

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 546 230**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **01.06.2010 E 10782820 (4)**

97 Fecha y número de publicación de la concesión europea: **22.07.2015 EP 2438193**

54 Título: **Métodos para evaluar el riesgo de cáncer de mama**

30 Prioridad:

01.06.2009 US 182809 P
05.11.2009 US 258420 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
21.09.2015

73 Titular/es:

GENETIC TECHNOLOGIES LIMITED (100.0%)
60-66 Hanover Street
Fitzroy, VIC 3065, AU

72 Inventor/es:

HINDS, DAVID A. y
WALSER, BRYAN

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

Observaciones :

Véase nota informativa (Remarks) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 546 230 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Métodos para evaluar el riesgo de cáncer de mama

Campo de la invención

5 La presente invención se relaciona con métodos y sistemas para evaluar el riesgo total de una mujer para desarrollar un fenotipo de cáncer de mama. En particular, la presente invención se relaciona con la combinación de la evaluación del riesgo clínico y la evaluación del riesgo genético para mejorar los análisis de riesgo.

Antecedentes de la invención

10 El cáncer de mama, al igual que otros cánceres comunes, muestra un agrupamiento familiar. Numerosos estudios epidemiológicos han demostrado que, en general, la enfermedad es aproximadamente dos veces más común en familiares de primer grado de consanguineidad de pacientes con cáncer de mama. Estudios de familia, y en particular estudios con gemelos, sugieren que la mayoría si no es que la totalidad de este agrupamiento tiene una base genética. Por ejemplo, Peto y Mack (2000) estimaron que el riesgo de cáncer de mama en gemelos MZ de una mujer afectada era aproximadamente cuatro veces mayor que el riesgo para una hermana de un caso.

15 Ya se han identificado diversos genes de susceptibilidad de cáncer de mama, de forma más importante BRCA1 y BRCA2. Las mutaciones en estos genes confieren un alto riesgo de cáncer de mama (del orden del 65% y del 45%, respectivamente, a los 70 años). La detección de mutaciones de una serie de casos de cáncer de mama basada en la población ha mostrado que sólo aproximadamente el 15% del riesgo familiar de cáncer de mama se puede explicar por las mutaciones en estos genes. Los otros genes conocidos de susceptibilidad al cáncer de mama (TP53, PTEN, ATM, CHEK2) sólo hacen pequeñas contribuciones al riesgo familiar (debido a que las mutaciones de predisposición son raras y/o confieren sólo un pequeño riesgo). Por lo tanto, en total, se ha estimado que los genes conocidos de susceptibilidad al cáncer de mama representan no más del 20% del riesgo familiar.

25 La variación genética en riesgo puede resultar de mutaciones raras altamente penetrantes (tales como aquellas en BRCA1 y BRCA2) o de variantes que confieren riesgos más moderados. Diversas líneas de evidencia sugieren fuertemente que las mutaciones de alta penetración no son contribuyentes principales para el riesgo familiar residual de cáncer de mama. En primer lugar, la detección de mutaciones en familias con múltiples casos ha encontrado que la gran mayoría de los casos con una historial familiar muy fuerte (por ejemplo, cuatro o más familiares afectados) alberga mutaciones en BRCA1 o BRCA2. En segundo lugar, a pesar de los grandes esfuerzos durante los últimos nueve años, los estudios de vinculación genética no han identificado ningún loci adicional relacionado. En tercer lugar, los análisis de segregación de grandes series de familias con cáncer de mama no han encontrado evidencia, después del ajuste para BRCA1 y BRCA2, de un alelo adicional de susceptibilidad al cáncer de mama dominante significativo. En un gran análisis, Antoniou y colaboradores (2009) encontraron que el modelo más parsimonioso para cáncer de mama fue un modelo poligénico, equivalente a un gran número de loci de pequeño efecto que se combinan en forma multiplicativa.

35 Lo que se necesita en la técnica, por lo tanto, son más métodos para evaluar la susceptibilidad al cáncer de mama en sujetos femeninos.

Resumen de la invención

40 La presente invención se relaciona con un método para evaluar el riesgo total de un sujeto femenino humano para desarrollar un fenotipo de cáncer de mama. Sorprendentemente, los autores han encontrado que combinar polimorfismos de un solo nucleótido con una evaluación de riesgo clínico proporciona una ventaja con respecto a la utilización de la evaluación de riesgo clínico en forma aislada. De esta forma, en un primer aspecto, la presente invención proporciona un método para evaluar el riesgo total de un sujeto femenino humano de desarrollar un fenotipo de cáncer de mama que comprende:

realizar una evaluación del riesgo clínico del sujeto femenino,

45 realizar una evaluación del riesgo genético del sujeto femenino, en donde la evaluación del riesgo genético implica detectar, en una muestra biológica derivada del sujeto femenino, la presencia de al menos cinco polimorfismos de un solo nucleótido que se sabe que están asociados con un fenotipo de cáncer de mama, seleccionado de entre el grupo que consiste de rs2981582, rs3803662, rs889312, rs123387042, rs13281615, rs4415084, rs3817198, rs4973768, rs6504950 y rs11249433; y

50 combinar la evaluación del riesgo clínico con la evaluación del riesgo genético para obtener el riesgo total de un sujeto femenino humano para desarrollar un fenotipo de cáncer de mama, en donde al menos cinco polimorfismos de un solo nucleótido comprenden al menos rs2981582, rs3803662, rs889312, rs13387042 y rs4415084.

La presente invención también es útil para reclasificar a sujetos que han sido analizados utilizando una evaluación del riesgo clínico. Por lo tanto, en un primer aspecto, la presente invención proporciona un método para determinar si el riesgo total de un sujeto femenino humano de desarrollar un fenotipo de cáncer de mama analizado mediante la evaluación del riesgo clínico se debe reclasificar, comprendiendo el método:

- 5 combinar la evaluación del riesgo clínico con una evaluación del riesgo genético para obtener el riesgo total de un sujeto femenino humano de desarrollar un fenotipo de cáncer de mama, en donde la evaluación del riesgo genético implica detectar, en una muestra biológica derivada del sujeto femenino, la presencia de al menos cinco polimorfismos de un solo nucleótido que se sabe que están asociados con un fenotipo de cáncer de mama
- 10 seleccionado de entre el grupo que consiste de rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198, rs4973768, rs6504950 y rs11249433, y
- determinar si la combinación de la evaluación del riesgo clínico con la evaluación del riesgo genético da como resultado una reclasificación del riesgo total del sujeto de desarrollar un fenotipo de cáncer de mama,
- en donde al menos cinco polimorfismos de un solo nucleótido comprenden al menos rs2981582, rs3803662, rs889312, rs13387042 y rs4415084.
- 15 En una realización, la mejora neta de reclasificación (NRI) del método es mayor a 0.01, más preferiblemente mayor a 0.05, y aún más preferiblemente mayor a 0.08. En una realización adicional, la NRI es de aproximadamente 0.09.
- En una realización, el riesgo en 5 años determinado por la evaluación del riesgo clínico está entre aproximadamente 1.5% hasta 2%. En esta realización, se prefiere que la mejora neta de reclasificación (NRI) del método sea mayor a 0.1, más preferiblemente mayor a 0.12, y aún más preferiblemente mayor a 0.15. En una realización adicional, la
- 20 NRI es de aproximadamente 0.195.
- Los ejemplos de procedimientos adecuados para la evaluación del riesgo clínico incluyen, pero no se limitan a, el modelo de Gail, el modelo de Claus, las tablas de Claus, BOADICEA, el modelo de Jonker, la fórmula extendida de Claus, el modelo de Tyrer-Cuzick, y el sistema de puntuación de Manchester. De preferencia, la evaluación del riesgo clínico se obtiene utilizando el modelo de Gail.
- 25 En una realización, la evaluación del riesgo clínico incluye obtener información de la mujer en uno o más de los siguientes tópicos: historial médico de cáncer de mama, carcinoma ductal o carcinoma lobular, edad, edad del primer periodo menstrual, edad a la cual tuvo su primer alumbramiento, historial familiar de cáncer de mama, resultados de biopsias anteriores de mama y raza/origen étnico.
- 30 El método comprende detectar la presencia de al menos cinco, seis, siete, ocho, nueve o diez polimorfismos de un solo nucleótido que se sabe que están asociados con un fenotipo de cáncer de mama. Más preferiblemente, el método comprende detectar la presencia de siete, ocho, nueve o diez polimorfismos de un solo nucleótido que se sabe que están asociados con un fenotipo de cáncer de mama.
- Los polimorfismos de un solo nucleótido se seleccionan de un grupo que consiste en rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198, rs4973768, rs6504950 y rs11249433. Se prefiere que el
- 35 método comprende detectar la presencia de al menos cinco, seis, siete, ocho, nueve o diez de los polimorfismos de un solo nucleótido, más preferiblemente siete, ocho, nueve o diez de los polimorfismos de un solo nucleótido.
- En una realización, los polimorfismos de un solo nucleótido se prueban individualmente por la asociación con cáncer de mama mediante regresión logística bajo un modelo aditivo logarítmico sin covariables.
- 40 En otra realización, los polimorfismos de un solo nucleótido se prueban en combinación por la asociación con cáncer de mama.
- Preferiblemente, la combinación de la evaluación del riesgo clínico con la evaluación del riesgo genético comprende multiplicar las evaluaciones de riesgo.
- La hembra puede ser de cualquier raza, tal como caucásica, negroide, australoide o mongoloide. En una realización preferida, la hembra es caucásica. Preferiblemente, la hembra es postmenopáusica.
- 45 Sorprendentemente, los inventores han encontrado que la combinación de análisis de polimorfismo de un solo nucleótido con evaluación de riesgo clínico proporciona un claro beneficio en aquellas mujeres que han tenido una biopsia previa de la mama. Por lo tanto, en una forma preferida de realización, la hembra ha tenido una biopsia de mama.

- Los inventores también han encontrado que la combinación de análisis de polimorfismo de un solo nucleótido con la evaluación del riesgo clínico proporciona un claro beneficio en la posible reclasificación de las mujeres cuya evaluación de riesgo clínico indica que tienen un riesgo de desarrollar cáncer de mama. Por lo tanto, en otra realización preferida, los resultados de la evaluación del riesgo clínico indican que la hembra debe ser sometida a detección más frecuente y/o terapia profiláctica contra el cáncer.
- En una realización adicional, si se determina que el sujeto tiene riesgo de desarrollar cáncer de mama utilizando un método de la invención, es más probable que el sujeto sea sensible a la inhibición de estrógeno, y no que no sea sensible.
- También se describe en este documento un sistema para evaluar el riesgo global de un sujeto femenino humano de desarrollar un fenotipo de cáncer de mama que comprende:
- instrucciones del sistema para llevar a cabo una evaluación del riesgo clínico del sujeto femenino;
- instrucciones del sistema para llevar a cabo una evaluación del riesgo genético del sujeto femenino; y
- instrucciones del sistema para combinar la evaluación del riesgo clínico con la evaluación del riesgo genético para obtener el riesgo total de un sujeto femenino humano de desarrollar un fenotipo cáncer de mama.
- El sistema puede ser un sistema implementado por ordenador.
- El sistema puede comprender además:
- un conjunto de sondas o cebadores marcadores configurado para detectar, en una muestra biológica derivada del sujeto femenino, la presencia de al menos dos polimorfismos de un solo nucleótido que se sabe que están asociados con un fenotipo de cáncer de mama;
- un detector que está configurado para detectar una o más salidas de señal del conjunto de sondas o cebadores marcadores, o un amplificación producido a partir del conjunto de sondas o cebadores marcadores, identificando de este modo la presencia o ausencia de al menos dos polimorfismos de un solo nucleótido que se sabe que están asociados con un fenotipo cáncer de mama.
- El detector puede detectar una o más emisiones de luz, en donde la emisión de luz es indicativa de la presencia o ausencia del polimorfismo de un solo nucleótido.
- El sistema puede comprender una muestra biológica derivada del sujeto femenino. En una realización adicional, la muestra comprende ADN genómico, ADN genómico amplificado, ADNc, ADNc amplificado, ARN o ARN amplificado.
- En aún otro aspecto, la presente invención proporciona un kit que comprende al menos cinco conjuntos de cebadores para amplificar cinco o más ácidos nucleicos, en donde los cinco o más ácidos nucleicos comprenden independientemente uno de al menos cinco polimorfismos de un solo nucleótido seleccionados de un grupo que consiste en rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198, rs4973768, rs6504950 y rs11249433, en donde al menos cinco polimorfismos de un solo nucleótido comprenden al menos rs2981582, rs3803662, rs889312, rs13387042 y rs4415084.
- Los ejemplos de cebadores adecuados para un kit de la invención se proporcionan en la Tabla 6.
- El kit puede comprender además otros reactivos necesarios para realizar una reacción de amplificación tal como un regulador, nucleótidos y/o una polimerasa, así como los reactivos para la extracción de ácidos nucleicos de una muestra.
- En un aspecto adicional, la presente invención proporciona una matriz genética que comprende al menos cinco ácidos nucleicos que comprenden independientemente uno de al menos cinco polimorfismos de un solo nucleótido seleccionado de un grupo que consiste en rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198, rs4973768, rs6504950 y rs11249433, en donde los al menos cinco polimorfismos de un solo nucleótido, comprenden al menos rs2981582, rs3803662, rs889312, rs13387042 y rs4415084.
- Preferiblemente, la matriz comprende ácidos nucleicos que representan a cada alelo conocido de cada SNP.
- En un aspecto adicional, la presente invención proporciona un método para determinar la necesidad de una prueba diagnóstica de rutina de un sujeto femenino humano para cáncer de mama que comprende la evaluación del riesgo global del sujeto de desarrollar un fenotipo de cáncer de mama usando un método de la invención.

En aún otro aspecto, la presente invención proporciona un método de detección para el cáncer de mama en un sujeto femenino humano, comprendiendo el método evaluar el riesgo total del sujeto a desarrollar un fenotipo del cáncer de mama utilizando el método de la invención, y la detección de rutina para el cáncer de mama en el sujeto si se evalúan por tener un riesgo de desarrollar cáncer de mama.

- 5 Los ejemplos de métodos adecuados de detección del cáncer de mama incluyen, pero no se limitan a, el examen manual de la mama y la mamografía.

En un aspecto adicional, la presente invención proporciona un método para determinar la necesidad de un sujeto humano femenino de una terapia profiláctica contra el cáncer de mama que comprende la evaluación del riesgo global del sujeto de desarrollar un fenotipo de cáncer de mama utilizando el método de la invención.

- 10 En un aspecto aún adicional, la presente invención proporciona un método para prevenir el cáncer de mama en un sujeto femenino humano, comprendiendo el método la evaluación del riesgo global del sujeto de desarrollar un fenotipo de cáncer de mama utilizando el método de la invención, y la administración de una terapia contra el cáncer de mama al sujeto si se evalúan como si tuvieran riesgo de desarrollar cáncer de mama.

- 15 Los ejemplos de terapias adecuadas contra el cáncer de mama incluyen, pero no se limitan a, la administración de tamoxifeno o raloxifeno.

En una realización, la terapia inhibe al estrógeno.

- 20 En un aspecto adicional, la presente invención proporciona un método para estratificar un grupo de sujetos femeninos humanos para un ensayo clínico de una terapia candidata, comprendiendo el método la evaluación del riesgo global individual de los sujetos de desarrollar un fenotipo de cáncer de mama utilizando el método de la invención, y utilizando los resultados de la evaluación para seleccionar los sujetos más propensos a ser sensibles a la terapia.

- 25 Como será evidente, por lo menos algunas de las características de los métodos, kits y sistemas se pueden utilizar de manera combinada. Por ejemplo, los sistemas para la detección de moduladores se pueden utilizar para llevar a cabo los métodos de detección del modulador. Se pueden utilizar sistemas para identificar correlaciones entre la susceptibilidad del fenotipo de cáncer de mama y polimorfismos para llevar a la práctica los métodos en la presente invención. Los kits se pueden utilizar para la práctica de los métodos en el presente documento. Por lo tanto, se pueden aplicar las características descritas de los sistemas, métodos y kits a los diferentes sistemas, métodos y kits de la presente invención.

- 30 A lo largo de esta memoria descriptiva, la palabra "comprende", o variaciones tales como "que comprende", se entenderá que implican la inclusión de un elemento, número entero o etapa, o grupo de elementos, números enteros o etapas establecidos, pero no la exclusión de cualquier otro elemento, número entero o etapa, o grupo de elementos, números enteros o etapas.

La invención se describe a continuación por medio de los siguientes ejemplos no limitantes y con referencia a las figuras acompañantes.

- 35 Breve descripción de los dibujos acompañantes

La figura 1 representa los resultados de una prueba de Hosmer-Lemeshow para evaluar la calibración de los puntajes de riesgo 10-SNP.

La figura 2 representa las curvas características de funcionamiento del receptor para los puntajes de riesgo 7-SNP y 10-SNP.

- 40 Descripción detallada de la invención

Técnicas generales y definiciones

- 45 A menos que específicamente se defina otra cosa, todos los términos técnicos y científicos utilizados en la presente invención se entenderán con el mismo significado como lo entiende comúnmente alguien ordinariamente capacitado en la técnica (por ejemplo, en cultivo celular, análisis de cáncer de mama, genética molecular, inmunología, inmunohistoquímica, química de proteínas y bioquímica).

