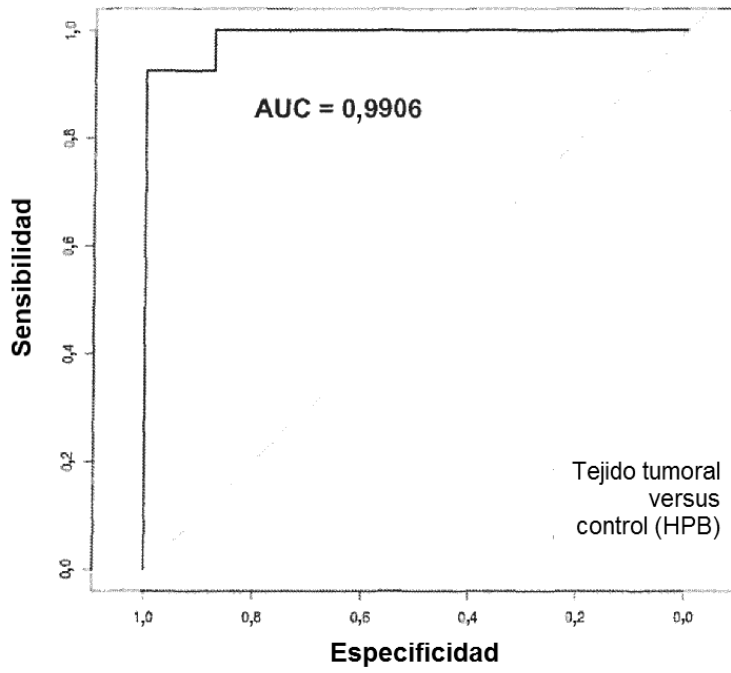
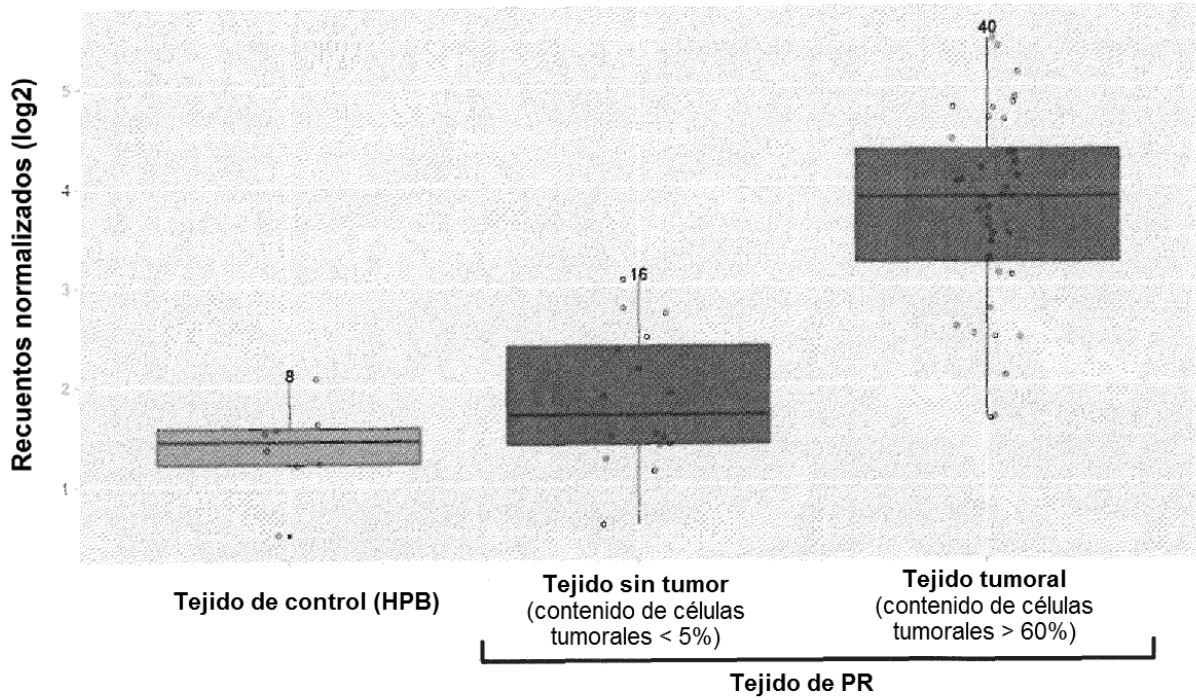


