

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 631 193**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.03.2012 PCT/EP2012/001477**

87 Fecha y número de publicación internacional: **04.10.2012 WO12130478**

96 Fecha de presentación y número de la solicitud europea: **30.03.2012 E 12712077 (2)**

97 Fecha y número de publicación de la concesión europea: **04.01.2017 EP 2691535**

54 Título: **Marcadores tumorales**

30 Prioridad:

31.03.2011 EP 11160657

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

29.08.2017

73 Titular/es:

**QUEEN MARY AND WESTFIELD COLLEGE
UNIVERSITY OF LONDON (100.0%)**

**Mile End Road
London E1 4NS, GB**

72 Inventor/es:

LORINCZ, ATTILA

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 631 193 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

Marcadores tumorales

Campo de la invención

5 La presente invención se encuentra en el campo de la biología y la química. En particular, la invención se encuentra en el campo de la biología molecular. Más particularmente, la invención hace referencia al análisis del estado de metilación de las regiones genómicas. Más especialmente, la descripción se encuentra en el campo del diagnóstico y/o pronóstico del cáncer de mama y próstata, pero también del cáncer en general.

Antecedentes

10 La metilación reversible de las citosinas es una modificación epigenética importante en organismos multicelulares y se encuentra en muchas enfermedades de seres humanos, incluido el cáncer. Se ha descubierto que los epigenomas del cáncer se hipometilan globalmente con hipermetilaciones específicas para el promotor. Además, la metilación de la citosina provoca represión transcripcional, la cual, en el caso de genes supresores tumorales, los genes apoptóticos, los genes de reparación del ADN y los factores que controlan los puntos de control del ciclo celular, conduce a la progresión tumoral.

15 El cáncer de próstata (PCa, por su sigla en inglés) es la tercera causa más común de muerte por cáncer en el sexo masculino en los países desarrollados. Si se diagnostica en una etapa temprana, el PCa es una enfermedad curable. Los enfoques de gestión clínica dependen de la extensión y gravedad del cáncer y, en el cáncer de bajo grado en etapa temprana, pueden consistir, en su mayoría, en la observación atenta (también denominado tratamiento expectante), mientras que, en cánceres avanzados o agresivos, los tratamientos pueden incluir la prostatectomía radical, la terapia de radiación u hormonal. Sin embargo, debido al resultado mayormente impredecible del cáncer, los pacientes se tratan a menudo sin ningún beneficio evidente y existe un problema mayor reconocido de tratamiento excesivo sustancial en muchos países.

20 El antígeno prostático específico (PSA, por su sigla en inglés) se utiliza como un marcador biológico para analizar a los hombres en busca de posibles desarrollos tumorales. No obstante, la especificidad baja y la sensibilidad moderada conducen a muchos cánceres de próstata con diagnóstico equivocado. En particular, el PSA elevado puede ser también el resultado de una inflamación o ecografía transrectal precedente, es decir, la invención dentro del estado de la técnica carece de un diagnóstico de PCa inequívoco.

Por lo tanto, es claro que ha existido y continúa hasta hoy una necesidad constante de marcadores de pronóstico confiables y precisos.

30 Los últimos años han llevado a una extensión marcada de nuestra comprensión sobre la base somática del cáncer de próstata. Con una a tres mutaciones por megabase, la frecuencia de mutación es similar a la que se observó en la leucemia mieloide aguda y el cáncer de mama invasivo (IBC, por su sigla en inglés) y yace dentro del intervalo inferior de las mutaciones del cáncer. De acuerdo con la frecuencia y el hecho de que principalmente un conjunto de genes diversos se ve afectado, las alteraciones genómicas principales parecen ser reestructuraciones y cambios genómicos en la estructura epigenética del ADN.

35 La metilación del ADN anormal cumple un papel importante en el desarrollo del cáncer de próstata y parece ser uno de los eventos más tempranos en la tumorigénesis. El gen metilado de manera diferenciada más prominente en el cáncer de próstata es glutatión-S-transferasa pi 1 (GSTP1).

40 Otros genes con cambios en la metilación del promotor incluyen la proteína resistente a múltiples fármacos 1 (MDR1), la O6-metilguanina-ADN metiltransferasa (MGMT), el miembro de la familia del dominio de asociación Ras 1 (RASSF1), el receptor de ácido retinoico beta (RARβ), la poliposis cólica adenomatosa (APC), el receptor androgénico (AR), los inhibidores de la cinasa dependientes de la ciclina 2A (CDKN2A), la cadherina-E (CDH1) y CD44, pero algunos de estos genes muestran niveles variables de metilación y a veces ninguna metilación del ADN en estudios diferentes.

45 El cáncer de mama (neoplasia maligna de mama) es el cáncer que se origina del tejido mamario, más comúnmente del revestimiento interno de los conductos lácteos o los lóbulos que suministran leche a los conductos. Los cánceres que se originan de los conductos se denominan carcinomas ductales y los que se originan de los lóbulos se denominan carcinomas lobulares.

50 El tamaño, la etapa, la velocidad de crecimiento y otras características del tumor determinan los tipos de tratamiento. El tratamiento puede incluir cirugía, fármacos (terapia hormonal y quimioterapia), radiación y/o inmunoterapia. La remoción quirúrgica del tumor proporciona el único beneficio más grande, siendo la cirugía sola capaz de producir una cura en muchos casos. Para aumentar de alguna forma la probabilidad de supervivencia sin enfermedad a largo plazo, se proporcionan comúnmente diversos regímenes de quimioterapia además de la cirugía. La mayoría de las formas de quimioterapia matan células que se dividen rápidamente en alguna parte del cuerpo, y como resultado
55 provocan la pérdida temporal de cabello y malestares digestivos, y cardiotoxicidad de forma ocasional. Puede

agregarse radiación para matar cualquier célula cancerosa en la mama que no hubiera sido detectada por la cirugía, lo cual extiende usualmente la supervivencia de alguna manera, si bien la exposición a la radiación para el corazón puede provocar insuficiencia cardíaca en el futuro. Algunos cánceres de mama son sensibles a las hormonas tales como el estrógeno y/o la progesterona, lo cual hace posible tratarlos bloqueando los efectos de estas hormonas, por ejemplo, a través del uso de antiestrógenos como los inhibidores de aromatasa o tamoxifeno que bloquean la síntesis de estrógeno del cuerpo.

El pronóstico y la tasa de supervivencia varían en mayor medida dependiendo del tipo de cáncer y la clasificación en estadios.

Además, para el IBC existe la necesidad de marcadores de pronóstico y diagnóstico confiables adicionales.

10 **Compendio de la invención**

La descripción suministra una solución para el problema mencionado anteriormente, al proporcionar un método para el diagnóstico y/o pronóstico del cáncer, que comprende las etapas de (a) analizar en una muestra de un sujeto el estado de metilación del ADN de una región genómica de al menos un miembro del grupo de (i) SFN según la SEQ ID NO. 1, (ii) SLIT2 según la SEQ ID NO. 2, (iii) SERPINB5 según la SEQ ID NO. 3 (iv) o TWIST1; en donde, si (i) SFN muestra un valor límite de metilación superior al 80 % y/o, (ii) SLIT2 muestra un valor límite de metilación superior al 45 % y/o, SERPINB5 muestra un valor límite de metilación superior al 70 %, la muestra se categoriza como una muestra de un paciente con cáncer con un pronóstico desalentador.

La descripción hace referencia también a ácidos nucleicos, composiciones y kits respectivos.

Definiciones

20 Las siguientes definiciones se proporcionan para términos específicos que se utilizan a continuación.

Como se utiliza en la presente, el término «amplificado», cuando se aplica a una secuencia de ácido nucleico, hace referencia a un proceso a través del cual se generan una o más copias de una secuencia de ácido nucleico particular a partir de una secuencia de plantilla de ácido nucleico, preferiblemente, mediante el método de reacción en cadena de la polimerasa. La mezcla de reacción comprende dNTP (cada uno de los cuatro desoxirribonucleótidos dATP, dCTP, dGTP y dTTP), cebadores, amortiguadores, polimerasa de ADN y modelo de ácido nucleico. La reacción PCR puede comprender (a) proporcionar un «par de cebadores» en donde un primer cebador contiene una secuencia complementaria a la cadena sentido de la secuencia de ácido nucleico diana y prepara la síntesis de una segunda cadena de ADN complementaria y un segundo cebador contiene una secuencia complementaria a la cadena antisentido de la secuencia de ácido nucleico y prepara la síntesis de una cadena de ADN complementaria a la cadena antisentido, y (b) amplificar la secuencia de plantilla de ácido nucleico que emplea una polimerasa de ácido nucleico. Generalmente, se utiliza una polimerasa Tag para amplificar una plantilla de ácido nucleico en la reacción PCR. Otros métodos de amplificación incluyen, de modo no taxativo, la reacción en cadena de la ligasa (LCR, por su sigla en inglés), la amplificación que se basa en polinucleótidos específicos o cualquier otro método conocido en la técnica.

35 Como se utiliza en la presente, la expresión «marcador biológico» hace referencia a (a) una región genómica que se somete a metilación de forma diferencial o (b) un gen que se expresa de forma diferencial, o (c) una mutación de una secuencia de ADN o polimorfismo de un solo nucleótido (SNP, por su sigla en inglés) que puede asociarse con sujetos que padecen cáncer o un estadio del cáncer en comparación con aquellos que no padecen cáncer.

40 Como se utiliza en la presente, el término «composición» hace referencia a cualquier mezcla. Puede ser una solución, una suspensión, líquido, polvo, una pasta, acuosa o no acuosa o cualquier combinación de estos.

La expresión «posición CpG», como se utiliza en la presente, hace referencia a las regiones del ADN donde se ubica un nucleótido de citosina en la posición adyacente 5' con respecto a un nucleótido de guanina en la secuencia lineal de bases a lo largo de su longitud. «CpG» es la abreviatura de «C-fosfato-G», es decir, la citosina y la guanina separadas por un fosfato, lo cual une a los dos nucleósidos en el ADN. Las citosinas en los dinucleótidos CpG pueden metilarse para formar la 5-metilcitosina.

Como se utiliza en la presente, el término «diagnóstico» hace referencia a la identificación del cáncer (por ejemplo, el cáncer de próstata (PCa) o el cáncer de mama invasivo (IBC) en cualquier etapa de su desarrollo e incluye también la determinación de la predisposición de un sujeto a desarrollar la enfermedad.

50 El término «pronóstico», como se utiliza en la presente, hace referencia a una predicción del curso y resultado probable de una enfermedad o afección. Un pronóstico de un paciente se realiza generalmente evaluando los factores, los marcadores y/o los síntomas de una enfermedad que son indicativos de un curso o resultado favorable o desfavorable de la enfermedad.

La frase «determinar el pronóstico», como se utiliza en la presente, hace referencia al proceso a través del cual el experto puede predecir el curso o resultado de una afección en un paciente. El término «pronóstico» no hace

referencia a la capacidad de predecir el curso o resultado de una afección con 100 % de precisión. En vez de eso, el experto comprenderá que el término «pronóstico» hace referencia a una mayor probabilidad de que se produzca un determinado curso o resultado, es decir, que es más probable que ocurra un curso o resultado en un paciente que muestra una afección dada, en comparación con aquellos individuos que no muestran la afección. Un pronóstico puede expresarse como la cantidad de tiempo que se espera que un paciente sobreviva. De manera alternativa, un pronóstico puede hacer referencia a la probabilidad de que la enfermedad entre en remisión o a la cantidad de tiempo que se puede esperar que la enfermedad permanezca en remisión. El pronóstico puede expresarse de diversas maneras, por ejemplo, el pronóstico puede expresarse como una probabilidad porcentual de que un paciente sobreviva un año, cinco años, diez años o similares. De manera alternativa, el pronóstico puede expresarse como la cantidad de años, en promedio, que se espera que un paciente pueda sobrevivir como resultado de una afección o enfermedad. El pronóstico de un paciente puede considerarse como una expresión de relativismo que tiene muchos factores que provocan el resultado final. Por ejemplo, para pacientes con determinadas afecciones, el pronóstico puede expresarse adecuadamente como la probabilidad de que una afección pueda ser tratada o curada o la probabilidad de que una afección entre en remisión, mientras que para los pacientes con pronóstico de afecciones más graves puede expresarse más adecuadamente como la probabilidad de supervivencia durante un período de tiempo específico.

Como se utiliza en la presente, la expresión «metilación diferencial» hace referencia a la diferencia en el nivel de metilación del ADN/citosina en una muestra de cáncer positiva en comparación con el nivel de metilación del ADN en una muestra de cáncer negativa. También puede hacer referencia a la diferencia en los niveles entre los pacientes que padecen de recidiva del cáncer luego de la cirugía en comparación con los pacientes que no sufren de recidiva. La metilación diferencial y los niveles específicos o patrones de la metilación del ADN pueden utilizarse como marcadores biológicos de pronóstico y predictivos una vez que se han definido las características predictivas o de corte correctas. El «estado de metilación del ADN» es intercambiable con la expresión «nivel de metilación del ADN» y puede ser evaluado al determinar la relación del ADN metilado y no metilado para una región genómica o una parte de esta y se expresa en porcentaje. El estado de metilación se clasifica en la presente ya sea como aumentado o disminuido y puede estar relacionado con una persona con recidiva del cáncer en comparación con una persona de control que sí experimentó una recidiva durante un período de observación similar. De manera alternativa, la metilación del ADN puede aumentar o disminuir en genes marcadores determinados en una persona con un cáncer de alto grado diagnosticado patológicamente que es sabido que tiene mayor riesgo de avance en comparación con una persona con un cáncer de bajo grado que es probable que tenga un resultado favorable.

En la presente, un «valor límite» se define de la siguiente manera: un nivel de metilación del ADN específico por encima del cual los resultados se consideran positivos (o negativos para un gen con una asociación inversa) comparado con cuando el nivel de metilación se encuentra por debajo del límite, los resultados se consideran negativos (o positivos para un gen con asociación inversa). Para explicar la variabilidad biológica que se sabe que es típica de todos los sistemas biológicos vivos tales como los organismos humanos y otros, es razonable considerar los intervalos de valores y, por lo tanto, todos los valores límites en la presente pueden variar en más menos 15 %, más menos 10 % o preferiblemente solo más menos 5 %. Esto depende además del montaje experimental.

La expresión «analizar el estado de metilación», como se utiliza en la presente, hace referencia a los medios y métodos útiles para evaluar el estado de metilación. Métodos útiles son los métodos a base de bisulfito, tales como la espectrometría de masas a base de bisulfito o los métodos de secuenciación a base de bisulfito.

La expresión «cebadores específicos de la región genómica», como se utiliza en la presente, hace referencia a un par de cebadores complementarios a una secuencia que se encuentra cerca de una región genómica según la invención, que puede producirse mediante los métodos de amplificación de ADN de cadena doble complementaria a una región genómica de la invención.