A menos que se indique otra cosa, la proteína recombinante, el cultivo celular y las técnicas inmunológicas utilizadas en la presente invención son procedimientos estándar, bien conocidos por aquellos capacitados en la técnica. Estas

- técnicas se describen y explican a todo lo largo de la literatura en fuentes tales como, J. Perbal, A Practical Guide to Molecular Cloning, John Wiley and Sons (1984), J. Sambrook y colaboradores, Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory Press (1989), T.A. Brown (editor), Essential Molecular Biology: A Practical Approach, Volúmenes 1 y 2, IRL Press (1991), D.M. Glover y B.D. Hames (editores), DNA Cloning: A Practical Approach, volúmenes 1-4, IRL Press (1995 y 1996), y F.M. Ausubel y colaboradores (editores), Current Protocols in Molecular Biology, Greene Pub. Associates and Wiley-Interscience (1988, incluyendo todas las actualizaciones hasta la fecha), Ed Harlow y David Lane (editores) Antibodies: A Laboratory Manual, Cold Spring Harbour Laboratory, (1988), y J.E. Coligan y colaboradores (editores) Current Protocols in Immunology, John Wiley & Sons (incluyendo todas las actualizaciones hasta la fecha).
- Se debe entender que esta invención no está limitada a realizaciones particulares, las cuales, por supuesto, pueden variar. También se debe entender que la terminología utilizada en la presente invención tiene el fin de describir únicamente realizaciones particulares, y no pretenden ser limitante. Como se utiliza en esta memoria descriptiva y en las reivindicaciones anexas, los términos en singular y las formas en singular “un”, “uno, una” y “el, la”, por ejemplo, incluyen opcionalmente referencias al plural a menos que el contenido claramente dicte otra cosa. De esta forma, por ejemplo, la referencia a “una sonda” incluye opcionalmente una pluralidad de moléculas sonda; de manera similar, dependiendo del contexto, el uso del término “un ácido nucleico” incluye opcionalmente, como una cuestión práctica, muchas copias de esa molécula de ácido nucleico.
- Como se utiliza en la presente invención, “fenotipo de cáncer de mama” se refiere a un fenotipo que muestra una predisposición a desarrollar cáncer de mama en un individuo. Un fenotipo que muestra una predisposición al cáncer de mama, puede, por ejemplo, mostrar una mayor probabilidad de que el cáncer se desarrollará en un individuo con el fenotipo que en miembros de una población general relevante bajo un conjunto dado de condiciones ambientales (dieta, régimen de actividad física, ubicación geográfica, etc.).
- Como se utiliza en la presente invención, “muestra biológica” se refiere a cualquier muestra que pueda ser de o derivarse de un paciente humano, por ejemplo, líquidos corporales (sangre, saliva, orina, etc.), biopsia, tejido, y/o desechos del paciente. Por lo tanto, las biopsias de tejido, heces, esputo, saliva, sangre, linfa, lágrimas, sudor, orina, secreciones vaginales, o similares se pueden detectar fácilmente para los SNP, al igual que esencialmente cualquier tejido de interés que contenga los ácidos nucleicos apropiados. Estas muestras típicamente se toman, después del consentimiento informado, de un paciente mediante métodos de laboratorio médico estándar. La muestra puede estar en una forma tomada directamente del paciente, o puede estar al menos parcialmente procesada (purificada) para eliminar al menos algo del material que no sea ácido nucleico.
- Un “polimorfismo” es un locus que es variable, es decir, dentro de una población, la secuencia de nucleótidos en un polimorfismo tiene más de una versión o alelo. Un ejemplo de un polimorfismo es un “polimorfismo de un solo nucleótido”, que es un polimorfismo en una posición de un solo nucleótido en un genoma (el nucleótido en la posición específica varía entre individuos o poblaciones).
- Como se utiliza en la presente invención, el término “SNP” o “polimorfismo de un solo nucleótido” se refiere a una variación genética entre individuos; por ejemplo, una posición de una sola base nitrogenada en el ADN de organismos que sea variable. Como se utiliza en la presente invención, “SNPs” es el plural de SNP. Por supuesto, cuando se hace referencia al ADN en la presente invención, esta referencia puede incluir los derivados del ADN, tales como amplicones, transcritos de ARN de los mismos, etc.
- El término “alelo” se refiere a una de dos o más secuencias diferentes de nucleótidos que se presentan o se codifican en un locus específico, o dos o más secuencias diferentes de polipéptidos codificadas por este locus. Por ejemplo, un primer alelo se puede presentar en un cromosoma, mientras que un segundo alelo se presenta en un segundo cromosoma homólogo, por ejemplo, como ocurre para diferentes cromosomas de un individuo heterocigoto, o entre diferentes individuos homocigotos o heterocigotos en una población. Un alelo se correlaciona “positivamente” con un rasgo cuando está relacionado con el mismo y cuando la presencia del alelo es un indicador de que el rasgo o la forma del rasgo se presentará en un individuo que contiene el alelo. Un alelo se correlaciona “negativamente” con un rasgo cuando está relacionado con el mismo y cuando la presencia del alelo es un indicador de que un rasgo o forma del rasgo no se presentará en un individuo que contiene el alelo.
- Un polimorfismo marcador o alelo se “correlaciona” o “asocia” con un fenotipo específico (susceptibilidad al cáncer de mama, etc.) cuando se puede relacionar estadísticamente (positiva o negativamente) con el fenotipo. Los métodos para determinar si un polimorfismo o alelo está estadísticamente relacionado, son conocidos por aquellos capacitados en la técnica. Es decir, el polimorfismo específico se presenta de manera más común en una población particular (por ejemplo, pacientes con cáncer de mama) que en una población control (por ejemplo, individuos que no tengan cáncer de mama). A menudo se concluye que esta correlación es de naturaleza causal, aunque no es necesario que esté relacionada genéticamente de forma simple (asociación con) un locus para que un rasgo que depende del fenotipo sea suficiente para que se presente la correlación/asociación.

La frase “desequilibrio de relación” (LD) se utiliza para describir la correlación estadística entre dos genotipos polimórficos vecinos. Típicamente, el LD se refiere a la correlación entre los alelos de un gameto aleatorio en los dos loci, asumiendo un equilibrio de Hardy-Weinberg (independencia estadística) entre gametos. El LD se cuantifica ya sea con un parámetro de asociación de Lewontin (D') o con un coeficiente de correlación de Pearson (r) (Devlin y Risch, 1995). Se dice que dos loci con un valor LD de 1 están en LD completa. En el otro extremo, dos loci con un valor LD de 0 se dice que están en equilibrio de relación. El desequilibrio de relación se calcula después de la aplicación del algoritmo de maximización de expectativa (EM) para la estimación de las frecuencias haplotípicas (Slatkin y Excoffier, 1996). Los valores de LD de acuerdo con la presente invención para genotipos/loci vecinos se seleccionan por encima de 0.1, preferiblemente por encima de 0.2, más preferiblemente por encima de 0.5, más preferiblemente por encima de 0.6, aún más preferiblemente por encima de 0.7, preferiblemente por encima de 0.8, más preferiblemente por encima de 0.9, idealmente aproximadamente 1.0.

“Frecuencia de alelos” se refiere a la frecuencia (proporción o porcentaje) en la cual está presente un alelo en un locus dentro de un individuo, dentro de una línea o dentro de una población de líneas. Por ejemplo, para un alelo “A”, individuos diploides del genotipo “AA”, “Aa”, o “aa” tienen frecuencias alélicas de 1.0, 0.5, o 0.0, respectivamente. Se puede estimar la frecuencia de alelos dentro de una línea o población (por ejemplo, casos o controles) promediando las frecuencias de alelos de una muestra de individuos de esa línea o población. De manera similar, se puede calcular la frecuencia de alelos dentro de una población de líneas al promediar las frecuencias de alelos de líneas que constituyen la población.

Un individuo es “homocigoto” si el individuo tiene sólo un tipo de alelo en un locus determinado (por ejemplo, un individuo diploide tiene una copia del mismo alelo en un locus para cada uno de los dos cromosomas homólogos). Un individuo es “heterocigoto” si está presente más de un tipo de alelo en un locus determinado (por ejemplo, un individuo diploide con una copia de cada uno de los dos alelos diferentes). El término “homogeneidad” indica que los miembros de un grupo tienen el mismo genotipo en uno o más loci específicos. Por el contrario, el término “heterogeneidad” se utiliza para indicar que los individuos dentro del grupo difieren en el genotipo en uno o más loci específicos.

Un “locus” es una posición o región cromosómica. Por ejemplo, un locus polimórfico es una posición o región donde se localiza un ácido nucleico polimórfico, un determinante de un rasgo, un gen o un marcador. En un ejemplo adicional, un “locus génico” es una ubicación (región) específica del cromosoma en el genoma de una especie donde se puede encontrar un gen específico.

Un “marcador”, “marcador molecular” o “ácido nucleico marcador” se refiere a una secuencia de nucleótidos o un producto codificado por la misma (por ejemplo, una proteína) utilizada como un punto de referencia para identificar un locus o un locus relacionado. Un marcador se puede derivar de una secuencia genómica de nucleótidos o de secuencias expresadas de nucleótidos (por ejemplo, de un ARN, ARNn, ARNm, un ADNc, etc.), o de un polipéptido codificado. El término también se refiere a secuencias de ácido nucleico complementarias o que flanquean las secuencias marcadoras, tal como los ácidos nucleicos utilizados como sondas o pares de cebadores capaces de amplificar la secuencia marcadora. Una “sonda marcadora” es una secuencia de ácido nucleico o molécula que se pueda utilizar para identificar la presencia de un locus marcador, por ejemplo, una sonda de ácido nucleico que sea complementaria a una secuencia del locus marcador. Los ácidos nucleicos son “complementarios” cuando se hibridan específicamente en solución, por ejemplo, de acuerdo con las reglas de apareamiento de bases de Watson-Crick. Un “locus marcador” es un locus que se puede utilizar para rastrear la presencia de un segundo locus relacionado, por ejemplo, un locus ligado o correlacionado, que codifica o contribuye a la variación de la población de un rasgo fenotípico. Por ejemplo, se puede utilizar un locus marcador para controlar la segregación de alelos en un locus, tal como un QTL, que están genética o físicamente relacionados con el locus marcador. De esta forma, un “alelo marcador”, alternativamente un “alelo de un locus marcador” es uno de una pluralidad de secuencias polimórficas de nucleótidos encontradas en un locus marcador en una población que sea polimórfica para el locus marcador. En un aspecto, la presente invención proporciona los loci marcadores que se correlacionan con un fenotipo de interés, por ejemplo, la susceptibilidad/resistencia al cáncer de mama. Se espera que cada uno de los marcadores identificados esté en proximidad cercana física y genética (dando como resultado en un enlazamiento físico y/o genético) con un elemento genético, por ejemplo, un QTL, que contribuye al fenotipo relevante. Los marcadores correspondientes a los polimorfismos genéticos entre los miembros de una población se pueden detectar mediante métodos bien establecidos en la técnica. Estos incluyen, por ejemplo, métodos de amplificación específica de la secuencias con base en una PCR, detección de los polimorfismos de restricción de la longitud del fragmento (RFLP), detección de marcadores de isoenzimas, detección de la hibridación específica del alelo (ASH), detección de la extensión de un solo nucleótido, detección de las secuencias variables amplificadas del genoma, detección de la replicación de secuencias autosostenida, detección de repeticiones sencillas de secuencia (SSR), detección de polimorfismos de un solo nucleótido (SNP), o detección de polimorfismos de longitud amplificada de fragmento (AFLP).

El término “amplificación” en el contexto de la amplificación de ácido nucleico es cualquier proceso mediante el cual se producen copias adicionales de un ácido nucleico seleccionado (o una forma transcrita del mismo). Los métodos típicos de amplificación incluyen diversos métodos de replicación con base en la polimerasa, incluyendo la reacción

en cadena de la polimerasa (PCR), los métodos mediados por ligasa tales como la reacción en cadena de la ligasa (LCR) y los métodos de amplificación con base en la ARN polimerasa (por ejemplo, mediante transcripción).

5 Un "amplicón" es un ácido nucleico amplificado, por ejemplo, un ácido nucleico que se produce por amplificación de un ácido nucleico molde mediante cualquier método de amplificación disponible (por ejemplo, PCR, LCR, transcripción, o similares).

Un ácido nucleico específico se "deriva de" un ácido nucleico determinado cuando se construye utilizando la secuencia dada de ácido nucleico, o cuando se construye el ácido nucleico específico utilizando el ácido nucleico dado.

10 Un "gen" es una o más secuencias de nucleótidos en un genoma que conjuntamente codifican una o más moléculas expresadas, por ejemplo, un ARN, o polipéptido. El gen puede incluir secuencias codificadoras que se transcriben en el ARN que luego se pueden traducir en una secuencia de polipéptido, y puede incluir secuencias estructurales o reguladoras asociadas que ayudan en la replicación o expresión del gen.

15 Un "genotipo" es la constitución genética de un individuo (el grupo de individuos) en uno o más loci genéticos. El genotipo se define por el(los) alelo(s) de uno o más loci conocidos del individuo, típicamente, la recopilación de alelos heredados de sus padres.

Un "haplotipo" es el genotipo de un individuo en una pluralidad de loci genéticos en una sola cadena de ADN. Típicamente, los loci genéticos descritos por un haplotipo se ligan física y genéticamente, es decir, en la misma cadena del cromosoma.

20 Un "conjunto" de marcadores, sondas o cebadores se refiere a una colección o grupo de sondas marcadoras, cebadores, o los datos derivados de los mismos, utilizados para un fin común, por ejemplo, identificar a un individuo con un genotipo específico (por ejemplo, en riesgo de desarrollar cáncer de mama). Con frecuencia, los datos correspondientes a los marcadores, sondas o cebadores, o derivados de su uso, se almacenan en un medio electrónico. Aunque cada uno de los miembros de un conjunto posee utilidad con respecto al fin específico, los marcadores individuales seleccionados del conjunto así como los subconjuntos que incluyen algunos, pero no todos los marcadores, también son efectivos para alcanzar el fin específico.

25 Los polimorfismos y genes, y las sondas marcadoras correspondientes, amplicones o cebadores descritos anteriormente se pueden incorporar en cualquier sistema en la presente invención, ya sea en la forma ácidos nucleicos físicos, o en la forma de instrucciones del sistema que incluyen información de secuencia para los ácidos nucleicos. Por ejemplo, el sistema puede incluir cebadores o amplicones correspondientes a (o que amplifican una porción de) un gen o polimorfismo descritos en la presente invención. Al igual que en los métodos anteriores, el conjunto de sondas marcadoras o cebadores detecta opcionalmente una pluralidad de polimorfismos en una pluralidad de dichos genes o loci genéticos. De esta forma, por ejemplo, el conjunto de sondas marcadoras o cebadores detecta al menos un polimorfismo en cada uno de estos polimorfismos o genes, o cualquier otro polimorfismo, gen o locus definido en la presente invención. Cualquiera de estas sondas o cebadores puede incluir una secuencia de nucleótidos de cualquiera de estos polimorfismos o genes, o un ácido nucleico complementario del mismo, o un producto transcrito del mismo (por ejemplo, un ARNn o una forma de ARNm producida a partir de una secuencia genómica, por ejemplo, mediante transcripción o empalme).

30 Como se utiliza en la presente invención, "regresión logística" se refiere a los métodos para predecir la probabilidad de ocurrencia de un evento al ajustar los datos a una curva logística. Una persona capacitada en la técnica entenderá la forma de emplear estos métodos en el contexto de la invención.

35 Como se utiliza en la presente invención, "prueba de Hosmer-Lemeshow" se refiere a un método estadístico para medir la falta de ajuste. Los métodos para utilizarla son bien conocidos en la técnica (Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: Wiley; 1989: Sección 5.2.2).

40 Como se utiliza en la presente invención "curvas características para la operación del receptor" se refieren a una representación gráfica de la sensibilidad versus (1 - especificidad) para un sistema clasificador binario a medida que se varía su umbral de discriminación. Las ROC también se puede representar en forma equivalente al graficar la fracción de positivos verdaderos (TPR = índice de positivos verdaderos) versus la fracción de positivos falsos (FPR = índice de positivos falsos). También se conoce como una curva característica de operación relativa, debido a que es una comparación de dos características de operación (TPR y FPR) a medida que cambia el criterio. El análisis ROC proporciona las herramientas para seleccionar los modelos posiblemente óptimos y, desechar aquellos por debajo del óptimo independientemente (y antes de la especificación) del contexto del coste o la distribución de clases. El análisis ROC se relaciona de una forma directa y natural con el análisis de costo/beneficio para tomar decisiones de diagnóstico. La curva ROC fue desarrollada primero por ingenieros eléctricos e ingenieros de radares durante la Segunda Guerra Mundial para detectar objetivos enemigos en los campos de batalla, también conocida como la

teoría de detección de señales, y fue pronto introducida en psicología para tomar en cuenta la detección perceptiva de señales. Desde entonces se ha utilizado el análisis ROC en medicinas, radiología y otras áreas durante muchas décadas, y ha sido introducido relativamente recientemente en otras áreas como el aprendizaje de máquina y la minería de datos. Los métodos de uso en el contexto de la invención serán claros para aquellos capacitados en la técnica.

Como se utiliza en la presente invención el término "combinación de la evaluación del riesgo clínico con la evaluación del riesgo genético para obtener el riesgo total" se refiere a cualquier análisis matemático adecuado que dependa de los resultados de las dos evaluaciones. Por ejemplo, los resultados de la evaluación del riesgo clínico y la evaluación del riesgo genético se pueden sumar, preferiblemente multiplicar.

Como se utiliza en la presente invención, los términos "detección rutinaria del cáncer de mama" y "detección más frecuente" son términos relativos, y se basan en una comparación del nivel de detección recomendado para un sujeto que no tenga un riesgo identificado de desarrollar cáncer de mama.

Evaluación del riesgo clínico

Cualquier procedimiento adecuado de evaluación del riesgo clínico puede ser utilizado en la presente invención. Preferiblemente, la evaluación del riesgo clínico no implica la genotipificación de la mujer en uno o más loci. En una realización, el procedimiento para evaluación del riesgo clínico incluye obtener la información de la mujer en uno o más de los siguientes aspectos: historial médico de cáncer de mama, carcinoma ductal o carcinoma lobular, edad, historial menstrual tal como la edad del primer periodo menstrual, la edad en la cual tuvo su primer alumbramiento, historial familiar de cáncer de mama u otro cáncer, incluyendo la edad del familiar en el momento del diagnóstico, los resultados de biopsias de mama anteriores, el uso de anticonceptivos orales, el índice de masa corporal, el historia de consumo de alcohol, el historial como fumador, el historial de ejercitación, dieta y raza/origen étnico. Los ejemplos de procedimientos de evaluación del riesgo clínico incluyen, pero no se limitan a, el modelo de Gail (Gail y colaboradores; 1989, 1999 y 2007; Costantino y colaboradores, 1999; Rockhill y colaboradores, 2001), el modelo de Claus (Claus y colaboradores, 1994 y 1998), las tablas de Claus, BOADICEA (Antoniou y colaboradores, 2002 y 2004), el modelo de Jonker (Jonker y colaboradores, 2003), la formula extendida de Claus (van Asperen y colaboradores, 2004), el modelo de Tyrer-Cuzick (Tyrer y colaboradores, 2004), el sistema de puntuación de Manchester (Evans y colaboradores, 2004), y similares. En una realización preferida, el procedimiento para evaluación del riesgo clínico es el modelo de Gail.

El modelo de Gail es un modelo estadístico que constituye la base de una herramienta para evaluación del riesgo de cáncer de mama, denominada después Dr. Mitchell Gail, investigador principal de la rama bioestadística de la división de epidemiología del cáncer y genética del NCI. El modelo utiliza el historial médico personal de una mujer (número de biopsias de mama anteriores y la presencia de hiperplasia atípica en cualquier espécimen previo de biopsia de mama), su propio historial reproductivo (edad al inicio de la menstruación y edad en el primer alumbramiento de un niño vivo), y el historial de cáncer de mama entre sus familiares en primer grado (madre, hermanas, hijas) para estimar su riesgo de desarrollar cáncer de mama invasivo durante periodos de tiempo específicos. Los datos del Breast Cancer Detection Demonstration Project (BCDDP), que fue un estudio conjunto de detección de cáncer de mama del NCI y del American Cancer Society que implicó a 280,000 mujeres en edades de 35 a 74 años, y del programa Surveillance, Epidemiology, and End Results (SEER) del NCI, fueron utilizados para desarrollar el modelo. Los estimados para mujeres afroamericanas se basaron en los datos del estudio de experiencias contraceptivas y reproductivas de las mujeres (CARE) y de los datos de SEER. Las participantes de CARE incluyeron 1,607 mujeres con cáncer de mama invasivo y 1,637 sin el mismo.