FIG. 5

Secuencias de ARN de la SEQ ID. NO 11:
A) Diagramas de caja



B) Curva ROC

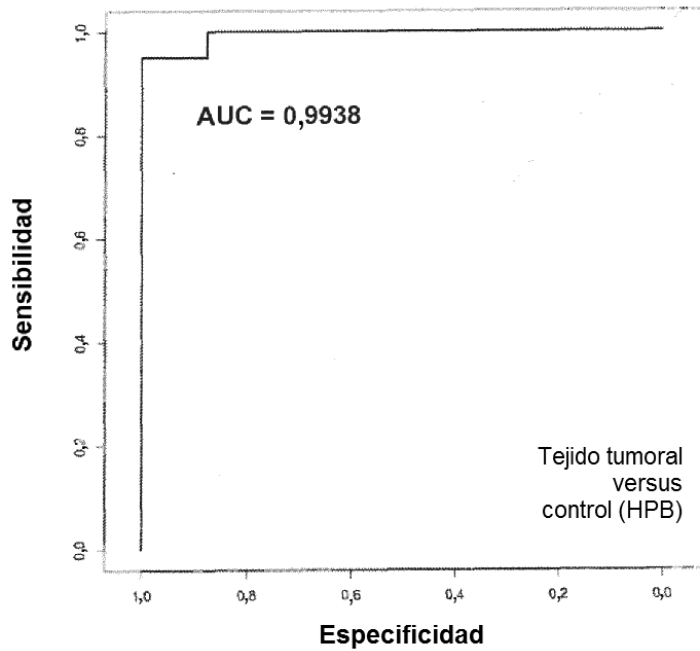