La expresión «sonda específica de la región genómica», como se utiliza en la presente, hace referencia a una sonda que se hibrida selectivamente con un producto de ADN de una región genómica. En una realización, una sonda específica de la región genómica puede ser una sonda etiquetada, por ejemplo, con un fluoróforo y un inactivador tal como una sonda TaqMan® o sondas de balizas moleculares.

Como se utiliza en la presente, las expresiones «que se hibrida con» e «hibridación» se usan de manera intercambiable con la expresión «específico para» y hacen referencia a las interacciones de unión no covalente específicas para la secuencia con un ácido nucleico complementario, por ejemplo, interacciones entre una secuencia de ácido nucleico diana y un cebador o sonda de ácido nucleico específico diana. En una realización preferida, un ácido nucleico, que hibrida, es uno que se hibrida con una selectividad superior al 70 %, superior al 80 %, superior al 90 % y lo más preferiblemente de 100 % (es decir, la hibridación cruzada con otra especie de ADN se produce preferiblemente en menos del 30 %, menos del 20 %, menos del 10 %). Como un experto en la técnica lo comprendería, es posible determinar un ácido nucleico que se «hibrida» con un producto de ADN de una región genómica de la invención, tomando en cuenta la longitud y la composición.

Como se utiliza en la presente, «aislado/a», cuando se utiliza en referencia a un ácido nucleico significa que una secuencia de origen natural ha sido eliminada de su entorno celular (p. ej., cromosómico) normal o se sintetizó en un entorno no natural (p. ej., se sintetizó de forma artificial). Por lo tanto, una secuencia «aislada» puede estar en una

solución libre de células o puede colocarse en un entorno celular diferente.

Como se utiliza en la presente, un «kit» es una combinación de paquetes que incluye opcionalmente instrucciones para usar la combinación y/u otras reacciones y componentes para dicho uso.

5 Como se utiliza en la presente, «ácido/s nucleico/s» o «molécula de ácido nucleico» hace referencia generalmente a cualquier ácido ribonucleico o ácido desoxirribonucleico que puede ser ADN o ARN modificado o no modificado. Los «ácidos nucleicos» incluyen, de modo no taxativo, los ácidos nucleicos de cadena simple y doble. Como se utiliza en la presente, la expresión «ácido/s nucleico/s» incluye también ADN como se describió anteriormente que contiene una o más bases modificadas. Por lo tanto, el ADN con estructuras modificadas en busca de estabilidad u otros motivos son «ácidos nucleicos». La expresión «ácido/s nucleico/s», como se utiliza en la presente, comprende
10 dichas formas químicas, enzimáticas o metabólicamente modificadas de ácidos nucleicos, así como también las formas químicas del ADN características de virus y células, incluso, por ejemplo, de células simples y complejas.

El término «cebador», como se utiliza en la presente, se refiere a un ácido nucleico, ya sea de origen natural como en una digestión restrictiva purificada o producido sintéticamente, que es capaz de actuar como punto de inicio de la síntesis cuando se lo coloca en condiciones en las cuales se induce la síntesis de un producto de extensión de cebadores que es complementario a una cadena de ácido nucleico, es decir, en presencia de nucleótidos y un agente de inducción tal como la polimerasa de ADN y a una temperatura y pH adecuados. El cebador puede ser de cadena simple o cadena doble y deber ser lo suficientemente largo para preparar la síntesis del producto de extensión deseado en presencia del agente de inducción. Las longitudes exactas del cebador dependerán de muchos factores, que incluyen la temperatura, la fuente del cebador y el método utilizado. Por ejemplo, para
15 aplicaciones de diagnóstico, dependiendo de la complejidad de la secuencia diana, el cebador de ácido nucleico contiene típicamente entre 15 y 25 nucleótidos o más, aunque este puede contener menos nucleótidos.

Como se utiliza en la presente, el término «sonda» significa ácido nucleico y análogos de este, y hace referencia a un intervalo de especies químicas que reconocen las secuencias diana de polinucleótidos a través de interacciones de enlace de hidrógeno con las bases de nucleótidos de las secuencias diana. La sonda o las secuencias diana pueden ser de ADN de cadena simple o doble. Una sonda tiene una longitud de al menos 8 nucleótidos y menos que la longitud de una secuencia diana de polinucleótidos completa. Una sonda puede tener una longitud de 10, 20, 30, 50, 75, 100, 150, 200, 250, 400, 500 y hasta 10 000 nucleótidos. Las sondas pueden incluir ácidos nucleicos modificados a efectos de obtener una o más etiquetas que puedan ser detectadas mediante fluorescencia, quimioluminiscencia y similares («sonda etiquetada»). La sonda etiquetada puede modificarse también para que
25 tenga una o más etiquetas modificables y una o más moléculas inactivadoras, por ejemplo, las sondas Taqman® y Molecular Beacon®. El ácido nucleico y sus análogos pueden ser ADN o análogos del ADN, denominados comúnmente oligómeros antisentido o ácido nucleico antisentido. Dichos análogos de ADN comprenden, de modo no taxativo, modificaciones de azúcar 2'-O-alquilo, metilfosfonato, fosforotioato, fosforoditioato, formacetal, 3'-tioformacetal, sulfona, sulfamato y modificaciones a la estructura principal del nitrógeno y análogos en donde se han modificado los restos de base. Además, los análogos de oligómeros pueden ser polímeros en los que el resto de azúcar ha sido modificado o reemplazado por otro resto adecuado, lo que cual da como resultado polímeros que incluyen, de modo no taxativo, análogos de morfolino y análogos de ácido nucleico peptídico (PNA, por su sigla en inglés) (Egholm, et al. Peptide Nucleic Acids (PNA)- Oligonucleotide Analogues with an Achiral Peptide Backbone, (1992)).

40 El término «muestra» se utiliza en la presente para hacer referencia a un tejido propiamente dicho, tejido canceroso, posible tejido canceroso, tejido mamario o prostático, sangre, orina, semen, secreciones prostáticas, leche, exudados mamarios, punciones con aguja o células mamarias o prostáticas aisladas, células que provienen de un sujeto, preferiblemente, del tejido prostático, tejido mamario, secreciones prostáticas, secreciones mamarias o células mamarias o células prostáticas aisladas, lo más preferiblemente, tejido prostático o tejido mamario.

45 La expresión «secuenciación con bisulfito» hace referencia a un método conocido por el experto en la técnica que comprende las etapas de: (a) tratar el ADN de interés con bisulfito, convirtiendo así las citosinas no metiladas en uracilos y dejando intactas las citosinas metiladas y (b) secuenciar el ADN tratado, en donde se revela la existencia de una citosina metilada mediante la detección de una citosina no convertida y se revela la ausencia de una citosina metilada mediante la detección de un uracilo.

50 Como se utiliza en la presente, los expertos en la técnica conocen las «condiciones rigurosas para la hibridación» y se pueden encontrar en *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y., 6.3.1- 6.3.6, 1991. Las condiciones rigurosas se definen como equivalentes a la hibridación en 6X de cloruro de sodio/citrato de sodio (SSC) a 45 °C, seguido de un lavado en 0,2 x SSC, 0,1 % de SDS a 65 °C.

55 Como se utiliza en la presente, los términos «sujeto» y «paciente» se usan de manera intercambiable para hacer referencia a un ser humano o animal (p. ej., un mamífero, un pez, un anfibio, un reptil, un ave y un insecto). En una realización específica, un sujeto es un mamífero (p. ej., un mamífero no humano y un ser humano). En otra realización, un sujeto es una mascota (p. ej., un perro, un gato, un cobayo, un mono y un ave), un animal de granja (p. ej., un caballo, una vaca, un cerdo, una cabra y una gallina) o un animal de laboratorio (p. ej., un ratón y una rata). En otra realización, un sujeto es un primate (p. ej., un chimpancé y un ser humano). En otra realización, un

sujeto es un ser humano. En otra realización, el sujeto es un ser humano de sexo masculino o un ser humano de sexo femenino.

5 Como se utiliza en la presente, la expresión «en las proximidades de una región genómica» hace referencia a una posición fuera o dentro de dicha región genómica. Como un experto en la técnica lo entendería, la posición puede tener una distancia de hasta 1000 nt, preferiblemente hasta 500 nt, más preferiblemente hasta 200 nt del extremo 5' o 3' de la región genómica. Incluso más preferiblemente, la posición se ubica en el extremo 5' o 3' de dicha región genómica. En otra realización de la invención, la posición se encuentra dentro de dicha región genómica.

Descripción detallada de la invención

10 La invención como se describe en la presente identifica las regiones genómicas que son útiles para diagnosticar un cáncer agresivo. Estas son también marcadores de pronóstico.

15 Por definición, las regiones genómicas identificadas son marcadores biológicos para el cáncer agresivo. A efectos de usar estas regiones genómicas (como marcadores biológicos), la invención enseña el análisis del estado de metilación del ADN de dichas regiones genómicas. La invención abarca adicionalmente los ácidos nucleicos específicos de la región genómica. La invención contempla además el uso de dichos ácidos nucleicos específicos de la región genómica para analizar el estado de metilación de una región genómica, ya sea directa o indirectamente mediante métodos que el experto conoce y que se explican en la presente. La invención describe también una composición y kit que comprende dichos ácidos nucleicos para el diagnóstico del PCa y IBC.

20 Los inventores descubrieron regiones genómicas que están sujetas a un estado de metilación anómalo. Se descubrieron asociaciones tumorales. Por lo tanto, la invención enseña el análisis de estas regiones genómicas que se metilan de manera diferencial en muestras de pacientes que padecen cáncer. Mejor que los métodos de diagnóstico actuales, la invención describe regiones genómicas, en donde de forma más sorprendente una combinación de hasta cuatro regiones genómicas funciona muy bien.

25 Por ende, la descripción hace referencia a un método para diagnosticar y/o pronosticar un cáncer que comprende las etapas de: (a) analizar, en una muestra de un sujeto, el estado de metilación del ADN de una región genómica de al menos un miembro del grupo de (i) SFN según la SEQ ID NO. 1, (ii) SLIT2 según la SEQ ID NO. 2, (iii) SERPINB5 según la SEQ ID NO. 3; en donde, si (i) SFN muestra un valor límite de metilación por encima del 80 % y/o, (ii) SLIT2 muestra un valor límite de metilación por encima del 45 % y/o, SERPINB5 muestra un valor límite de metilación por encima del 70 %, la muestra se categoriza como una muestra de un paciente con un cáncer de pronóstico desalentador.

Nombre de la región	<i>Regiones genómicas desde el extremo 5' del cebador F al extremo 5' del cebador R</i>
SFN (SEQ ID NO. 1)	GGAGAAGGTGGAGACTGAGCTCCAGGGCGTGTGCGACACCG TGCTGGGCCTGCTGGACAGCCACCTCATCAAGGAGGCCGGG
SLIT2 (SEQ ID NO. 2)	AGATCCCCTCTTCTGTCTTGTACCTTCGCCACTGGCATCGG ATTTGCAGAAGCGTGCGTGGGATCAGAGACCGCCCTCCCC
SERPINB5 (SEQ ID NO. 3)	CAAGAGGCTTGAGTAGGAGAGGAGTGCCGCCGAGGCGGGGC GGGGCGGGGCGTGGAGCTGGGCTGGCAGTG
TWIST 1 (SEQ ID NO. 4)	TCCTCCTGCTCTCTCCTCCGCGGGCCGCATGCCCCGGGCCG GCGCCGCGCGGGGGAAGCTGGCGGGCTGAGGCGCCCCGC

Los números de acceso se muestran a continuación:

Gen	Número de acceso NCBI Secuencia de referencia	Número de acceso UCSC clasificador de gen humano
SFN	NM 006142 (ARNm)	ucOOlbnc.1
SLIT2	NT 006316.16	uc003gpr.1
SERPINB5	NT 025028.14	uc0021iz.3
TWIST 1	NM 000474,3 (ARNm)	uc003sum.2

Por consiguiente, la invención se refiere a un método para diagnosticar el cáncer de próstata (PCa).

5 La invención se refiere a un método para diagnosticar el cáncer de próstata que comprende las etapas de analizar, en una muestra de un sujeto, el estado de metilación del ADN de las siguientes regiones genómicas, (i) SFN según la SEQ ID NO. 1, (ii) SLIT2 según la SEQ ID NO. 2 y (iii) SERPINB5 según la SEQ ID NO. 3, en donde, si (i) SFN muestra un valor límite de metilación superior al 80 % y/o, (ii) SLIT2 muestra un valor límite de metilación superior al 45 % y/o (iii) SERPINB5 muestra un valor límite de metilación superior al 70 %, y/o (iv) TWIS1 muestra un valor límite de metilación inferior al 15 %, preferiblemente inferior al 10 %, y lo más preferiblemente inferior al 5 %, la muestra se categoriza como una muestra de un paciente que padece cáncer o con un diagnóstico desalentador. En esta realización, se analizan las cuatro regiones en sus límites respectivos y se escala la metilación y los datos se utilizan en una fórmula de riesgo aditiva para producir un punto de corte clínico que se utiliza para categorizar una muestra como de pronóstico desalentador o no.

10 En una realización preferida de la descripción, el diagnóstico de cáncer se produce antes de la manifestación de los síntomas. Los sujetos con un mayor riesgo de desarrollar una enfermedad agresiva son motivo de especial preocupación. El método de diagnóstico de la descripción permite también confirmar el cáncer en un sujeto que se sospecha que padece cáncer o cáncer agresivo.

15 El método de la descripción es particularmente útil para el diagnóstico temprano del cáncer de próstata (PCa) o el cáncer de mama (IBC). El método es útil para diagnosticar también pacientes que tienen una masa prostática identificada o síntomas que se asocian con el cáncer de próstata, p. ej., niveles de PSA anormalmente altos. El método es útil además como una prueba de seguimiento para mujeres que han sido diagnosticadas con cáncer de mama o quienes tienen mamografías anormales y puede proporcionar biopsias, tejidos, aspirado u otros fluidos provenientes de tejidos para evaluar el estado de metilación de los genes indicados. El método de la presente descripción puede ser de uso específico con pacientes que posean mayor riesgo de desarrollar cáncer de mama o próstata (p. ej., pacientes que tienen antecedentes familiares de cáncer de mama o próstata identificados por tener oncogenes mutantes u otros factores de riesgo). El método de la presente descripción puede ser de especial utilidad para monitorear la eficacia del tratamiento de un cáncer de mama o próstata (p. ej. la eficacia de la quimioterapia).

20 En una realización del método de la descripción, la muestra comprende células obtenidas de un paciente. Las células pueden encontrarse en una muestra de tejido prostático o una muestra mamaria recolectada, por ejemplo, por medio de una biopsia de tejido o corte histológico. En otra realización, la muestra del paciente es un fluido corporal que se asocia con las mamas o la próstata. Dichos fluidos incluyen, por ejemplo, fluidos sanguíneos, linfáticos, de orina, fluido prostático, semen, aspirado de mamas, o exudados, o leche y pueden incluir células cancerosas aisladas separadas de las muestras clínicas humanas heterogéneas mediante el uso de métodos de separación y purificación tales como las selecciones inmunitarias, clasificación de flujo u otros métodos para enriquecer las células cancerosas deseadas. El ADN puede aislarse de la muestra por medio de la disrupción celular o lisis celular mediante sonicación y/o digestión enzimática, tratamiento con detergentes para quitar los lípidos de membrana y precipitación del ADN usando alcohol. El ADN sin células puede recolectarse también y purificarse a partir de los fluidos que contienen pocas células, si es que las hay, tal como, por ejemplo, el suero. Luego el ADN puede ser reenviado al método de análisis.