El modelo de Gail ha sido probado en grandes poblaciones de mujeres blancas y ha demostrado que proporciona estimados precisos de riesgo de cáncer de mama. En otras palabras, el modelo ha sido "validado" para mujeres blancas. También se ha probado en los datos de la iniciativa de salud para mujeres afroamericanas, y el modelo se desempeña bien, aunque puede subestimar el riesgo en mujeres afroamericanas con biopsias anteriores. El modelo todavía necesita ser validado para mujeres hispanas, mujeres asiáticas, y otros subgrupos, y los resultados deben ser interpretados por un suministrador de cuidados para la salud para mujeres con factores de riesgo especiales, tales como mujeres tratadas por la enfermedad de Hodgkin con radiación al pecho y portadoras de mutaciones génicas que aumentan el riesgo de cáncer de mama. Los investigadores están llevando acabo estudios adicionales, incluyendo estudios con poblaciones menores, para reunir más datos y probar y mejorar el modelo.

Evaluación del riesgo genético

La evaluación del riesgo genético se realiza mediante el análisis del genotipo del sujeto en dos o más loci para los polimorfismos de un solo nucleótido que se sabe están asociados con un fenotipo de cáncer de mama. Como apreciará el experto, cada SNP tiene una proporción desigual de asociación con el cáncer de mama mayor a 1.0, más preferiblemente mayor a 1.02. Los ejemplos de estos SNP incluyen, pero no se limitan a, aquellos designados como rsZ981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198, rs4973768, rs6504950, rs11249433, rs10941679, rs2180341, rs3744448, rs16943468, rs1011970, rs2380205, rs10995190, rs704010 y

rs614367 (véase la Tabla 1). También se describe en la presente invención, un polimorfismo de un solo nucleótido en el desequilibrio de enlazamiento con uno o más de los mismos.

Tabla 1. Ejemplos de los SNP asociados con un fenotipo de cáncer de mama

Identificación de los rs de la base de datos del SNP	Gen	Ubicación	OR (95% CI) ²	Referencia
rs2981582	FGFR2	10q	1.26 (1.23-1.30)	Easton y colaboradores, 2007
rs3803662	TNRC9	16q	1.20 (1.16-1.24)	Easton y colaboradores, 2007
rs889312	MAP3K1	5q	1.13 (1.10-1.16)	Easton y colaboradores, 2007
rs13387042	(ninguno)	2q35	1.20 (1.14-1.26)	Stacey y colaboradores, 2007
rs13281615	(ninguno)	8q24	1.08 (1.05-1.11)	Easton y colaboradores, 2007
rs4415084	FGF10	5p	1.16 (1.10-1.21)	Stacey y colaboradores, 2008
rs3817198	LSP1	11p	1.07 (1.04-1.11)	Easton y colaboradores, 2007
rs4973768	SLC4A7	3p24	1.11 (1.08-1.13)	Ahmed y colaboradores, 2009
rs6504950	COX11	17q23.2	1.05 (1.03-1.09)	Ahmed y colaboradores, 2009
rs11249433	FCGR1B	1p11.2	1.14 (1.10-1.19)	Thomas y colaboradores, 2009
rs10941679		5p12	1.19	Stacey, y colaboradores 2008
rs2180341	ECHDC1	6q22	1.41	Gold, y colaboradores 2008
rs3744448	TCF1	17q21	1.2 (1.00-1.50)	Keleman, y colaboradores 2009
rs16943468	YPEL2	17	1.30 (1.10-1.80)	Keleman, y colaboradores 2009
rs1011970	CDKN2A	9p21	1.00 (0.89-1.12)	Turnbull, y colaboradores 2010
rs2380205	ANKRD16	10p15	1.14 (1.00-1.30)	Turnbull, y colaboradores 2010
rs10995190	ZNF365	10	1.01 (0.87-1.16)	Turnbull, y colaboradores 2010
rs704010		10q22	1.03 (0.90-1.17)	Turnbull, y colaboradores 2010
rs614367		11q13	1.17 (1.05-1.31)	Turnbull, y colaboradores 2010

- 5 Los polimorfismos de un solo nucleótido se seleccionan de un grupo que consiste de rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198, rs4973768, rs6504950 y rs11249433.

10 El experto puede identificar fácilmente los SNP en desequilibrio de enlazamiento con aquéllos específicamente mencionados en la presente invención. Los ejemplos de estos SNP incluyen rs1219648 y rs2420946, que están en fuerte desequilibrio de enlazamiento con rs2981582 (otros posibles ejemplos son proporcionados en la Tabla 2), rs12443621 y rs8051542 que están en fuerte desequilibrio de enlazamiento con el SNP rs3803662 (otros posibles ejemplos son proporcionados en la Tabla 3), y rs10941679 que están en fuerte desequilibrio de enlazamiento con el SNP rs4415084 (otros posibles ejemplos son proporcionados en la Tabla 4). Además, los ejemplos de los SNP en desequilibrio de enlazamiento con rs13387042 son proporcionados en la Tabla 5.

Tabla 2. Marcadores sustitutos para el SNP rs2981582. Se seleccionaron marcadores con un r^2 mayor a 0.05 para rs2981582 en el conjunto de datos HAPMAP (<http://hapmap.ncbi.nlm.nih.gov>) en un intervalo de 1 Mpb que flanquea al marcador. Se muestra el nombre del SNP correlacionado, los valores para r^2 y D' para rs2981582 y el valor LOD correspondiente, así como la posición del marcador sustituto en NCB Build 36.

Identificación de los rs de la base de datos del SNP	Position	SNP correlacionado	Ubicación	D'	r^2	LOD
rs2981582	123342307	rs3135715	123344716	1.000	0.368	15.02
rs2981582	123342307	rs7899765	123345678	1.000	0.053	2.44
rs2981582	123342307	rs1047111	123347551	0.938	0.226	9.11
rs2981582	123342307	rs1219639	123348302	1.000	0.143	6.53
rs2981582	123342307	rs10886955	123360344	0.908	0.131	5.42
rs2981582	123342307	rs1631281	123380775	0.906	0.124	5.33
rs2981582	123342307	rs3104685	123381354	0.896	0.108	4.58
rs2981582	123342307	rs1909670	123386718	1.000	0.135	6.12
rs2981582	123342307	rs7917459	123392364	1.000	0.135	6.42
rs2981582	123342307	rs17102382	123393846	1.000	0.135	6.42
rs2981582	123342307	rs10788196	123407625	1.000	0.202	9.18
rs2981582	123342307	rs2935717	123426236	0.926	0.165	7.30
rs2981582	123342307	rs3104688	123426455	0.820	0.051	2.07
rs2981582	123342307	rs4752578	123426514	1.000	0.106	5.15
rs2981582	123342307	rs1696803	123426940	0.926	0.168	7.33
rs2981582	123342307	rs12262574	123428112	1.000	0.143	7.39
rs2981582	123342307	rs4752579	123431182	1.000	0.106	5.15
rs2981582	123342307	rs12358208	123460953	0.761	0.077	2.46
rs2981582	123342307	rs17102484	123462020	0.758	0.065	2.39
rs2981582	123342307	rs2936859	123469277	0.260	0.052	1.56
rs2981582	123342307	rs10160140	123541979	0.590	0.016	0.40

Tabla 3. Marcadores sustitutos para el SNP rs3803662. Se seleccionaron marcadores con un r^2 mayor a 0.05 para rs3803662 en el conjunto de datos HAPMAP (<http://hapmap.ncbi.nlm.nih.gov>) en un intervalo de 1 Mpb que flanquea al marcador. Se muestra el nombre del SNP correlacionado, los valores para r^2 y D' para rs3803662 y el valor LOD correspondiente, así como la posición del marcador sustituto en NCB Build 36.

Identificación de los rs de la base de datos del SNP	Position	SNP correlacionado	Ubicación	D'	r^2	LOD
rs3803662	51143842	rs4784227	51156689	0.968	0.881	31.08
rs3803662	51143842	rs3112572	51157948	1.000	0.055	1.64
rs3803662	51143842	rs3104747	51159425	1.000	0.055	1.64
rs3803662	51143842	rs3104748	51159860	1.000	0.055	1.64
rs3803662	51143842	rs3104750	51159990	1.000	0.055	1.64
rs3803662	51143842	rs3104758	51166534	1.000	0.055	1.64
rs3803662	51143842	rs3104759	51167030	1.000	0.055	1.64
rs3803662	51143842	rs9708611	51170166	1.000	0.169	4.56
rs3803662	51143842	rs12935019	51170538	1.000	0.088	4.04
rs3803662	51143842	rs4784230	51175614	1.000	0.085	4.19
rs3803662	51143842	rs11645620	51176454	1.000	0.085	4.19
rs3803662	51143842	rs3112633	51178078	1.000	0.085	4.19
rs3803662	51143842	rs3104766	51182036	0.766	0.239	7.55
rs3803662	51143842	rs3104767	51182239	0.626	0.167	4.88
rs3803662	51143842	rs3112625	51183053	0.671	0.188	5.62
rs3803662	51143842	rs12920540	51183114	0.676	0.195	5.84
rs3803662	51143842	rs3104774	51187203	0.671	0.188	5.62
rs3803662	51143842	rs7203671	51187646	0.671	0.188	5.62
rs3803662	51143842	rs3112617	51189218	0.666	0.177	5.44
rs3803662	51143842	rs11075551	51189465	0.666	0.177	5.44
rs3803662	51143842	rs12929797	51190445	0.676	0.19	5.87
rs3803662	51143842	rs3104780	51191415	0.671	0.184	5.65
rs3803662	51143842	rs12922061	51192501	0.832	0.631	19.14
rs3803662	51143842	rs3112612	51192665	0.671	0.184	5.65
rs3803662	51143842	rs3104784	51193866	0.666	0.177	5.44
rs3803662	51143842	rs12597685	51195281	0.671	0.184	5.65

Identificación de los rs de la base de datos del SNP	Position	SNP correlacionado	Ubicación	D'	r ²	LOD
rs3803662	51143842	rs3104788	51196004	0.666	0.177	5.44
rs3803662	51143842	rs3104800	51203877	0.625	0.17	4.99
rs3803662	51143842	rs3112609	51206232	0.599	0.163	4.86
rs3803662	51143842	rs3112600	51214089	0.311	0.016	0.57
rs3803662	51143842	rs3104807	51215026	0.302	0.014	0.52
rs3803662	51143842	rs3112594	51229030	0.522	0.065	1.56
rs3803662	51143842	rs4288991	51230665	0.238	0.052	1.53
rs3803662	51143842	rs3104820	51233304	0.528	0.069	1.60
rs3803662	51143842	rs3104824	51236594	0.362	0.067	1.93
rs3803662	51143842	rs3104826	51237406	0.362	0.067	1.93
rs3803662	51143842	rs3112588	51238502	0.354	0.062	1.80

Tabla 4. Marcadores sustitutos para el SNP rs4415084. Se seleccionaron marcadores con un r² mayor a 0.05 para rs4415084 en el conjunto de datos HAPMAP (<http://hapmap.ncbi.nlm.nih.gov>) en un intervalo de 1 Mpb que flanquea al marcador. Se muestra el nombre del SNP correlacionado, los valores para r² y D' para rs4415084 y el valor LOD correspondiente, así como la posición del marcador sustituto en NCB Build 36

Identificación de los rs de la base de datos del SNP	Position	SNP correlacionado	Ubicación	D'	r ²	LOD
rs4415084	44698272	rs12522626	44721455	1.000	1.0	47.37
rs4415084	44698272	rs4571480	44722945	1.000	0.976	40.54
rs4415084	44698272	rs6451770	44727152	1.000	0.978	44.88
rs4415084	44698272	rs920328	44734808	1.000	0.893	39.00
rs4415084	44698272	rs920329	44738264	1.000	1.0	47.37
rs4415084	44698272	rs2218081	44740897	1.000	1.0	47.37
rs4415084	44698272	rs16901937	44744898	1.000	0.978	45.06
rs4415084	44698272	rs11747159	44773467	0.948	0.747	28.79
rs4415084	44698272	rs2330572	44776746	0.952	0.845	34.31
rs4415084	44698272	rs994793	44779004	0.952	0.848	34.49
rs4415084	44698272	rs1438827	44787713	0.948	0.749	29.76
rs4415084	44698272	rs7712949	44806102	0.948	0.746	29.19

Identificación de los rs de la base de datos del SNP	Position	SNP correlacionado	Ubicación	D'	r ²	LOD
rs4415084	44698272	rs11746980	44813635	0.952	0.848	34.49
rs4415084	44698272	rs16901964	44819012	0.949	0.768	30.54
rs4415084	44698272	rs727305	44831799	0.972	0.746	27.65
rs4415084	44698272	rs10462081	44836422	0.948	0.749	29.76
rs4415084	44698272	rs13183209	44839506	0.925	0.746	28.55
rs4415084	44698272	rs13159598	44841683	0.952	0.848	34.19
rs4415084	44698272	rs3761650	44844113	0.947	0.744	28.68
rs4415084	44698272	rs13174122	44846497	0.971	0.735	26.70
rs4415084	44698272	rs11746506	44848323	0.973	0.764	29.24
rs4415084	44698272	rs7720787	44853066	0.952	0.845	34.31
rs4415084	44698272	rs9637783	44855403	0.948	0.748	29.16
rs4415084	44698272	rs4457089	44857493	0.948	0.762	29.70
rs4415084	44698272	rs6896350	44868328	0.948	0.764	29.46
rs4415084	44698272	rs1371025	44869990	0.973	0.785	30.69
rs4415084	44698272	rs4596389	44872313	0.948	0.749	29.76
rs4415084	44698272	rs6451775	44872545	0.948	0.746	29.19
rs4415084	44698272	rs729599	44878017	0.948	0.748	29.16
rs4415084	44698272	rs987394	44882135	0.948	0.749	29.76
rs4415084	44698272	rs4440370	44889109	0.948	0.748	29.16
rs4415084	44698272	rs7703497	44892785	0.948	0.749	29.76
rs4415084	44698272	rs13362132	44894017	0.952	0.827	34.09
rs4415084	44698272	rs1438821	44894208	0.951	0.844	34.52

Tabla 5. Marcadores sustitutos para el SNP rs13387042. Se seleccionaron marcadores con un r² mayor a 0.05 para rs13387042 en el conjunto de datos HAPMAP (<http://hapmap.ncbi.nlm.nih.gov>) en un intervalo de 1 Mpb que flanquea al marcador. Se muestra el nombre del SNP correlacionado, los valores para r² y D' para rs13387042 y el valor LOD correspondiente, así como la posición del marcador sustituto en NCB Build 36

ES 2 546 230 T3

Identificación de los rs de la base de datos del SNP	Position	SNP correlacionado	Ubicación	D'	r ²	LOD
rs13387042	217614077	rs4621152	217617230	0.865	0.364	15.30
rs13387042	217614077	rs6721996	217617708	1.000	0.979	50.46
rs13387042	217614077	rs12694403	217623659	0.955	0.33	14.24
rs13387042	217614077	rs17778427	217631258	1.000	0.351	16.12
rs13387042	217614077	rs17835044	217631850	1.000	0.351	16.12
rs13387042	217614077	rs7588345	217632061	1.000	0.193	8.93
rs13387042	217614077	rs7562029	217632506	1.000	0.413	20.33
rs13387042	217614077	rs13000023	217632639	0.949	0.287	12.20
rs13387042	217614077	rs13409592	217634573	0.933	0.192	7.69
rs13387042	217614077	rs2372957	217635302	0.855	0.168	5.97
rs13387042	217614077	rs16856888	217638914	0.363	0.101	3.31
rs13387042	217614077	rs16856890	217639976	0.371	0.101	3.29
rs13387042	217614077	rs7598926	217640464	0.382	0.109	3.60
rs13387042	217614077	rs6734010	217643676	0.543	0.217	7.90
rs13387042	217614077	rs13022815	217644369	0.800	0.319	12.94
rs13387042	217614077	rs16856893	217645298	0.739	0.109	3.45
rs13387042	217614077	rs13011060	217646422	0.956	0.352	14.71
rs13387042	217614077	rs4674132	217646764	0.802	0.327	13.10
rs13387042	217614077	rs16825211	217647249	0.912	0.326	12.95
rs13387042	217614077	rs41521045	217647581	0.903	0.112	4.70
rs13387042	217614077	rs2372960	217650960	0.678	0.058	2.12
rs13387042	217614077	rs2372967	217676158	0.326	0.052	1.97
rs13387042	217614077	rs3843337	217677680	0.326	0.052	1.97
rs13387042	217614077	rs2372972	217679386	0.375	0.062	2.28
rs13387042	217614077	rs9677455	217680497	0.375	0.062	2.28
rs13387042	217614077	rs12464728	217686802	0.478	0.073	2.54

Los cebadores de amplificación para amplificar marcadores (por ejemplo, los loci marcadores) y las sondas adecuadas para detectar estos marcadores o para el genotipo de una muestra con respecto a múltiples alelos marcadores, se pueden utilizar en la presente invención. Por ejemplo, la selección del cebador para la PCR de largo alcance se describe en la solicitud de patente de los Estados Unidos No. 10/042,406 y la solicitud de patente de los Estados Unidos No. 10/236,480; para la PCR de corto alcance, la solicitud de patente de los Estados Unidos No. 10/341,832 proporciona una guía con respecto a la selección del cebador. También, existen programas disponibles al público tales como "oligo" disponible para el diseño del cebador. Con esta selección disponible de cebadores y un software de diseño, la secuencia del genoma humano disponible al público y las ubicaciones de polimorfismos, un experto puede construir cebadores para amplificar los SNP para llevar a cabo la invención. Además, se debe apreciar que la sonda precisa que se utiliza para la detección de un ácido nucleico que comprende un SNP (por ejemplo, un amplicón que comprende el SNP) puede variar, por ejemplo, cualquier sonda que pueda identificar la región de un amplicón marcador que sea detectada se puede utilizar junto con la presente invención. Además, la configuración de las sondas de detección, por supuesto, puede variar. Por lo tanto, la invención no se limita a las secuencias mencionadas en la presente invención.

De hecho, se apreciará que la amplificación no es un requisito para la detección del marcador, por ejemplo se puede detectar directamente un ADN genómico no amplificado simplemente al realizar una transferencia tipo Southern sobre una muestra de ADN genómico. Los procedimientos para realizar la transferencia tipo Southern, la amplificación estándar (PCR, LCR, o similares) y muchos otros métodos para detección de ácido nucleico están bien establecidos y se enseñan, por ejemplo, en Sambrook y colaboradores (ver más arriba).

También se pueden omitir las sondas de detección separadas en los métodos de amplificación/detección, por ejemplo, mediante la realización de una reacción de amplificación en tiempo real que detecte la formación del producto mediante la modificación del cebador de amplificación relevante después la incorporación en un producto, la incorporación de nucleótidos marcados en un amplicón, o mediante el control de cambios en las propiedades de rotación molecular de los amplificados en comparación con precursores no amplificados (por ejemplo, mediante polarización de fluorescencia).

Típicamente, los marcadores moleculares se detectan mediante cualquier método establecido disponible en la técnica, incluyendo, pero sin limitarse a, hibridación específica de alelos (ASH), detección de la extensión de nucleótidos individuales, hibridación en arreglos (incluyendo opcionalmente ASH), u otros métodos para detectar polimorfismos de un solo nucleótido (SNP), la detección de polimorfismo de la longitud del fragmento amplificado (AFLP), la detección de la secuencia variable amplificada, la detección de ADN polimórfico amplificado aleatoriamente (RAPD), la detección del polimorfismo de la longitud del fragmento de restricción (RFLP), la detección de la replicación de la secuencia autosostenida, la detección de la repetición de una sola secuencia (SSR), la detección de polimorfismos de conformación de una sola cadena (SSCP), la detección del marcador de isoenzima, análisis tipo Northern (donde los niveles de expresión se utilizan como marcadores), amplificación cuantitativa de ARNm o ADNc, o similares.