FIG. 6

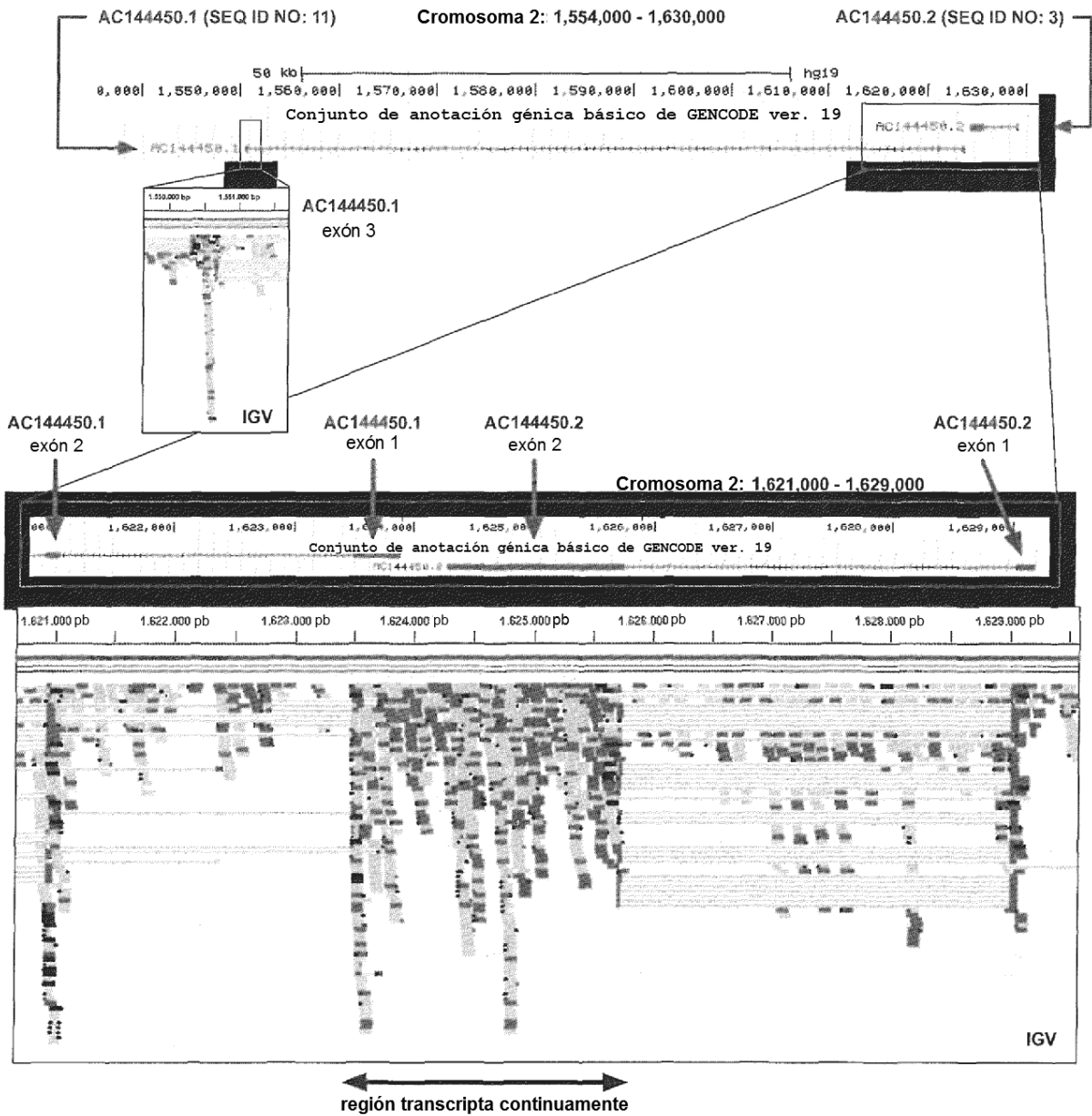
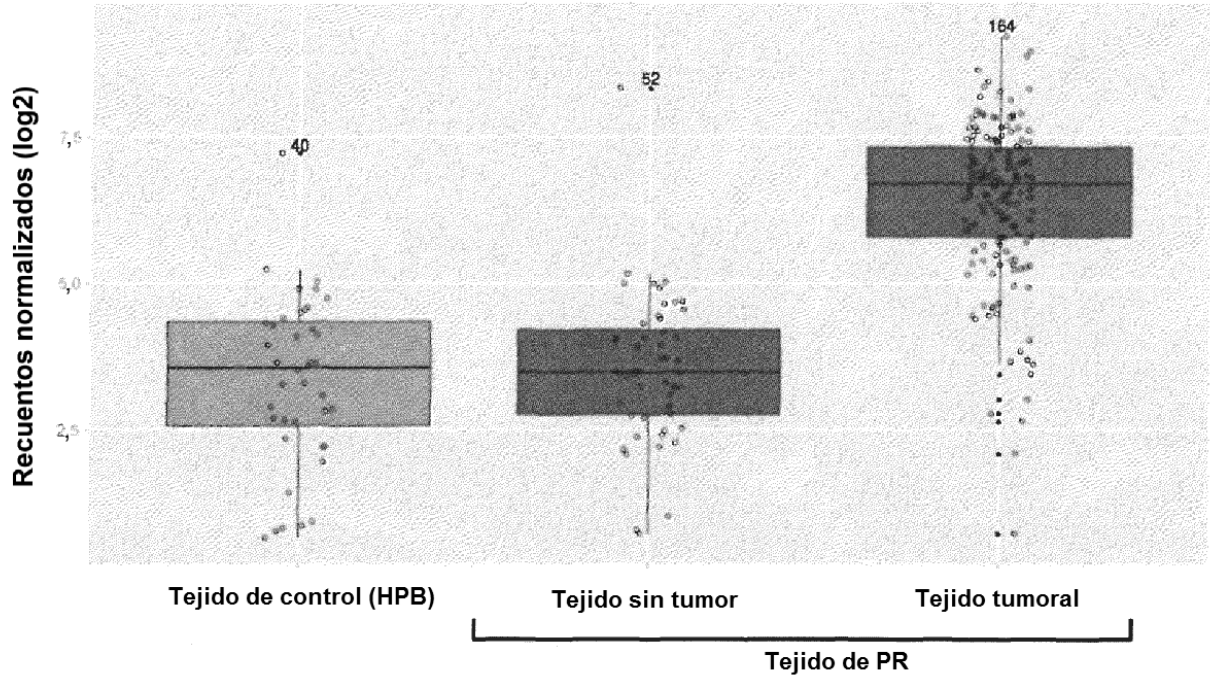