30 A efectos de analizar el estado de nivel de metilación de una región genómica, se pueden utilizar tecnologías cuantitativas o semicuantitativas convencionales.

En primer lugar, el ADN de interés extraído puede enriquecerse, por ejemplo, mediante la inmunoprecipitación del ADN metilado (MeDIP, por su sigla en inglés). Luego, el estado de metilación del ADN puede analizarse ya sea directamente o luego del tratamiento con bisulfito.

45 En una realización, se utilizan enfoques basados en bisulfito para preservar la información de metilación. Por lo tanto, el ADN se trata con bisulfito, convirtiendo así los residuos de citosina no metilada en uracilo mientras que las citosinas metiladas permanecen intactas. La conversión selectiva hace que la metilación se pueda detectar y cuantificar fácilmente mediante métodos clásicos que revelan la existencia o ausencia de la metilación del ADN (citosina) del ADN de interés. El ADN de interés puede amplificarse antes de la detección si fuera necesario. Dicha

5 detección puede realizarse mediante espectrometría de masas. Preferiblemente, se secuenciar el ADN de interés. Los métodos de secuenciación adecuados son la secuenciación directa y la pirosecuenciación. En otra realización de la invención, el ADN de interés se detecta mediante una sonda específica para la región genómica que es cuantitativamente selectiva para la secuencia en la cual una citosina se convirtió o no se convirtió. Otras técnicas
 10 que se pueden aplicar luego del tratamiento con bisulfito son el análisis de conformación de cadena simple sensible a la metilación (MS-SSCA, por su sigla en inglés), el análisis de fusión de alta resolución (HRM, por su sigla en inglés), la extensión de cebadores de un solo nucleótido sensibles a la metilación (MS-SnuPE, por su sigla en inglés) y la escisión específica de bases y los métodos epiTYPER. Un método adicional se muestra en «*Sensitive digital quantification of DNA methylation in clinical samples*» Nat Biotechnol. Setiembre de 2009; 27(9): 858-863. A este se le denomina también *Methyl-BEAMing* (bolillas de metal, emulsión, amplificación de ADN y magnetismo de metilo).

En una realización alternativa, el estado de metilación del ADN se analiza sin tratamiento con bisulfito, tal como la escisión por medio de enzimas que son sensibles a los niveles de metilación del ADN seguido de la PCR específica para la metilación o mediante el uso de una sonda específica de la región genómica que pueden seleccionarse para esa secuencia en la cual una citosina está metilada o no metilada como se indica mediante la escisión.

15 Para traducir los datos sin procesar que se generaron mediante el ensayo de detección (p. ej., una secuencia de nucleótidos) en datos de valor predictivo para un médico clínico, se puede utilizar un programa de análisis de riesgo informático.

Los datos de descripción pueden prepararse en un formato adecuado para que un médico tratante los interprete. Por ejemplo, en vez de proporcionar datos de secuencia de nucleótidos sin procesar o estado de metilación, el formato
 20 preparado puede representar un diagnóstico o evaluación de riesgo (p. ej., la probabilidad de que se encuentre presente un cáncer o el subtipo de cáncer) para el sujeto, junto con las recomendaciones para las opciones de tratamiento particulares.

En algunas realizaciones, los resultados se utilizan en un entorno clínico para determinar los procedimientos a seguir. En otras realizaciones, los resultados se utilizan para determinar los procedimientos a seguir para el
 25 tratamiento (p. ej., la elección de terapias o la observación atenta).

Preferiblemente, se analiza el estado de metilación de una región genómica adicional y/o un marcador biológico adicional que puede seleccionarse del grupo de:

- APC según la SEQ ID NO. 5.
- HLAa según la SEQ ID NO. 6 y/o NKX2-5 según la
 30 SEQ ID NO. 7

La muestra se categoriza como se describió anteriormente y los límites son:

- 3 % para APC según la SEQ ID NO. 5,
- 35 % para HLAa según la SEQ ID NO. 6 y/o 5 % para NKX2-5 según la SEQ ID NO. 7

35 Los pacientes con muestras que tienen porcentaje de metilación por encima de los límites para la APC y HLAa se considera que tienen mayor riesgo de desarrollar cáncer agresivo y, por otra parte, se considera que los pacientes con porcentajes de metilación por debajo del límite de NKX2-5 tienen mayor riesgo de desarrollar un cáncer agresivo.

APC (SEQ ID NO. 5)	GGGCTAGGGCTAGGCAGGCTGTGCGGTTGGGCGGGGCCCTG TGCCCCACTGCGGAGTGCGGGTCGGGAAGCGGAGAGAGAAG CAGCTGTGTAATC
HLAa (SEQ ID NO. 6)	GGGCCCTGGCCCTGACCCAGACCTGGGCGGGTGAGTGCGGG GTCGGGAGGGAAACCGCCTCTGCGGGGAGAAGCAAGGGGCC CTCCTG
NKX2-5 (SEQ ID NO. 7)	CCTTCTCAGTCAAAGACATCCTAACCTGGAACAGCAGCAG CGCAGCCTGGCTGCCGCCGAGAGCTCTTGCCCGCCTGGA GGCGACCCTGGCGCCCTCCTCCTGCATGCTGGCC

Los números de acceso se muestran a continuación:

Gen	Número de acceso NCBI Secuencia de referencia	Número de acceso UCSC clasificador de gen humano
APC	NT 034772.6	uc003kpy.3
HLAa	NT 113891.2	uc003nol.2
NKX2- 5	NM 001166176,1 (ARNm)	uc003mcm.1

5 La invención se refiere preferiblemente a cualquier combinación de regiones según las SEQ ID NO. 1, 2 y 3, y las SEQ ID NO. 4, 5, 6 y/o 7. Por ejemplo, las SEQ ID NO. 1, 2, 3 y 4; las SEQ ID NO. 1, 2, 3 y 5; las SEQ ID NO. 1, 2, 3 y 6; las SEQ ID NO. 1, 2, 3 y 7; las SEQ ID NO. 1, 2, 3, 4 y 5; las SEQ ID NO. 1, 2, 3, 4 y 6; y se puede utilizar cualquier otra combinación, siempre que las SEQ ID NO. 1, 2, 3 y 4 estén presentes.

10 Analizar el estado de metilación de una región genómica significa analizar el estado de metilación de al menos una posición CpG por región genómica.

Los inventores descubrieron sorprendentemente que el estado de metilación dentro de una región genómica según la invención es casi constante, lo que conduce a niveles de metilación de distribución uniforme dentro de dicha región genómica. En una realización de la invención, se analizan todas las posiciones CpG de una región genómica. En una realización específica, es posible analizar las posiciones CpG en las proximidades de la región genómica.

15 En una realización alternativa, se analiza un subconjunto de posiciones CpG de una región genómica. Idealmente, se analizan 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 posiciones CpG de una región genómica. Por lo tanto, una realización preferida de la invención se refiere a un método, en donde analizar el estado de metilación de una región genómica significa analizar el estado de metilación de al menos una posición CpG por región genómica.

20 De manera significativa, los inventores descubrieron que un mínimo de una región genómica es suficiente para discriminar de forma adecuada entre maligno agresivo y maligno no agresivo, así como también se pueden utilizar tejidos benignos en el pronóstico. La extensión con regiones genómicas adicionales aumenta incluso el potencial discriminatorio del conjunto de marcadores. Por ende, en otra realización, la invención hace referencia a un método, en donde se analiza el estado de metilación de una región genómica adicional y/o un marcador biológico adicional.

25 En una realización de la invención, se analiza adicionalmente un marcador biológico conocido. Para el PCa estos pueden ser, p. ej., GSTP1, la proteína resistente a múltiples fármacos 1 (MDR1), la O-6-metilguanina-ADN metiltransferasa (MGMT), el miembro de la familia del dominio de asociación Ras 1 (RASSF1), el receptor de ácido retinoico beta (RARβ), la poliposis cólica adenomatosa (APC), el receptor androgénico (AR), los inhibidores de la

cinasa dependientes de la ciclina 2A (CDKN2A), la cadherina-E (CDH1) y/o CD44. Tales marcadores biológicos pueden basarse también en la expresión génica, p. ej., de dichos genes codificadores. En una realización preferida, la concentración o actividad del antígeno prostático específico (PSA) se determina por medio de un inmunoensayo. El análisis de los marcadores biológicos dentro de este contexto puede ser el análisis del estado de metilación, el análisis de la expresión génica (ARNm) o el análisis de la cantidad o concentración o actividad de la proteína.

Para el cáncer de mama estos pueden ser típicamente marcadores mamarios tales como ER, PR, HER2, marcadores de proliferación tales como Ki67 o paneles de ARN tales como OncotypeDx, Mammaprint o PAM-50.

En otra realización, se analiza una región genómica adicional según la invención.

El estado de metilación se analiza, por ejemplo, mediante métodos que se basan en la PCR específica para la no metilación, métodos que se basan en la metilación o métodos que se basan en micromatrices.

En una realización preferida de la invención, se puede utilizar la pirosecuenciación para el análisis del estado de metilación. Pueden preferirse también los métodos epiTYPER o HRM ya que son relativamente simples y cuantitativos.

Como se mencionó anteriormente, el análisis del estado de metilación del ADN puede requerir cebadores específicos de la región genómica para la amplificación de una secuencia en las proximidades de una región genómica. Asimismo, pueden ser necesarias las sondas específicas de la región genómica con el fin de detectar el estado de metilación directa o indirectamente. Se pueden utilizar dichos ácidos nucleicos en técnicas tales como la PCR en tiempo real cuantitativa que usa, por ejemplo, SYBR®Green o que utiliza las técnicas TaqMan® o Molecular Beacon, en donde los ácidos nucleicos se usan en forma de cebadores específicos de la región genómica o sondas específicas de la región genómica, tal como una sonda etiquetada TagMan o una sonda etiquetada Molecular Beacon. Dentro del contexto de la invención, el ácido nucleico se hibrida selectivamente en las proximidades de la región genómica como se definió anteriormente. Lo más preferiblemente, este se hibrida selectivamente dentro de la región genómica de interés. En una realización de la invención, se utiliza un único ácido nucleico específico de la región genómica. En una realización preferida, los ácidos nucleicos se utilizan como un par, en donde el primer ácido nucleico es específico para la cadena antisentido y el segundo ácido nucleico para la cadena antisentido de la secuencia de ADN en las proximidades de la región genómica.

Por lo tanto, la descripción se refiere a una molécula de ácido nucleico que se hibrida en condiciones rigurosas en las proximidades de una de las regiones genómicas según la SEQ ID NO. 1 a la SEQ ID NO: 7, en donde dicha proximidad se refiere a una posición como se define en la presente.

En una realización, dicho ácido nucleico tiene una longitud de 15 a 100 nt. En una realización preferida, dicho ácido nucleico tiene una longitud de 15 a 50 nt, en una realización más preferida, de 15 a 40 nt.

En una realización, dicho ácido nucleico es un cebador.

Dentro del contexto de la invención, los cebadores específicos de la región genómica se utilizan para amplificar selectivamente la secuencia de ADN en las proximidades de una de las regiones genómicas mediante PCR como parte del método de análisis del estado de metilación del ADN. Por lo tanto, en una realización de la presente invención, el cebador es específico para una de las regiones genómicas y se hibrida en las proximidades dentro del significado que se especifica en la presente.

El estado de metilación de una región genómica puede detectarse directa o indirectamente mediante el uso de una sonda específica para la región genómica. Por ende, en una realización de la presente invención, dicho ácido nucleico es una sonda que se hibrida en las proximidades de dicha región. La sonda puede ser específica para el sitio de metilación.

En una realización preferida de la presente invención, se etiqueta la sonda.

Los métodos actuales para el análisis del estado de metilación requieren de un tratamiento con bisulfito a priori, convirtiendo así las citosinas no metiladas en uracilos. Para asegurar la hibridación del ácido nucleico específico de la región genómica de la invención en el ADN tratado con bisulfito, es posible adaptar la secuencia de nucleótidos del ácido nucleico. Por ejemplo, si se desea diseñar ácidos nucleicos que sean específicos para una secuencia, en donde se considera que una citosina se metila de forma diferencial, ese ácido nucleico específico de la región genómica puede tener dos secuencias: la primera que contiene una adenina, la segunda que contiene una guanina en esa posición que es complementaria al nucleótido de citosina en la secuencia de la región genómica. Las dos formas pueden utilizarse en un ensayo para analizar el estado de metilación de una región genómica de modo tal que estas sean capaces de discriminar entre citosinas metiladas y no metiladas. Dependiendo del método de análisis y el tipo de ácido nucleico (cebador/sonda) se puede utilizar solamente una forma o ambas formas del ácido nucleico específico de la región genómica dentro del ensayo. Por lo tanto, en una realización alternativa de la presente invención, el ácido nucleico se hibrida en condiciones rigurosas en tal proximidad de una de las regiones genómicas luego del tratamiento con bisulfito. Por lo tanto, este puede hibridarse en una posición CpG luego del tratamiento con bisulfito. En una alternativa a la versión metilada en la otra alternativa a la versión no metilada de

dicha posición.

Los medios y métodos de la presente invención comprenden el uso de ácidos nucleicos específicos de la región genómica para el diagnóstico y/o pronóstico de cáncer agresivo, preferiblemente PCa o IBC.

5 El análisis puede realizarse también mediante secuenciación. Históricamente han existido dos enfoques exitosos para la determinación de secuencias de ADN: el método enzimático de terminación de cadena, p. ej., Sanger et al, Proc. Nati. Acad. Sci., 74: 5463-5467 (1977); y el método de degradación química, p. ej., Maxam et al, Proc. Nati. Acad. Sci., 74: 560-564 (1977). Sin embargo, el deseo de mayores rendimientos y métodos de secuenciación más económicos ha llevado a una cantidad de métodos de secuenciación de última generación los cuales se examinan, por ejemplo, en Metzker, *Genome Research* 15:1767-1776 (2005) y *Next-Genration Genome Sequencing*, editado por M. Janitz, Wiley- VCH Verlag, Weinheim/Alemania, 2008.

10 Algunos de los métodos más recientes se basan en un enfoque de «secuenciación mediante síntesis». Dichos métodos incluyen, por ejemplo, la tecnología de pirosecuenciación ("454 Sequencing™", Roche Diagnostics) que se basa en la liberación de pirofosfato, su conversión en ATP y la producción de luz visible mediante luciferasa de luciérnagas (EP 0 932 700 B1; Ronaghi, *Genome Research* 11:3-11 (2001)). Se puede utilizar.