En la Tabla 6 se proporcionan ejemplos de cebadores oligonucleótidos útiles para amplificar ácidos nucleicos que comprenden SNP que se sabe que están asociados con un fenotipo de cáncer de mama. Como apreciará la persona capacitada en la materia, la secuencia de la región genómica con la cual se hibridan estos oligonucleótidos se puede utilizar para diseñar cebadores que son más largos en el extremo 5' y/o 3', posiblemente más cortos en el extremo 5' y/o 3' (siempre y cuando la versión truncada aún pueda ser utilizada para la amplificación), que tienen una o unas cuantas diferencias de nucleótidos (aunque no obstante aún se puedan utilizar para la amplificación), o que no compartan ninguna similitud de secuencia con aquellas proporcionadas pero que están diseñadas con base en secuencias genómicas cercanas a donde los oligonucleótidos proporcionados específicamente se hibridan y que aún se pueden utilizar para la amplificación.

Ejemplos de técnicas para la detección de marcadores

Los marcadores correspondientes a los polimorfismos genéticos entre los miembros de una población se pueden detectar mediante numerosos métodos bien establecidos en la técnica (por ejemplo, amplificación específica de secuencias con base en la PCR, polimorfismos de longitud de fragmentos de restricción (RFLP), marcadores de isoenzima, análisis tipo Northern, hibridación específica de alelo (ASH), hibridación con base en arreglos, secuencias variables amplificadas del genoma, replicación de secuencias autosostenida, repetición simple de secuencias simples (SSR), polimorfismo de un solo nucleótido (SNP), ADN polimórfico amplificado en forma aleatoria ("RAPD") o polimorfismos de longitud del fragmento amplificado (AFLP). En una realización adicional, se determina la presencia o ausencia de un marcador molecular simplemente a través de la secuenciación de nucleótidos de la región del marcador polimórfico. Cualquiera de estos métodos se adapta fácilmente a un análisis de alto rendimiento.

Tabla 6. Ejemplos de cebadores oligonucleótidos útiles para la invención

Nombre	Secuencia
rs889312_for	TATGGGAAGGAGTCGTTGAG (SEQ ID NO: 1)
rs6504950_for	CTGAATCACTCCTTGCCAAC (SEQ ID NO:2)
rs4973768_for	CAAAATGATCTGACTACTCC (SEQ ID NO:3)
rs4415084_for	TGACCAGTGCTGTATGTATC (SEQ ID NO:4)
rs3817198_for	TCTCACCTGATACCAGATTC (SEQ ID NO:5)
rs3803662_for	TCTCTCCTTAATGCCTCTAT (SEQ ID NO:6)
rs2981582_for	ACTGCTGCGGGTTCCTAAAG (SEQ ID NO:7)
rs13387042_for	GGAAGATTCGATTCAACAAGG (SEQ ID NO:8)
rs13281615_for	GGTAACTATGAATCTCATC (SEQ ID NO:9)
rs11249433_for	AAAAAGCAGAGAAAGCAGGG (SEQ ID NO: 10)
rs889312_rev	AGATGATCTCTGAGATGCCC (SEQ ID NO:11)
rs6504950_rev	CCAGGGTTTGTCTACCAAAG (SEQ ID NO: 12)
rs4973768_rev	AATCACTTAAACAAGCAG (SEQ ID NO: 13)
rs4415084_rev	CACATACCTCTACCTTAGC (SEQ ID NO: 14)
rs3817198_rev	TTCCCTAGTGAGCAGTGG (SEQ ID NO:15)
rs3803662_rev	CTTCTTCGCAAATGGGTGG (SEQ ID NO:16)
rs2981582_rev	GCACTCATCGCCACTTAATG (SEQ ID NO: 17)
rs13387042_rev	GAACAGCTAAACCAGAACAG (SEQ ID NO:18)
rs13281615_rev	ATCACTCTTATTCTCCCCC (SEQ ID NO: 19)
rs11249433_rev	TGAGTCACTGTGCTAAGGAG (SEQ ID NO:20)

Algunas técnicas para detectar marcadores genéticos utilizan la hibridación de ácido nucleico sonda con los ácidos nucleicos correspondientes al marcador genético (por ejemplo, los ácidos nucleicos amplificados producidos utilizando ADN genómico como molde). Los formatos de hibridación, que incluyen, pero no se limitan a: ensayos en fase en solución, en fase sólida, en fase mixta, o *in situ* son útiles para la detección de alelos. Una guía amplia para la hibridación de ácidos nucleicos se encuentra en Tijssen (1993) Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes Elsevier, Nueva York, así como en Sambrook y colaboradores (ver más arriba).

Por ejemplo, los marcadores que comprenden polimorfismos de longitud de fragmentos de restricción (RFLP) se detectan, por ejemplo, mediante la hibridación de una sonda que típicamente es un subfragmento (o un oligonucleótido sintético correspondiente a un subfragmento) del ácido nucleico que va a ser detectado con el ADN genómico digerido por restricción. La enzima de restricción se selecciona para proporcionar fragmentos de restricción de al menos dos longitudes alternativas (o polimórficas) en diferentes individuos o poblaciones. La determinación de una o más enzimas de restricción que produzca fragmentos informativos para cada alelo de un marcador es un procedimiento simple, bien conocido en la técnica. Después de la separación por longitud en una

matriz adecuada (por ejemplo, agarosa o poliacrilamida) y la transferencia a una membrana (por ejemplo, nitrocelulosa, nailon, etc.), se hibrida la sonda marcada bajo condiciones que den por resultado un enlazamiento en equilibrio de la sonda con el objetivo seguido por la eliminación del exceso de sonda mediante lavado.

Las sondas de ácido nucleico para los loci marcadores se pueden clonar y/o sintetizar. Se puede utilizar cualquier marcador adecuado con una sonda para utilizarse en la invención. Los marcadores detectables adecuados para utilizarse con las sondas de ácido nucleico, incluyen, por ejemplo, cualquier composición que se pueda detectar por medios espectroscópicos, radioisotópicos, fotoquímicos, bioquímicos, inmunoquímicos, eléctricos, ópticos o químicos. Los marcadores útiles incluyen biotina para tinción con un conjugado marcado de estreptavidina, perlas magnéticas, tintes fluorescentes, marcadores radioactivos, enzimas y marcadores calorimétricos. Otros marcadores incluyen ligandos que se unen a los anticuerpos marcados con fluoróforos, agentes quimioluminiscentes y enzimas. Una sonda también puede constituir cebadores para PCR marcados radioactivamente que se utilizan para generar un amplicón marcado en forma radioactiva. Las estrategias de marcación para marcar ácidos nucleicos y las estrategias de detección correspondientes se pueden encontrar, por ejemplo, en Haugland (2003) Handbook of Fluorescent Probes and Research Chemicals, novena edición por Molecular Probes, Inc. (Eugene Oreg.). Los detalles adicionales con relación a la estrategia para detección del marcador se encuentran a continuación.

Métodos de detección con base en amplificación

PCR, RT-PCR y LCR en particular son ampliamente utilizadas como métodos de amplificación y de detección de la amplificación para amplificar ácidos nucleicos de interés (por ejemplo, aquellos que comprenden loci marcadores), que facilitan la detección de los ácidos nucleicos de interés. Los detalles relacionados con el uso de estos y otros métodos de amplificación se pueden encontrar en cualquiera entre una variedad de textos estándar, incluyendo, por ejemplo, Sambrook y colaboradores (ver más arriba). Muchos textos de biología disponibles también tienen análisis extensos que se relacionan con métodos de PCR y métodos de amplificación relacionados. Un experto apreciará que esencialmente cualquier ARN se puede convertir en un ADN bicatenario adecuado para digestión por restricción, expansión y secuenciación mediante PCR utilizando transcriptasa inversa y una polimerasa ("PCR por transcripción inversa, o "RT-PCR").

Métodos de amplificación/detección en tiempo real

En un aspecto, se realiza una PCR o LCR en tiempo real sobre las mezclas de amplificación descritas en la presente invención, por ejemplo, utilizando balizas moleculares o sondas TaqMan^{MR}. Una baliza molecular (MB) es un oligonucleótido o PNA el cual, bajo condiciones de hibridación adecuadas, se autohibrida para formar una estructura de vástago y de bucle. La MB tiene un marcador y un atenuador en los terminales del oligonucleótido o PNA; por lo tanto, bajo condiciones que permitan la hibridación intramolecular, típicamente se inactiva el marcador (o al menos se altera en su fluorescencia) mediante el atenuador. Bajo condiciones donde la MB no exhiba una hibridación intramolecular (por ejemplo, cuando se una a un ácido nucleico objetivo, por ejemplo, a una región de un amplicón durante la amplificación), el marcador de la MB está sin inactivar. Los detalles que se relacionan con los métodos estándar para elaborar y utilizar las MB están bien establecidos en la literatura y las MB están disponibles a través de diversas fuentes comerciales de reactivos (véase también, por ejemplo, Leone y colaboradores, 1995; Tyagi y Kramer, 1996; Blok y Kramer, 1997; Hsuih y colaboradores, 1997; Kostrikis y colaboradores, 1998; Sokol y colaboradores, 1998; Tyagi y colaboradores, 1998; Bonnet y colaboradores, 1999; Fang y colaboradores, 1999; Marras y colaboradores, 1999; y Vet y colaboradores, 1999). Los detalles adicionales que se relacionan con la construcción de MB y su uso se encuentran en la literatura de patentes, por ejemplo, la patente de los Estados Unidos No. 5,925,517, la patente de los Estados Unidos No. 6,150,097 y la patente de los Estados Unidos No. 6,037,130.

La detección mediante PCR utilizando sondas de oligonucleótidos fluorogénicas doblemente marcadas, comúnmente denominadas como sondas "TaqMan^{MR}", también se pueden realizar de acuerdo con la presente invención. Estas sondas están compuestas de oligodesoxinucleótidos cortos (por ejemplo, 20-25 bases) que están marcadas con dos diferentes tintes fluorescentes. En el terminal 5' de cada sonda está un colorante reportero, y sobre el terminal 3' de cada sonda se encuentra un colorante atenuador. La secuencia de la sonda de oligonucleótido es complementaria a una secuencia objetivo interna presente en un amplicón PCR. Cuando la sonda está intacta, se presenta transferencia de energía entre los dos fluoróforos y la emisión proveniente del reportero se atenúa mediante el atenuador por FRET. Durante la fase de extensión de PCR, la sonda se escinde mediante la actividad de la nucleasa 5' de la polimerasa utilizada en la reacción, liberando así el reportero del atenuador del oligonucleótido y produciendo un aumento en la intensidad de emisión del reportero. Por consiguiente, las sondas TaqMan^{MR} son oligonucleótidos que tienen una etiqueta y un atenuador, donde la etiqueta se libera durante la amplificación mediante la acción de la exonucleasa de la polimerasa utilizada en la amplificación. Esto proporciona una medición en tiempo real de la amplificación durante la síntesis. Una variedad de reactivos TaqMan^{MR} se encuentran comercialmente disponibles, por ejemplo, a través de Applied Biosystems (Division Headquarters en Foster City, California), así como de una variedad de proveedores de especialidades, tales como Biosearch Technologies (por ejemplo, sondas atenuadoras de agujero negro). Los detalles adicionales que se relacionan con las estrategias de sondas de marca dual se pueden encontrar, por ejemplo, en el documento WO 92/02638.

Otros métodos similares incluyen, por ejemplo, transferencia de energía de resonancia de fluorescencia entre dos sondas híbridas adyacentes, por ejemplo, utilizando el formato "LightCycler®" descrito en la patente de los Estados Unidos No. 6.174.670.

Detección del marcador con base en una matriz

5 La detección con base en una matriz se puede realizar utilizando matrices comercialmente disponibles, por ejemplo, de Affymetrix (Santa Clara, California) u otros fabricantes. Las revisiones que se relacionan con la operación de las matrices de ácido nucleico incluyen Sapolsky y colaboradores, 1999; Lockhart, 1998; Fodor, 1997a; Fodor, 1997b y Chee y colaboradores, 1996. La detección con base en la matriz es un método preferido para marcadores de identificación de la invención en muestras, debido a la naturaleza inherente de alto rendimiento de la detección con base en una matriz.

10 Una variedad de matrices de sondas se han descrito en la literatura y se pueden utilizar en el contexto de la presente invención para la detección de marcadores que se pueden ser correlacionados con los fenotipos observados en la presente invención. Por ejemplo, se usan chips de matrices de sondas de ADN u obleas de matrices de sondas de ADN más grandes (a partir de las cuales se obtendrían chips individuales por rompimiento de la oblea) en una realización de la invención. Las obleas de matrices de sondas de ADN generalmente comprenden obleas de vidrio sobre las cuales se han colocado las matrices de alta densidad de sondas de ADN (segmentos cortos de ADN). Cada uno de estas obleas puede soportar, por ejemplo, aproximadamente 60 millones de sondas de ADN que se utilizan para reconocer secuencias más largas de ADN de la muestra (por ejemplo, de individuos o poblaciones, por ejemplo, que contienen marcadores de interés). El reconocimiento del ADN de la muestra por el conjunto de sondas de ADN sobre la oblea de vidrio tiene lugar a través de la hibridación de ADN. Cuando se hibrida una muestra de ADN con una matriz de sondas de ADN, la muestra se une a aquellas sondas que son complementarias a la secuencia de ADN de la muestra. Por medio de la evaluación con cuáles sondas el ADN de la muestra para un individuo se hibrida más fuertemente, es posible determinar si una secuencia conocida de ácido nucleico está presente o no en la muestra, determinando así si un marcador encontrado en el ácido nucleico está presente. También se puede llevar a cabo este enfoque para realizar ASH, controlando las condiciones de hibridación para permitir la discriminación de un solo nucleótido, por ejemplo, para identificación del SNP y para determinación del genotipo de una muestra para uno o más SNP. Las matrices proporcionan una realización conveniente para detectar múltiples marcadores polimórficos simultáneamente (o en serie). Por ejemplo, se pueden construir matrices de detección de susceptibilidad de cáncer de mama en las cuales se detectan simultáneamente cualquiera o todos los polimorfismos observados en la presente invención (o polimorfismos vinculados a los mismos) para asignar un fenotipo de susceptibilidad de cáncer de mama. Desde luego, se puede utilizar en forma similar cualquier tecnología de detección (PCR, LCR, PCR en tiempo real, etc.), por ejemplo, con reacciones de amplificación/detección múltiples, o simplemente corriendo diferentes reacciones por separado, por ejemplo, simultáneamente o en serie.

35 El uso de matrices de sondas de ADN para obtener información de alelos involucra típicamente las siguientes etapas generales: diseño y fabricación de matrices de sondas de ADN, preparación de la muestra, hibridación de la muestra de ADN con la matriz, detección de eventos de hibridación y análisis de datos para determinar la secuencia. Las obleas preferidas se fabrican utilizando un proceso adaptado de la fabricación de semiconductores para lograr efectividad de costos y alta calidad, y se encuentran disponibles, por ejemplo, a través de Affymetrix, Inc. de Santa Clara, California.

40 Por ejemplo, las matrices de sondas se pueden fabricar mediante procesos de síntesis química dirigida por luz, que combinan síntesis química en fase sólida con técnicas de fabricación fotolitográficas como las empleadas en la industria de semiconductores. Usando una serie de máscaras fotolitográficas para definir los sitios de exposición del chip, seguido por etapas específicas de síntesis química, el proceso construye matrices de alta densidad de oligonucleótidos, con cada sonda en una posición predefinida en la matriz. Se pueden sintetizar simultáneamente múltiples matrices de sonda en una oblea de vidrio grande. Este proceso paralelo mejora la reproducibilidad y ayuda a lograr economías de escala.

45 Una vez fabricadas, se pueden utilizar las matrices de sondas de ADN para obtener datos relacionados con la presencia y/o niveles de expresión para marcadores de interés. Las muestras de ADN se pueden etiquetar con biotina y/o un grupo reportero fluorescente mediante métodos bioquímicos estándar. Las muestras marcadas se incuban con una matriz, y los segmentos de las muestras se unen, o hibridan, con secuencias complementarias sobre la matriz. La matriz puede ser lavada y/o teñida para producir un patrón de hibridación. La matriz es luego escaneada y se detectan los patrones de hibridación mediante la emisión de luz proveniente de los grupos reporteros fluorescentes. Los detalles adicionales relacionados con estos procedimientos se encuentran en los ejemplos más adelante. Ya que la identidad y la posición de cada sonda sobre la matriz es conocida, se puede determinar la naturaleza de las secuencias de ADN en la muestra aplicada a la matriz. Cuando se utilizan estas matrices para experimentos de determinación de genotipo, se pueden denominar como matrices para la determinación de genotipo.

La muestra de ácido nucleico que va a ser analizada se aísla, se amplifica y típicamente, se marca con biotina y/o un grupo reportero fluorescente. La muestra de ácido nucleico marcada es luego incubada con la matriz utilizando una estación de fluidos y un horno de hibridación. La matriz puede ser lavada y/o teñida o contrastada, según sea conveniente con el método de detección. Después de la hibridación, el lavado y la coloración, se inserta la matriz en un escáner, donde se detectan los patrones de hibridación. Los datos de hibridación se recolectan como luz emitida a partir de los grupos reporteros fluorescentes ya incorporados en el ácido nucleico marcado, que se enlaza ahora a la matriz de la sonda. Las sondas que más claramente coinciden con el ácido nucleico marcado producen señales más fuertes que aquellas que no tienen coincidencias. Ya que la secuencia y posición de cada sonda sobre la matriz se conocen por complementariedad, se puede identificar la identidad de la muestra de ácido nucleico aplicada a la matriz de la sonda.

En una realización, se pueden marcar en forma diferencial dos muestras de ADN e hibridar con un solo conjunto de las matrices diseñadas para la determinación del genotipo. De esta forma, se pueden obtener dos conjuntos de datos a partir de las mismas matrices físicas. Las etiquetas que se pueden utilizar incluyen, pero no se limitan a, citocromo, fluoresceína o biotina (más tarde teñida con ficoeritrina-estreptavidina después de la hibridación). Se describe la marcación con dos colores en la patente de los Estados Unidos No. 6,342,355. Cada matriz puede ser escaneada de tal manera que la señal de ambas etiquetas se detecta simultáneamente, o se puede escanear dos veces para detectar cada señal por separado.

Los datos de intensidad se recolectan por el escáner para todos los marcadores para cada uno de los individuos que se analizan por la presencia del marcador. Las intensidades medidas son una medida indicativa de la cantidad de un marcador particular presente en la muestra para un individuo dado (nivel de expresión y/o número de copias del alelo presente en un individuo, dependiendo de que se analicen o no los ácidos nucleicos genómicos o expresados). Esto se puede utilizar para determinar si el individuo es homocigótico o heterocigótico para el marcador de interés. Los datos de intensidad se procesan para proporcionar la información del marcador correspondiente para las diferentes intensidades.