FIG. 7

Análisis de la SEQ ID. NO 3 mediante microarreglos:
A) Diagramas de caja



B) Curva ROC

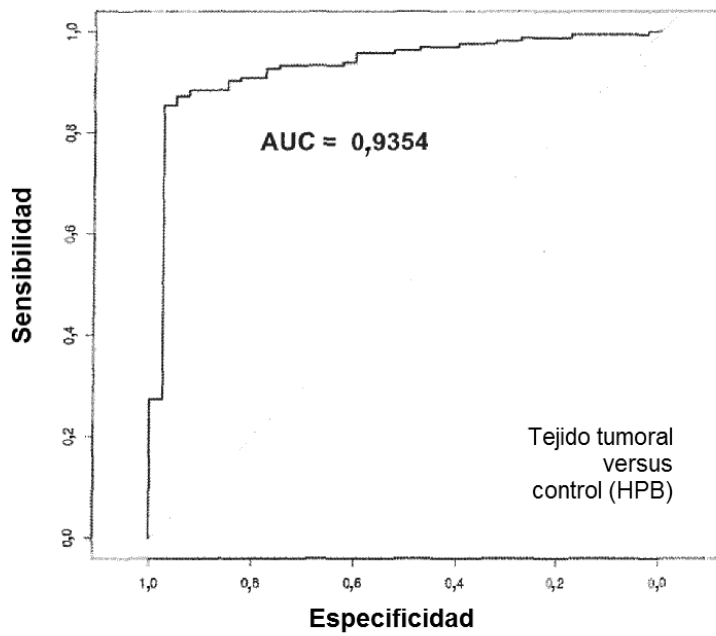
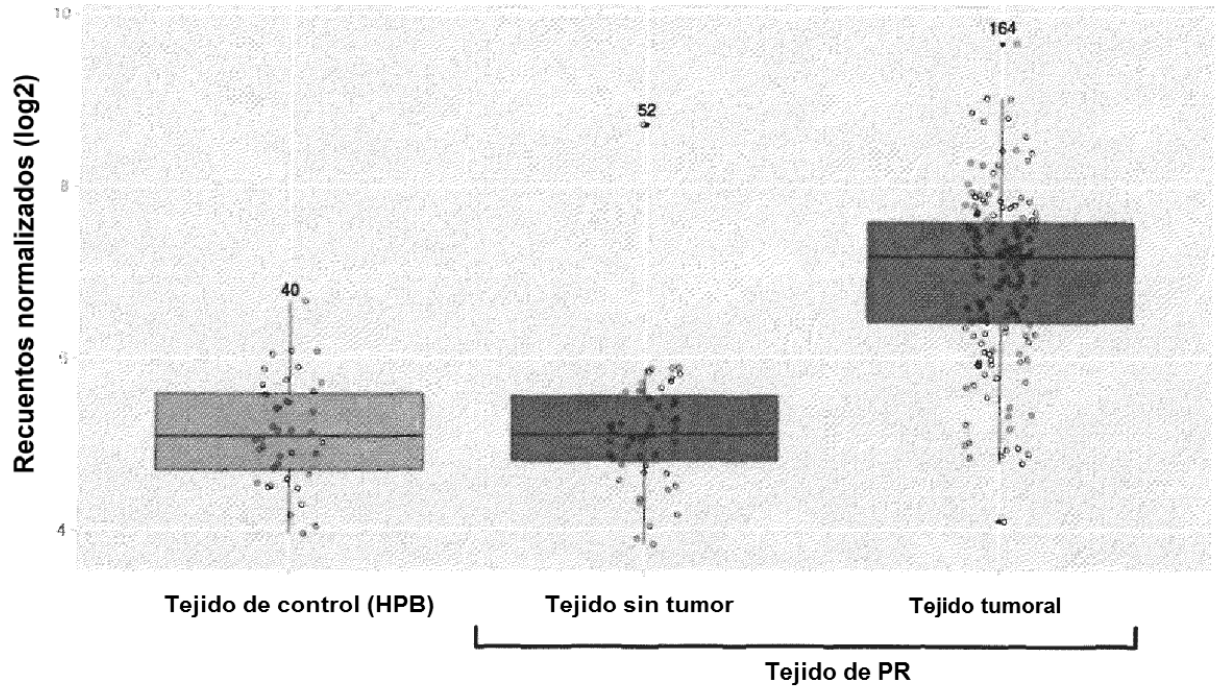


FIG. 8

Análisis de la SEQ ID. NO 11 mediante microarreglos:

A) Diagramas de caja



B) Curva ROC

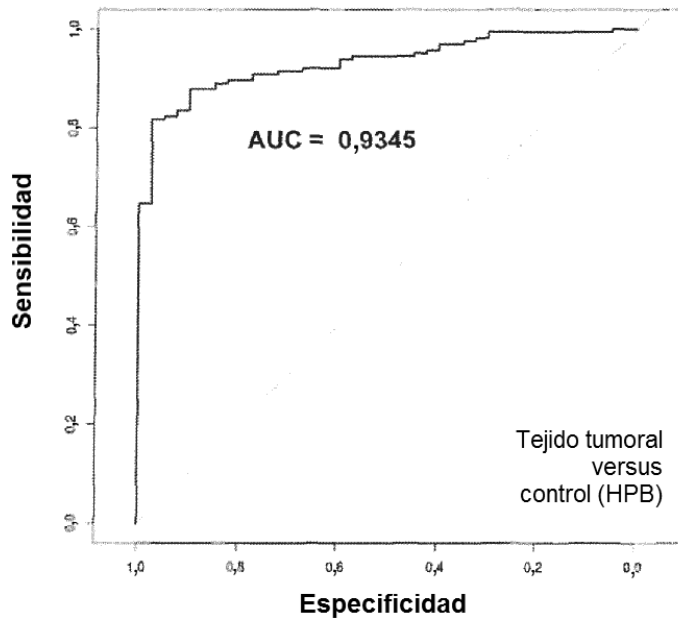


FIG. 9