15 El método SOLiDTM («Secuenciación por ligación y detección de oligonucleótidos») (Life Technologies; WO 06/084132 A2) se basa en el acoplamiento de fragmentos amplificados con PCR de ácidos nucleicos modelo mediante secuencias de adaptadores universales a perlas magnéticas y la posterior detección de secuencias de fragmento por ligación de sondas etiquetadas a cebadores hibridados con las secuencias de adaptadores. Para la lectura se utiliza un conjunto de cuatro sondas de dos bases etiquetadas de forma fluorescente. Luego de la lectura, las partes de las sondas se escinden y se llevan a cabo nuevos ciclos de ligación, detección y escisión. Debido al uso de sondas de dos bases, se deben realizar dos rondas de secuenciación para cada secuencia modelo. Esto también puede utilizarse y se prefiere en la presente.

Otro método de secuenciación basado en la ligación se denomina CycLiC (Ligación cíclica y escisión). CycLiC utiliza bibliotecas de oligonucleótidos en las cuales se degeneran todos salvo un nucleótido.

25 El método implica ciclos de extensión de cebadores iterativos, el crecimiento de cadena base por base mediante etapas de detección y ligación sucesiva utilizando oligonucleótidos etiquetados (Mir et al, *Nucleic Acids Research* 37(1), e5 (2009)). Este también se puede utilizar.

30 El método de secuenciación Illumina Solexa® se basa en la química de secuenciación por síntesis (Bentley, *Curr Opin Genet Dev.* 16(6):545-552 (2006)). Se generan grandes cantidades de «colonias» únicas (colonias generadas por la polimerasa) que pueden secuenciarse de forma simultánea. Estas reacciones paralelas se producen en la superficie de una «celda de flujo» que proporciona una gran área de superficie para muchos miles de reacciones químicas paralelas. Esto se prefiere también.

35 WO 2007/133831 describe un método para secuenciar un ácido nucleico que comprende la etapa de ligación y una biblioteca de adaptadores. El modelo de secuenciación es una repetición concatémica grande de una única secuencia. Esto es una opción.

US-A1 6 013 445 hace referencia a un método de secuenciación adicional.

El ácido nucleico para llevar a cabo el método según la invención se formula de forma ventajosa en una composición estable. Por consiguiente, la presente invención se refiere a una composición para el diagnóstico y/o pronóstico del cáncer agresivo que comprende dicho ácido nucleico.

40 La composición puede incluir también otras sustancias, por ejemplo, estabilizadores, p. ej., EDTA, portadores de ácido nucleico protectores, etc.

La descripción abarca también un kit para diagnosticar y/o pronosticar un cáncer agresivo preferiblemente PCa o IBC, que comprende el ácido nucleico inventivo como se describió anteriormente.

45 El kit puede incluir un recipiente para un primer conjunto de cebadores específicos de la región genómica. En una realización preferida, el kit puede incluir un recipiente para un segundo conjunto de cebadores específicos de la región genómica. En una realización adicional, el kit puede incluir además un recipiente para un tercer conjunto de cebadores específicos de la región genómica. En una realización adicional, el kit puede incluir además un recipiente para un cuarto conjunto de cebadores específicos de la región genómica y así sucesivamente.

50 El kit puede incluir también un recipiente para el bisulfito, que se puede utilizar para tratar la región genómica de interés con bisulfito.

El kit puede incluir además sondas específicas de la región genómica.

El kit puede incluir recipientes de sustancias para realizar una reacción de amplificación, tal como recipientes que comprenden dNTP (cada uno de los cuatro desoxirribonucleótidos dATP, dCTP, dGTP y dTTP), amortiguadores y

polimerasa de ADN.

El kit puede incluir también modelo/s de ácido nucleico para una reacción de control positivo y/o control negativo y viales con diferentes concentraciones de ácidos nucleicos conocidos para posibilitar la construcción de una curva de respuesta según la dosis a efectos de permitir la cuantificación precisa de los objetivos. En una realización, se utiliza una polimerasa para amplificar un modelo de ácido nucleico en la reacción PCR. Otros métodos de amplificación incluyen, de modo no taxativo, la reacción en cadena de la ligasa (LCR, por su sigla en inglés), la amplificación basada en polinucleótidos específicos (NSBA, por su sigla en inglés) o cualquier otro método conocido en la técnica.

El kit puede incluir también recipientes de sustancias para realizar una reacción de secuenciación, por ejemplo, la pirosecuenciación, tal como la polimerasa de ADN, la ATP sulfurilasa, luciferasa, apirasa, los cuatro desoxinucleótidos trifosfatos (dNTP) y, p. ej., luceferina.

Tabla 1: Análisis de supervivencia de una sola variable de los cuatro genes de riesgo y grupos de riesgo principales que provienen de la dicotomización en los límites, como se define en la presente, en grupos de metilación baja o alta; la cantidad total de mujeres se indica por medio de la N y la cantidad de mujeres con recidiva de cáncer de mama se muestra como (Eventos). Los valores p se basan en la prueba de rango logarítmico (todos los valores p indicados resultaron significativos con un alfa inferior a 0,05).

Gen	Metilación baja N (Eventos)	Metilación alta N (Eventos)	N faltante (Eventos)	Valor p
SFN	87(12)	21 (8)	15(5)	0,007
TWIST1	45(15)	55 (9)	10(1)	0,024
SERPINB5	67 (9)	42 (13)	12(3)	0,024
SLIT2	107(19)	12(6)	2(0)	0,005

Leyendas de las figuras

Fig. 1:

Valores de metilación de los casos con puntuación de Gleason alta (> 7, triángulo) y Gleason baja (<= 7, círculo) del PCa para tres pares de genes. Las líneas de separación muestran los límites de discriminación lineal que se ajustaron para clasificar la puntuación de Gleason baja y alta. El objetivo era separar estas dos categorías lo más posible. Los genes graficados son A) SFN y SERPINB5, B) SFN y SLIT2, C) SERPINB5 y SLIT2. Una puntuación de Gleason alta en patología (se considera como una puntuación de 8 o mayor) es un marcador sustituto para un cáncer de próstata agresivo.

Fig. 2: (ejemplo comparativo)

Una gráfica de caja y bigotes de 1 a 10 genes metilados de forma diferencial que muestra el intervalo de porcentajes de metilación en un estudio de 121 mujeres con cáncer tratado quirúrgicamente a quienes se les hizo un seguimiento durante al menos 5 años y en quienes se registraron posteriormente eventos (E) de recidiva invasiva o sin recidiva (N) de los cánceres de mama tratados originalmente. Cada caja vertical ligeramente sombreada muestra los percentiles 25 y 75 del porcentaje de metilación del ADN de los genes nombrados para los no eventos N y, de manera similar, la caja con sombra oscura muestra los porcentajes de metilación del ADN entre los eventos E. Las líneas horizontales gruesas en la mitad de las cajas muestran el 50 % de los valores. Los percentiles 5 y 95 se muestran por medio de barras horizontales en la parte inferior o superior de los bigotes respectivamente y se muestran los valores atípicos individuales como círculos por debajo (atípicos bajos) o por encima (atípicos altos) de los intervalos especificados.

Figura 3: (ejemplo comparativo)

La Fig. 3 muestra una gráfica de caja y bigotes de 11 a 22 genes metilados de forma diferencial que muestra el intervalo de los porcentajes de metilación en un estudio de 121 mujeres con cáncer tratado de forma quirúrgica al que se hizo seguimiento durante 5 años.

Fig. 4: (ejemplo comparativo)

Las curvas de Kaplan Meier que muestran la capacidad del TWIST 1 (un gen de asociación inversa) de separar las 121 mujeres de la cohorte de seguimiento del cáncer de mama en dos grupos con diferentes riesgos de recidiva. Las mujeres con valores de metilación del ADN por encima del límite TWIST 1 (TWIST 1+; curva superior) tuvieron

una tasa inferior de recidiva que las mujeres por debajo del límite TWIST1 (TWIST 1-; curva inferior). Por ejemplo, a los 5 años, el 68 % de las mujeres con TWIST- no presentó una recidiva mientras que, en la misma fecha de evaluación, el 83 % de mujeres con TWIST 1+ no presentó una recidiva. Por lo tanto, ser positivo para este marcador indicó un 15 % de riesgo absoluto inferior de recidiva de IBC y esta diferencia resultó significativa en el nivel $p=0,024$ con la prueba de rango logarítmico.

Fig. 5ⁱ (ejemplo comparativo)

Las curvas de Kaplan Meier que muestran la capacidad del SLIT2 (un gen de asociación directa) de separar las 121 mujeres de la cohorte de seguimiento del cáncer de mama en dos grupos con diferentes riesgos de recidiva. Las mujeres con valores de metilación del ADN por encima del límite SLIT2 (SLIT2+; curva inferior) presentaron una mayor tasa de recidiva que las mujeres por debajo del límite SLIT2 (SLIT2-; curva superior). Por ejemplo, a los 5 años, el 70 % de las mujeres con SLIT2+ no presentó una recidiva mientras que, en la misma fecha de evaluación, el 83 % de las mujeres con SLIT2- no presentó una recidiva. Por lo tanto, ser positivo para este marcador indicó un 13 % de riesgo absoluto mayor de recidiva de IBC y esta diferencia resultó significativa en el nivel $p=0,044$ con la prueba de rango logarítmico.

Fig. 6ⁱ (ejemplo comparativo)

Las curvas de Kaplan Meier que muestran la capacidad del SFN de separar las 121 mujeres de la cohorte de seguimiento del cáncer de mama en dos grupos con diferentes riesgos de recidiva.

Fig. 7: (ejemplo comparativo)

Las curvas de Kaplan Meier que muestran la capacidad del SERPINB5 de separar las 121 mujeres de la cohorte de seguimiento del cáncer de mama en dos grupos con diferentes riesgos de recidiva.

Fig. 8ⁱ (ejemplo comparativo)

Las curvas de Kaplan Meier que muestran la capacidad de una combinación de dos genes SFN y TWIST 1 de separar las 121 mujeres de la cohorte de seguimiento del cáncer de mama en dos grupos con diferentes riesgos de recidiva. Esta figura muestra los efectos aditivos (sinérgicos) de combinar los dos genes. Por ejemplo, a los 5 años, de las mujeres que resultaron negativas para SFN y positivas para TWIST1 (SFN-, TWIST1+), el 89 % no presentó una recidiva de IBC. Por el contrario, de las mujeres que resultaron positivas para SFN y negativas para TWIST 1 (SFN+, TWIST 1-), el 33 % no presentó una recidiva de IBC. La diferencia absoluta entre estos dos grupos es del 56 % y el valor p para la diferencia con la prueba de rango logarítmico resultó sumamente significativo en $p < 0,001$.

Ejemplos

Muestras de tejido prostático humano

El conjunto de estudios incluyó tejido prostático recién congelado de 77 pacientes de los cuales 48 fueron diagnosticados con cáncer y 29 como BPH. Las muestras se recolectaron ya sea luego de la prostatectomía radical, la resección transuretral de la próstata (RTUP) o RTUP en pacientes con cáncer (canal RTUP). Las muestras se recolectaron de tres lugares diferentes, el Hospital Changhai en Shanghai, China, el Hospital Whipps Cross, Londres y el Hospital de San Bartolomé, Londres durante el período 1996-2008. Todas las muestras se examinaron a nivel central para confirmar el diagnóstico por medio de patólogos genitourinarios expertos. El sistema de puntuación de Gleason se realizó según los criterios estandarizados modernos.

Se obtuvo el consentimiento informado de todos los pacientes. Se obtuvo la aprobación nacional de Reino Unido del *Northern Multi-Research Ethics Committee* (Comité de ética de investigación múltiple del Norte), seguido de la aprobación del comité de ética local de cada una de las fundaciones hospitalarias colaboradoras. La aprobación de ética del Comité de ética del Hospital Changhai se obtuvo para las muestras chinas.

Extracción de ADN y conversión de bisulfito

El ADN genómico se extrajo de 2-3 partes de 10 nm del material recién congelado usando un mini kit de ADN QIAamp (Qiagen Inc., Hilden, Alemania) y se cuantificó por absorción UV, proporcionando típicamente en total $>1 \mu\text{g}$ de ADNg por muestra. Se utilizaron 120-300 ng de ADNg en las reacciones de conversión de bisulfito en donde las citosinas no metiladas se convirtieron en uracilo con el kit de bisulfito EpiTect (Qiagen) según las instrucciones del fabricante. En resumen, el ADN se mezcló con agua, tampón de protección del ADN y mezcla de bisulfito, y la conversión se realizó en un termociclador (Biometra, Goettingen, Alemania) en las condiciones de ciclado recomendadas. El ADN convertido se purificó en una columna de centrifugación y se eluyó dos veces en un total de 40 μl de amortiguador EB.

PCR y pirosecuenciación: Se analizaron veintiocho genes de metilación del ADN candidatos en el estudio. Se utilizaron conjuntos de cebadores con un cebador etiquetado con biotina para amplificar el ADN convertido con bisulfito. Se diseñaron nuevos cebadores para cada uno de los 28 genes (los genes se muestran a continuación, los nombres de los genes siguen el sistema de denominación de genes de la UCSC <http://genome.ucsc.edu/>) que utiliza

la versión 2.0.1.15 del programa de diseño de ensayos PyroMark (Qiagen); en donde se diseñan posibles cebadores para mantener cortos los amplicones con longitudes entre 90 y 140 pares de bases (bp) para facilitar estudios posteriores de muestras embebidas en parafina fijadas con formalina (FFPE, por su sigla en inglés). El tamaño de los amplicones se limitó a un máximo de 210 bp. Todos los cebadores se ubicaron en el promotor o islas CpG del primer exón identificadas mediante MethPrimer, dependiendo de donde el diseño del ensayo permitió cebadores óptimos. Se puso especial cuidado para evitar que cualquier cebador se superponga a la díada de CG para prevenir desviaciones de la amplificación. El tamaño medio de todos los amplicones fue de 104 bp. Para los genes, investigados previamente por medio de otros métodos, los cebadores se colocaron para investigar los mismos CG o aquellos a muy poca distancia. Para algunos genes, p. ej., CDH1, GSTP1, examinamos dos sitios diferentes dentro de la isla CpG separados por varios cientos de pares de bases. Para proporcionar el control interno para la conversión total de bisulfito se incluyó una citosina no CG en la región para pirosecuenciar cuando sea posible.