Detalles adicionales relacionados con la amplificación del ácido nucleico

Según se observa, las técnicas de amplificación de ácido nucleico tales como PCR y LCR son bien conocidas en el arte y se pueden aplicar a la presente invención para amplificar y/o detectar los ácidos nucleicos de interés, tales como los ácidos nucleicos que comprenden los loci marcadores. Los ejemplos de técnicas suficientes para dirigir a las personas capacitadas a través de tales métodos *in vitro*, incluyen la reacción en cadena de polimerasa (PCR), la reacción en cadena de ligasa (LCR), la amplificación por Q β -replicasa y otras técnicas mediadas por ARN polimerasa (por ejemplo,

NASBA), se encuentran en las referencias anotadas anteriormente, por ejemplo, Sambrook y colaboradores. Los detalles adicionales se encuentran en la patente de los Estados Unidos No. 4,683,202; Kwoh y colaboradores, 1989; Guatelli y colaboradores, 1990; Landegren y colaboradores, 1988; Van Brunt, 1990; Wu y Wallace, 1989; Barringer y colaboradores, 1990; y Sooknanan y Malek, 1995. Los métodos mejorados para amplificación de ácidos nucleicos grandes mediante PCR, que es útil en el contexto de clonación posicional de genes vinculados con los polimorfismos de la presente invención, se resumen adicionalmente en Cheng y colaboradores (1994), y las referencias citadas allí, en las cuales se generan amplicones mediante PCR de hasta 40 kb. Los métodos para PCR de gran intervalo se describen, por ejemplo, en la solicitud de patente de los Estados Unidos No. 10/042,406; la solicitud de patente de los Estados Unidos No. 10/236,480; y la patente de los Estados Unidos No. 6,740,510. La solicitud de patente de los Estados Unidos No. 10/341,832 también proporciona detalles relacionados con los métodos para selección de cebadores para la realización de PCR de corto intervalo.

Antes de la amplificación y/o detección de un ácido nucleico que comprende un marcador, se purifica opcionalmente el ácido nucleico de las muestras mediante cualquier método disponible, por ejemplo, aquellos enseñados en Berger y Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volumen 152, Academic Press, Inc., San Diego, California, Sambrook y colaboradores (ver más arriba), y/o Ausubel y colaboradores (ver más arriba). También se encuentran comercialmente disponibles una cantidad de kits para la purificación de ácidos nucleicos provenientes de células u otras muestras (véase, por ejemplo, EasyPrep^{MR} ElexiPrep^{MR}, ambos de Pharmacia Biotech; StrataClean^{MR}, de Stratagene; y QIAprep^{MR} de Qiagen). Alternativamente, las muestras pueden simplemente ser directamente sometidas a amplificación o detección, por ejemplo, después de tomar una alícuota y/o dilución.

Los ejemplos de marcadores pueden incluir polimorfismos, polimorfismos de un solo nucleótido, presencia de uno o más ácidos nucleicos en una muestra, ausencia de uno o más ácidos nucleicos en una muestra, presencia de una o más secuencias de ADN genómico, ausencia de una o más secuencias de ADN genómico, presencia de uno o más ARNm, ausencia de uno o más ARNm, niveles de expresión de uno o más ARNm, presencia de una o más proteínas, niveles de expresión de una o más proteínas, y/o datos derivados de cualquiera de los anteriores o combinaciones de los mismos. Esencialmente, se puede detectar cualquier número de marcadores, utilizando métodos disponibles, por ejemplo, usando tecnologías de matriz que proporcionen alta densidad, mapeo de

marcadores de alto rendimiento. De esta forma, al menos aproximadamente 10, 100, 1,000, 10,000 o incluso 100,000 más marcadores genéticos, se pueden probar simultáneamente o en forma serial (o combinación de los mismos), por correlación con un fenotipo relevante, en la primera y/o segunda población. Las combinaciones de marcadores también se pueden probar en forma deseable, por ejemplo, para identificar combinaciones genéticas o combinaciones de patrones de expresión en poblaciones que se correlacionan con el fenotipo.

Métodos de síntesis de la sonda/cebador

En general, los métodos de síntesis para la elaboración de oligonucleótidos, incluyendo sondas, cebadores, balizas moleculares, PNA, LNA (ácidos nucleicos bloqueados), etc., son bien conocidos. Por ejemplo, los oligonucleótidos se pueden sintetizar químicamente de acuerdo con el método de la fosforamidita triéster en fase sólida descrito por Beaucage y Caruthers (1981), por ejemplo, utilizando un sintetizador automatizado comercialmente disponible, por ejemplo, como se describe en Needham-VanDevanter y colaboradores (1984). Los oligonucleótidos, incluyendo oligonucleótidos modificados también se pueden ordenar a partir de una variedad de fuentes comerciales conocidas por los expertos en la técnica. Existen muchos proveedores comerciales de servicios de síntesis de oligos, y por lo tanto esta es una tecnología muy accesible. Cualquier ácido nucleico se puede ordenar a partir de cualquiera de una variedad de fuentes comerciales, tales como; The Midland Certified Reagent Company (mcrc@oligos.com), The Great American Gene Company (www.genco.com), ExpressGen Inc. (www.expressgen.com), Operon Technologies Inc. (Alameda, California) y muchos otros. En forma similar, los PNA se pueden ordenar a partir cualquiera de una variedad de fuentes, tales como PeptideGenic (pkim@ccnet.com), HTI Bio-products, Inc. (htibio.com), BMA Biomedicals Ltda (Reino Unido), Bio-Synthesis, Inc., y muchos otros.

Cebadores de amplificación para la detección de marcadores

En algunas realizaciones preferidas, los SNP se detectan utilizando un método de detección adecuado con base en PCR, en donde el tamaño de secuencia del amplicón PCR es indicativo de ausencia o presencia del SNP. En estos tipos de métodos, los cebadores PCR se hibridan con las regiones conservadas que flanquean la región polimórfica.

Los cebadores adecuados que se utilizan con la invención se pueden diseñar utilizando cualquier método adecuado. No se pretende que la invención se limite a cualquier cebador particular o par de cebadores. Por ejemplo, los cebadores se pueden diseñar utilizando cualquier programa de software adecuado, tal como LASERGENE®, por ejemplo, tomando en cuenta la información de secuencias disponibles al público. Las secuencias de flaqueo para los polimorfismos identificados en la presente invención están disponibles al público; por consiguiente, los cebadores de amplificación adecuados se pueden construir con base en las reglas bien conocidas de apareamiento de bases. La secuencia de cualquier amplicón se puede detectar como ya se ha analizado anteriormente, por ejemplo, mediante hibridación, hibridación de la matriz, PCR, PCR en tiempo real, LCR, o similares.

En algunas realizaciones, los cebadores de la invención están marcados en forma radiactiva, o marcados por cualquier medio adecuado (por ejemplo, utilizando una etiqueta fluorescente no radioactiva), para permitir una visualización rápida de los amplicones de diferente tamaño después de una reacción de amplificación sin ninguna etapa adicional de marcación o etapa de visualización. En algunas realizaciones, los cebadores no se marcan, y los amplicones se visualizan después de su resolución de tamaño, por ejemplo, después de una electroforesis en gel de agarosa o acrilamida. En algunas realizaciones, la tinción con bromuro de etidio de los amplicones de PCR después de la resolución por tamaño permite la visualización de los amplicones de diferente tamaño.

No se pretende que los cebadores de la invención se limiten a la generación de un amplicón de cualquier tamaño particular. Por ejemplo, los cebadores utilizados para amplificar los loci marcadores y alelos en la presente invención no se limitan a amplificar la región completa del locus relevante, o cualquier subregión del mismo. Los cebadores pueden generar un amplicón de cualquier longitud adecuada para detección. En algunas realizaciones, la amplificación del marcador produce un amplicón de al menos 20 nucleótidos de longitud, o alternativamente, al menos 50 nucleótidos de longitud, o alternativamente al menos 100 nucleótidos de longitud, o alternativamente, al menos 200 nucleótidos de longitud. Los amplicones de cualquier tamaño se pueden detectar utilizando las diversas tecnologías descritas en la presente invención. Las diferencias en la composición o tamaño de las bases se pueden detectar mediante métodos convencionales tales como electroforesis.

Marcadores de correlación para fenotipos

Estas correlaciones se pueden realizar mediante cualquier método que pueda identificar una relación entre un alelo y un fenotipo, o una combinación de alelos y una combinación de fenotipos. Por ejemplo, los alelos en los genes o loci definidos en la presente invención se pueden correlacionar con uno o más fenotipos de cáncer de mama. Más típicamente, estos métodos implican hacer referencia a una tabla de consulta que comprenda las correlaciones entre los alelos del polimorfismo y el fenotipo. La tabla puede incluir los datos para múltiples relaciones alelo-fenotipo y puede tomar en cuenta los efectos aditivos u otros de orden superior de múltiples relaciones alelo-fenotipo. Por

ejemplo, a través del uso de herramientas estadísticas tales como análisis de componentes principales, algoritmos heurísticos, etc.

La correlación de un marcador con un fenotipo incluye opcionalmente realizar una o más pruebas estadísticas para la correlación. Se conocen muchas pruebas estadísticas, y la mayoría se implementan en ordenador para facilidad de análisis. Para la presente invención se conocen una variedad de métodos estadísticos para determinar asociaciones/correlaciones entre rasgos fenotípicos y marcadores biológicos. Hartl (1981) *A Primer of Population Genetics* Washington University, Saint Louis Sinauer Associates, Inc. Sunderland, Mass. ISBN: 0-087893-271-2. Se describen una variedad de modelos estadísticos adecuados en Lynch y Walsh (1998) *Genetics and Analysis of Quantitative Traits*, Sinauer Associates, Inc. Sunderland Mass. ISBN 0-87893-481-2. Estos modelos pueden, por ejemplo, proporcionar correlaciones entre valores genotípicos y fenotípicos, caracterizar la influencia de un locus sobre un fenotipo, clasificar la relación entre el ambiente y el genotipo, determinar la dominancia o penetrancia de los genes, determinar efectos maternos y otros efectos epigenéticos, determinar los componentes principales en un análisis (a través del análisis del componente principal, o "PCA"), y similares. Las referencias citadas en estos textos proporcionan considerables detalles adicionales sobre los modelos estadísticos para correlacionar marcadores y fenotipo.

Además de los métodos estadísticos estándar para determinar la correlación, se pueden usar otros métodos que determinan correlaciones mediante reconocimiento de patrones y entrenamiento, tales como el uso de algoritmos genéticos, para determinar correlaciones entre marcadores y fenotipos. Esto es particularmente útil cuando se identifican correlaciones de orden superior entre múltiples alelos y múltiples fenotipos. Para ilustrar, se pueden acoplar enfoques de red neuronal con programación de tipo algoritmo genético para desarrollo heurístico de un modelo espacial de datos de función-estructura que determina correlaciones entre información genética y resultados fenotípicos. Por ejemplo, NNUGA (Neural Network Using Genetic Algorithms) es un programa disponible (por ejemplo, en la Red Mundial como cs.bgu.ac.il/about.omri/NNUGA que acopla redes neuronales y algoritmos genéticos. Una introducción a las redes neurales se puede encontrar, por ejemplo, en Kevin Gurney, *An Introduction to Neural Networks*, UCL Press (1999) y en la Red Mundial como shef.ac.uk/Psychology/gurney/notes/index.html. Las Redes Neuronales útiles adicionales incluyen aquellas observadas anteriormente en relación con algoritmos genéticos y, por ejemplo, Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press (1995), y Ripley y colaboradores, *Pattern Recognition and Neural Networks*, Cambridge University Press (1995).

Las referencias adicionales que son útiles para comprender las aplicaciones de análisis de datos para utilizar y establecer correlaciones, los componentes principales de un análisis, modelación de redes neuronales, y similares, incluyen, por ejemplo, Hinchliffe, *Modelling Molecular Structures*, John Wiley and Sons (1996), Gibas y Jambeck, *Bioinformatics Computer Skills*, O'Reilly (2001), Pevzner, *Computational Molecular Biology and Algorithmic Approach*, The MIT Press (2000), Durbin y colaboradores, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press (1998), y Rashidi y Buehler, *Bioinformatic Basics: Applications in Biological Science and Medicine*, CRC Press LLC (2000).

En cualquier caso, se puede aplicar esencialmente cualquier prueba estadística en un modelo implementado en ordenador, mediante métodos de programación estándar, o al utilizar cualquiera de una variedad de paquetes de software "fuera de anaquel" que realicen estos análisis estadísticos, incluyendo, por ejemplo, aquellos observados anteriormente y aquellos que están disponibles comercialmente, por ejemplo, de Partek Incorporated (St. Peters, Missouri; www.partek.com), por ejemplo, que proporcionan un software para el reconocimiento de patrones (por ejemplo, que proporciona Partek Pro 2000 Pattern Recognition Software) que se puede aplicar a algoritmos genéticos para análisis de datos multivariantes, visualización interactiva, selección de variables, modelación de redes neuronales y estadística, etc. Las relaciones se pueden analizar, por ejemplo, mediante diagramas de dispersión y bi-diagramas mapeados para análisis de componentes principales (PCA), diagramas de dispersión mapeados de escalamiento multidimensional (MDS), gráficos de dispersión mapeados de escalamiento multidimensional (MDS), gráficos estrella, etc. El software disponible para realización del análisis de correlación incluye SAS, R y MathLab.

El(Los) marcador(es), ya sea polimorfismos o patrones de expresión, se pueden utilizar para cualquiera de una variedad de análisis genéticos. Por ejemplo, una vez que los marcadores han sido identificados, como en el presente caso, se pueden utilizar en varios análisis diferentes para estudios de asociación. Por ejemplo, las sondas se pueden diseñar para microarreglos que interrogan estos marcadores. Otros ejemplos de ensayo incluyen, por ejemplo, los ensayos TaqMan y ensayos de baliza molecular descritos más arriba, así como técnicas convencionales de PCR y/o secuenciación.

Detalles adicionales relacionados con los estudios de asociación se pueden encontrar en la solicitud de patente de los Estados Unidos No. US 10/106,097, la solicitud de patente de los Estados Unidos No. 10/042,819, la solicitud de patente de los Estados Unidos No.10/286,417, la solicitud de patente de los Estados Unidos No.10/768,788, la solicitud de patente de los Estados Unidos No.10/447,685, la solicitud de patente de los Estados Unidos No.10/970,761, y la patente de los Estados Unidos No.7,127,355.

En algunas realizaciones, los datos del marcador se utilizan para realizar los estudios de asociación para mostrar las correlaciones entre marcadores y fenotipos. Esto se puede lograr mediante la determinación de las características de los marcadores en individuos con el fenotipo de interés (es decir, individuos o poblaciones que muestran el fenotipo de interés) y la comparación de la frecuencia de alelos u otras características (niveles de expresión, etc.) de los marcadores en estos individuos con la frecuencia de alelos u otras características en un grupo de individuos de control. Tales determinaciones de marcadores se pueden realizar con base en el genoma, o se pueden enfocar en regiones específicas del genoma (por ejemplo, bloques de haplotipos de interés). En una realización, los marcadores que están enlazados a los genes o loci definidos aquí se evalúan por la correlación con uno o más fenotipos de susceptibilidad específicos de cáncer de mama.

Además de las otras realizaciones de los métodos de la presente invención expuestas en la presente invención, los métodos, adicionalmente permiten la "disección" de un fenotipo. Es decir, un fenotipo particular puede (y típicamente lo hace) resultar de dos o más bases genéticas diferentes. Por ejemplo, un fenotipo de susceptibilidad en un individuo puede ser el resultado de un "defecto" (o simplemente un "defecto" de un alelo particular con respecto a un fenotipo de susceptibilidad depende del contexto, por ejemplo, ya sea que el fenotipo sea conveniente o no conveniente en el individuo en un entorno determinado) en un gen definido en la Tabla 1, mientras que el mismo fenotipo básico en un individuo diferente puede ser el resultado de múltiples "defectos" en múltiples genes. De esta forma, el escaneo de una pluralidad de marcadores (por ejemplo, como en el escaneo de genomas o de bloques de haplotipos) permite la disección de diferentes bases genéticas para fenotipos similares (o graduados).

Como se describió en el párrafo anterior, un método para llevar a cabo estudios de asociación es comparar la frecuencia de alelos (o nivel de expresión) de los marcadores en individuos con un fenotipo de interés ("grupo de casos") con la frecuencia de alelos en un grupo de control de individuos. En un método, se usan SNP informativos para realizar la comparación del patrón de haplotipos del SNP (un "SNP informativo" es un marcador genético de SNP tal como un SNP o un subconjunto (más de uno) de los SNP en un genoma o bloque de haplotipo que tiende a distinguir un SNP o genoma o patrón de haplotipos de otros SNP, genomas o patrones de haplotipos). El enfoque de utilizar SNP informativos tiene una ventaja sobre otros métodos de escaneo o determinación del genotipo del genoma completo conocidos en la técnica, en lugar de leer la totalidad de 3 mil millones de bases de cada genoma del individuo, o incluso leer los 3-4 millones de los SNP comunes que se pueden encontrar, sólo es necesario detectar los SNP informativos a partir de una población de muestras. La lectura de estos SNP informativos, particulares, proporciona suficiente información para permitir que se extraigan datos de asociación estadísticamente precisos de las poblaciones experimentales específicas, como se describió anteriormente.

De esta forma, en una realización de un método para determinar las asociaciones genéticas determinantes, la frecuencia de alelos de los SNP informativos se determina para los genomas de una población de control que no exhiba el fenotipo. La frecuencia de alelos de los SNP informativos también se determina para los genomas de una población que no exhiba el fenotipo. Se comparan las frecuencias de alelos de los SNP informativos. Las comparaciones de frecuencia de alelos se pueden realizar, por ejemplo, al determinar la frecuencia del alelo (número de casos de un alelo particular en una población dividida entre el número total de alelos) en cada ubicación del SNP informativo en cada población y comparando estas frecuencias de alelos. Los SNP informativos que exhiben diferencia entre la frecuencia de alelo de ocurrencia en las poblaciones/grupos de control versus de casos, se seleccionan para análisis. Una vez se seleccionan los SNP informativos, se identifican el(los) bloque(s) de haplotipos de SNP que contienen los SNP informativos, que a su vez identifica una región genómica de interés que se correlaciona con el fenotipo. Las regiones genómicas se pueden analizar por métodos genéticos o biológicos conocidos en la técnica, por ejemplo, para uso como objetivos de descubrimiento de fármacos o como marcadores de diagnóstico.

Los sistemas para realizar las correlaciones anteriores también son una característica de la invención. Típicamente, el sistema incluirá las instrucciones del sistema que correlaciona la presencia o ausencia de un alelo (ya sea detectado directamente o por ejemplo, a través de los niveles de expresión) con un fenotipo predicho. Opcionalmente, las instrucciones del sistema también pueden incluir un software que acepte la información de diagnóstico asociada con cualquier información del alelo detectado, por ejemplo, un diagnóstico de que un sujeto con el alelo relevante tiene un fenotipo particular. Este software puede ser de naturaleza heurística, utilizando tales asociaciones imputadas para mejorar la precisión de las tablas de consulta y/o la interpretación de las tablas de consulta por el sistema. Anteriormente se describieron una variedad de tales enfoques, incluyendo redes neuronales, modelación de Markov, y otros análisis estadísticos.

La invención proporciona los módulos para obtención de datos para detectar uno o más marcador(es) genético(s) detectable(s) (por ejemplo, una o más matrices que comprenden una o más sondas biomoleculares, detectores, manipuladores de fluidos, o similares). Las sondas biomoleculares de este módulo para obtención de datos pueden incluir cualquiera que sean adecuados para detectar el marcador biológico, por ejemplo, sondas de oligonucleótido, proteínas, aptámeros, anticuerpos, etc. Estos pueden incluir manipuladores de muestras (por ejemplo, manipuladores de fluidos), sistemas robóticos de microfluidos, ácido nucleico o módulos para purificación de proteínas, matrices (por ejemplo, matrices de ácido nucleico), detectores, termocicladores o combinaciones de los mismos, por ejemplo, para adquisición de muestras, dilución o toma de alícuotas de muestras, purificación de

materiales marcadores (por ejemplo, ácidos nucleicos o proteínas), amplificación de ácidos nucleicos marcadores, detección ácidos nucleicos marcadores amplificados, y similares.

Por ejemplo, los dispositivos automáticos que se pueden incorporar en los sistemas en la presente invención han sido usados para evaluar una variedad de fenómenos biológicos, incluyendo, por ejemplo, niveles de expresión de genes en respuesta a estímulos seleccionados (Service, 1998a), determinación de genotipo de ADN de alto rendimiento (Zhang y colaboradores, 1999) y muchos otros. En forma similar, también se encuentran disponibles sistemas integrados para la realización de experimentos de mezclado, amplificación de ADN, secuenciación de ADN y similares (Service, 1998b). Están disponibles una variedad de componentes de sistemas automatizados, por ejemplo, a través de Caliper Technologies (Hopkinton, Mass.), que utiliza diferentes sistemas Zymate, que incluyen típicamente, por ejemplo, módulos de robótica y para manipulación de fluidos. En forma similar, el robot ORCAR^{MR} común, que se utiliza en una variedad de sistemas de laboratorios, por ejemplo, para manipulación de bandejas de microtitulación, también se encuentra comercialmente disponible, por ejemplo, a través de Beckman Coulter, Inc. (Fullerton, California). En forma similar, los sistemas microfluidos comercialmente disponibles que se pueden utilizar como componentes del sistema en la presente invención incluyen aquellos de Agilent Technologies y Caliper Technologies. Además, la patentes y la literatura técnica incluye numerosos ejemplos de sistemas de microfluidos, que incluyen aquellos que pueden interconectarse directamente con placas de micropozos para manipulación automática de fluidos.

En los sistemas de la presente invención se pueden utilizar cualquiera entre una variedad de configuraciones de manipulación de líquidos y/o matrices. Un formato común para uso en los sistemas de la presente invención es una placa de microtitulación, en la cual el manipulador de líquidos o de la matriz incluye una bandeja microtitulación. Tales bandejas están comercialmente disponibles y se pueden ordenar en una variedad de tamaños y cantidad de pozos por bandeja, así como con cualquiera entre una variedad de superficies con adición de grupos funcionales para enlazamiento de componentes del ensayo o de la matriz. Las bandejas comunes incluyen la placa ubicua de 96 pozos, con placas de 384 y 1536 pozos también de uso común. Las muestras se pueden procesar en tales bandejas, con todas las etapas de procesamiento siendo realizadas en las bandejas. Las muestras también se pueden procesar en aparatos de microfluidos, o combinaciones de aparatos de microtitulación y microfluidos.

Además de las matrices en fase líquida, se pueden almacenar o analizar los componentes en matrices de fase sólida. Estas matrices fijan los materiales en un patrón espacialmente accesible (por ejemplo, una rejilla de filas y columnas) sobre un sustrato sólido tal como una membrana (por ejemplo, nailon o nitrocelulosa), una superficie polimérica o cerámica, una superficie de vidrio o de sílice modificada, una superficie metálica, o similares. Los componentes se pueden evaluar, por ejemplo, mediante hibridación, por rehidratación local (por ejemplo, utilizando una pipeta u otro elemento para manipulación de fluidos) y transferencia de fluidos, o por raspado de la matriz o cortando los sitios de interés sobre la matriz.

El sistema también puede incluir aparatos de detección que se utilizan para detectar información de alelos, utilizando cualquiera de los enfoques observados en la presente invención. Por ejemplo, se puede incorporar en el sistema un detector configurado para detectar productos PCR en tiempo real (por ejemplo, un detector de luz, tal como un detector de fluorescencia) o un lector de matrices. Por ejemplo, el detector se puede configurar para detectar una emisión de luz de una reacción de hibridación o amplificación que comprende un alelo de interés, en donde la emisión de luz es indicativa de la presencia o ausencia del alelo. Opcionalmente, se proporciona un enlace operable entre el detector y un ordenador que contiene las instrucciones del sistema observadas anteriormente, permitiendo la entrada automática de información detectada específica del alelo al ordenador, que puede, por ejemplo, almacenar la información en la base de datos y/o ejecutar las instrucciones del sistema para comparar la información específica del alelo detectado con la tabla de consulta.

También se pueden incorporar dentro del sistema, sondas que se utilizan para generar información detectada por el detector, junto con cualquier otro hardware o software para utilizar las sondas para detectar el amplicón. Estas pueden incluir elementos termocicladores (por ejemplo, para realizar la amplificación PCR o LCR del alelo que será detectado por las sondas), matrices después de lo cual se organizan y/o hibridan las sondas, o similares. Los elementos para manipulación de fluidos mencionados anteriormente para procesamiento de muestras, se pueden utilizar para mover materiales de la muestra (por ejemplo, ácidos nucleicos molde y/o proteínas que se van a detectar) cebadores, sondas, amplicones, o similares, en contacto entre sí. Por ejemplo, el sistema puede incluir un conjunto de sondas marcadoras o cebadores configurados para detectar al menos un alelo de uno o más genes o loci ligados asociados con un fenotipo. El módulo detector se configura para detectar una o más salidas de señal desde el conjunto de sondas marcadoras o cebadores, o un amplicón producido a partir del conjunto de sondas marcadoras o cebadores, identificando así la presencia o ausencia del alelo.

La muestra que se va a analizar es opcionalmente parte del sistema, o se puede considerar separada del mismo. La muestra incluye opcionalmente por ejemplo, ADN genómico, ADN genómico amplificado, ADNc, ADNc amplificado, ARN, ARN amplificado, proteínas, etc., como se mencionó aquí.

Opcionalmente, se proporcionan componentes del sistema para interacción con el usuario. Por ejemplo, los sistemas pueden incluir una pantalla para el usuario para observar la salida de instrucciones del sistema implementadas en el ordenador, dispositivos de entrada para el usuario (por ejemplo, teclados o dispositivos señaladores tales como un ratón), para el ingreso de comandos por el usuario y activación del sistema, etc. Típicamente, el sistema de interés incluye un ordenador, en donde las diferentes instrucciones del sistema implementadas en el ordenador se incluyen en el software del ordenador, por ejemplo, se almacenan en un medio legible para el ordenador.

Las aplicaciones estándar de escritorio tales como un software para procesamiento de palabra (por ejemplo, Microsoft Word^{MR} o Corel WordPerfect^{MR}) y un software para base de datos (por ejemplo, software de hoja de cálculo tal como Microsoft Excel^{MR}, Corel Quattro Pro^{MR} o programas de bases de datos tales como Microsoft Access^{MR}, Sequel^{MR}, Oracle^{MR}, Paradox^{MR}) se pueden aplicar a la presente invención mediante la entrada de una cadena de caracteres correspondiente a un alelo en la presente invención, o una asociación entre un alelo y un fenotipo. Por ejemplo, los sistemas pueden incluir un software que tiene la información adecuada de cadena de caracteres, por ejemplo, utilizada junto con la interfaz del usuario (por ejemplo, un GUI en un sistema operativo estándar, tal como un sistema Windows, Macintosh o LINUX) para manipular las cadenas de caracteres. Los programas especializados de alineación de secuencias tales como BLAST también se pueden incorporar en los sistemas de la invención para el alineamiento de ácidos nucleicos o proteínas (o las cadenas de caracteres correspondientes), por ejemplo, para identificar y relacionar múltiples alelos.

Como se observa, los sistemas pueden incluir un ordenador con una base de datos adecuada y una secuencia o correlación de alelos de la invención. El software para alineación de secuencias, así como los conjuntos de datos ingresados en el sistema del software que comprende cualquiera de las secuencias de la invención pueden ser una característica de la invención. La computadora puede ser, por ejemplo, una PC ((Intel x86 o DOS^{MR} compatible con el chip Pentium, OS2^{MR}, WINDOWS^{MR}, WINDOWS NT^{MR}, WINDOWS95^{MR}, WINDOWS98^{MR}, WINDOWS2000^{MR}, WINDOWSME, o una máquina con base en LINUX, una MACINTOSH^{MR}, Power PC, o una máquina con base en UNIX (por ejemplo, una estación de trabajo SUN^{MR} o una máquina con base en LINUX) u otro ordenador común en el comercio que es conocido por el experto. El software para ingresar y alinear o bien manipular secuencias está disponible, por ejemplo, BLASTP y BLASTN, o se puede ser construido fácilmente por alguien entrenado utilizando un lenguaje de programación estándar como VisualBasic, Fortran, Basic, Java o similares.

La invención provee métodos para determinar el perfil polimórfico de un individuo en uno o más de los SNP de la invención. Los SNP incluyen aquellos mostrados en las Tablas 1 a 5.

El perfil polimórfico constituye las formas polimórficas que ocupan los diferentes sitios polimórficos en un individuo. En un genoma diploide, dos formas polimórficas, iguales o diferentes entre si, por lo general ocupan cada sitio polimórfico. De esta forma, el perfil polimórfico en los sitios X y Y se puede representar en la forma X (x1, x1) y Y (y1, y2), en donde x1, x1 representa dos copias del alelo x1 que ocupan el sitio X y y1, y2 representa alelos heterocigotos que ocupan el sitio Y.

El perfil polimórfico de un individuo puede recibir una puntuación mediante comparación con las formas polimórficas asociadas con la resistencia o susceptibilidad a los fenotipos de cáncer de mama que se presentan en cada sitio. La comparación se puede realizar en al menos, por ejemplo, 1, 2, 5, 10, 25, 50, o todos los sitios polimórficos, y, opcionalmente, otros en el desequilibrio de enlace con los mismos. Los sitios polimórficos se pueden analizar en combinación con otros sitios polimórficos. Sin embargo, el número total de sitios polimórficos analizados es por lo general menor que 10,000, 1000, 100, 50 ó 25 y pueden ser de aproximadamente 10 o menos, aproximadamente 5 o menos, o aproximadamente 2 o menos.

El número de alelos de resistencia o susceptibilidad presentes en un individuo particular se puede combinar aditivamente o como una relación para proporcionar una puntuación total para la propensión genética del individuo a los fenotipos de cáncer de mama (véase, la solicitud de patente de los Estados Unidos No. 60/566,302, la solicitud de patente de los Estados Unidos No.60/590,534, la solicitud de patente de los Estados Unidos No.10/956,224, y el documento WO/2005/086770). Los alelos de resistencia pueden recibir cada uno arbitrariamente puntajes como +1 y alelos de susceptibilidad como -1 (o viceversa). Por ejemplo, si un individuo se clasifica en 100 sitios polimórficos de la invención y es homocigoto para resistencia en todos ellos, se le puede asignar un puntaje de 100% de propensión genética para resistencia a fenotipos de cáncer de mama o 0% de propensión para susceptibilidad a fenotipos de cáncer de mama. Lo inverso aplica si el individuo es homocigoto para todos los alelos de susceptibilidad. Más típicamente, un individuo es homocigoto para alelos de resistencia en algunos loci, homocigotos para alelos de susceptibilidad en algunos loci, y heterocigoto para alelos de resistencia/susceptibilidad en otros loci. Tal propensión genética del individuo para fenotipos de cáncer de mama se le puede adjudicar un puntaje asignándole a todos los alelos de resistencia un puntaje de +1, y a todos los alelos de susceptibilidad un puntaje de -1 (o viceversa) y combinando los puntajes. Por ejemplo, si un individuo tiene 102 alelos de resistencia y 204 alelos de susceptibilidad, se el individuo puede recibir un puntaje de propensión genética del 33% para resistencia y de propensión genética del 67% para susceptibilidad. Alternativamente, los alelos de resistencia de homocigotos se les puede asignar un puntaje de +1, los alelos heterocigotos un puntaje de cero, y alelos de susceptibilidad homocigotos un puntaje de -1. Los números relativos de alelos de resistencia y alelos de susceptibilidad también se pueden expresar como

porcentaje. De esta forma, un individuo que es homocigoto para alelos de resistencia en 30 sitios polimórficos, homocigoto para alelos de susceptibilidad en 60 sitios polimórficos, y heterocigotos en los 63 sitios restantes se le asigna una propensión genética del 33% para resistencia. Como una alternativa adicional, la homocigosidad para susceptibilidad puede tener un puntaje de +2, heterocigosidad, como +1, y homocigosidad para resistencia como 0.

5 La puntuación del individuo, y la naturaleza del perfil polimórfico son útiles en el pronóstico o diagnóstico de la susceptibilidad del individuo a un fenotipo de cáncer de mama. Opcionalmente, un paciente puede ser informado de susceptibilidad a un fenotipo de cáncer de mama indicado por el perfil genético. La presencia de una alta propensión genética a fenotipos de cáncer de mama se puede ser tratada como una advertencia para iniciar un tratamiento profiláctico o terapéutico. Por ejemplo, los individuos con riesgo elevado de desarrollar un fenotipo de cáncer de mama pueden ser controlados en forma diferente (por ejemplo, una mamografía más frecuente), o pueden ser tratados profilácticamente (por ejemplo, con uno o más fármacos). La presencia de una alta propensión a un fenotipo de cáncer de mama también indica la utilidad de realizar análisis secundarios, tales como una biopsia y otros métodos conocidos en la técnica.

10 El perfilado polimórfico se utiliza, por ejemplo, en la selección de agentes para afectar el tratamiento o la profilaxis de los fenotipos de cánceres de mama en un individuo determinado. Los individuos que tienen perfiles polimórficos similares probablemente respondan a agentes de una forma similar.

15 El perfilado polimórfico también es útil para estratificar individuos en ensayos clínicos de agentes que son probados por la capacidad para tratar fenotipos de cáncer de mama o condiciones relacionadas. Tales ensayos se realizan sobre poblaciones tratadas o de control que tienen perfiles polimórficos similares o idénticos (véase, el documento EP 99965095.5), por ejemplo, un perfil polimórfico que indica que un individuo tiene un mayor riesgo de desarrollar un fenotipo de cáncer de mama. El uso de poblaciones que coinciden genéticamente elimina o reduce la variación en el resultado del tratamiento debido a factores genéticos, lo que conduce a una evaluación más precisa de la eficacia de un fármaco potencial. Los algoritmos implementados en un ordenador se pueden utilizar para identificar subpoblaciones genéticamente más homogéneas en las cuales el tratamiento o la profilaxis tiene un efecto significativo no obstante que el tratamiento o profilaxis es ineficaz en las poblaciones mayores más heterogéneas. En tales métodos, se proporcionan los datos para una primera población con un fenotipo de cáncer de mama tratado con un agente, y una segunda población también con el fenotipo de cáncer de mama, pero tratada con un placebo. El perfil polimórfico de individuos en las dos poblaciones se determina en al menos un sitio polimórfico en o dentro de 100 kb o 50 kb o 20 kb de una región definida por los SNP suministrados en las Tablas 1 a 5. También se proporcionan los como si cada paciente en las poblaciones alcanza un punto final deseado indicativo de exitoso en el tratamiento o la profilaxis. Las subpoblaciones de cada una de la primera y segunda poblaciones se seleccionan luego de tal forma que los individuos en las subpoblaciones tengan mayor similitud de perfiles polimórficos entre si que los individuos en las poblaciones primeras y segundas originales. Existen muchos criterios mediante los cuales se puede evaluar la similitud. Por ejemplo, un criterio es solicitar que los individuos en las subpoblaciones tengan al menos un alelo de susceptibilidad en cada uno de al menos uno de los genes anteriores. Otro criterio es que los individuos en las subpoblaciones tengan al menos un 75% de alelos de susceptibilidad para cada uno de los sitios polimórficos en los cuales se determina el perfil polimórfico. Sin importar los criterios utilizados para evaluar la similitud, los datos de punto final de las subpoblaciones se comparan para determinar si el tratamiento o profilaxis ha alcanzado un resultado estadísticamente significativo en las subpoblaciones. Como resultado de una implementación en ordenador, se pueden analizar miles de millones de criterios para similitud para identificar una o unas cuantas subpoblaciones que muestren significancia estadística.

20 El perfilado polimórfico también es útil para excluir individuos sin predisposición a los fenotipos de cáncer de mama de los ensayos clínicos. La inclusión de estos individuos en los ensayos incrementa el tamaño de la población necesaria para alcanzar un resultado estadísticamente significativo. Los individuos sin predisposición a fenotipos de cáncer de mama se pueden identificar al determinar los números de alelos de resistencia y susceptibilidad en un perfil polimórfico como se describió anteriormente. Por ejemplo, si un sujeto se somete a la determinación de genotipo en diez sitios en diez genes de la invención asociados con los fenotipos de cáncer de mama, se determinan veinte alelos en total. Si más del 50% y preferiblemente más del 60% o el 75% de estos son genes de resistencia, es poco probable que el individuo desarrolle un fenotipo de cáncer de mama y se puede excluir del ensayo.

25 En otras realizaciones, la estratificación de individuos en ensayos clínicos se puede llevar a cabo utilizando perfilado polimórfico en combinación con otros métodos de estratificación, incluyendo, pero no sin limitarse a, historia familiar, modelos de riesgo (por ejemplo, puntuación de Gail, Modelo de Claus), fenotipos clínicos (por ejemplo, lesiones atípicas), y biomarcadores de candidatos específicos.

30 Los perfiles polimórficos también se pueden utilizar después de terminado el ensayo clínico para elucidar las diferencias en la respuesta a un tratamiento dado. Por ejemplo, el conjunto de polimorfismos se puede utilizar para estratificar los pacientes inscritos en los subtipos o clases de enfermedad. También es posible utilizar los polimorfismos para identificar subconjuntos de pacientes con perfiles polimórficos similares que tengan una respuesta inusual (alta o baja) al tratamiento o que no respondan en absoluto (que no responden). De esta forma, la información acerca de los factores genéticos subyacentes que influyen en la respuesta al tratamiento se puede

utilizar en muchos aspectos del desarrollo del tratamiento (estos varían desde la identificación de nuevos objetivos, a través del diseño de nuevas pruebas para la marcación del producto y el direccionamiento del paciente). Adicionalmente, se pueden utilizar los polimorfismos para identificar los factores genéticos implicados en la respuesta adversa al tratamiento (eventos adversos). Por ejemplo los pacientes que muestran una respuesta adversa pueden tener perfiles polimórficos más similares de lo que se esperaría por casualidad. Esto permite la pronta identificación y exclusión de estos individuos del tratamiento. También se proporciona información que se puede utilizar para comprender las causas biológicas de los eventos adversos y modificar el tratamiento para evitar estos resultados.

Los perfiles polimórficos también se pueden utilizar para otros fines, incluyendo prueba de paternidad y análisis forenses como se describe en la patente de los Estados Unidos No. 6,525,185. En los análisis forenses, el perfil polimórfico a partir de una muestra en la escena de un crimen se compara con el de un sospechoso. Una coincidencia entre los dos es evidencia de que el sospechoso, de hecho, cometió el crimen, mientras que la falta de una coincidencia excluye al sospechoso. Los presentes sitios polimórficos se pueden utilizar en estos métodos, al igual que se pueden utilizar otros sitios polimórficos en el genoma humano.