Las PCR se realizaron usando un equivalente de ADN_g convertido de 200 células que emplea el kit PCR PyroMark (Qiagen). Los equivalentes del genoma celular de los cálculos de ADN supusieron 6 pg de ADN por célula diploide. En resumen, 12,5 µl de mezcla maestra, 2,5 µl de rojo coral, 5 pmol de cada cebador, 7 µl de agua y 2 µl de muestra se mezclaron para cada reacción y se ejecutaron en condiciones de ciclado térmico: 95 °C durante 15 minutos y luego 45 ciclos: 30 seg a 94 °C; 30 seg a la temperatura de hibridación específica para el cebador optimizado; 30 seg a 72 °C y una extensión final durante 10 minutos a 72 °C. El ADN amplificado se confirmó mediante electroforesis en un gel de agarosa al 2 % con punto de fusión bajo (Sigma- Aldrich, Steinheim, Alemania) en amortiguador TBE o por medio del instrumento de electroforesis capilar QiaExel (Qiagen).

Se aplicó un protocolo de preparación de muestras con pirosecuenciación estándar. 3 µl de perlas de estreptavidina (GE Healthcare, Buckinghamshire, Reino Unido), 37 µl de solución amortiguadora de unión de PyroMark (Qiagen), 20 µl de producto de PCR y 20 µl de agua se mezclaron e incubaron durante 10 minutos en una tabla de agitación a 1300 rpm. Utilizando la estación de trabajo de vacío Q96 de Biotage, los amplicones se separaron, desnaturalizaron y agregaron a 45 µl de amortiguador de hibridación que contiene 0,33 µM de cebador de pirosecuenciación. La hibridación de cebadores se realizó incubando las muestras a 80 °C durante 2 minutos y se dejó que enfriaran a temperatura ambiente antes de someterlas a la pirosecuenciación. Los reactivos PyroGold se utilizaron para la reacción de pirosecuenciación y se analizó la señal usando el sistema 96MA PSQ (Biotage, Uppsala, Suecia). Se evaluaron los CG diana mediante el programa de instrumento (PSQ96MA 2.1) que convierte los pirogramas en valores numéricos para las alturas de los picos y calcula la proporción de metilación en cada base como una relación C/T. Todas las ejecuciones contenían curvas estándar que comprendían un intervalo de ADN metilado de control (0 %, 25 %, 50 %, 75 % y 100 %) para permitir las comparaciones directas estandarizadas entre los distintos conjuntos de cebadores. Para las curvas estándar, se mezcló un total de 300 ng de ADN no metilado (Qiagen) e hipermetilado (Millipore, Billerica, MA, EUA) para obtener las distintas relaciones de metilación del ADN y luego se convirtió con bisulfito como se describió anteriormente.

Análisis estadísticos:

Los análisis principales se basaron en los valores medios de todos los CG analizados. La cantidad de CG analizados osciló entre dos y seis en cada gen según lo permitieron los parámetros definidos del programa. Para limitar la cantidad de pruebas y costos, se investigaron los genes que no mostraron ningún potencial al diferenciar entre la BPH y el cáncer o entre la puntuación de Gleason baja y alta en pocas muestras.

Se examinaron las diferencias de metilación entre los tejidos por medio de la prueba de Mann-Whitney. Para explicar la gran cantidad de genes sometidos a prueba en los mismos datos, se aplicó el procedimiento de incrementos de Benjamín y Hochberg para controlar la tasa de falso descubrimiento (FDR, por su sigla en inglés) con una FDR de 1 %.

Para explorar la relación entre la metilación de los genes y la edad, se normalizó la metilación mediante puntuaciones z, en donde la metilación bruta menos la media muestral se dividió por la desviación estándar de la muestra. La asociación entre la metilación y la edad se exploró a través de la prueba de Spearman mientras que para la metilación en comparación con la puntuación de Gleason, se utilizó la prueba de tendencia de Cuzick. Para los casos con una medición del PSA, se realizó la asociación con la metilación usando la prueba de rango de Spearman. Asimismo, se calcularon los coeficientes de Spearman, que se basan en los ordenamientos de rango de la metilación de genes bruta en todos los cánceres para explorar la corrección en la metilación entre genes. Los límites que se eligieron para presentar las tasas de verdaderos positivos (TP, proporción de cánceres correctamente clasificados) y las tasas de falsos positivos (FP, proporción de no cánceres identificados de forma incorrecta) por gen se eligieron usando la misma función de costo para todos los genes, a saber, para minimizar la FP - TP.

Para ayudar a investigar la metilación asociada con las puntuaciones de Gleason altas, se aplicó un algoritmo de clasificación de selvas aleatorias. Las gráficas se utilizaron para inspeccionar los genes identificados a través de las selvas aleatorias, con la adición de límites de clasificación a partir de los análisis discriminantes lineales. Los intervalos de confianza (CI, por su sigla en inglés) de 95 % de precisión de la clasificación de puntuaciones de Gleason se basaron en un método de muestreo con reemplazamiento no paramétrico con 1000 muestreos nuevos. Todos los cálculos estadísticos se realizaron usando el software R, versión 2.9.2. El rechazo de la hipótesis nula se

supuso en un $<0,05$.

Estadística descriptiva de la metilación del gen candidato

La reproducibilidad del método PSQ se investigó al principio midiendo la metilación del GSTP1 en tres ocasiones distintas. La diferencia de metilación media entre la lectura más alta y más baja para la misma muestra fue de 7 % para los casos de BPH y de 13 % para los cánceres; la correlación de Pearson para las ejecuciones 1 vs. 2 fue de 0,90 mientras que para las ejecuciones 1 y 3 fue de 0,97. Esta concordancia fue considerada aceptable y todos los datos posteriores se basaron en mediciones individuales.

Los datos de metilación se ajustaron para la preferencia del promotor a través de la reclasificación de cada una de las mediciones de metilación del gen mediante la curva estándar mediana que se obtuvo para cada conjunto de cebadores. El impacto de aplicar estas correcciones a los genes tuvo pequeños efectos en las diferencias de metilación mediana, pero permitió la comparación entre los distintos genes. De los 28 genes estudiados, la metilación de 20 genes: RARB, HIN1, BCL2, GSTP1, CCND2, EGFR5, APC, RASSF1A, MDR1, NKX2-5, CDH13, DPYS, PTGS2, EDNRB, MAL, PDLIM4, SERPINB5, HLAa, ESR1 y TIG1 podría distinguir el cáncer de próstata del tejido de BPH a una FDR del 1 %. Se calcularon los niveles límite para evaluar el diagnóstico potencial de las diferencias de metilación. Esto permitió la dicotomización de los datos, en donde un límite de 21 % de metilación de RARB separó todos los cánceres de BPH con 100 % de precisión, es decir, TP=100 %, FP = 0 %.

En las muestras de BPH, el EGFR5, DPYS, ESR1, MDR1, SERPINB5 y SFN mostraron metilación mediana por encima del 10 % mientras que la mayoría de los otros genes no se sometieron a metilación (~ 2 % de metilación mediana). En particular, SERPINB5 y SFN se metilaron a aproximadamente 50 % en BPH. Además, SERPINB5 fue el único gen con una metilación considerablemente mayor en la BPH que en los cánceres ($p < 0,001$). MCAM, CDKN2A, THRB, TWIST 1, CDH1 y DAPK1 se metilaron por debajo del 10 % tanto en la BPH como en el PCa.

Asociación entre las covariables demográficas y clínicas y la metilación del gen.

Se analizaron la relación entre los niveles de metilación del gen, la edad, la puntuación de Gleason y los niveles PSA en el PCa. Hubo una asociación positiva de la puntuación de Gleason con la edad ($p < 0,001$) y el PSA ($p = 0,0013$), pero ninguna asociación entre el PSA y la edad ($p = 0,22$).

Hubo una tendencia positiva entre la edad y los valores de metilación media estandarizada en todos los genes, similar al estado de metilación global para cada caso (correlación de Pearson 0,52, $p < 0,0001$). Asimismo, la inspección de la distribución de los valores p sugirió un efecto moderado de la edad común a la metilación de todos los genes, mientras que la puntuación de Gleason pareció afectar solamente a los subconjuntos de genes. Los niveles de metilación de NKX2-5 y APC ($p = 0,009$), TIG1, ESR1, GSTP1 ($p = 0,01$), CDH13, EGFR5 ($p = 0,02$), MCAM ($p = 0,03$) y SLIT2 ($p = 0,04$) mostraron una asociación positiva con la edad. La prueba de tendencia de Cuzick mostró que la metilación de SFN ($p = 0,01$), TIG1 ($p = 0,02$), PDLIM4, APC y SERPINB5 ($p = 0,04$) se asoció con la puntuación de Gleason. Asimismo, según la clasificación de selvas aleatorias, la alta metilación del SFN, SLIT2 y SERPINB5 separó los cánceres con puntuación de Gleason baja y alta (Figura 1). Los límites de discriminación lineal describieron la estructura que se encontró por medio de la clasificación de selvas aleatorias. La medición compuesta del nivel de metilación de SFN y SERPINB5 se clasificó correctamente 81 % (95 % CI 56-91) de puntuaciones altas de Gleason mientras que se clasificaron erróneamente las puntuaciones bajas de Gleason. De manera similar, la metilación de SFN y SLIT2, detectó 62 % (47-81) de puntuaciones de Gleason altas, no clasificadas y bajas 12 % (9 - 47) mientras que la metilación de SERPINB5 y SLIT2 detectó 62 % (47-81) de puntuaciones de Gleason altas, no clasificadas y bajas 13 % (3 - 31). Los niveles de metilación de 17 genes: HIN1, TWIST1, GSTP1, RARB ($p < 0,001$), HLAa, BCL2, APC, PDLIM4, PTGS2, DPYS, CDH13 ($p < 0,01$) and RASSF1A, MDR1, EGFR5, EDNRB, TIG1, CCND2 ($p < 0,05$) se asociaron positivamente con el PSA. Además, los niveles de metilación de la mayoría de los genes podrían distinguir entre la BPH y el cáncer, p. ej., RARB, APC, EGFR5, HIN1, RASSF1A, PTGS2 y CDH13 se correlacionaron de forma moderada.

Aparte del PSA, con todas sus limitaciones, se encuentran disponibles actualmente marcadores biológicos validados y no aceptados generalmente para el pronóstico o predicción terapéutica en el cáncer de próstata. Aunque se están considerando varios marcadores nuevos, por ejemplo, PCA3, TMPRSS-ERG, Ki-67, HSP27 y otros, estos no han sido validados para un uso extendido y, por lo tanto, la puntuación de Gleason y el PSA en el contexto de otra información clínica siguen siendo la base del proceso decisorio en el PCa.

Veinte de los genes investigados, a saber: RARB, HIN1, BCL2, GSTP1, CCND2, EGFR5, APC, RASSF1A, MDR1, NKX2-5, CDH13, DPYS, PTGS2, EDNRB, MAL, PDLIM4, SERPINB5, HLAa, ESR1 y TIG1 fueron mucho más metilados en cánceres que en el tejido de BPH mientras que el riesgo de falso descubrimiento (FDR) fue inferior al 1 %. Según nuestros conocimientos, la metilación de MAL, HLAa, SERPINB5, THRB, TWIST1 y SLIT2 se demostró aquí por primera vez en el tejido prostático. Si bien THRB y TWIST 1 no fueron metilados en general, HLAa y MAL mostraron metilación baja con clara capacidad de discriminar entre los tejidos con diferencia mediana de 15 % y 7 %, respectivamente. El estado de metilación de HLAa se relacionó con el nivel de PSA, pero no con la edad, la puntuación de Gleason o la metilación de otros genes. La metilación de SLIT2 resultó baja, no obstante, la metilación mediana se elevó en los cánceres ~ 6 % en comparación con ~ 2 % en BPH y además el nivel de

metilación de SLIT2 podría separar los cánceres con puntuación de Gleason baja y alta. La metilación promedio de SERPINB5 y SFN fue inferior en los cánceres que en la BPH, 15 % ($p < 0,001$) y 12 % ($p = 0,05$), respectivamente.

MCAM resultó no metilado en las muestras investigadas pese a que se informó previamente la metilación anómala del promotor MCAM en el PCa.

- 5 Los informes de metilación de DAPK1, CDH1 y CDKN2A han sido contradictorios. Observamos una metilación igualmente baja (mediana < 10 %) de estos genes en la BPH y los tejidos cancerosos, si bien no se observó ninguna diferencia considerable de pequeña magnitud. Para el CDH1, esta observación resultó verdadera para ambas regiones del promotor informadas previamente para mostrar las diferencias en la metilación. Las contradicciones pueden deberse a la amplificación de diferencias no significativas mediante un método
- 10 semicuantitativo que provoca proporciones distorsionadas de la metilación. Además, estos genes parecen que tienen poca asociación con la puntuación de Gleason o los niveles de PSA.

- Los datos del PSA de referencia estuvieron disponibles solamente para 35 cánceres, sin embargo, la metilación del HIN1 se relacionó positivamente con el PSA ($p < 0,001$), sin que haya evidencia de una asociación coincidente con la edad o la puntuación de Gleason. Salvo por TWIST 1, 16 de los genes que muestran diferencia entre el cáncer y la BPH se asociaron positivamente también con el PSA ($p < 0,05$).
- 15

Pacientes con cáncer de mama, muestras de patología y tratamiento (ejemplo comparativo):

- El estudio incluyó 124 pacientes con cáncer de mama que se presentaron en el Centro Médico de la Universidad Dong-A, Busan, República de Corea desde enero de 2004 a diciembre de 2006. Se incluyeron en el estudio todos los pacientes con cáncer de mama operable consecutivo que aceptaron proporcionar muestras de tejido normal
- 20 adyacente y de tumor mamario reciente (que mide aproximadamente 0,5 x 0,5 cm cada uno) en el momento de la cirugía y en quienes fue posible dicha recolección de muestras y congelación inmediata de muestras y se mantuvieron en un congelador a -70 sin poner en peligro el diagnóstico clínico y el tratamiento. Se incluyeron los pacientes que recibieron quimioterapia neoadyuvante (NACT, por su sigla en inglés). Se excluyeron los pacientes con datos clínico-patológicos faltantes para 2 o más variables ($n = 3$), lo que dio como resultado 121 participantes que cumplen los requisitos adecuados.
- 25

- Todas las muestras fueron examinadas a nivel central para confirmar el diagnóstico por parte del patólogo mamario (DCK). La evaluación histopatológica e inmunohistoquímica para ER (DAKO, Clone 1D5, 1:50), PgR (DAKO, PgR 636, 1:100) y HER2 (Neomarkers, Clone e2-4001/3B5, Fremont, CA, EUA, 1:200) se realizó de acuerdo con los protocolos recomendados de fábrica, los cuales cumplen con los estándares internacionales. Se determinó el grado histológico mediante la Modificación Nottingham de la puntuación de Richardson-Bloom (RBS). Se evaluó la expresión de HER2 mediante inmunohistoquímica (IHC) y se realizó la confirmatoria HER2 FISH (kit de sonda del ADN de HER2 de PathVision®, Vysis Inc., Downers Grove, IL, EUA) en pacientes que tuvieron 2+ o 3+ (membrana fuerte que se tiñe sobre el 10 % de las células tumorales) en la IHC, solamente estas mujeres confirmadas se consideraron positivas para HER2. Las variables de tratamiento y clínico-patológicas que incluyen la edad, el tamaño del tumor (solamente patológico), el tipo, el grado histológico, el estado de los ganglios linfáticos, ER, PgR, HER2, el tipo de cirugía, los detalles del tratamiento con adyuvante/neoadyuvante (p. ej., el tipo de quimioterapia, la cantidad de ciclos, la cantidad de fracciones de radioterapia, etc.) se registraron en la base de datos del estudio para todos los pacientes que cumplieran con los requisitos adecuados. Los datos sobre el tamaño del tumor fueron una mezcla del tamaño del tumor clínico y patológico y algunos pacientes también recibieron NACT. En vista de los datos heterogéneos en esta variable, todos los análisis se llevaron a cabo excluyendo la variable del tamaño tumoral. A los pacientes se les hizo un seguimiento periódico cada 6 meses.
- 30
- 35
- 40

Este estudio contó con la aprobación del Consejo de Revisión Institucional del Centro Médico de la Universidad Dong-A y todos los pacientes otorgaron su consentimiento informado por escrito.