Ejemplos

Ejemplo 1 - Evaluación de un panel de SNP para evaluación del riesgo de cáncer de mama en un estudio jerarquizado de control del Women's Health Initiative Rationale

Es poco claro el grado en el cual la variación genética asociada con el cáncer de mama recientemente descubierto (BCa) puede ayudar en la evaluación del riesgo de BCa. Se evaluó el efecto de la adición de la información de riesgo de un panel de 7 SNP asociados con BCa sobre la estratificación del riesgo ofrecida por el modelo de Gail.

Métodos

Se examinaron 1,664 mujeres que desarrollaron BCa después de un sorteo aleatorio en el ensayo clínico WHI y 1,636 coincidieron con los controles libres de BCa. Los controles se igualaron con base en la edad, origen étnico reportado por la persona, participación en el ensayo clínico, años desde el sorteo aleatorio, y estado de histerectomía. Siete de los SNP de la literatura publicada publicada que cumplían con criterios rigurosos de significancia amplia de genoma y replicación se seleccionaron para la determinación del genotipo (panel 7-SNP). Para el riesgo de SNP modelo a través de 7 SNP del panel seleccionado, se utilizaron los estimados de tamaño efectivo reportado previamente junto con un modelo multiplicativo de riesgo relativo. Para producir un riesgo clínico/genético combinado, los estimados de riesgo absoluto del modelo de Gail se multiplicaron por los riesgos relativos combinados de SNP para producir un riesgo clínico/genético combinado. Se evaluó el desempeño de la clasificación utilizando tablas de reclasificación para cuantificar el índice de reclasificación neta (NRI), y las curvas características de operación del receptor (ROC).

Resultados

Las asociaciones individuales de SNP con BCa para mujeres blancas fueron generalmente consistentes con reportes previos. El riesgo absoluto de 5 años de Gail y el estimado de riesgo relativo de 7-SNP fueron ambos asociados con la incidencia de BCa cuando se probaron por regresión logística; hubo una correlación estadísticamente significativa ($P = 0,01$), aunque muy débil ($r = 0,044$), entre el riesgo de Gail y el riesgo de SNP.

En esta cohorte, un incremento de dos veces en el riesgo de Gail produce un aumento menor a dos veces en la incidencia de cáncer, lo que sugiere que el riesgo de Gail no está tan bien calibrado en este conjunto de datos, lo cual es consistente con un reporte anterior (Chiebowski y colaboradores, 2007). El predictor combinado estaba más fuertemente asociado ya sea con el riesgo de Gail o los componentes de SNP solos. En el análisis de la curva ROC, la combinación de riesgo de Gail y de SNP tenía un área bajo la curva (AUC) de 0,594 (95% de IC: 0,576-0,612) comparado con 0,556 (95% IC: 0,537-0,575) para el riesgo de Gail solamente. La diferencia en AUC fue estadísticamente significativo (95% CI: 0,025-0,050, P empírico $<0,001$). En el análisis de la tabla de reclasificación, se seleccionaron umbrales de riesgo de 5 años que son potencialmente y clínicamente significativos en el contexto para tomar una decisión acerca de la prevención primaria: $<1,5\%$ (por debajo del riesgo promedio), $1,5\text{-}2\%$ (riesgo moderado) y $>2\%$ (riesgo elevado). La reclasificación en mujeres blancas se evaluó para el puntaje combinado SNP X Gail versus el riesgo de Gail solo. El NRI en este contexto fue de 0,09 ($Z = 4,5$, $P = 0,033$), con una mejora de la clasificación para el 6,4% de los casos y el 2,6% de los controles.

Las mujeres con biopsias de mama previas parecen ser un subgrupo con beneficio particular a partir de la reclasificación utilizando el puntaje combinado SNP X Gail.

En este subgrupo, el modelo de Gail tuvo una AUC de 0,514 (95% IC: 0,471 a 0,561), indistinguible de la casualidad. Comparado con la cohorte total, la diferencia en AUC para el modelo de Gail estaba significativamente en el límite

(95% IC: -0,002 a 0,081, P empírico = 0,06). El modelo combinado tuvo un AUC de 0,571 (95% IC: 0,526 a 0,614). Un estimado de arranque de la diferencia en AUC para el modelo combinado versus Gail solamente también fue significativo (95% IC: 0,029 a 0,085, P <0,001).

5 La métrica de reclasificación en el subconjunto de biopsias indicó que el NRI es 0,18, lo cual es muy significativo a pesar del menor número de casos ($Z = 3,9$, $P = 4.9 \times 10^{-5}$). Aquí, la clasificación mejoró para el 14.8% de los controles ($P = 1.5 \times 10^{-5}$), pero solamente el 2.8% de los casos ($p = 0.16$; Tablas 7 y 8). El nuevo muestreo de arranque indicó que la diferencia en NRI entre la cohorte completa y el subconjunto de la biopsia fue estadísticamente significativo. Con base en 1000 replicas de arranque, un intervalo de confianza del 95% para la mejora en NRI en el subgrupo de biopsia se extendió de 0.02 hasta 0.16, con un P empírico = 0.03.

10 Tabla 7. Pruebas de regresión logística de asociación con subtipos de receptores de cáncer de mama

Predictor	Tumores positivos para ER		Tumores negativos para ER	
	β (95% de IC)	P	β (95% de IC)	P
log(riesgo de 5 años de Gail)	0.55 (0.37 a 0.72)	1.1×10^{-9}	-0.03 (-0.37 a 0.32)	0.89
log(riesgo de SNP)	1.20 (0.92 a 1.47)	1.7×10^{-18}	0.56 (0.03 a 1.09)	0.04
log(riesgo de SNP X Gail)	0.72 (0.57 a 0.87)	2.4×10^{-22}	0.14 (-0.14 a 0.43)	0.32

Tabla 8. Índice de probabilidad combinado para el panel de 7-SNP que indica un valor predictivo mejorado para el cáncer de mama ER+ por oposición al cáncer de mama ER-

Identificación de los rs de la base de datos del SNP	Gen	Frecuencia del alelo de riesgo	Índice de probabilidad de cáncer de mama		
			Todo	ER+	ER-
rs2981582	FGFR2	0.38	1.26	1.3	1.035
rs3803662	TNRC9	0.25	1.2	1.275	1.105
rs889312	MAP3K1	0.28	1.13	1.12	1.07
rs13387042	(ninguno)	0.5	1.2	1.22	1.06
rs13281615	(ninguno)	0.4	1.08	1.13	1.03
rs4415084	FGF10	0.44	1.16	1.23	0.98
rs3817198	LSP1-H19	0.3	1.07	1.07	1.04
Índice de probabilidad combinado			2.7484	3.3682	1.2939

15 Discusión

Una estrategia que combina tanto los factores de riesgo clínico (riesgo de Gail) como los factores de riesgo genético comunes bien validados (panel 7-SNP) da como resultado una mejora en la clasificación de los riesgos BCa en mujeres blancas posmenopáusicas. Esto puede tener implicaciones significativas para informar las estrategias primarias de prevención y/o detección.

20 Ejemplo 2 - Uso de los SNP para la evaluación del riesgo de cáncer de mama : modelo 10-SNP

Determinación del genotipo de SNP

Se identificaron diez (10) SNP que se ha reportado que están asociados con cáncer de mama con alta significancia estadística a través de múltiples conjuntos de muestras y se muestran en la Tabla 9.

En el Ejemplo 1 se describe la determinación del genotipo de 7 de 10 SNP. Se determinó el genotipo de los 3 SNP restantes (rs4973768, rs6504950, rs11249433) sobre la plataforma Sequenom MassArray. Se diseñaron dos ensayos en orientaciones opuestas para cada uno de los SNP. Dos de estos SNP (rs6504950 y rs11249433) habían sido previamente sometidos a la determinación del genotipo sobre las mismas muestras en matrices de oligonucleótidos. Las muestras que se habían desempeñado deficientemente en la determinación de genotipo por Sequenom también se desempeñaron deficientemente en este conjunto de datos, con el acuerdo de menos del 90% en estos dos SNP, comparado con el 99.9% para las otras muestras. Como resultado, se excluyeron los datos para este mismo conjunto de muestras problemáticas.

La Tabla 10 resume los resultados de la determinación del genotipo para los 10 SNP asociados con cáncer de mama.

Tabla 9. loci replicados asociados con cáncer de mama invasivo

Identificación de los rs de la base de datos del SNP	Gen	Ubicación	Frecuencia ¹	OR (95% CI) ²	Referencia
rs2981582	FGFR2	10q	0.38	1.26 (1.23-1.30)	Easton y colaboradores, 2007
rs3803662	TNRC9	16q	0.25	1.20 (1.16-1.24)	Easton y colaboradores, 2007
rs889312	MAP3K1	5q	0.28	1.13(1.10-1.16)	Stacey y colaboradores, 2007
rs13387042	(ninguno)	2q35	0.50	1.20 (1.14-1.26)	Easton y colaboradores, 2007
rs13281615	(ninguno)	2q35	0.50	1.08 (1.05-1.11)	Stacey y colaboradores, 2008
rs4415084	(ninguno)	8q24	0.40	1.08 (1.05-1.11)	Easton y colaboradores, 2007
rs3817198	FGF10	5p	0.44	1.16 (1.10-1.21)	Ahmed y colaboradores, 2009
rs4973768	LSP1	11p	0.30	1.07 (1.04-1.11)	Ahmed y colaboradores, 2009
rs6504950	SLC4A7	3p24	0.46	1.11 (1.08-1.13)	Thomas y colaboradores, 2009
rs11249433	COX11	17q23.2	0.73	1.05 (1.03-1.09)	
	FCGR1B	1p11.2	0.39	1.14 (1.10-1.19)	

¹Frecuencia de alelo de alto riesgo.

² Índice de probabilidad (e intervalo de confianza) por copia del alelo de alto riesgo.

Tabla 10. Desempeño de la determinación del genotipo de los loci de cáncer de mama

Identificación de los rs de la base de datos del SNP	Plataforma	Intervalo de confianza	Frecuencia ¹	P_{HWE} ²
rs2981582	Matriz	1.000	0.406	0.47
rs3803662	Matriz	1.000	0.289	0.46
rs889312	Matriz	0.996	0.285	0.68
rs13387042	Sequenom	0.971	0.512	0.68

Identificación de los rs de la base de datos del SNP	Plataforma	Intervalo de confianza	Frecuencia ¹	P_{HWE} ²
rs13281615	Matriz	1.000	0.420	0.29
rs4415084	Matriz	0.997	0.401	0.06
rs3817198	Matriz	1.000	0.324	0.48
rs4973768	Sequenom	0.007	0.496	0.14
rs6504950	Sequenom	0.998	0.727	0.79
rs11249433	Sequenom	0.995	0.426	0.82

¹Frecuencia del alelo de alto riesgo reportado anteriormente en mujeres blancas.

²Valor P para el equilibrio de Hardy-Weinberg, a partir de una prueba de relación probable, en mujeres blancas.

Se excluyeron 264 muestras de los análisis que tuvieron más de dos genotipos perdidos fuera de los 10 SNP; para las muestras restantes incluidas en los análisis, 96% tuvieron datos completos para los 10 SNP.

5 Cada SNP fue analizado individualmente por la asociación con cáncer de mama invasivo mediante regresión logística bajo un modelo aditivo logarítmico sin covarianzas. Los resultados para las mujeres blancas fueron generalmente consistentes con los reportes previos, como se muestra en la Tabla 11.

Tabla 11. Ensayos de SNP individuales de asociación con cáncer de mama invasivo en mujeres blancas

Identificación de los rs de la base de datos del SNP	OR (95% de IC) ¹	P
rs2981582	1.37 (1.23-1.52)	2.5310 ⁻⁹
rs3803662	1.21 (1.08-1.35)	1.0310 ⁻³
rs889312	1.26 (1.13-1.41)	4.9310 ⁻⁵
rs13387042	1.16 (1.05-1.29)	0.0032
rs13281615	1.11 (1.00-1.23)	0.042
rs4415084	1.18 (1.06-1.31)	0.0020
rs3817198	1.12 (1.00-1.24)	0.040
rs4973768	1.03 (0.94-1.14)	0.51
rs6504950	1.08 (0.97-1.21)	0.18
rs11249433	1.10 (0.99-1.21)	0.076
¹ Índice de probabilidad (e intervalo de confianza) por copia del alelo de alto riesgo reportado anteriormente.		

10 Ninguno de los SNP nuevos se asociaron significativamente con cáncer de mama, pero todas las pruebas tendieron en la dirección esperada y los intervalos de confianza abarcaron los índices de probabilidad anteriormente

reportados. No se detectaron interacciones significativas por pares entre los 10 SNP (45 ensayos distintos produjeron dos ensayos con $P < 0.05$, y ninguno con $P < 0.01$).

El puntaje compuesto de riesgo de SNP

5 Para el modelo de riesgo de SNP usando 10 SNP seleccionados asociados con cáncer de mama, se usaron los estimados de tamaños de efecto previamente reportados. Se utilizó un modelo multiplicador para el riesgo relativo a través de los SNP, donde los valores de riesgo para cada SNP se escalaron para tener una población promedio de 1 con base en las frecuencias esperadas de los tres posibles genotipos diploides. A los genotipos perdidos también se les asignó un riesgo relativo de 1.

10 El riesgo absoluto de 5 años de Gail y el estimado de riesgo relativo de 10-SNP para la asociación con la incidencia de cáncer de mama mediante regresión logística con estimados de riesgo transformados por logaritmo se analizaron en forma separada. Ambos estaban fuertemente asociados, como se muestra en la Tabla 12.

Tabla 12. Pruebas de regresión logística de la asociación con cáncer de mama invasivo en mujeres blancas

Puntaje de riesgo	β (95% de IC) ¹	OR por 2 X riesgo ²	P
log(riesgo de 5 años de Gail)	0.51 (0.34 a 0.68)	1.42 (1.26-1.60)	3.8×10^{-9}
log(riesgo de SNP)	1.07 (0.83 a 1.30)	2.09 (1.78-2.46)	1.8×10^{-19}
log(riesgo de SNP X Gail)	0.68 (0.54 a 0.82)	1.61 (1.46-1.77)	3.6×10^{-23}

¹Coefficiente de regresión logístico para el puntaje de riesgo.

²Índice de probabilidad ajustado correspondiente a 2 veces el incremento en el puntaje de riesgo.

15 El riesgo de Gail y el riesgo de SNP se correlacionaron débilmente aunque fueron significativos ($r = 0.043$, $P = 0.010$). El predictor combinado formado al multiplicar el riesgo absoluto de Gail por el riesgo relativo de SNP se asoció más fuertemente que cualquier componente solo. La inclusión del riesgo de Gail y el riesgo de SNP como términos separados mejoró adicionalmente el ajuste ($P = 1.8 \times 10^{-4}$). Un término de interacción no mejoró la predicción del estado de cáncer de mama ($P = 0.50$). El puntaje combinado de 10-SNP X Gail parece ser ligeramente más informativo que el puntaje de 7-SNP X Gail: en un modelo con ambos puntajes, el puntaje 10-SNP X Gail mejoró el ajuste con respecto al puntaje de 7-SNP solo (prueba de índice de probabilidad: $P = 0.060$) pero no viceversa ($P = 0.47$).

20

Se utilizó el ensayo de Hosmer-Lemeshow para evaluar la calibración de los puntajes de riesgo 10-SNP. Como con los puntajes de 7-SNP, los puntajes de 10-SNP parecen estar bien calibrados ($P = 0.98$, Figura 1).

Desempeño de la clasificación

25 El desempeño de la clasificación se evaluó utilizando las curvas características de operación del receptor (ROC). El AUC para los puntajes de 7-SNP X Gail y 10-SNP X Gail permaneció esencialmente sin cambio (0.599 y 0.600, respectivamente), y la diferencia no es significativa (figura 2).

30 También se evaluó la precisión de la clasificación utilizando tablas de reclasificación (Cook y colaboradores, 2006), y se cuantificaron las diferencias en la clasificación mediante "la mejora neta de reclasificación" o NRI (Pencina y colaboradores, 2008). Umbrales de riesgo de cinco años de 1.5% (para riesgos por debajo del promedio) y el 2% (para riesgo elevado) y reclasificación evaluada para el puntaje combinado de SNP X Gail versus el riesgo de Gail solo, en mujeres blancas (Tabla 13).

Tabla 13. Tabla de reclasificación para el riesgo de SNP X Gail versus el riesgo de Gail en mujeres blancas

Riesgo de 5 años de Gail	Riesgo de SNP X riesgo de 5 años de Gail			
	< 1.5%	1.5% - 2.0%	> 2.0%	Total
< 1.5%				
Mujeres	942	245	86	1273
Eventos	400	131	57	588
Sin eventos	542	114	29	685
Proporción	0.425	0.535	0.663	0.462
1.5% - 2.0%				
Mujeres	336	283	280	899
Eventos	147	139	169	455
Sin eventos	189	145	111	445
Proporción	0.438	0.489	0.604	0.506
> 2.0%				
Mujeres	49	134	724	907
Eventos	23	67	425	515
Sin eventos	26	67	299	392
Proporción	0.469	0.500	0.587	0.568
Total				
Mujeres	1327	662	1090	3079
Eventos	570	337	651	1558
Sin eventos	757	326	439	1522
Proporción	0.430	0.508	0.597	0.506

La NRI para esta tabla es 0.095 ($Z = 4.4$, $P = 6.1 \times 10^{-6}$). La clasificación mejoró en el 7.7% de los casos ($P = 4.2 \times 10^{-7}$), y 1.8% de los controles ($P = 0.11$). Estos resultados fueron ligeramente mejores que pero no significativamente diferentes de, los resultados del modelo 7-SNP.

También se evaluó directamente la reclasificación bajo los modelos 10-SNP versus 7-SNP. La NRI para el modelo 10-SNP fue de 0.011, que no fue significativo ($Z = 0.72$, $P = 0.23$). La reclasificación neta de los casos mejoró en un 2.6% ($P = 0.007$), pero empeoró para el 1.6% de los controles ($P = 0.94$). Si la NRI se evalúa con un gran número de valores de umbral, entonces mejoran las estadísticas ya que más mujeres tienen cambios en los puntajes que son suficientemente grandes para moverlas a una nueva categoría. Por ejemplo, con 25 umbrales con cuantiles de 4% de riesgo, la NRI para los modelos 10-SNP versus 7-SNP es de 0.056 ($Z = 1.8$, $P = 0.04$), y con 100 umbrales, la NRI mejora hasta 0.069 ($Z = 2.0$, $P = 0.02$).