- Se llevó a cabo una macrodissección simple de las partes del tejido antes de la extracción de ADN para aumentar las áreas del cáncer, el método empleado es rápido (aproximadamente 10 minutos por caso) y el técnico de laboratorio promedio lo domina fácilmente. Se obtuvieron cinco secciones consecutivas por muestra (espesor de 10 μm) mediante el crioseccionamiento de los tejidos cancerosos y la tinción de la primera y quinta sección mediante H&E para el examen de histopatología para confirmar las áreas del cáncer y guiar las disecciones de las tres secciones centrales. Se extrajo del ADN genómico de las tres partes del material tisular usando un mini kit de ADN QIAamp (Qiagen Inc., Hilden, Alemania) y se cuantificó por absorción UV (Nanodrop, Thermo Scientific, Wilmington, Delaware, EUA), una mayoría de secciones que producen un >1 μg combinado de ADNg por muestra. Se utilizaron 120 - 300 ng de ADNg en las reacciones de conversión con bisulfito en donde las citosinas no metiladas se convirtieron en uracilo con el kit de bisulfito EpiTect (Qiagen) según las instrucciones del fabricante. En resumen, el ADN se mezcló con agua, amortiguador de protección de ADN y mezcla de bisulfito, y la conversión se ejecutó en un termociclador (Biometra, Goettingen, Alemania) en las condiciones de ciclado recomendadas. El ADN convertido se purificó y eluyó en 2 etapas en un total de amortiguador EB de 40 μl y se diluyó además en 20 μl de alícuotas de 100 células-equivalentes/ μl . (los cálculos celulares supusieron 6 pg de ADN por célula diploide).
- 45
- 50
- 55

Se utilizaron conjuntos de cebadores con un cebador etiquetado con biotina para amplificar las muestras de ADN

convertido con bisulfito. Se identificaron treinta genes de la bibliografía como genes candidatos para este estudio como se describió previamente. Los nuevos cebadores para cada uno de los 30 genes fueron diseñados usando el programa de diseño de ensayos PyroMark, versión 2.0.1.15 (Qiagen) con el objetivo de hacer que los amplicones se mantengan cortos con longitudes de 90 a 140 pares de bases (bp) para facilitar estudios posteriores en las muestras de FFPE. El tamaño máximo permitido de los amplicones fue de 210 bp. Todos los cebadores se ubicaron en el promotor o islas CpG del primer exón identificados mediante MethPrimer, dependiendo de donde el diseño del ensayo permitió cebadores óptimos. No se permitieron díadas de CG en ninguna posición de cebadores de secuenciación, inversa o directa para prevenir toda tendencia a la amplificación. El tamaño medio de todos los amplicones fue de 117 bp. Para los genes, investigados previamente por medio de otros métodos, los cebadores se colocaron para investigar los mismos CG o aquellos a muy poca distancia. Para proporcionar el control interno para la conversión total de bisulfito, se incluyó una citosina no CG en la región para pirosecuenciar cuando sea posible. Se investigaron de tres a seis posiciones CG en cada gen.

Las PCR se realizaron usando un equivalente de ADN_g convertido con bisulfito de 200 a 400 células empleando el kit de PCR PyroMark (Qiagen). En resumen, 12,5 µl de mezcla maestra, 2,5 µl de rojo coral, 5 pmol de cada cebador, 7 µl de agua y 2 µl de muestra se mezclaron para cada reacción y se ejecutaron en condiciones de ciclado térmico: 95 °C durante 15 minutos y luego 45 ciclos: 30 seg a 94 °C; 30 seg a la temperatura de hibridación específica para el cebador optimizado; 30 seg a 72 °C y una extensión final durante 10 minutos a 72 °C. El ADN amplificado correcto se confirmó mediante electroforesis en un gel de agarosa 2 % con punto de fusión bajo (Sigma-Aldrich, Steinheim, Alemania) en amortiguador TBE o por medio del instrumento de electroforesis capilar QiaExel (Qiagen). Se aplicó un protocolo de preparación de muestras con pirosecuenciación estándar. 3 µl de perlas de estreptavidina (GE Healthcare, Buckinghamshire, Reino Unido), 37 µl de solución amortiguadora de unión de PyroMark (Qiagen), 20 µl de producto de PCR y 20 µl de agua se mezclaron e incubaron durante 10 minutos en una tabla de agitación a 1300 rpm. Utilizando la estación de trabajo de vacío Q96 de Biotage, los amplicones se separaron, desnaturalizaron y agregaron a 45 µl de amortiguador de hibridación que contiene 0,33 µM de cebador de pirosecuenciación. La hibridación de cebadores se realizó incubando las muestras a 80 °C durante 2 minutos y se dejó que enfriaran a temperatura ambiente antes de someterlas a la pirosecuenciación. Los reactivos PyroGold se utilizaron para la reacción de pirosecuenciación y se analizó la señal usando el sistema 96MA PSQ (Biotage, Uppsala, Suecia). Se evaluaron los CG diana mediante el programa de instrumento (PSQ96MA 2.1) que convierte los pirogramas en valores numéricos para las alturas de los picos y calcula la proporción de metilación en cada base como una relación C/T. Todas las ejecuciones contenían curvas estándar que comprendían un intervalo de ADN metilado de control (0 %, 25 %, 50 %, 75 % y 100 %) para permitir las comparaciones directas estandarizadas entre los distintos experimentos. Para las curvas estándar, se mezcló un total de 300 ng de ADN no metilado (Qiagen) e hipermetilado (Millipore, Billerica, MA, EUA) para obtener las distintas relaciones de metilación del ADN y luego se convirtió con bisulfito como se describió anteriormente.

Se realizó una selección adicional de genes preferidos de los 30 genes candidatos iniciales luego de que se procesaron las primeras 30 muestras. Los genes (n = 20) que se correlacionan con (p < 0,1) cualquier edad, condición de los ganglios, grado histológico, ER, PgR, HER2 fueron seleccionados como preferidos. Los 10 genes restantes tuvieron niveles y frecuencia de metilación muy bajos y, por lo tanto, era improbable que tuvieran éxito como marcadores biológicos. Estos genes, por ende, no fueron investigados adicionalmente en este estudio; informamos los resultados en 20 genes preferidos seleccionados.

Todas las muestras fueron analizadas una vez salvo las muestras que no produjeron resultados aceptables en el primer ensayo. Dichas muestras fueron analizadas una vez más y los datos fueron registrados como faltantes si las muestras no tuvieron éxito en la segunda instancia tampoco. Las muestras fallidas constituyen una proporción muy baja de las muestras investigadas.

Los análisis principales convirtieron los datos de relación C/T individuales en valores medios de todos los CG analizados en un segmento de gen particular. La cantidad de CG analizados osciló entre dos y seis en cada gen. Los datos de metilación se ajustaron para la preferencia del promotor a través de la reclasificación de cada una de las mediciones de metilación del gen mediante la curva estándar mediana que se obtuvo usando mezclas de control para cada conjunto de cebadores.

Los análisis de correlación se llevaron a cabo usando la prueba de Spearman para investigar las asociaciones entre las diversas variables clínico-histopatológicas (categóricas) como la edad, la condición de los ganglios, el grado, ER, PgR, HER2 y el porcentaje de metilación (MeC %) de los genes (en una clasificación continua).

Los análisis monofactoriales que utilizan la prueba de rango logarítmico se realizaron con supervivencia libre de recidiva (RFS, por su sigla en inglés) como un criterio de valoración para examinar la relación entre la condición de los ganglios, el grado, ER, PgR y HER2. Se graficaron las estimaciones de supervivencia de Kaplan-Meier. Se utilizan diversos enfoques para la evaluación de los marcadores biológicos y la optimización del punto de corte basado en resultados, X-Tile es uno de dichos enfoques. La determinación de corte, sin embargo, es subjetiva en el enfoque X-Tile. Empleamos un enfoque similar con parámetros objetivos para determinar los puntos de corte. Se utilizaron diferentes puntos de corte de 0 % a 99 %, MeC%, en las etapas de 1 % en análisis monofactoriales mediante la prueba de rango logarítmico. No se consideraron los puntos de corte que producen un grupo inferior al 10 % de toda la población del estudio (es decir, <12 sujetos). Se eligieron los puntos de corte que producen el valor

p más bajo (punto de corte óptimo mediante el método de valor p). En el caso de múltiples puntos de corte que provocan el mismo valor p, se seleccionó un punto de corte que provocó la diferencia máxima en los números en dos grupos (un grupo por debajo del punto de corte y otro por encima); esencialmente se seleccionaron un punto de corte inferior con el mismo valor p si los puntos de corte equivalentes están por debajo del 50 % y un punto de corte mayor si están por encima del 50 %.

Los puntos de corte por debajo del 5 % y por encima del 90 % son sensibles a pequeñas diferencias en lecturas de ensayo y, por lo tanto, puede dar lugar a la clasificación incorrecta de muestras. Desde el punto de vista de la reproducibilidad, es esencial que los puntos de corte no se vean afectados por variaciones menores en diferentes ejecuciones de prueba. Por lo tanto, se aplicó una condición adicional, punto de corte $\geq 5\%$ y $\leq 90\%$, antes de que los genes pudieran considerarse para análisis multifactoriales y otros análisis adicionales; los genes con el mejor punto de corte por debajo del 5 % o por encima 90 % fueron considerados en otros análisis en este estudio.

Todos los cálculos estadísticos se realizaron usando el software R, versión 2.9.2 o SAS 10.0. Todos los valores p son bilaterales con un $\alpha < 0,05$ a menos que se especifique de otra manera.

Ciento veintitres pacientes cumplieron con los criterios de inclusión para este estudio. Todos los pacientes se sometieron a una cirugía para preservar la mama o a la mastectomía. La gestión axilar fue la eliminación axilar completa o biopsia del ganglio centinela (con eliminación completa en quienes SNB mostró ganglio/s metastásico/s). Dieciséis pacientes se sometieron a una cirugía para conservar la mama, 104 se sometieron a una mastectomía total y se realizó una mastectomía radical en 1 paciente. Cuarenta y nueve pacientes se sometieron a SNB como el procedimiento inicial de estadificación axilar. Trece pacientes recibieron alguna NACT, mientras que 99 pacientes recibieron quimioterapia solamente en el postoperatorio. Nueve pacientes no recibieron quimioterapia alguna, 46 recibieron quimioterapia a base de metotrexato (CMF, por su sigla en inglés), 19 recibieron quimioterapia a base de antraciclina (FAC o FEC) y 32 pacientes recibieron una combinación secuencial de antraciclinas y taxanos (AC/EC seguido de Paclitaxel o Docetaxel). Veintitres pacientes recibieron adicionalmente Doxifluridina luego de terminar la quimioterapia adyuvante estándar. Sesenta pacientes recibieron radioterapia (incluidos todos los pacientes sometidos a cirugía conservadora), 16 de ellos recibieron un refuerzo de la cama tumoral. Se les recetó Tamoxifen durante 5 años a todos los pacientes que resultaron positivos para el receptor de la hormona premenopáusicas mientras que, a los pacientes postmenopáusicos se les recetó inhibidores de aromatasa. Luego de un seguimiento medio de 5,1 años (4,87-5,4) 25 pacientes experimentaron recidivas mientras que 3 pacientes fallecieron. Un paciente falleció debido a las complicaciones relacionadas con la quimioterapia luego del primer ciclo de quimioterapia CMF, no se conoció la causa de muerte en los otros dos pacientes. Los pacientes que fallecieron dentro del período de estudio de 5 años fueron excluidos de nuestros análisis de modelado de marcadores biológicos.

La mayoría de los pacientes ($n = 112$) padeció carcinoma ductal invasivo (IDC, por su sigla en inglés), 10 de estos mostraron características medulares; 4 pacientes sufrieron de carcinoma lobular invasivo (ILC, por su sigla en inglés) y 5 tumores tuvieron otras morfologías como el carcinoma mucinoso o el carcinoma tubular.

Se investigaron las correlaciones entre los diversos factores hispatológicos, entre los genes y los factores clínico-patológicos y entre genes. ER mostró una fuerte correlación positiva con PR (Rho de Spearman = 0,79, $p < 0,001$), y correlación negativa con HER2 (Rho = -0,172, $p < 0,042$), y grado (menos ER positivamente con grado creciente; Rho = -0,361, $p < 0,001$). PgR mostró esperadamente correlaciones similares; HER2 (Rho = -0,252, $p = 0,004$), grado (Rho = -0,366, $p < 0,001$). HER2 y el grado no se correlacionaron entre sí y la condición de los ganglios no mostró correlación con ninguna otra variable.

La Figura 2 y 3 muestran las gráficas de caja y bigotes de los porcentajes de metilación del ADN para cada uno de los 20 genes que mostraron metilación diferencial razonablemente medible en el IBC. Ninguno de los genes se correlacionó con la edad. Solamente SERPINB5 y PDLIM4 se correlacionaron positivamente con el grado y la condición de los ganglios, respectivamente. HER2 se correlacionó positivamente con PDLIM4, RARp y RASSF1A. La mayoría de las correlaciones fueron entre genes y ER o PgR. ER o PgR correlacionaron positivamente con CDH13, EDNRB, EGFR5, HIN1, RASSF1A y negativamente con RARp y SERPINB5. ER solo se correlacionó positivamente con SLIT2. Mientras la mayoría de los genes mostraron correlaciones positivas moderadas a débiles, se observaron determinadas correlaciones negativas particularmente de SERPINB5, SFN and TWIST1. SERPINB5 se correlacionó negativamente con EGFR5, MAL, SLIT2, HIN1, CDH13, EDNRB y RASSF1A. Curiosamente, SFN, SERPINB5 y EGFR5 mostraron un % de MeC mucho mayor en comparación con los demás genes, mientras que aquellos para DAPK1, HLAa y RARp fueron muy bajos.

El análisis de supervivencia monofactorial analizó las relaciones entre la edad, el grado histológico, la condición de los ganglios, ER, PgR, HER2 y RFS. Solo la condición de los ganglios ($p = 0,004$) se asoció significativamente con la RFS. El grado histológico mostró una tendencia de asociación con la RFS ($p = 0,086$). Ambas variables fueron consideradas para el análisis multifactorial.