Conclusiones

- 5 En este conjunto de datos, se demuestra que el clasificador de 10-SNP es esencialmente tan predictivo y bien calibrado como se demostró para el clasificador 7-SNP. Los conjuntos de datos de WHI no parecen ser lo suficientemente grandes para demostrar efectivamente que el clasificador 10-SNP es mejor que el clasificador 7-SNP. Sin embargo, la mejora más pequeña esperada a partir de la adición de los SNP al modelo probablemente será clínicamente significativa en algunos contextos.
- Se apreciará por los expertos en la técnica que se pueden realizar muchas variaciones y/o modificaciones a la invención como se muestra en las realizaciones específicas sin apartarse del espíritu o alcance de la invención como se describió ampliamente. Por lo tanto, las realizaciones de la presente invención, se considerarán en todos los aspectos como ilustrativas y no restrictivas.
- 10 Todas las publicaciones discutidas y/o referenciadas en la presente invención se incorporan aquí en su totalidad.
- La presente solicitud reivindica la prioridad de la solicitud de patente de los Estados Unidos No. 61/182,809 presentada el 1 de junio de 2009, y la solicitud de patente de los Estados Unidos No. 61/258,420 presentada el 5 noviembre de 2009, cuyo contenido total de ambas se incorpora en la presente invención como referencia.
- 15 Cualquier discusión de los documentos, actas, materiales, dispositivos, artículos o similares que se hayan incluido en la presente memoria descriptiva tiene únicamente el propósito de proporcionar un contexto para la presente invención. No se debe tomar como una admisión de que cualquiera o todas estas cuestiones formen parte de la base de la técnica anterior o sean del conocimiento general común en el campo relevante para la presente invención, ya que existió antes de la fecha de prioridad de cada reivindicación de esta solicitud.
- Listado de secuencias
- 20 <110> Genetic Technologies Ltd.
- <120> Métodos para evaluar el riesgo de cáncer de mama
- <130> 509400
- <150> 61/258420 <151> 2009-11-05
- <150> 61/182,809 <151> 2009-06-01
- 25 <160> 20
- <170> PatentIn versión 3.5
- <210> 1
- <211> 20
- <212> ADN
- 30 <213> Secuencia artificial
- <220>
- <223> Cebador oligonucleótido
- <400> 1
- tatgggaagg agtcgttgag 20
- 35 <210> 2
- <211> 20
- <212> ADN
- <213> Secuencia artificial

<220>
 <223> Cebador oligonucleótido
 <400> 2
 ctgaatcact ccttgccaac 20
 5 <210> 3
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 10 <223> Cebador oligonucleótido
 <400> 3
 caaaatgatc tgactactcc 20
 <210> 4
 <211> 20
 15 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido
 <400> 4
 20 tgaccagtcg tgtatgtatc 20
 <210> 5
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 25 <220>
 <223> Cebador oligonucleótido
 <400> 5
 tctcacctga taccagattc 20
 <210> 6
 30 <211> 20
 <212> ADN

<213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido
 <400> 6
 5 tctctcctta atgcctctat 20
 <210> 7
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 10 <220>
 <223> Cebador oligonucleótido
 <400> 7
 actgctgcgg gttcctaaag 20
 <210> 8
 15 <211> 21
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido
 20 <400> 8
 ggaagattcg attcaacaag g 21
 <210> 9
 <211> 19
 <212> ADN
 25 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido
 <400> 9
 ggtaactatg aatctcatc 19
 30 <210> 10
 <211> 20

<212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido
 5 <400> 10
 aaaaagcaga gaaagcaggg 20
 <210> 11
 <211> 20
 <212> ADN
 10 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido
 <400> 11
 agatgatctc tgagatgcc 20
 15 <210> 12
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> Cebador oligonucleótido
 <400> 12
 ccagggtttg tctaccaaag 20
 <210> 13
 <211> 19
 25 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido
 <400> 13
 30 aatcacttaa aacaagcag 19
 <210> 14

<211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 5 <223> Cebador oligonucleótido
 <400> 14
 cacatacctc tacctctagc 20
 <210> 15
 <211> 19
 10 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido
 <400> 15
 15 ttccctagtg gagcagtgg 19
 <210> 16
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> Cebador oligonucleótido
 <400> 16
 ctttcttcgc aaatgggtgg 20
 <210> 17
 25 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido
 30 <400> 17
 gcactcatcg ccacttaatg 20

<210> 18
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 5 <220>
 <223> Cebador oligonucleótido
 <400> 18
 gaacagctaa accagaacag 20
 <210> 19
 10 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido
 15 <400> 19
 atcactctta ttctccccc 20
 <210> 20
 <211> 20
 <212> ADN
 20 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido
 <400> 20
 tgagtcactg tgctaaggag 20
 25

Referencias

- Ahmed y colaboradores (2009) *Nature Genetics* 41:585-590.
- Antoniou y colaboradores (2009) *Hum Mol Genet* 18:4442-4456.
- Barringer y colaboradores (1990) *Gene* 89:117-122.
- 5 Beaucage y Caruthers (1981) *Tetrahedron Letts.* 22:1859-1862.
- Blok y Kramer (1997) *Mol Cell Probes* 11:187-194.
- Bonnet y colaboradores (1999) *PNAS* 96:6171-6176.
- Chee y colaboradores (1996) *Science* 274:610-614.
- Cheng y colaboradores (1994) *Nature* 369:684-685.
- 10 Chiebowski y colaboradores (2007) *J Natl Cancer Inst* 99:1695-1705.
- Cook y colaboradores (2006) *Ann Intern Med* 145:21-29.
- Costantino y colaboradores (1999) *J Natl Cancer Inst* 91:1541-1548.
- Devlin y Risch (1995) *Genomics*. 29: 311-322.
- Easton y colaboradores (2007) *Nature* 447: 1087-1093.
- 15 Fang y colaboradores (1999) *J. Am. Chem. Soc.* 121:2921-2922.
- Fodor (1997a) *FASEB Journal* 11:A879.
- Fodor (1997b) *Science* 277: 393-395.
- Gail y colaboradores (1989) *J Natl Cancer Inst* 81:1879-1886.
- Gail y colaboradores (1999) *J Natl Cancer Inst* 91:1829-1846.
- 20 Gail y colaboradores (2007) *J Natl Cancer Inst* 99(23):1782-1792.
- Gold y colaboradores (2008) *PNAS* 105: 4340-4345.
- Guatelli y colaboradores (1990) *PNAS* 87:1874-1878.
- Hsuih y colaboradores (1997) *J Clin Microbiol* 34:501-507.
- Kelemen y colaboradores (2009) *Cancer Epidemiol Biomarkers Prev* 18:1864-1868.
- 25 Kostrikis y colaboradores (1998) *Science* 279:1228-1229.
- Kwoh y colaboradores (1989) *PNAS* 86:1173-1177.
- Landegren y colaboradores (1988) *Science* 241:1077-1080.
- Leone y colaboradores (1995) *Nucleic Acids Res.* 26:2150-2155.
- Lockhart (1998) *Nature Medicine* 4:1235-1236.
- 30 Marras y colaboradores (1999) *Genet. Anal. Biomol. Eng.* 14:151-156.
- Needham Van Devanter y colaboradores (1984) *Nucleic Acids Res.* 12:6159-6168.

- Pencina y colaboradores (2008) Stat Med 27:157-172.
- Peto y Mako (2000) Nat Genet 26:411-414.
- Rockhill y colaboradores (2001) J Natl Cancer Inst 93(5):358-366.
- Sapolsky y colaboradores (1999) Genet Anal: Biomolec Engin 14:187-192.
- 5 Service (1998a) Science 282:396-399.
- Service (1998b) Science 282: 399-401.
- Slatkin y Excoffier (1996) Heredity 76: 377-383.
- Sokol y colaboradores (1998) PNAS 95:11538-11543.
- Sooknanan y Malek (1995) Biotechnology 13: 563-564.
- 10 Stacey y colaboradores (2007) Nature Genetics 39:965-869.
- Stacey y colaboradores (2008) Nature Genetics 40:703-706.
- Thomas y colaboradores (2009) Nature Genetics 41:579-584.
- Turnbull y colaboradores (2010) Nature Genetics 42: 504-509.
- Tyagi y Kramer (1996) Nature Biotechnology 14:303-308.
- 15 Tyagi y colaboradores (1998) Nature Biotechnology 16:49-53.
- Van Brunt (1990) Biotechnology 8:291-294.
- Vet y colaboradores (1999) PNAS 96:6394-6399.
- Wu y Wallace (1989) Gene 4:560-569.
- Zhang y colaboradores (1999) Anal. Chem. 71:1138-1145.
- 20

REIVINDICACIONES

1. Un método para evaluar el riesgo total de un sujeto femenino humano para desarrollar un fenotipo de cáncer de mama que comprende:

5 realizar una evaluación del riesgo clínico del sujeto femenino,

realizar una evaluación del riesgo genético del sujeto femenino, en donde la evaluación del riesgo genético implica detectar, en una muestra biológica derivada del sujeto femenino, la presencia de al menos cinco polimorfismos de un solo nucleótido que se sabe que están asociados con un fenotipo de cáncer de mama, seleccionado de entre el grupo que consiste de rs2981582, rs3803662, rs889312, rs123387042, rs13281615, rs4415084, rs3817198, rs4973768, rs6504950 y rs11249433; y

combinar la evaluación del riesgo clínico con la evaluación del riesgo genético para obtener el riesgo total de un sujeto femenino humano para desarrollar un fenotipo de cáncer de mama, en donde al menos cinco polimorfismos de un solo nucleótido comprenden al menos rs2981582, rs3803662, rs889312, rs13387042 y rs4415084.

2. Un método para determinar si el riesgo total de un sujeto femenino humano para desarrollar un fenotipo del cáncer de mama analizado mediante evaluación del riesgo clínico debe ser reclasificado, comprendiendo el método

combinar la evaluación del riesgo clínico con una evaluación del riesgo genético para obtener el riesgo total de un sujeto femenino humano de desarrollar un fenotipo de cáncer de mama, en donde la evaluación del riesgo genético implica detectar, en una muestra biológica derivada del sujeto femenino, la presencia de al menos cinco polimorfismos de un solo nucleótido que se sabe que están asociados con un fenotipo de cáncer de mama seleccionado de entre el grupo que consiste de rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198, rs4973768, rs6504950 y rs11249433, y

determinar si la combinación de la evaluación del riesgo clínico con la evaluación del riesgo genético da como resultado una reclasificación del riesgo total del sujeto de desarrollar un fenotipo de cáncer de mama,

en donde al menos cinco polimorfismos de un solo nucleótido comprenden al menos rs2981582, rs3803662, rs889312, rs13387042 y rs4415084.

3. El método de la reivindicación 1 o la reivindicación 2, en donde la combinación de la evaluación del riesgo clínico con la evaluación del riesgo genético comprende multiplicar las evaluaciones de riesgos.

4. El método de la reivindicación 2 o la reivindicación 3, en donde la mejora de la reclasificación neta (NRI) del método es mayor que 0,01, o es mayor que 0,05.

5. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 4, que tiene una o más de las siguientes características;

i) la realización de la evaluación del riesgo clínico utiliza un modelo seleccionado del grupo que consiste de el modelo de Gail, el Modelo de Claus, las Tablas de Claus, BOADICEA, el Modelo de Jonker, la Fórmula extendida de Claus, el Modelo de Tyrer-Cuzick, y el sistema de puntuación de Manchester,

ii) la realización de la evaluación del riesgo clínico incluye la obtención de información del sujeto femenino en uno o más de los siguiente aspectos: la historia clínica de cáncer de mama, el carcinoma ductal o carcinoma lobular, edad, edad del primer período de menstruación, la edad a la que dio a luz por primera vez, antecedentes familiares de cáncer de mama, resultados de las biopsias de mama anteriores y raza/etnia,

iii) que comprende la detección de la presencia de al menos seis, siete, ocho, nueve o diez polimorfismos de un solo nucleótido que se sabe están asociados con un fenotipo de cáncer de mama,

iv) los polimorfismos de un solo nucleótido se prueban individualmente para la asociación con el cáncer de mama mediante regresión logística bajo un modelo aditivo logarítmico sin covariables,

v) el sujeto femenino es caucásico,

vi) el sujeto femenino ha tenido una biopsia de mama,

vii) los resultados de la evaluación del riesgo clínico indican que el sujeto femenino debe ser sometido a una detección más frecuente y/o terapia profiláctica contra cáncer de mama, y

viii) si se determina que el sujeto tiene un riesgo de desarrollar cáncer de mama, es más probable que el sujeto sea sensible que no sensible a la inhibición de estrógenos.

5 6. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 5, que comprende

i) detectar al menos siete polimorfismos de un solo nucleótido, en donde los polimorfismos de un solo nucleótido comprenden al menos rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084 y rs3817198,

ii) detectar al menos siete polimorfismos de un solo nucleótido, en donde los polimorfismos de un solo nucleótido comprenden al menos rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084 y rs3817198,

10 iii) detectar, al menos, diez polimorfismos de un solo nucleótido, en donde los polimorfismos de un solo nucleótido, comprenden al menos rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198, rs4973768,

rs6504950 y rs11249433,

15 iv) detectar al menos diez polimorfismos de un solo nucleótido, en donde los polimorfismos de un solo nucleótido, comprenden al menos rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198, rs4973768, rs6504950 y rs11249433.

7. El método de la reivindicación 5 o la reivindicación 6, en donde la evaluación del riesgo clínico usa el modelo de Gail.

20 8. Un kit que comprende al menos cinco conjuntos de cebadores para amplificar cinco o más ácidos nucleicos, en donde los cinco o más ácidos nucleicos comprenden independientemente uno de al menos cinco polimorfismos de un solo nucleótido seleccionado de entre un grupo que consiste de rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198, rs4973768, rs6504950 y rs11249433, en donde al menos cinco polimorfismos de un solo nucleótido comprenden al menos rs2981582, rs3803662, rs889312, rs13387042 y rs4415084.

25 9. Una matriz genética que comprende al menos cinco ácidos nucleicos que comprenden independientemente uno de al menos cinco polimorfismo de un solo nucleótido seleccionados de entre el grupo que consiste de rs2981582, rs3803662, rs889312, rs13387042, rs13281615, rs4415084, rs3817198, rs4973768, rs6504950 y rs11249433, en donde al menos cinco de los polimorfismos de un solo nucleótido comprenden al menos rs2981582, rs3803662, rs889312, rs13387042 y rs4415084.

30 10. Un método para determinar la necesidad de pruebas de diagnóstico de rutina de un sujeto femenino humano de cáncer de mama que comprende la evaluación del riesgo global del sujeto de desarrollar un fenotipo de cáncer de mama utilizando el método de acuerdo con una cualquiera de las reivindicaciones 1 a 7.

35 11. Un método de detección de cáncer de mama en un sujeto femenino humano, comprendiendo el método la evaluación del riesgo total del sujeto de desarrollar un fenotipo de cáncer de mama utilizando el método de acuerdo con una cualquiera de las reivindicaciones 1 a 7, y de detectar rutinariamente cáncer de mama en el sujeto si se evalúan por tener un riesgo de desarrollar cáncer de mama.

12. Un método para determinar la necesidad de sujeto femenino humano por una terapia profiláctica contra cáncer de mama que comprende evaluar el riesgo total del sujeto de desarrollar un fenotipo de cáncer de mama utilizando el método de acuerdo con cualquiera de las reivindicaciones 1 a 7.

40 13. Una terapia contra el cáncer de mama para uso en la prevención del cáncer de mama en un sujeto femenino humano en riesgo del mismo, en donde el sujeto se evalúa por tener un riesgo de desarrollar cáncer de mama de acuerdo con el método de cualquiera de las reivindicaciones 1 a 7.

45 14. Un método para estratificar un grupo de sujetos femeninos humanos para un ensayo clínico de una terapia candidata, comprendiendo el método la evaluación del riesgo total individual de los sujetos de desarrollar un fenotipo de cáncer de mama utilizando el método de acuerdo con cualquiera de las reivindicaciones 1 a 7, y usar los resultados de la evaluación para seleccionar los sujetos que sean más probablemente sensibles a la terapia.

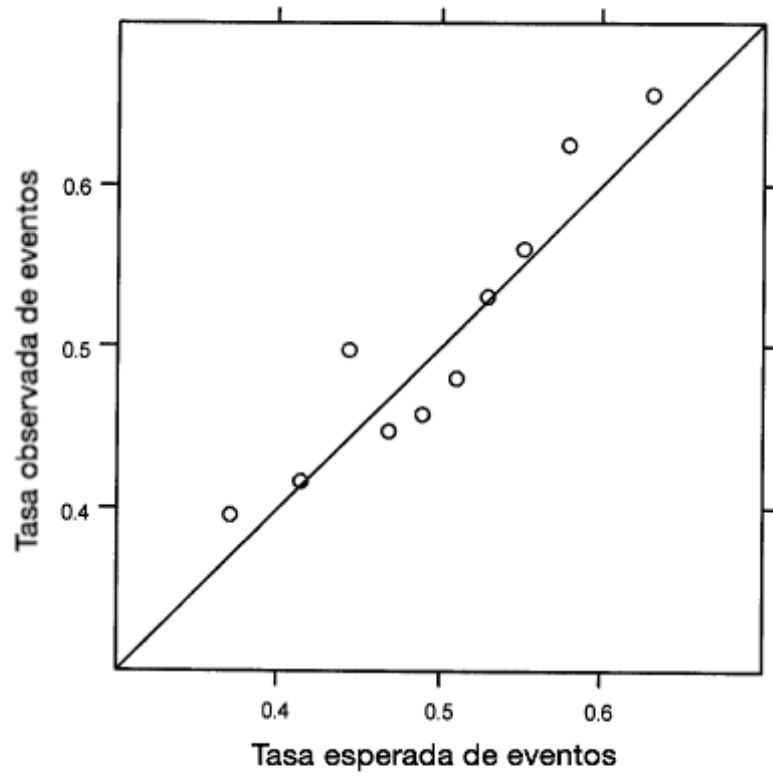


Figura 1

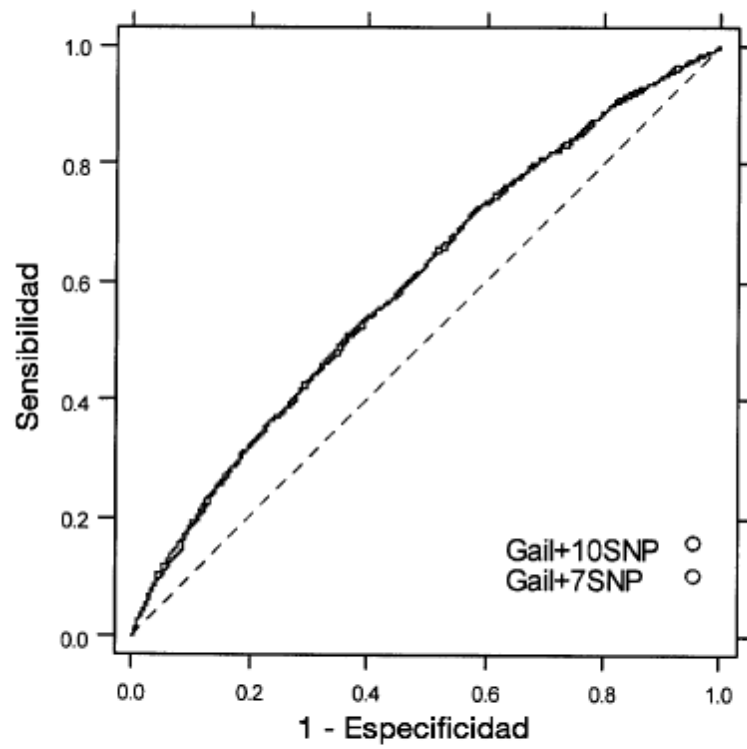


Figura 2