Se determinaron los puntos de corte para los 20 genes de acuerdo con el análisis de supervivencia monofactorial usando el método de minimización de valor p de deslizamiento descrito. Once genes mostraron asociación significativa con la RFS en puntos de corte elegidos, se seleccionaron 7 genes (HLAa, NKX2-5, APC, SERPINB5,

5 SFN, SLIT2 y TWIST 1) para análisis adicionales porque tuvieron las características estadísticas más prometedoras, principalmente de acuerdo con los valores p más bajos sino también considerando la solidez de los datos en el ensayo PSQ. Las gráficas de supervivencia de Kaplan-Meier que muestran las tasas de recidiva del IBC en distintos grupos de mujeres se muestran para TWIST1 SLIT2, SFN, SERPINB5 y una combinación de SFN y TWIST 1 en las Figuras 4, 5, 6, 7 y 8 respectivamente.

Listado de secuencias

<110> Queen Mary and westfield college, university of London

10 <120> Marcadores de cáncer de próstata

<130> B051-0085WO01

<150> EP 11 16 0657.0

15 <151> 31-03-2011

<160> 7

<170> PatentIn versión 3.5

20 <210> 1

<211> 121

<212> ADN

<213> Homo sapiens

25 <400> 1

ggagaagggtg gagactgagc tccagggcgt gtgcgacacc gtgctgggcc tgctggacag 60

ccacctcatc aaggaggccg gggacgccga gagccggggtc ttctacctga agatgaaggg 120

t 121

30 <210> 2

<211> 112

<212> ADN

<213> Homo sapiens

35 <400> 2

agatccccctc ttctgtcttg taccttcgcc actggcatcg gatttgcaga agcgtgcgtg 60

ggatcagagg accgccctcc ccacaacaac cggcccctgc atcttagcag cc 112

40 <210> 3

<211> 71

<212> ADN

<213> Homo sapiens

<400> 3

caagaggctt gagtaggaga ggagtgccgc cgaggcgggg cggggcgggg cgtggagctg 60

45 ggctggcagt g 71

50 <210> 4

<211> 98

<212> ADN

<213> Homo sapiens

<400> 4

tcctcctgct ctctcctccg cgggccgcat cgcccgggcc ggcgccgcgc gcgggggaag 60

55 ctggcgggct gaggcgcccc gctcttctcc tctgcccc 98

ES 2 631 193 T3

<210> 5
<211> 95
<212> ADN
<213> Homo sapiens
5
<400> 5
gggctagggc taggcaggct gtgcggttgg gcggggcct gtgccccact gcggagtgcg 60
10 ggtcgggaag cggagagaga agcagctgtg taatc 95
<210> 6
<211> 41
<212> ADN
15 <213> Homo sapiens
<400> 6
gggccctggc cctgaccag acctgggagg gtgagtgcgg g 41
20
<210> 7
<211> 116
<212> ADN
<213> Homo sapiens
25 <400> 7
ccttctcagt caaagacatc ctaaacctgg aacagcagca gcgcagcctg gctgccgccg 60
gagagctctc tgcccgcctg gaggcgacct tggcgcctc ctctgcatg ctggcc 116
30

REIVINDICACIONES

1. Un método para pronosticar el cáncer de próstata que comprende las etapas de analizar, en una muestra de un sujeto, el estado de metilación del ADN de las siguientes regiones genómicas,
- 5 (i) SFN según la SEQ ID NO. 1,
(ii) SLIT2 según la SEQ ID NO. 2 y
(iii) SERPINB5 según la SEQ ID NO. 3 y
- en donde, si al menos dos de las siguientes características aplican, la muestra se categoriza como una muestra de un paciente con un pronóstico desalentador:
- 10 (i) SFN muestra un valor límite de metilación por encima del 80 % y/o,
(ii) SLIT2 muestra un valor límite de metilación por encima del 45 % y/o,
(iii) SERPINB5 muestra un valor límite de metilación por encima del 70 %.
2. Un método según cualquiera de las reivindicaciones precedentes, en donde se analiza el estado de metilación de una región genómica adicional y/o un marcador biológico adicional que puede seleccionarse del grupo de
- 15 i. APC según la SEQ ID NO. 5
ii. HLAa según la SEQ ID NO. 6 y/o
iii. NKX2-5 según la SEQ ID NO. 7.
3. Un método según cualquiera de las reivindicaciones precedentes, en donde analizar el estado de metilación de una región genómica significa analizar el estado de metilación de al menos una posición CpG por región genómica.
- 20 4. Un método según cualquiera de las reivindicaciones precedentes, en donde se analiza el estado de metilación mediante métodos basados en la PCR específica para la no metilación, los métodos basados en la metilación, la secuenciación de última generación o los métodos basados en micromatrices.
- 25 5. Un método según la reivindicación 4, en donde el método basado en la PCR específica para la no metilación es la pirosecuenciación.
6. El uso *in vitro* de un ácido nucleico que se hibrida en condiciones rigurosas en las proximidades de una de las regiones genómicas según la SEQ ID NO. 1 a la SEQ ID NO: 7, en donde dicha proximidad es cualquier posición que tenga una distancia de hasta 1000 nt desde el extremo 3' o 5' de dicha región genómica y en donde dicha proximidad incluye la propia región genómica para el pronóstico del cáncer de próstata según el método de la reivindicación 1.
- 30 7. El uso según la reivindicación 6, en donde el ácido nucleico tiene una longitud de 15 a 100 nt.
8. El uso según la reivindicación 6 o 7, en donde el ácido nucleico es un cebador.
9. El uso según la reivindicación 8, en donde el cebador es específico para una de las regiones genómicas de la reivindicación 1.
- 35 10. El uso según la reivindicación 6 o 7, en donde el ácido nucleico es una sonda.
11. El uso de acuerdo con la reivindicación 10, en donde se etiqueta la sonda.
12. El uso según cualquiera de las reivindicaciones 6 a 11, en donde el ácido nucleico se hibrida en condiciones rigurosas en dicha proximidad de una de las regiones genómicas luego de tratar la región genómica con bisulfito.

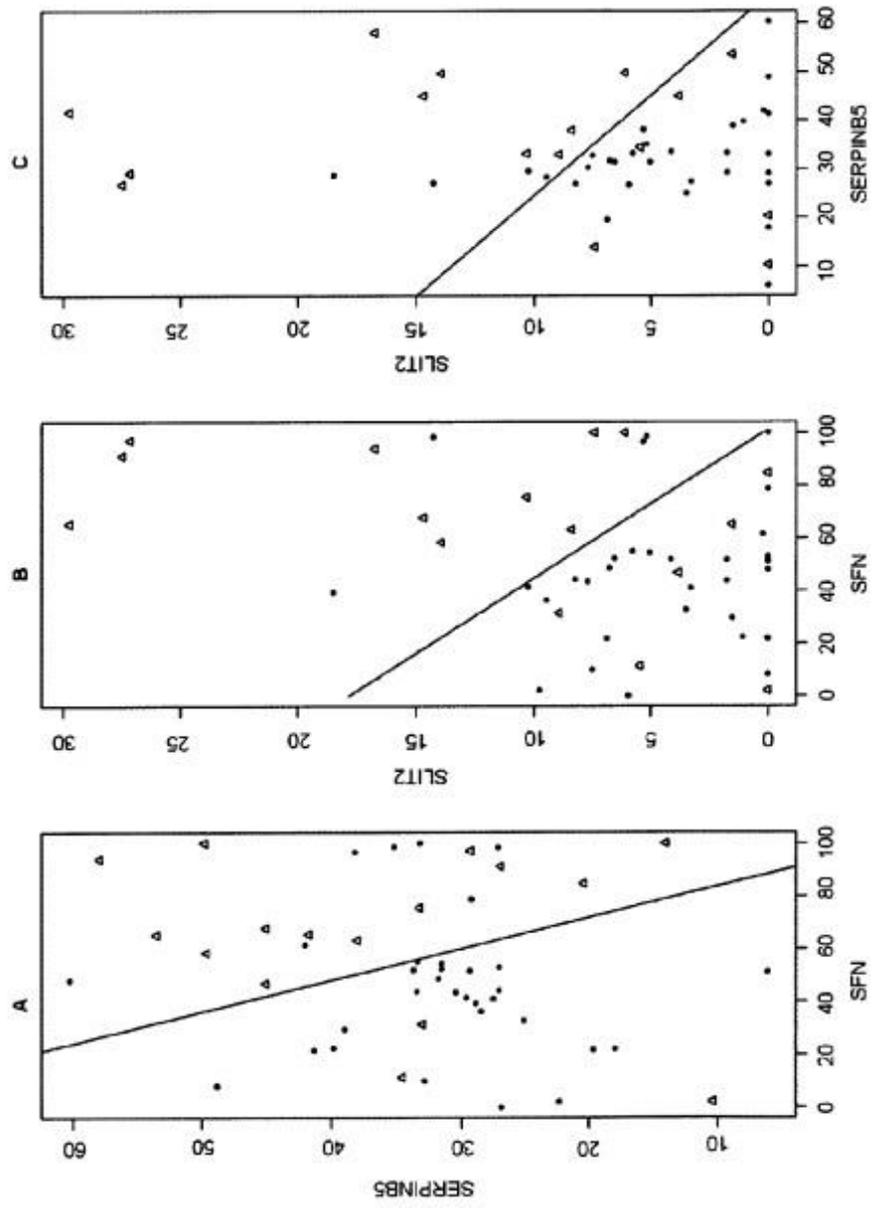


FIG 1:

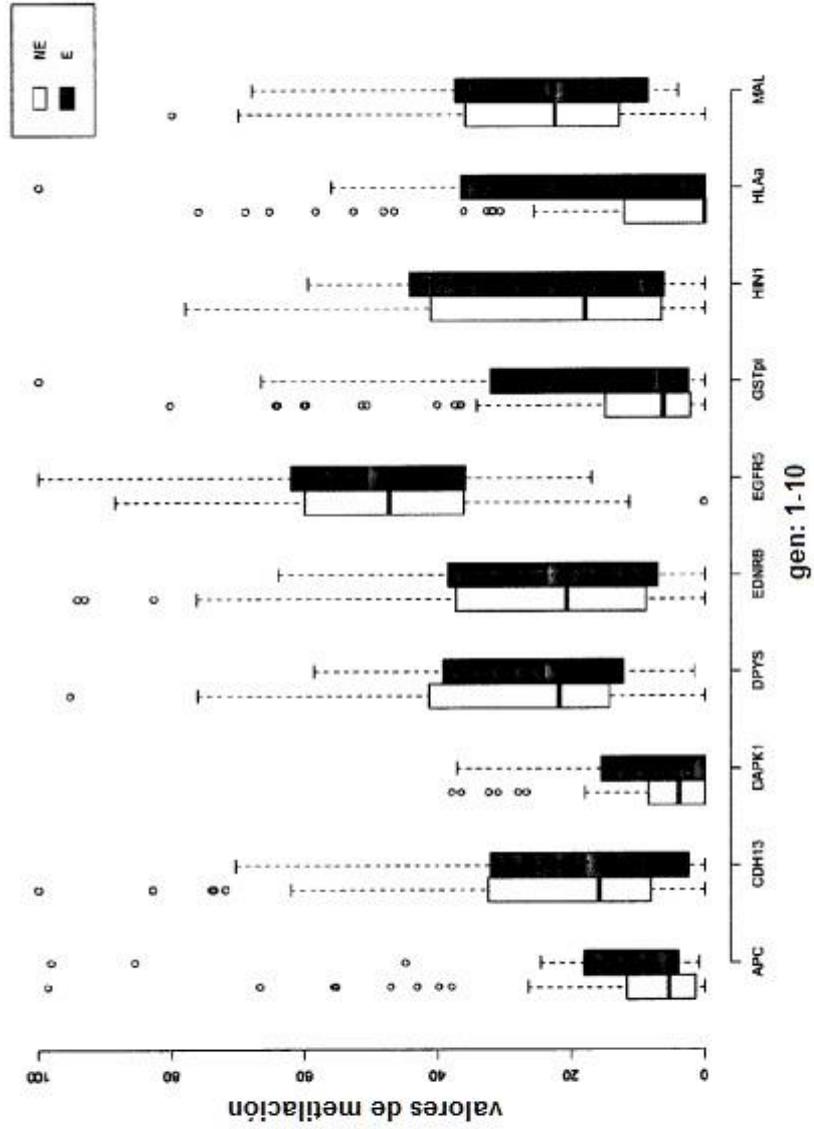


FIG. 2:

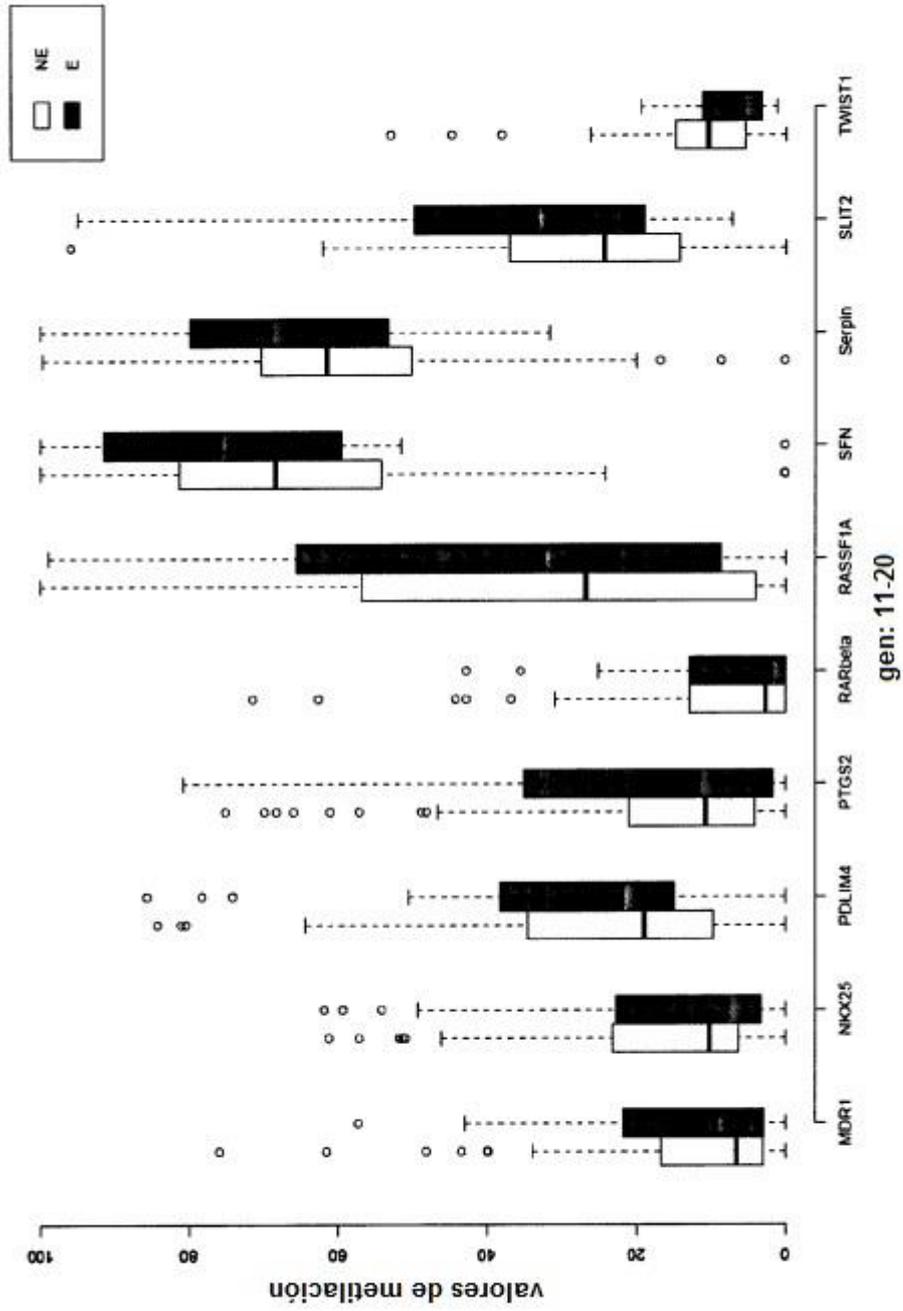


FIG. 3:

FIG 4:

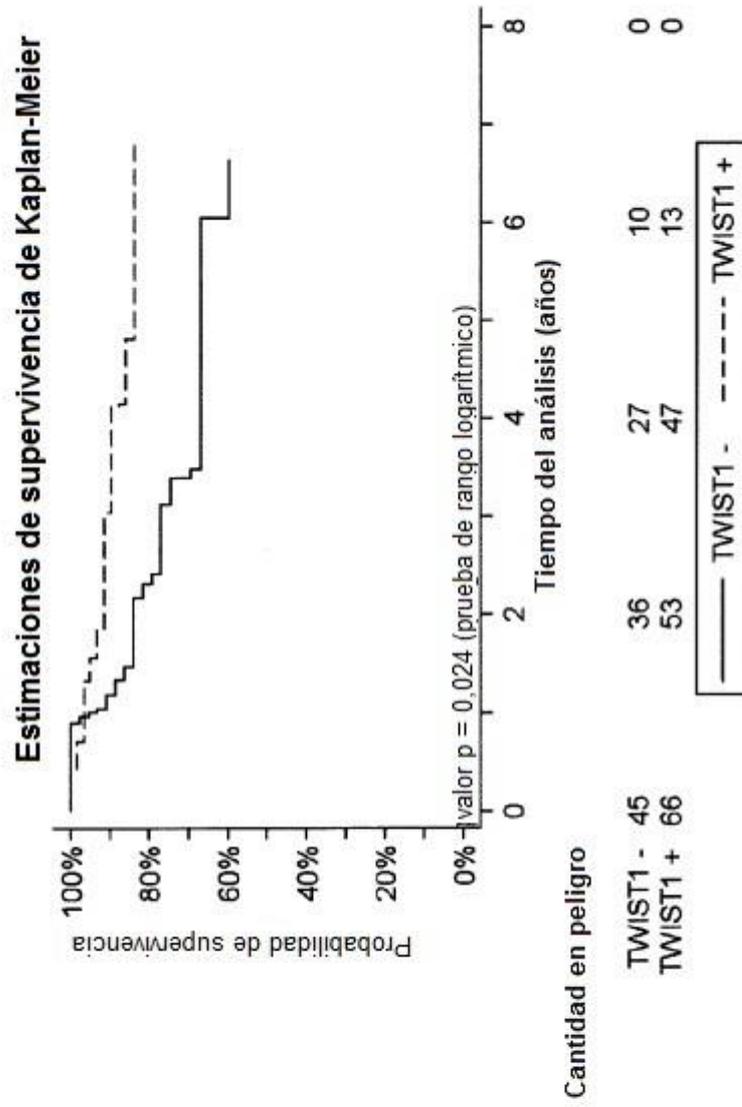


FIG 5:

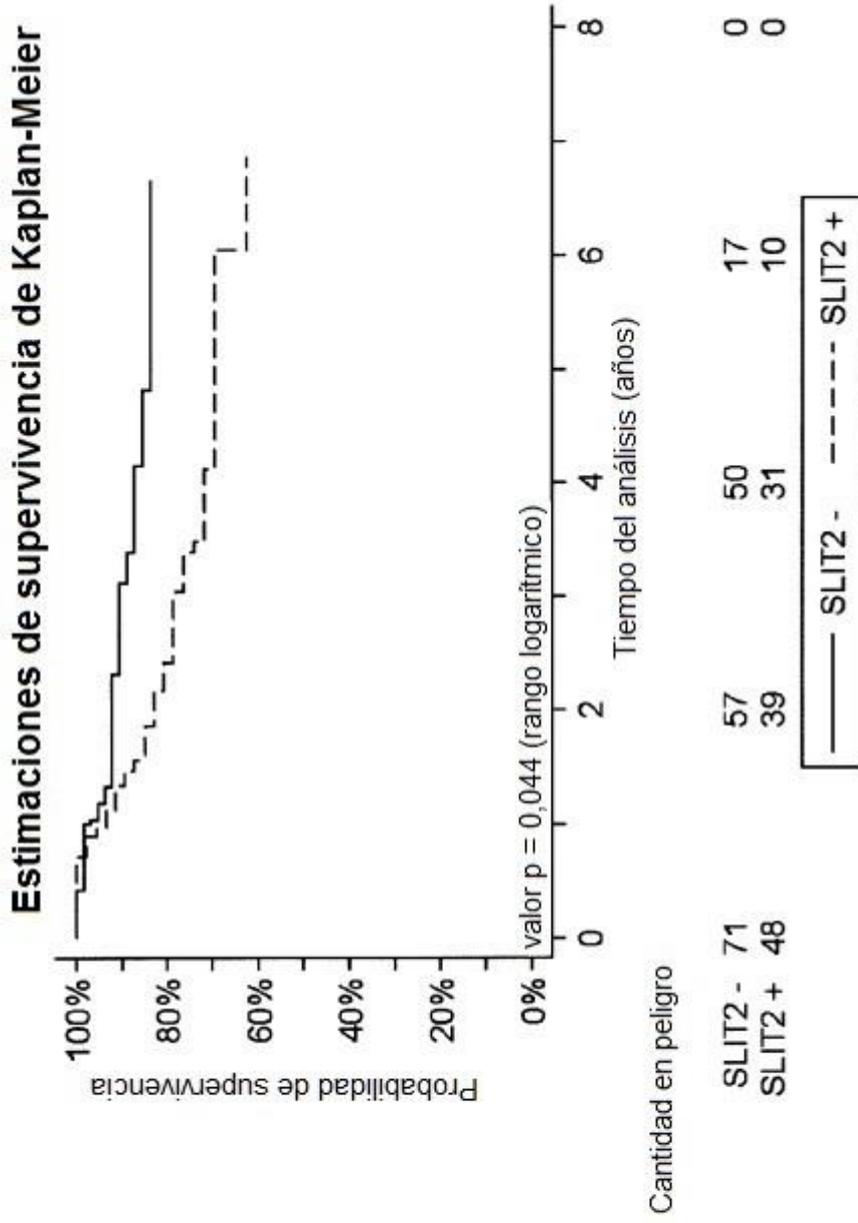


FIG 6:

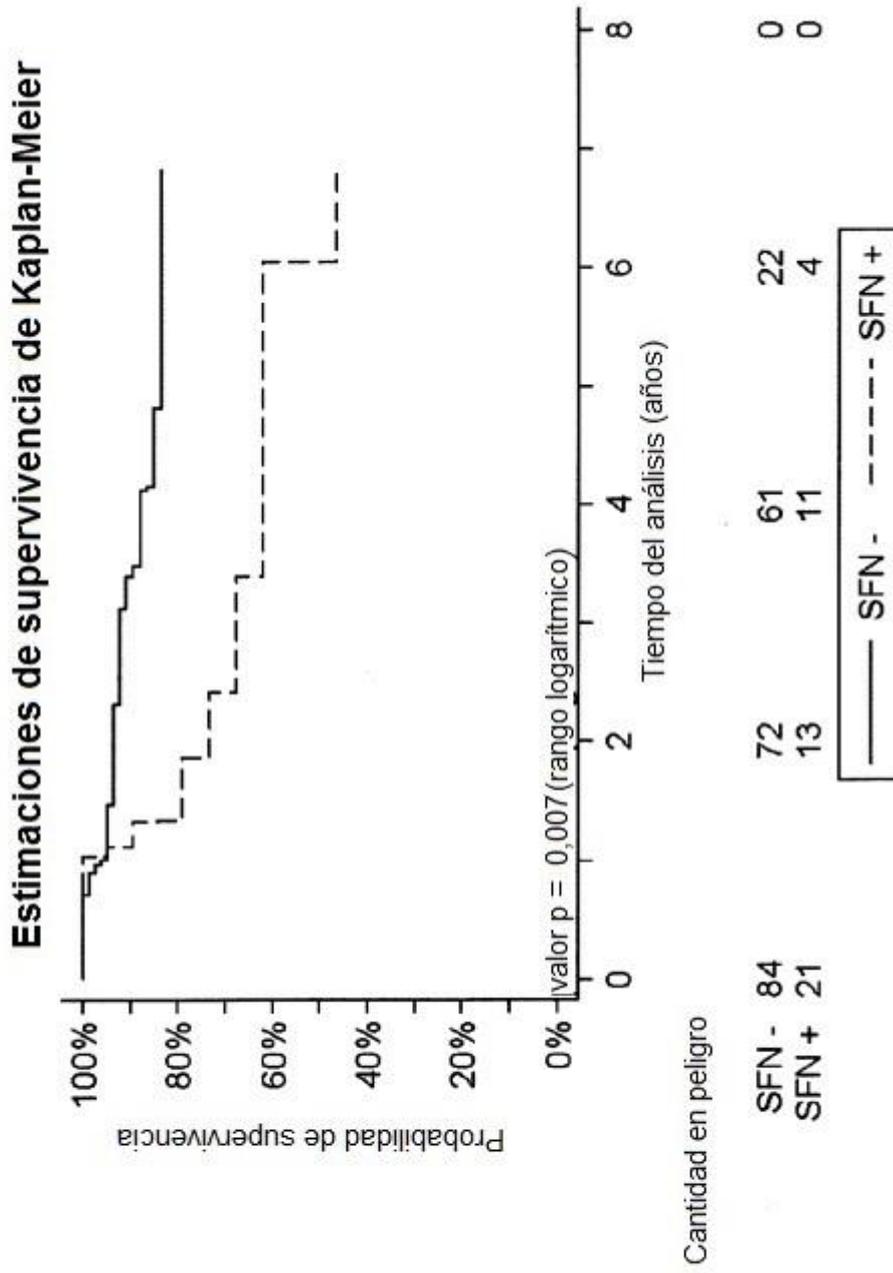


FIG 7:

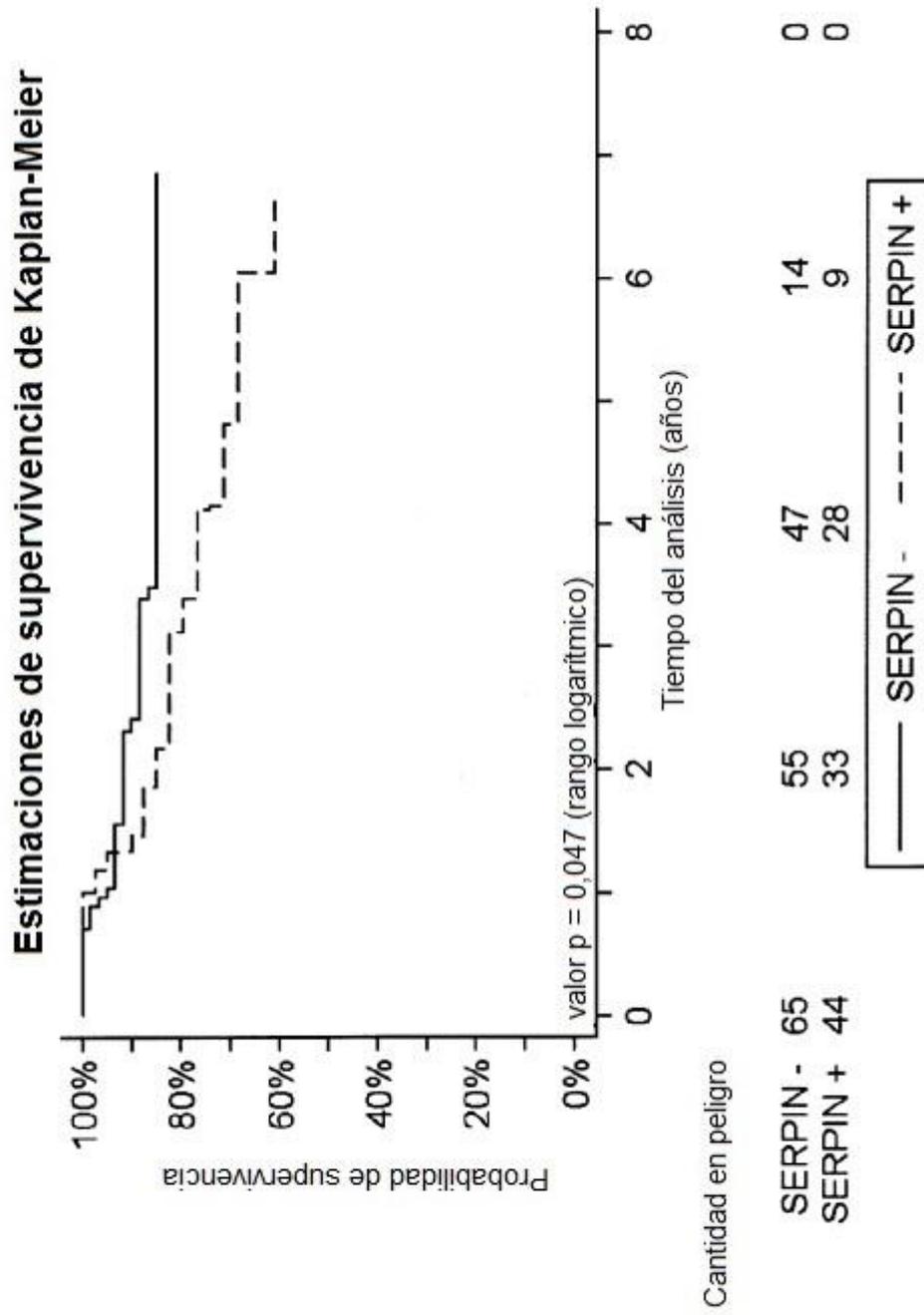


FIG 8:

