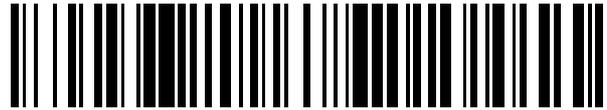


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 671 119**

51 Int. Cl.:

**C12Q 1/68** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **24.09.2008** **E 14170551 (7)**

97 Fecha y número de publicación de la concesión europea: **28.02.2018** **EP 2772553**

54 Título: **Métodos de análisis genéticos**

30 Prioridad:

**27.09.2007 US 995564 P**

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

**05.06.2018**

73 Titular/es:

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

72 Inventor/es:

**COX, DAVID R. y**  
**MCCAMISH, MARK A.**

74 Agente/Representante:

**CARVAJAL Y URQUIJO, Isabel**

**ES 2 671 119 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 de análisis genéticos

**Campo de la invención**

5 La presente invención se refiere a métodos para determinar el riesgo de un individuo de desarrollar o presentar una afección médica multifactorial. Los métodos implican determinar una puntuación de un individuo basándose en la información genotípica del individuo y comparar la puntuación con al menos un valor umbral, en donde el resultado de la comparación es indicativo del riesgo del individuo de desarrollar o presentar la afección médica multifactorial.

**Antecedentes de la invención**

10 El ADN que forma parte de los cromosomas humanos proporciona las instrucciones que dirigen la producción de todas las proteínas del cuerpo. Estas proteínas llevan a cabo las funciones vitales de la vida. Las variaciones en la secuencia de ADN que codifica una proteína, producen variaciones o mutaciones en las proteínas codificadas, lo que afecta a la función normal de las células. Aunque el medio ambiente a menudo juega un papel importante en las enfermedades, las variaciones y mutaciones en el ADN de un individuo están directamente relacionadas con casi todas las enfermedades humanas, incluidas las enfermedades cardiovasculares, metabólicas e infecciosas, el  
15 cáncer y los trastornos inmunitarios. Además, el conocimiento de la genética, particularmente de la genética humana, ha llegado a la conclusión de que muchas enfermedades son el resultado de interacciones complejas de varios genes o sus productos. Por ejemplo, la diabetes de los tipos I y II se ha relacionado con genes múltiples, cada uno de ellos con su propio patrón de mutaciones.

20 Por otro lado, el conocimiento de la genética humana ha llevado a una comprensión limitada de las variaciones entre los individuos cuando se trata de una respuesta a los fármacos en el campo de la farmacogenética. Hace más de medio siglo, las respuestas adversas a los fármacos se correlacionaron con las variaciones de aminoácidos en dos enzimas metabolizadoras de fármacos, la colinesterasa plasmática y la glucosa-6 fosfato deshidrogenasa. Desde entonces, análisis genéticos cuidadosos han vinculado los polimorfismos de secuencia (variaciones) en más de 35  
25 enzimas del metabolismo de fármacos, 25 dianas farmacológicas y 5 transportadores de fármacos con niveles comprometidos de eficacia o seguridad del fármaco (Evans y Relling, Science 296:487-91 (1999)). En la clínica, dicha información, se utiliza para prevenir la toxicidad de los fármacos; por ejemplo, los pacientes se examinan habitualmente por diferencias genéticas en el gen de la tiopurina metiltransferasa que causa una disminución del metabolismo de 6-mercaptopurina o azatioprina. Sin embargo, solo un pequeño porcentaje de las toxicidades observadas de los fármacos, se ha explicado adecuadamente mediante el conjunto de marcadores  
30 farmacogenéticos validados hasta la fecha. Incluso más habitual que los problemas de toxicidad pueden ser los casos en los que se ha demostrado que los fármacos son seguros y/o eficaces para algunos individuos que tienen una eficacia terapéutica insuficiente o efectos secundarios imprevistos en otros individuos.

Debido a que dos seres humanos son 99,9 % similares en cuanto a su composición genética, la mayor parte de la secuencia del ADN de sus genomas es idéntica. Sin embargo, hay variaciones en la secuencia de ADN entre los  
35 individuos. Por ejemplo, hay deleciones de muchos tramos de bases de ADN, inserción de tramos de ADN, variaciones en el número de elementos de ADN repetitivos en regiones codificantes o no codificantes, y cambios en posiciones de bases nitrogenadas individuales en el genoma denominados "polimorfismos mononucleotídicos" (SNP, del inglés *single nucleotide polymorphisms*). La variación en la secuencia del ADN humano representa una gran fracción de diferencias observadas entre los individuos, incluida la susceptibilidad o resistencia a la enfermedad y  
40 cómo responderá un individuo a un régimen terapéutico o de tratamiento en particular.

Los rasgos multifactoriales, o rasgos complejos, están influenciados por múltiples factores, tales como los genes, los factores ambientales y sus interacciones. A menudo, más de una combinación de factores genéticos y/o ambientales dará como resultado el mismo rasgo multifactorial, y esta complejidad dificulta determinar quién desarrollará dicho rasgo. Además, la contribución de cada factor generalmente no es idéntica a las contribuciones de cualquier otro  
45 factor. Es decir, por ejemplo, algunos factores pueden tener una contribución muy fuerte, mientras que otros pueden tener una contribución muy débil. Para complicar aún más la base biológica de los rasgos multifactoriales, las contribuciones de un factor pueden ser aditivas, sinérgicas o completamente independientes de la contribución de cualquier otro factor. Algunos rasgos complejos manifiestan enfermedades comunes, tales como enfermedades cardiovasculares, diabetes, obesidad y colesterol alto. Otros rasgos complejos incluyen fenotipos tales como la  
50 forma en la que un individuo responde a un fármaco o a otro régimen de tratamiento médico.

Hasta hace poco, la investigación sobre la base genética de la enfermedad ha dado como resultado el desarrollo de algunas pruebas genéticas para enfermedades. Sin embargo, estas pruebas genéticas no serán útiles para predecir la probabilidad de una persona sana de desarrollar una enfermedad multifactorial común. Muchos argumentan que las pruebas genéticas para los rasgos multifactoriales comunes (por ejemplo, enfermedades) no serán útiles en la  
55 práctica debido a la penetración incompleta y a la baja contribución individual de cada gen involucrado (Holtzman

and Marteau, 2000; Vineis et al. 2001). Sin embargo, estos argumentos se basan en gran parte en el uso de locus únicos para predecir si un individuo presentará o no el rasgo (Beaudet 1999; Evans et al. 2001). Lo que se necesita es un enfoque confiable para determinar el riesgo de un individuo de desarrollar o presentar un rasgo multifactorial que se base en el genotipo del individuo en una pluralidad de locus, cada uno de los cuales son factores en la manifestación del rasgo multifactorial. Además, en la técnica también se necesitan métodos para tratar individuos que presentan una afección médica relacionada con el genotipo de un individuo en una pluralidad de locus.

El documento US 2005/196770 A1 (Cox David et al, 2005-09-08) desvela métodos para determinar el riesgo de un individuo de desarrollar o presentar un rasgo multifactorial determinando una puntuación para el individuo basándose en el genotipo del individuo en una pluralidad de locus polimórficos bialélicos.

El documento US2003/162207 A1 (Comings David et al, 2003-08-28) desvela un método para determinar si un grupo de genes contribuye conjuntamente a un rasgo poligenético.

### Resumen

La presente invención proporciona un método para determinar el riesgo de un individuo a desarrollar o presentar una afección médica multifactorial, que comprende:

a) determinar una puntuación para el individuo basándose en información genotípica obtenida de una muestra del mismo, en el que la información genotípica es la presencia o ausencia de alelos asociados a la afección médica en una pluralidad de locus polimórficos, en el que a un alelo asociado se le asigna un valor diferente para determinar la puntuación de otros alelos asociados, y en el que la magnitud del efecto sobre el riesgo de los alelos se utiliza para determinar la puntuación, y en el que la determinación de la puntuación para el individuo se basa además en la información no genética; y

b) comparar la puntuación con al menos un valor umbral, en el que el resultado de dicha comparación es indicativo del riesgo del individuo a desarrollar o presentar la afección médica multifactorial.

La presente solicitud desvela métodos para determinar el riesgo de un individuo a desarrollar o presentar un rasgo multifactorial determinando una puntuación para el individuo basándose al menos en el genotipo del individuo en una pluralidad de locus polimórficos bialélicos, y comparando esa puntuación con al menos un valor umbral. En determinadas realizaciones, para cada uno de los locus polimórficos, el genotipo del individuo puede ser homocigoto para un alelo asociado, homocigoto para un alelo no asociado o heterocigoto. En alguna de las realizaciones, además de los genotipos, también se puede utilizar información no genotípica tal como información clínica, demográfica, epidemiológica y/u otros factores no genéticos para determinar la puntuación y opcionalmente el valor o valores umbral. Si la puntuación del individuo es mayor que un valor umbral, entonces se puede considerar que el individuo está en riesgo de desarrollar o presentar el rasgo multifactorial, y si la puntuación del individuo es igual o menor a un valor umbral, entonces se puede considerar que el individuo no está en riesgo de desarrollar o presentar el rasgo multifactorial. Si la puntuación del individuo es mayor que un valor umbral, pero menor o igual que otro valor umbral, entonces se puede considerar que el individuo tiene un riesgo intermedio de desarrollar o presentar el rasgo multifactorial.

La presente solicitud desvela además métodos para identificar alelos de locus polimórficos bialélicos que están asociados a un rasgo multifactorial, denominados en el presente documento "alelos asociados". Los métodos implican la realización de un estudio de asociación en el que se compara la composición genética de un grupo de individuos que presentan el rasgo multifactorial (grupo de casos) con la composición genética de un grupo de individuos que no presentan el rasgo multifactorial ("grupo de control"), y se identifican con alelos asociados aquellos alelos que son significativamente más prevalentes en la composición genética del grupo de casos que la composición genética del grupo de control. En determinadas realizaciones, los alelos asociados identificados en un primer estudio de asociación con un primer grupo de casos y un primer grupo de control se verifican realizándose un estudio de asociación con un segundo grupo de casos y un segundo grupo de control.

La presente solicitud también desvela métodos para determinar un valor umbral para su uso en una prueba de evaluación de riesgos. En un aspecto, un umbral se determina analizando una serie de valores límite de riesgo que se basan en un conjunto de puntuaciones de un grupo de casos y en un conjunto de puntuaciones de un grupo de control. La determinación de un valor umbral puede implicar el uso de información que incluye, pero sin limitación, sensibilidad, especificidad, VPP, VPN, precisión, CV+ y CV- para una prueba de evaluación de riesgos utilizando cada valor de límite de riesgo como un valor umbral; teniendo en cuenta datos clínicos, demográficos o epidemiológicos del rasgo multifactorial, las posibles opciones del tratamiento y el individuo que se está evaluando; y la participación de al menos una agencia reguladora.

La presente divulgación también proporciona métodos para tratar a un individuo que presenta una afección médica que implica determinar una puntuación para el individuo basándose al menos en información genotípica del

- individuo, comparar la puntuación con al menos un valor umbral y proporcionar un tratamiento si la comparación es indicativa de que el individuo experimentará una respuesta beneficiosa al tratamiento. La puntuación y/o valor o valores umbral se basan en información genotípica, opcionalmente en combinación con información no genotípica, tal como información clínica, epidemiológica, demográfica, ambiental y de otro tipo. La información genotípica puede proceder de estudios de asociación que determinan locus genéticos asociados con la predisposición a la afección médica o respuesta al tratamiento. El tratamiento puede incluir cambios en el estilo de vida, administración de un fármaco, uso de un dispositivo médico, psicoterapia, fisioterapia, inclusión de un ensayo clínico, exclusión de un ensayo clínico y/o intervención quirúrgica.
- 5
- En diversos aspectos, la información genotípica incluye genotipos para el individuo en una pluralidad de locus polimórficos bialélicos, en el que cada uno de la pluralidad tiene un alelo asociado y un alelo no asociado y además en el que cada uno de los genotipos se selecciona del grupo que incluye homocigotos para el alelo asociado, heterocigotos y homocigotos para el alelo no asociado. En varios aspectos, los locus polimórficos son SNP.
- 10
- En diversos aspectos, el método implica además identificar los alelos asociados y los no asociados para una pluralidad de locus polimórficos bialélicos realizando un estudio de asociación con un grupo de casos de individuos y un grupo de control de individuos, determinando así un conjunto de alelos de los locus polimórficos que son significativamente más abundantes en el grupo de casos que el grupo de control, en el que el conjunto de alelos o un subconjunto de los mismos son los alelos asociados.
- 15
- En diversos aspectos, los individuos en el grupo de control presentan una respuesta eficaz al tratamiento y los individuos en el grupo de casos no presentan la respuesta eficaz. En otros aspectos, los individuos en el grupo de casos presentan un acontecimiento adverso en respuesta al tratamiento y los individuos en el grupo de control no presentan el acontecimiento adverso. En otros aspectos, los individuos en el grupo de casos presentan una respuesta eficaz al tratamiento y los individuos en el grupo de control no presentan una respuesta eficaz.
- 20
- En diversos aspectos, los individuos en un grupo de casos tienen la afección médica y los individuos en el grupo de control no tienen la afección médica.
- 25
- En diversos aspectos, al menos uno del grupo de casos y el grupo de control comprende al menos 200 individuos o al menos 1000 individuos.
- En diversos aspectos, los grupos de casos y controles se emparejan antes de realizar el estudio de asociación.
- En diversos aspectos, realizar un estudio de asociación implica además a) genotipificar el grupo de casos y el grupo de control como un conjunto de locus polimórficos que incluye la pluralidad de locus polimórficos bialélicos, b) calcular una frecuencia alélica relativa para cada uno del conjunto de locus polimórficos para cada uno de los grupos de casos y de control, c) para cada uno del conjunto de locus polimórficos, comparando la frecuencia alélica relativa calculada para el grupo de casos con la frecuencia alélica relativa calculada para el grupo de control, identificando así un subconjunto del conjunto de locus polimórficos, en el que cada uno de los subconjuntos tiene una frecuencia alélica relativa que es significativamente diferente para el grupo de casos que para el grupo de control y d) la determinación de un alelo para cada uno de los subconjuntos que es más abundante en el grupo de casos que el grupo de control, en el que el alelo es uno de los alelos asociados.
- 30
- 35
- En diversos aspectos, el conjunto de locus polimórficos incluye al menos aproximadamente 500 locus polimórficos. En otros aspectos, el conjunto de locus polimórficos incluye locus polimórficos de cada cromosoma en el genoma del individuo.
- 40
- En diversos aspectos, los métodos de la invención implican, además, la validación de alelos asociados realizando un segundo estudio de asociación con un segundo grupo de casos y un segundo grupo de control, determinando de ese modo cuáles de los alelos asociados son significativamente más abundantes en el segundo grupo de casos que en el segundo grupo de control, en el que los alelos asociados que son significativamente más abundantes en el segundo grupo de casos que en el segundo grupo de control son los alelos asociados validados.
- 45
- En diversos aspectos, los métodos de la invención implican además determinar uno de dicho al menos un valor umbral mediante un método que implica a) calcular una puntuación para cada miembro del grupo de casos y el grupo de control, b) seleccionar una serie de valores de límite de riesgo, c) computar un conjunto de valores para cada una de las series de valores de límite de riesgo, en el que el conjunto de valores comprende al menos uno de una sensibilidad, una especificidad, un VPP, un VPN, una precisión, un riesgo relativo, una CV+, una CV-, e información clínica y d) elegir uno de los valores de límite de riesgo como uno de al menos un valor umbral basándose en el conjunto de valores, determinando por lo tanto uno del al menos un valor umbral.
- 50
- En diversos aspectos, el cálculo de una puntuación para cada miembro del grupo de casos y el grupo de control implica a) determinar un genotipo para cada miembro en la pluralidad de locus polimórficos bialélicos, en el que el

genotipo se selecciona del grupo que consiste en homocigotos para un alelo asociado, heterocigotos y homocigotos para un alelo no asociado, b) asignar un primer valor a cada uno de los locus polimórficos que tiene un genotipo que es homocigoto para un alelo que no es el alelo asociado, c) asignar un segundo valor a cada uno de los locus polimórficos que tiene un genotipo que es heterocigoto, d) asignar un tercer valor a cada uno de los locus polimórficos que tiene un genotipo que es homocigoto para el alelo asociado y e) sumar los valores determinados en las etapas a) a c) para todos los locus polimórficos, con lo que se calcula una puntuación para cada miembro del grupo de casos y el grupo de control.

En diversos aspectos, seleccionar una serie de valores límite de riesgo implica identificar una puntuación más alta de las puntuaciones calculadas para cada un miembro del grupo de casos y el grupo de control, determinando un intervalo de límite de riesgo, en el que el intervalo varía de 1 a la puntuación más alta y seleccionando una serie de valores de todo el intervalo de límite de riesgo, seleccionando así una serie de valores de límite de riesgo.

En diversos aspectos, seleccionar la serie de valores del intervalo de límite de riesgo implica un método seleccionado del grupo que incluye seleccionar cada valor dentro del intervalo de límite de riesgo, seleccionando cada enésimo valor dentro del intervalo de límite de riesgo, dividiendo el intervalo de límite de riesgo en porcentajes y seleccionando un valor en cada enésimo porcentaje del intervalo de límite de riesgo, seleccionando un mayor número de valores de una parte central del intervalo de límite de riesgo que de una parte superior o inferior del intervalo de límite de riesgo y seleccionando un mayor número de valores de una parte superior o inferior del intervalo de límite de riesgo que de una parte central del intervalo de límite de riesgo.

En diversos aspectos, la determinación del uno del al menos un valor umbral implica además el uso de una curva ROC (por las siglas del inglés *receiver operating characteristic*, característica operativa del receptor) basada en dicha sensibilidad y dicha especificidad computada según la invención, en el que una representación gráfica de la curva ROC recibe el nombre de gráfico.

En diversos aspectos, la elección como uno de al menos un valor umbral un valor de límite de riesgo correspondiente a un punto de datos en la curva ROC que está más cerca de una esquina superior izquierda del gráfico que cualquier otro punto de datos en la curva ROC, en el que cada punto de datos en la curva ROC corresponde a un valor de límite de riesgo diferente.

En diversos aspectos, los métodos implican además a) determinar una posición en la curva ROC más cercana a la esquina superior izquierda del gráfico y determinar una sensibilidad y una especificidad que corresponde a la posición y b) analizar las puntuaciones de cada miembro del grupo de casos y del grupo control para identificar un valor de límite de riesgo cuya sensibilidad y especificidad están más próximas a la sensibilidad y especificidad que corresponden a la posición, en el que el valor de límite de riesgo cuya sensibilidad y especificidad están más próximas a la sensibilidad y especificidad que corresponden a la posición es uno de al menos un valor umbral.

En diversos aspectos, donde para un valor de límite de riesgo dado, el riesgo relativo se computa mediante un método que implica a) determinar un porcentaje de los miembros del grupo de casos que tiene una puntuación que es al menos tan grande como el valor de límite de riesgo dado, b) determinar un porcentaje de los miembros del grupo de control que tiene una puntuación que es al menos tan grande como el valor de límite de riesgo dado y c) dividir el porcentaje determinado en a) entre el porcentaje determinado en b) para computar el riesgo relativo.

En diversos aspectos, la puntuación se basa en información genotípica.

En diversos aspectos de la invención, la determinación de una puntuación para el individuo implica adicionalmente a) determinar un genotipo para el individuo en la pluralidad de locus polimórficos bialélicos, en el que el genotipo se selecciona del grupo que consiste en homocigoto para un alelo asociado, heterocigoto y homocigoto para un alelo no asociado, b) asignar un primer valor a cada uno de los locus polimórficos que tiene un genotipo que es homocigoto para un alelo que no es el alelo asociado, c) asignar un segundo valor a cada uno de los locus polimórficos que tiene un genotipo que es heterocigoto, d) asignar un tercer valor a cada uno de los locus polimórficos que tiene un genotipo que es homocigoto para el alelo asociado y e) sumar los valores predeterminados en las etapas a) a c) para todos los locus polimórficos, determinando así una puntuación para el individuo.

La presente divulgación desvela además un ensayo de diagnóstico o pronóstico que comprende sondas de ácido nucleico diseñadas para detectar los alelos asociados en una muestra biológica. En determinadas realizaciones, las sondas del ensayo de diagnóstico o pronóstico están unidas a un sustrato sólido.

## 50 Breve descripción de las figuras

La figura 1 ilustra un ejemplo de una curva característica operativa del receptor que establece un valor umbral para una prueba de evaluación de riesgos.

**Descripción detallada**Generalidades

5 Determinadas realizaciones de la presente invención proporcionan métodos para determinar con un alto grado de certeza la predisposición de un individuo a desarrollar o presentar un rasgo multifactorial, que puede ser, por ejemplo, el desarrollo de una enfermedad u otro trastorno, o una respuesta positiva o negativa a un fármaco. Esta determinación se basa al menos en el genotipo del individuo en una pluralidad de locus genéticos, cada uno de los cuales es un factor genético involucrado en la manifestación del rasgo multifactorial. Los métodos proporcionan además el beneficio de hacer tal determinación sin saber el grado al cual, o la forma en la que, cada factor genético influye en la manifestación del rasgo multifactorial. Los métodos de la invención en cambio se basan en los efectos acumulativos de múltiples factores genéticos y permiten que un experto realice una predicción precisa de la probabilidad de que un individuo desarrolle o presente el rasgo multifactorial basado en el genotipo del individuo en una pluralidad de locus genéticos que se ha determinado que están asociados a la incidencia del rasgo multifactorial. La predisposición del individuo al rasgo multifactorial se determina utilizando factores genéticos en combinación con otra información, tal como datos clínicos, demográficos y/o epidemiológicos.

15 Los rasgos multifactoriales están influenciados por una pluralidad de factores genéticos, factores ambientales e interacciones entre ellos. Además, la contribución de cada factor generalmente no es idéntica a las contribuciones de cualquier otro factor. Es decir, por ejemplo, algunos factores pueden tener una fuerte contribución, mientras que otros pueden tener una contribución débil. Para complicar aún más la base biológica de los rasgos multifactoriales, las contribuciones de un factor pueden ser aditivas, sinérgicas o completamente independientes de la contribución de cualquier otro factor. En determinadas realizaciones, los métodos presentados en este documento no se basan en las magnitudes del efecto que tiene cada factor en el rasgo multifactorial, ni dependen de si los efectos de los factores son aditivos, sinérgicos o independientes. En dichas realizaciones, los métodos no requieren que se tenga en cuenta la magnitud del efecto de cada factor al calcular el "riesgo" de un individuo (por ejemplo, probabilidad, verosimilitud) de desarrollar dicho rasgo multifactorial. En determinadas realizaciones, los métodos no requieren saber el perfil de expresión de ningún gen, ARN o proteína que comprende, y está genéticamente relacionado con, o son productos de, un factor genético. En determinadas realizaciones, los métodos no requieren saber cuáles son factores ambientales que pueden influir en el rasgo multifactorial. En determinadas realizaciones, los métodos presentados en este documento se basan en un conjunto de suposiciones de que la contribución individual de cada factor genético es igual a la contribución de cualquier otro factor genético, de que las contribuciones individuales son simplemente aditivas en todos los factores genéticos subyacentes al rasgo multifactorial y de que el riesgo de un individuo puede evaluarse si no se tiene conocimiento de la contribución de los factores ambientales a la manifestación del rasgo multifactorial. En otras realizaciones, los métodos tienen en cuenta características adicionales de los factores, tales como, por ejemplo, las magnitudes del efecto que cada factor tiene sobre el rasgo multifactorial y/o si los factores son aditivos, sinérgicos, antagónicos, independientes o no, o de otra manera epistáticos, cuando se calcula el riesgo de un individuo de desarrollar un rasgo multifactorial. En algunas realizaciones, también pueden considerarse datos de expresión. En realizaciones adicionales, el conocimiento de los factores ambientales que pueden influir en el rasgo multifactorial, se tiene en cuenta al calcular el riesgo de un individuo de desarrollar un rasgo multifactorial. En otras realizaciones adicionales, también se tienen en cuenta factores no genotípicos adicionales, tales como información clínica, demográfica y/o epidemiológica, al calcular el riesgo de un individuo de desarrollar un rasgo multifactorial.

45 Determinadas realizaciones de la presente invención proporcionan métodos para realizar un estudio de asociación para identificar un conjunto de locus polimórficos asociado a un rasgo multifactorial. También se proporcionan métodos para determinar cuál del conjunto de los locus polimórficos asociado se debe incluir en una prueba de evaluación de riesgos (por ejemplo, una prueba genética o poligenética) para el rasgo multifactorial, así como medios para determinar ciertas características de dicha prueba, por ejemplo, sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, riesgo relativo, cociente de verosimilitud, precisión etc. Además, se proporcionan métodos para su uso en un conjunto de locus polimórficos asociados en una prueba de evaluación de riesgos para determinar la predisposición de un individuo a desarrollar o presentar ese rasgo multifactorial. En una realización, el rasgo multifactorial es una enfermedad y los individuos identificados como propensos a desarrollar esa enfermedad pueden someterse a tratamientos o a otras intervenciones médicas para tratar o prevenir el desarrollo de la enfermedad. En otra realización, los métodos de la presente invención se utilizan para predecir la eficacia de un tratamiento médico propuesto, en el que si es poco probable que el tratamiento sea eficaz, entonces no se administra a un paciente. En otra realización, el rasgo multifactorial es la exposición de un acontecimiento adverso en respuesta a un fármaco u otro tratamiento médico. Los individuos identificados como propensos a presentar el acontecimiento adverso pueden excluirse del régimen de tratamiento o, si se tratan (por ejemplo, como último recurso), se puede utilizar un control adicional o el método de administración (por ejemplo, dosificación, frecuencia, etc.) puede ajustarse anticipándose a, o para, reducir la incidencia del acontecimiento adverso. En otras realizaciones más, los métodos desvelados en la presente invención se usan para el desarrollo de tratamientos médicos (por ejemplo, fármacos, dispositivos médicos, cirugía, etc.) y específicamente para aumentar la eficacia y la seguridad de los tratamientos seleccionando pacientes apropiados para su inclusión en los estudios, o para desarrollar un diagnóstico que se utilizará en la clínica para identificar a los pacientes con mayor probabilidad de

beneficiarse de la administración del tratamiento.

Como será muy obvio para los expertos en la técnica, los métodos de la presente invención se deben utilizar como herramientas de ayuda en la identificación de individuos que tienen, o que están en riesgo de desarrollar un rasgo multifactorial de interés, y que los métodos aquí presentados se pueden utilizar junto con información clínica sobre el rasgo, el individuo o individuos que se están evaluando y la población de la que se selecciona el individuo o individuos, así como otras pruebas clínicas e incluso la "intuición" clínica del médico tratante. Las pruebas de evaluación de riesgos generalmente se utilizan para ayudar a los médicos, no para reglamentar la toma de decisiones clínicas. Básicamente, es el médico quien debe determinar cómo utilizar una prueba de diagnóstico o pronóstico utilizando, por ejemplo, el conocimiento clínico del rasgo (por ejemplo, enfermedad) y las posibles opciones de tratamiento, las características de la prueba de diagnóstico, la población con la que se desarrolló la prueba y el paciente específico que se está evaluando, equilibrando los riesgos con los individuos identificados incorrectamente mediante la prueba y los beneficios los con individuos identificados correctamente. En otro aspecto, un médico también puede considerar los riesgos para individuos identificados incorrectamente como "positivos" por la prueba en comparación con los riesgos para individuos identificados incorrectamente como "negativos" por la prueba (por ejemplo, ¿negar un tratamiento a un paciente que necesite dicho tratamiento causa más daño que administrar el tratamiento a un paciente que no lo necesite?).

A continuación se hará referencia en detalle a diversas realizaciones y aplicaciones particulares de la invención. Aunque la invención se describirá junto con las diversas realizaciones y aplicaciones, se entenderá que dichas realizaciones y aplicaciones no pretenden limitar la invención.

## 20 Estudios de asociación

En una realización de la presente invención, se identifica un conjunto de locus polimórficos asociado a la manifestación de un rasgo multifactorial y los alelos asociados que corresponden a esos locus polimórficos, realizando un estudio de asociación, y los alelos asociados se utilizan adicionalmente para determinar si un individuo que no es un miembro de los grupos de casos o control está genéticamente predispuesto a desarrollar o presentar el rasgo multifactorial. Un rasgo multifactorial puede ser cualquier tipo de rasgo fenotípico, tal como presentación de, susceptibilidad a, o resistencia a, una enfermedad u otro trastorno médico, una respuesta a un fármaco u otro régimen de tratamiento médico, u otra característica física o mental. Por ejemplo, en una realización, el rasgo multifactorial es una enfermedad y un estudio de asociación compara la composición genética de un grupo de individuos que presenta la enfermedad (casos) con la composición genética de un grupo de individuos que no presenta la enfermedad (controles). Los ejemplos de enfermedades que son multifactoriales incluyen, pero sin limitación, asma y otras enfermedades pulmonares, soriasis, artritis, dislexia, infertilidad, gota, cataratas, obesidad, diabetes, trastornos neurodegenerativos, (por ejemplo, enfermedad de Alzheimer, enfermedad de Parkinson, etc.), trastornos gastrointestinales (por ejemplo, enfermedad de Crohn), cáncer, enfermedad cardiovascular, ictus, hipertensión, síndrome metabólico y trastornos relacionados, trastorno por déficit de atención, esquizofrenia, depresión maniaca, osteoporosis, trastornos del sistema inmunitario, esclerosis múltiple, aterosclerosis y epilepsia. Determinadas anomalías del desarrollo también se incluyen en esta categoría, tal como el labio/paladar leporino, defectos cardíacos congénitos y defectos del tubo neural. En otra realización, el rasgo multifactorial es una respuesta a un tratamiento médico (por ejemplo, un fármaco, un dispositivo médico) y un estudio de asociación compara la composición genética de un grupo de individuos que muestra una respuesta particular al tratamiento (casos) con la composición genética de un grupo de individuos que no muestra la respuesta particular al tratamiento (controles). En un aspecto, la respuesta puede estar relacionada con la eficacia del tratamiento. Por ejemplo, el tratamiento puede ser muy eficaz para individuos en el grupo de casos y tener poca eficacia para individuos en el grupo de control o viceversa. En otro aspecto, la respuesta puede estar relacionada con un acontecimiento adverso en respuesta a la administración del tratamiento. Por ejemplo, los individuos en el grupo de casos pueden presentar un acontecimiento adverso en respuesta al tratamiento y los individuos en el grupo de control pueden no presentar el acontecimiento adverso. Aunque en el presente documento se proporcionan diversos ejemplos que describen usos de los métodos de la presente invención en combinación con rasgos multifactoriales específicos, estos ejemplos no pretenden limitar el alcance de la invención, que abarca el uso de los métodos presentados en este documento junto con cualquier rasgo multifactorial cuya manifestación implica una pluralidad de locus genéticos.

Típicamente, al menos 50, y preferentemente al menos 100 individuos, están tanto en los grupos de casos como en los de control. En algunos estudios, hay al menos 200, o al menos 500, o al menos 1000 individuos en al menos uno de los grupos de casos y controles. A menudo, hay más individuos en el grupo de control que en el grupo de casos. En determinadas realizaciones, los individuos en los grupos de casos y de control son mamíferos, pero los grupos de casos y de control también pueden comprender individuos no mamíferos tales como, por ejemplo, bacterias, hongos, protistas, virus, arqueas y otros eucariotas tales como reptiles, anfibios, peces, aves, crustáceos, insectos y plantas. En algunas realizaciones, los individuos en los grupos de casos y control son seres humanos.

Típicamente, la composición de los grupos de casos y control debe ser similar con respecto a las características, aparte del rasgo multifactorial a tener en cuenta. Por ejemplo, en una realización, en cada grupo se seleccionarán números similares de hombres y mujeres de edades similares. En determinadas realizaciones, un factor de riesgo

ambiental puede influir en la composición de los grupos de casos y control. Por ejemplo, para un estudio sobre el cáncer de pulmón, solo pueden seleccionarse fumadores (o no fumadores) para que comprendan los grupos de casos y control. En algunas realizaciones de la presente invención, la pertenencia al grupo de casos y al grupo de control se ajusta de modo que las estructuras poblacionales de los dos grupos se "correlacionen" antes de realizar un estudio de asociación. La estructura poblacional (o "estratificación poblacional") se refiere a la heterogeneidad de la composición genética de los individuos dentro de una población. Por ejemplo, la estructura poblacional de un grupo de casos compuesto principalmente por italianos, es diferente de la de un grupo de control que está compuesto principalmente por mexicanos, debido a los diferentes orígenes étnicos entre los dos grupos. Si se realiza un estudio de asociación sin correlacionar los grupos, entonces los locus genéticos que están asociados con un ancestro italiano, pero no con un ancestro mexicano, pueden aparecer erróneamente asociados al rasgo multifactorial en estudio. Al correlacionar la estructura poblacional de los grupos de casos y control, un experto puede controlar las diferencias genéticas entre ambos grupos, que no están relacionadas con el rasgo multifactorial de interés. Por lo tanto, las diferencias genéticas entre los grupos que se identifican mediante el estudio de asociación posterior, tienen más probabilidad de ser locus que están causalmente relacionados con el rasgo multifactorial de interés. Los métodos para correlacionar grupos de casos y control antes de realizar un estudio de asociación se describen con detalle en la patente de Estados Unidos n.º 7.124.033, presentada el 17 de octubre de 2006, titulada "Method for Identifying Matched Groups"; y en la solicitud de patente provisional de Estados Unidos n.º 60/497,771, presentada el 26 de agosto de 2003 titulada "Matching Strategies for Genetic Association Studies in Structured Populations".

Para su uso en ensayos de genotipificación se recogen muestras de ácido nucleico de individuos de los grupos de casos y control. Las muestras de ácido nucleico pueden ser muestras de ADN o ARN y pueden obtenerse de diversas muestras biológicas, tales como, por ejemplo, sangre entera, semen, saliva, lágrimas, materia fecal, orina, sudor, muestras bucales, dérmicas y pilosas. En determinados aspectos, las muestras de ácido nucleico comprenden ADN genómico. Las muestras de ácido nucleico pueden prepararse para análisis utilizando cualquier técnica conocida por los expertos en la materia. Preferentemente, dichas técnicas dan como resultado la producción de una molécula de ácido nucleico suficientemente pura para determinar la presencia o la ausencia de uno o más polimorfismos en una o más posiciones en las moléculas de ácido nucleico. Dichas técnicas son comúnmente conocidas y se pueden encontrar, por ejemplo, en Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Nueva York) (2001), y en Ausubel, et al., *Current Protocols in Molecular Biology* (John Wiley and Sons, Nueva York).

Antes de determinar la presencia o la ausencia de uno o más polimorfismos en el ácido nucleico, uno o más ácidos nucleicos de interés pueden amplificarse y/o marcarse. Cualquier técnica de amplificación conocida por los expertos en la materia puede utilizarse junto con determinados métodos de la presente invención, que incluyen, pero sin limitación, técnicas de reacción en cadena de la polimerasa (PCR). La PCR se puede llevar a cabo utilizando materiales y métodos conocidos por los expertos en la materia. Véase, en general, la tecnología de PCR: *Principals and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, et al., Academic Press, San Diego, CA, 1990); Matilla et al., *Nucleic Acids Res.* 19: 4967 (1991); Eckert et al., *PCR Methods and Applications 1*: 17 (1991); PCR (eds. McPherson et al., IRL Press, Oxford); y la patente de Estados Unidos n.º 4.683.202. Otros métodos de amplificación adecuados incluyen la reacción en cadena de la ligasa (LCR) (véase Wu y Wallace, *Genomics* 4: 560 (1989) y Landegren et al., *Science* 241: 1077 (1988)), amplificación de la transcripción (Kwoh et al., *Proc. Natl. Acad. Sci. Estados Unidos* 86: 1173 (1989)), replicación de secuencia autosostenida (Guatelli et al., *Proc. Nat. Acad. Sci. Estados Unidos*, 87: 1874 (1990)) y amplificación de secuencia basada en ácido nucleico (NASBA, *Nucleic Acid Sequence Based Amplification*).

Además, los métodos desvelados en las solicitudes de patente de Estados Unidos pendientes de trámite n.º 10/106.097, presentada el 26 de marzo de 2002, titulada "Methods for Genomic Analysis"; 10/042.406, presentada el 9 de enero del 2002 y presentada como patente de Estados Unidos n.º 6,898,531 el 24 de mayo de 2005 titulada "Algorithms for Selection of Primer Pairs"; 10/042.492, presentada el 9 de enero de 2002 y presentada como patente de Estados Unidos n.º 6.740.510 el 25 de mayo de 2004 titulada "Methods for Amplification of Nucleic Acids"; 10/236.480, presentada el 5 de septiembre de 2002 y publicada como 2003-0108919 el 12 de junio de 2003 titulada "Methods for Amplification of Nucleic Acids"; 10/174.101, presentada el 17 de junio de 2002 titulada "Methods for Storage of Reaction Cocktails"; 10/447.685, presentada el 28 de mayo de 2003 y publicada como 2004-0241657 el 2 de diciembre de 2004, titulada "Liver Related Disease Compositions and Methods", 10/768.788, presentada el 4 de marzo de 2004 y publicada como 005-0019787 el 27 de enero de 2005 titulada "Apparatus and Methods for Analyzing and Characterizing Nucleic Acid Sequences"; 11/510.261 presentada el 25 de agosto de 2006 y publicada como 2007-0037198 el 15 de febrero del 2007 y 10/427.696, presentada el 30 de abril de 2003 y publicada como patente de Estados Unidos n.º 7,124,033 el 27 de octubre de 2006 titulada "Method for Identifying Matched Groups", son adecuadas para la amplificación, el marcaje o la manipulación adicional (es decir, fragmentación) de ácidos nucleicos para su uso en determinados métodos de la presente invención.

En un estudio de asociación, los locus genéticos que se sabe que son polimórficos (por ejemplo, los SNP) se genotipan para cada individuo en cada uno de los grupos de casos y control y se calcula una frecuencia alélica

relativa para cada uno de los locus para cada uno de los grupos según los genotipos presentes en los grupos. Es decir, si se genotifican 10 locus polimórficos, se determinan 20 frecuencias alélicas relativas, 10 por cada uno de los grupos de casos y control. Para un locus polimórfico determinado, la frecuencia alélica relativa del grupo de casos se compara con la del grupo de control, y si el locus polimórfico tiene una frecuencia alélica relativa significativamente diferente en el grupo de casos en comparación con el grupo de control, este se identifica como un locus que puede estar asociado con el rasgo multifactorial que diferencia a los grupos de casos y control ("locus asociado"). En determinadas realizaciones, una diferencia significativa en la frecuencia alélica relativa es una diferencia mayor que aproximadamente 5 %, o mayor que aproximadamente 8 %, o mayor que aproximadamente 10 %, o mayor que aproximadamente 12 %, o mayor que aproximadamente 15 %. El alelo que está presente con más frecuencia en la población de casos se puede denominar "alelo asociado" y alelo que está presente con más frecuencia en la población de control se puede denominar el "alelo no asociado". El número de locus asociados (y, por lo tanto, alelos asociados para los locus asociados bialélicos) identificado variará en gran medida dependiendo de cuántos locus polimórficos contribuyan al rasgo multifactorial (por ejemplo, enfermedad) en estudio o están en desequilibrio de ligamiento con locus que contribuyen. Por ejemplo, si la manifestación de una enfermedad implica 10 genes, entonces el número de locus asociados identificados dependerá de cuántos de los locus polimórficos que están genotificados en el estudio de asociación están en desequilibrio de ligamiento con los alelos de los genes que causan la enfermedad. Típicamente, el número de locus implicados en la manifestación de un rasgo multifactorial varía entre aproximadamente cinco a varios cientos, pero puede ser mayor o menor. Para una descripción detallada de métodos para realizar un estudio de asociación utilizando frecuencias alélicas relativas de un grupo de casos y un grupo de control, véanse las solicitudes de patente de Estados Unidos n.º 60/460.329, presentada el 3 de abril de 2003 y 10/768.788, presentada el 30 de enero de 2004 y publicada como 2005-0019787 el 27 de enero de 2005 ambas tituladas "Apparatus and Methods for Analyzing and Characterizing Nucleic Acid Sequences".

El genotipado de los individuos puede realizarse utilizando cualquier técnica conocida por los expertos en la materia. Las técnicas preferidas permiten realizar una determinación rápida y precisa de múltiples variaciones manipulando muy poco la muestra. Algunos ejemplos de técnicas adecuadas implican, pero sin limitación, secuenciación directa de ADN, electroforesis capilar, hibridación, sondas o cebadores específicos de alelos, análisis de polimorfismo de conformación monocatenaria, matrices de ácidos nucleicos, matrices de perlas, análisis de polimorfismos de longitud de fragmento de restricción, análisis de polimorfismos de longitud de fragmento escindible, ADN polimórfico amplificado al azar, reacción con detección ligasa, análisis de heteroduplex o fragmentos, secuenciación diferencial con espectrometría de masas, microscopía de fuerza atómica, pirosecuenciación, FRET (por ejemplo TaqMan (Applied Biosystems, Inc., Foster City, CA) y ensayos Molecular Beacon (Stratagene, La Jolla, CA)) y otras técnicas bien conocidas en la materia. Diversos métodos para la secuenciación de ADN son muy conocidos y generalmente están disponibles en la técnica. Véase, por ejemplo, Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Nueva York) (2001); Ausubel, et al., *Current Protocols in Molecular Biology* (John Wiley and Sons, Nueva York) (1997), Twyman, et al. (2003) "Techniques Patents for SNP Genotyping", *Pharmacogenomics* 4(1):67-79; and Kristensen, et al. (2001) "High-Throughput Methods for Detection of Genetic Variation", *BioTechniques* 30(2):318-332. Para detalles sobre el uso de matrices de ácidos nucleicos (microplacas de ADN) para la detección, por ejemplo, de SNP, véase la patente de Estados Unidos n.º 6.300.063 publicada por Lipshultz, et al., y la patente de Estados Unidos n.º 5.837.832 de Chee, et al., *HuSNP Mapping Assay*, reagent kit and user manual, Affymetrix parte N.º. 90094 (Affymetrix, Santa Clara, CA).

La frecuencia alélica relativa para un grupo de casos o de controles, puede determinarse directamente, genotificando individualmente todos los individuos de la población para determinar la cantidad exacta de cada alelo en cada individuo de la población. En la solicitud de patente de Estados Unidos 10/351.973, presentada el 27 de enero de 2003, titulada "Apparatus and Methods for Determining Individual Genotypes", solicitud de patente de Estados Unidos 10/786.475, presentada el 14 de febrero de 2004 y publicada como 2004-0210400 el 21 de octubre de 2004 titulada "Improvements to Analysis Methods for Individual Genotyping", solicitud de patente de Estados Unidos n.º 10/970.761, presentada el 21 de octubre de 2004, titulada "Analysis Methods and Apparatus for Individual Genotyping" y solicitud de patente de Estados Unidos n.º 11/173.809, presentada el 1 de julio de 2005, titulada "Algorithm for Estimating Accuracy of Genotype Assignment", se describen con detalle métodos para genotipificar individualmente una pluralidad de individuos. Como alternativa, el genotipado agrupado se puede utilizar para determinar una frecuencia alélica relativa para cada uno de los grupos de casos y controles. Para el genotipado agrupado, las muestras de ácidos nucleicos del grupo de casos se agrupan entre sí (grupo de casos) y las muestras de ácidos nucleicos del grupo de controles se agrupan entre sí (grupo de controles) y las frecuencias alélicas relativas para el grupo de casos y de controles se determinan analizando los grupos de casos y controles. En las solicitudes de patente de Estados Unidos n.º 60/460.329, presentada el 3 de abril de 2003 y 10/768.788, presentada el 30 de enero de 2004 y publicada como 2005-0019787 el 27 de enero del 2005, ambas tituladas "Apparatus and Methods for Analyzing and Characterizing Nucleic Acid Sequences", se analizan con detalle métodos de genotipado agrupado.

## 60 Locus genéticos

La expresión "SNP" o "polimorfismo mononucleotídico" se refiere a una variación genética entre individuos; por

ejemplo, una posición de una sola base nitrogenada en el ADN de organismos que es variable. Los SNP se encuentran en el genoma; gran parte de la variación genética entre individuos se debe a variación en locus de SNP y a menudo esta variación genética da como resultado una variación fenotípica entre individuos. Los SNP para su uso en la presente invención, y sus alelos respectivos, pueden proceder de cualquier cantidad de fuentes, tales como de bases de datos públicas (U.C. Santa Cruz Human Genome Browser Gateway (<http://genome.ucsc.edu/cgi-bin/hgGateway>) o del sitio web NCBI dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>), o puede determinarse experimentalmente como se describe en la patente de Estados Unidos n.º 6.969.589; y 10/284.444, presentada el 31 de octubre del 2002 y publicada como 2006-0188875 el 24 de agosto del 2006 titulada "Human Genomic Polymorphisms". Aunque el uso de los SNP se describe en alguna de las realizaciones presentadas en este documento, se entenderá que también se pueden utilizar otros marcadores genéticos bialélicos. Un marcador genético bialélico es uno que tiene dos formas, o alelos, polimórficos. Como se mencionó anteriormente, para un marcador genético bialélico que está asociado a un rasgo, el alelo que es más abundante en la composición genética de un grupo de casos en comparación con un grupo de controles se denomina "alelo asociado", y el otro alelo puede denominarse "alelo no asociado". Por lo tanto, para cada polimorfismo bialélico que está asociado a un rasgo determinado (por ejemplo, una enfermedad o una respuesta a un fármaco), existe un alelo asociado correspondiente. Otros polimorfismos bialélicos que se pueden utilizar con los métodos expuestos en el presente documento incluyen, pero sin limitación, cambios, inserciones, deleciones y translocaciones de multinucleótidos. Se apreciará además que las referencias al ADN en este documento pueden incluir ADN genómico, ADN mitocondrial, ADN episomal y/o derivados de ADN tales como amplicones, transcritos de ARN, ADNc, análogos de ADN, etc. Los locus polimórficos que se exploran en un estudio de asociación pueden estar en un estado diploide o haploide e, idealmente, procederían de sitios de todo el genoma.

En algunas realizaciones de la presente invención, un estudio de asociación implica la exploración de al menos aproximadamente 100 SNP, o de al menos aproximadamente 500 SNP, o de al menos aproximadamente 1000 SNP, o de al menos aproximadamente 10.000 SNP, o de al menos aproximadamente 100.000 SNP o de al menos aproximadamente 1.000.000 SNP. En determinadas realizaciones, se exploran los SNP que están localizados en una o en más partes del genoma que se cree que están asociadas con el rasgo multifactorial. Por ejemplo, los SNP en "regiones candidatas" previamente identificadas (por ejemplo, genes que se han identificado en la bibliografía como asociados o "ligados" al rasgo multifactorial) pueden incluirse en un estudio de asociación; en determinadas asociaciones, los SNP en regiones candidatas que se han identificado en estudios previos múltiples (por ejemplo, más de uno o dos o más estudios) se eligen para su inclusión en un estudio de asociación. En determinadas realizaciones, se exploran los SNP en uno o más cromosomas. En otras realizaciones más, se exploran los SNP de cada cromosoma en un genoma. En otras realizaciones, se exploran SNP múltiples de cada cromosoma en un genoma. En otras realizaciones, se exploran los SNP que están localizados en la región codificante o en la región reguladora de un gen. En realizaciones adicionales, se exploran los SNP que se ha encontrado que están asociados a la expresión alélica diferencial de un gen. (La expresión alélica diferencial se produce cuando un alelo de un gen se expresa a un nivel más alto que otro alelo del mismo gen en un heterocigoto, y se describe con detalle en la solicitud de patente de Estados Unidos n.º 10/438.184, presentada el 13 de mayo de 2003 y publicada como 2004-0229224 el 18 de noviembre de 2004 titulada "Allele-specific Expression Patterns"). En determinadas realizaciones, se exploran todos los SNP conocidos (aproximadamente 3 millones hasta ahora). En otras realizaciones, se explora un subconjunto de SNP que puede utilizarse para predecir la composición alélica de un subconjunto de SNP que no se explora. Dichos "SNP de etiqueta" se describen adicionalmente, por ejemplo, en Hinds, et al. ("Whole-Genome Patterns of Common DNA Variation in Three Human Populations" *Science* (2005) 307:1072-1079). Los SNP explorados mediante los métodos presentados en este documento pueden estar en un estado diploide o haploide en un individuo.

El número de SNP asociados (y, por lo tanto, de alelos asociados) identificados por los métodos presentados en este documento, depende de diversos criterios. Primero, depende del número de locus genéticos que están implicados en la manifestación de la enfermedad. Por ejemplo, si la base genética de una enfermedad multifactorial implica solo algunos locus, entonces el número de SNP asociados y de alelos asociados será típicamente menor que el encontrado para una enfermedad multifactorial cuya base genética implique cientos de locus. Además, el número de SNP asociados y de alelos asociados identificados depende de cuántos SNP se exploren en el estudio de asociación. Por ejemplo, un estudio de asociación que explore solo 100 SNP en los grupos de casos y controles tendrá menos probabilidades de encontrar una gran cantidad de SNP asociados que uno que explore un millón de SNP. Típicamente, los métodos presentados en este documento identificarán entre aproximadamente 10 y varios cientos de SNP asociados/alelos asociados, pero pueden identificarse más o menos.

#### 55 Validación del conjunto de alelos asociados

En una realización, para validar la identificación de los alelos asociados, el estudio de asociación se repite utilizando una segunda población de casos y una segunda población de controles. Este segundo estudio de asociación determina si los alelos asociados del primer estudio de asociación todavía se identifican como alelos asociados en función de las frecuencias alélicas relativas de un nuevo conjunto de casos y controles, y los que sí se "repiten" se validaron como SNP asociados. En determinadas realizaciones, los locus polimórficos que se exploraron en el primer estudio de asociación también se exploraron en el segundo de asociación de validación. En otras

realizaciones, un subconjunto de los locus polimórficos que se exploraron en el primer estudio de asociación se exploraron en el segundo estudio de asociación de validación. En una realización específica, el conjunto de locus polimórficos explorados en el segundo estudio de asociación comprende los locus polimórficos asociados que se identificaron mediante el primer estudio de asociación. Por ejemplo, si se identifican 30.000 SNP como asociados con la frecuencia de una enfermedad en un primer estudio de asociación, entonces esos 30.000 SNP se exploran posteriormente en un segundo estudio de asociación para el cual se selecciona un segundo grupo de casos de individuos que presenta la enfermedad y un segundo grupo de control que no presenta la enfermedad. En determinadas realizaciones, el segundo grupo de casos se selecciona según los mismos criterios que el primer grupo de casos, y el segundo grupo de control se selecciona según los mismos criterios que los del primer grupo de control. En un aspecto, el primer y segundo grupo de casos y el primer y segundo grupo de control no tiene miembros en común. El segundo estudio de asociación se puede realizar utilizando una metodología de genotipificación individual o agrupada.

En otros aspectos, si se realiza un estudio de asociación utilizando una genotipificación agrupada, el conjunto de alelos asociados determinado por la metodología de genotipificación agrupada puede validarse genotipificando individualmente el conjunto de SNP asociados en cada individuo en los grupos de casos y controles y volviendo a calcular y comparar las frecuencias alélicas relativas. Los alelos asociados que se identificaron basándose en el análisis de genotipificación agrupada inicial que tienen una frecuencia alélica significativamente más alta en el grupo de casos en comparación con el grupo de control basándose en los datos de genotipificación individuales se verifican de este modo como alelos asociados. Esta etapa de validación puede realizarse para un primer estudio de asociación que utiliza una metodología de genotipificación agrupada, o puede realizarse para un segundo estudio de asociación de validación que utiliza una metodología de genotipificación agrupada.

En un diseño de estudio para identificar un conjunto de SNP asociados puede utilizarse más de un método de validación. Por ejemplo, en una realización de la presente invención, se realiza un estudio de asociación inicial con una población de casos de individuos que presentan una enfermedad y una población de control de individuos que no presentan la enfermedad. Se utiliza una metodología de genotipificación agrupada para genotipificar los grupos de casos y control en aproximadamente 1,5 millones de locus de SNP para identificar aproximadamente 30.000 SNP con frecuencias alélicas relativas que difieren significativamente entre los grupos de casos y controles. En una siguiente etapa de validación, los grupos de casos y controles se genotipan individualmente en cada uno de los aproximadamente 30.000 SNP identificados en el estudio de asociación "agrupado" para identificar aproximadamente 300 SNP que tienen frecuencias alélicas relativas significativamente diferentes en el grupo de casos que en el grupo de control basándose en la metodología de genotipificación individual. Por lo tanto, estos aproximadamente 300 SNP han sido validados por genotipificación individual. En una etapa de validación adicional, se realiza un segundo estudio de asociación en el que los aproximadamente 300 SNP validados mediante la etapa de genotipificación individual se validan adicionalmente realizando un segundo estudio de asociación basado en una metodología de genotipificación individual con un segundo grupo de casos y un segundo grupo de control. Aquellos SNP que se repiten en el segundo estudio de asociación se clasifican como SNP asociados para la enfermedad y los alelos de los SNP asociados que son más abundantes en los grupos de casos que en los grupos de control se denominan alelos asociados.

En otras realizaciones, un estudio de asociación puede comprender además una resecuenciación de regiones genómicas que contienen variantes que se encuentran asociadas con el estado fenotípico de interés en una ronda más temprana del estudio de asociación (por ejemplo, una ronda de genotipificación agrupada o individual). Los datos generados por esta resecuenciación pueden utilizarse de varias maneras, por ejemplo, para validar las variantes previamente encontradas asociadas y/o para identificar variantes adicionales (por ejemplo, SNP, deleciones, inserciones, translocaciones, etc., habituales y/o poco frecuentes) asociadas con el estado fenotípico de interés. En algunas realizaciones, la resecuenciación se realiza en muestras agrupadas, y en otras realizaciones se realiza en muestras individuales. Las muestras a resecuenciar, ya sea agrupadas o individualmente, pueden amplificarse mediante métodos conocidos por los expertos habituales en la materia, por ejemplo, PCR, LCR, etc. La resecuenciación puede realizarse en muestras de todos los individuos de casos y controles hasta el momento analizadas en el estudio, o en una parte de las mismas. Como alternativa, la resecuenciación puede realizarse en muestras de individuos de casos y controles no analizados previamente en el estudio, solos o en combinación con muestras de individuos que se han analizado previamente. En algunas realizaciones, la resecuenciación se realiza en una secuencia genómica génica y no génica, y en otras realizaciones, solo se resecuencia la secuencia génica y no génica. En determinadas realizaciones específicas, solo se resecueñan los exones. Por ejemplo, se identifica un conjunto de variantes en una primera etapa o etapas de un estudio de asociación y estas variantes están en desequilibrio de ligamiento con un conjunto de genes. Estos genes (o partes de ellos, por ejemplo, exones) se someten a resecuenciación agrupada. En determinadas realizaciones, la resecuenciación se realiza de manera que cada cromosoma se representa al menos dos, tres, cuatro, siete o diez 10 veces en el conjunto de datos de los resultados de secuenciación final. El conjunto de datos de secuenciación se analiza para identificar variantes asociadas al estado fenotípico de interés, por ejemplo, SNP habituales, SNP poco frecuentes y/u otras variantes de secuencia. En determinadas realizaciones, los SNP no sinónimos son de particular interés, ya que diferentes alelos de dicho SNP codifican proteínas con diferentes secuencias de aminoácidos. También se pueden resecuenciar genes adicionales, por ejemplo, genes elegidos basándose en información en la bibliografía publicada que indica

una verosimilitud razonable de que el gen está asociado al estado fenotípico de interés. En realizaciones adicionales, las variantes identificadas como asociadas por una etapa de resecuenciación agrupada pueden someterse adicionalmente a resecuenciación individual, por ejemplo, en muestras de todos o de una parte de los individuos de casos y control analizados previamente en el estudio de asociación. Esta resecuenciación individual puede identificar locus adicionales asociados al estado fenotípico de interés, así como validar locus previamente identificados. Las tecnologías de resecuenciación son muy conocidas por los expertos en la técnica e incluyen la secuenciación de Maxam-Gilbert, los métodos de determinación de cadena, la pirosecuenciación (por ejemplo, de 454 Life Sciences (Branford, CT)), el sistema SOLiD™ (Applied Biosystems (Foster City, CA)), la secuenciación por hibridación (por ejemplo, de Perlegen Sciences, Inc. (Mountain View, CA)), la tecnología de amplificación Bridge™ (Illumina (San Diego, CA)), la secuenciación de una sola molécula (Helicos Biosciences (Cambridge, MA)), etc.

En las solicitudes de patente de Estados Unidos n.º 10/106.097 presentada el 26 de marzo de 2002 (ahora patente de Estados Unidos n.º 6.969.589); 10/042.819, presentada el 7 de enero de 2002 (ahora patente de Estados Unidos n.º 6.897.025); 10/286.417, presentada el 31 de octubre 2002 y publicada como 2004-0023237 el 5 de febrero de 2004; 60/648.957, presentada el 31 de enero de 2005 de las cuales la 11/344. 975 reivindica prioridad y fue publicada como 2006-0228728 en 12 de octubre de 2006; 10/447.685, presentada el 28 de mayo de 2003 y publicada como 2004-0241657 el 2 de diciembre de 2004; 10/691.069, presentada el 21 de octubre de 2003 y publicada como 2005-0086009 el 21 de abril de 2005; 10/427.696, presentada el 30 de abril de 2003 (ahora patente de Estados Unidos n.º 7,124,033); 60/572,533, presentada el 18 de mayo de 2004; 10/845. 316, presentada el 12 de mayo de 2004 y publicada como 2005-0003410 el 6 de enero de 2005; 10/940.410, presentada el 13 de septiembre de 2004 (ahora patente de Estados Unidos n.º 7,335.474); 11/043,689, presentada el 24 de enero de 2005 y publicada como 2006-0166224 el 27 de julio de 2006; 10/956.224, presentada el 30 de septiembre de 2004 (ahora patente de Estados Unidos n.º 7.127.355); y 60/643.006, presentada el 11 de enero de 2005 de las cuales la 11/299.298 reivindica prioridad y fue publicada como 2006-0177847; así como las solicitudes PCT n.º US04/016950, presentada el 27 de mayo de 2004; US04/13577, presentada el 30 de abril de 2004; US05/07375, presentada el 3 de marzo de 2005; y en "Genomewide association study of 14,000 cases of seven common diseases and 3,000 shared controls", Nature (2007) 447:661-678, se proporcionan más ejemplos de diseño de estudios de asociación.

En determinadas realizaciones, un estudio de asociación puede comprender además el uso de información fenotípica y genética para identificar un conjunto de características fenotípicas, así como un conjunto de locus genéticos (por ejemplo, SNP) que pueden utilizarse conjuntamente para predecir el riesgo de que un individuo tenga, o sea susceptible a, un rasgo multifactorial determinado. Al igual que los datos de genotipificación, se reciben datos sobre un conjunto de fenotipos de los individuos para los individuos tanto de casos como de control. Los datos en un conjunto de fenotipos incluyen preferiblemente datos sobre al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 fenotipos diferentes, o más preferentemente, sobre al menos 10, 25, 30, 35, 40, 45 o 50 fenotipos diferentes de los individuos en el estudio de asociación. Los datos en el conjunto de fenotipos se pueden recopilar antes, después de, o simultáneamente con, la recopilación/recogida de datos de genotipificación. Los datos de fenotipo recogidos pueden (al igual que los datos de genotipificación) también almacenarse en un medio legible por ordenador para su uso posterior. En un estudio de asociación para un rasgo multifactorial se utilizan simultáneamente tanto los datos de genotipo como los de fenotipo del grupo de individuos. Los resultados del estudio de asociación pueden comercializarse en cualquier forma de, por ejemplo, datos, kits y/o fármacos mejorados. Por ejemplo, en la solicitud de patente de Estados Unidos n.º 11/043.689, presentada el 24 de enero de 2005 y publicada como 2006-0166224 el 27 de julio de 2006 titulada "Associations Using Genotypes and Phenotypes", se proporciona una descripción adicional de dichas realizaciones.

#### Uso de alelos asociados para determinar los límites de riesgo

En determinadas realizaciones de la presente invención, los genotipos de los individuos en los grupos de casos y control en cada uno de los locus de SNP asociados a enfermedad, se utilizan para desarrollar una serie de valores límite que se utilizará para determinar la predisposición de un individuo para desarrollar el rasgo multifactorial que diferencia el grupo de casos del grupo de control.

En un aspecto, se recopilan los genotipos en cada posición de SNP asociado de todos los individuos en los grupos de casos y de control. Si la genotipificación individual se realizó durante el estudio de asociación, como se ha indicado anteriormente, entonces pueden utilizarse los datos de genotipificación recopilados para las posiciones de SNP asociados durante el estudio de asociación. Sin embargo, si no se han determinado los genotipos individuales, entonces cada miembro del grupo de casos y de control debe genotipificarse individualmente para el conjunto de SNP asociados. Por ejemplo, en el caso de un SNP bialélico, un individuo diploide puede tener uno de tres genotipos diferentes, homocigoto para el alelo asociado, homocigoto para el alelo no asociado y heterocigoto (que tiene un alelo asociado y un alelo no asociado). Los métodos presentados en este documento también se pueden aplicar a organismos haploides, o a locus haploides en organismos diploides (por ejemplo, locus del cromosoma Y en seres humanos). Para un locus haploide, solo habría dos genotipos, uno para cada posible alelo.

En otro aspecto, a cada individuo en los grupos de casos y de control se le asigna una puntuación basándose al menos en su genotipo en cada uno de los locus de SNP asociados. La puntuación de un individuo se puede calcular

de varias maneras. Por ejemplo, en una realización de la presente invención, cada alelo asociado se valora en un punto, de modo que cada genotipo de SNP que es homocigoto para el alelo asociado vale dos puntos, cada genotipo de SNP que es heterocigoto vale un punto y cada genotipo de SNP que es homocigoto para el alelo no asociado vale cero puntos. En otro ejemplo, cada alelo asociado se valora en un punto y cada alelo no asociado se valora en -1 punto, por lo que cada genotipo de SNP que es homocigoto para el alelo asociado vale dos puntos, cada genotipo de SNP que es heterocigoto vale cero puntos y cada genotipo de SNP que es homocigoto para el alelo no asociado vale -2 puntos. En otros ejemplos, cada alelo asociado vale cero puntos y cada alelo no asociado vale un punto, o cada alelo asociado vale -1 punto y cada alelo no asociado vale 1 punto. En una realización para un locus haploide, cada genotipo de SNP con el alelo asociado vale un punto y cada genotipo de SNP con el alelo no asociado vale cero puntos. En otra realización, para un locus haploide, cada genotipo de SNP con el alelo asociado vale dos puntos y cada genotipo de SNP con el alelo no asociado vale cero puntos. En determinadas realizaciones, si hay información adicional sobre el impacto biológico de ese alelo en el rasgo multifactorial, puede ajustarse un valor para un alelo particular. Por ejemplo, si se sabe que un alelo asociado particular confiere un mayor riesgo de desarrollar el rasgo multifactorial que otros alelos asociados, entonces al alelo asociado particular se le puede asignar un valor que es diferente al de los valores asignados a otros alelos asociados para explicar el mayor riesgo conferido. En otro ejemplo, a los alelos asociados (o en LD con) un gen previamente identificado como asociado con el rasgo multifactorial, se les asigna un valor diferente al de los valores asignados a otros alelos asociados; en algunas realizaciones, el valor puede ajustarse en función de cuánta información (por ejemplo, número de referencias en la bibliografía) indica una asociación con el rasgo. Un experto habitual en la técnica reconocerá fácilmente que el valor particular asignado a un alelo asociado o no asociado no es tan importante como la coherencia con la que se aplica ese valor en todos los alelos al calcular una puntuación para un individuo. Para simplificar, las siguientes descripciones corresponderán a realizaciones en las que se asigna un valor de uno a un alelo asociado y se asigna un valor de cero a un alelo no asociado, pero la invención no está de ninguna manera limitada a dicho método de valoración de alelos asociados y no asociados.

En otras realizaciones de la presente invención, cuando se genera una puntuación para un individuo, se tienen en cuenta otros factores tales como, por ejemplo, las magnitudes del efecto que cada factor tiene sobre el rasgo multifactorial y/o si los factores son, o no, aditivos, sinérgicos, antagónicos, independientes o, de otra manera, epistáticos cuando se calcula la puntuación de un individuo. Por ejemplo, un primer alelo puede tener una magnitud dos veces mayor que un efecto sobre el riesgo de un individuo que un segundo alelo, por lo que el valor asignado al primer alelo puede ser dos veces mayor que el valor asignado al segundo alelo en el cálculo de la puntuación del individuo. En algunas realizaciones, un conjunto dado de factores puede mostrar una combinación de diferentes interacciones epistáticas. En otras realizaciones, también pueden tenerse en cuenta datos de expresión. Por ejemplo, el nivel de expresión de un alelo particular de un gen (o un ARN o producto proteico del mismo) que es un factor que subyace al desarrollo de un cáncer, puede ser predictivo en un aumento o disminución del riesgo de desarrollar el cáncer. En realizaciones adicionales, el conocimiento de los factores ambientales que pueden influir en el rasgo multifactorial se tiene en cuenta cuando se calcula la puntuación de un individuo. En determinadas realizaciones, los puntos asociados con la presencia de uno o más factores ambientales se utilizan en el cálculo de la puntuación de un individuo, cuyos métodos son fácilmente obvios para los expertos en la técnica a partir de las descripciones de este documento.

En otras realizaciones, cuando se calcula la puntuación de un individuo, se tiene en cuenta la información clínica, demográfica y/o epidemiológica con respecto al paciente y/o al rasgo multifactorial. Por ejemplo, la edad, el tiempo que transcurre con un diagnóstico específico de la enfermedad, antecedentes de otras enfermedades asociadas, el uso de diversos fármacos, la insuficiencia renal o hepática, la presencia de anomalías fisiológicas, tales como hipertrofia ventricular izquierda, raza, peso, sexo y/o acontecimientos clínicos pasados, podrían combinarse con factores genéticos para predecir una respuesta clínica. En dichas realizaciones, los puntos asociados con dicha información se utilizan en el cálculo de la puntuación del individuo.

En determinadas realizaciones, para un individuo determinado, todos los puntos a través de todos los SNP asociados (y cualquier otro factor, como se indicó anteriormente) se suman para proporcionar una puntuación para ese individuo. Por ejemplo, si se genotipan 100 SNP asociados, la puntuación máxima para un individuo basada solo en los genotipos de esos SNP es de 200, lo que significa que el individuo tiene dos alelos asociados en cada posición de SNP asociada. En otras palabras, el individuo es homocigoto para el alelo asociado en cada posición de SNP. La puntuación mínima es 0, para un individuo que no tiene alelos asociados en ninguna de las posiciones de SNP asociados, o es homocigoto para el alelo no asociado en cada posición de SNP. Las puntuaciones se calculan para cada individuo en los grupos de casos y control. En un ejemplo en el que las puntuaciones se calculan usando solo genotipos de SNP, se examinan 100 SNP asociados para una población de casos de 102 individuos y una población de control de 405 individuos. La puntuación más baja en el grupo de casos es de 42 y la puntuación más alta es de 97; para el grupo de control, la puntuación más baja es 23 y la puntuación más alta es de 79. Este es un ejemplo de una realización para determinar una puntuación para un individuo que no se basa en las magnitudes del efecto que tiene cada factor sobre el rasgo multifactorial, ni depende de si los efectos de los factores son aditivos, sinérgicos o independientes. Además, esta realización no requiere conocer el perfil de expresión de ninguno de los genes, ARN o proteínas que comprenden, están genéticamente ligados con, o que son productos de, cualquier factor. Además, la realización no requiere conocer cuáles son los factores ambientales que pueden influir en el rasgo

multifactorial.

En otro aspecto, se determina una serie de valores de límite de riesgo. Los valores de límite de riesgo representan valores umbral hipotéticos para su uso en una prueba de evaluación de riesgos (por ejemplo, una prueba genética o poligénica) para identificar individuos que pueden desarrollar o presentar un rasgo multifactorial. Por ejemplo, los individuos que tienen una puntuación más alta que la de un valor umbral pueden diagnosticarse como propensos a presentar el rasgo multifactorial y aquellos que tienen una puntuación igual o inferior a la del umbral pueden ser diagnosticados como poco propensos a presentar el rasgo multifactorial. Como alternativa, para determinar el riesgo de un individuo de presentar el rasgo multifactorial pueden utilizarse umbrales múltiples. Por supuesto, dependiendo de cómo se calculen las puntuaciones, una puntuación baja puede ser indicativa de un riesgo alto y una puntuación alta puede ser indicativa de un riesgo bajo. Por claridad, las siguientes descripciones corresponden a realizaciones en las que una puntuación más alta indica un riesgo más alto.

La serie de valores de límite de riesgo abarca un intervalo que varía de 1 a la puntuación más alta calculada para un individuo en el estudio de asociación, independientemente de si es un miembro del grupo de casos o de control. En el ejemplo descrito anteriormente, la puntuación más alta para un individuo es de 97 puntos, por lo que el intervalo a partir del cual se determinan los valores de límite de riesgo (el intervalo de límite de riesgo) está entre 1 y 97. En determinados aspectos, los valores de límite de riesgo se seleccionan desde el intervalo de límite de riesgo, aunque la selección de valores de límite de riesgo particulares es algo arbitraria. En algunas realizaciones, se elige cada puntuación dentro del intervalo de límite de riesgo. En otras realizaciones, se elige cada enésima puntuación (cada 5 o 10, por ejemplo). En aún otras realizaciones, el intervalo se divide en porcentajes y se elige cada enésimo porcentaje. En algunas realizaciones, se seleccionan más valores de límite de riesgo de la parte central del intervalo de puntuaciones completo que de la parte superior o inferior del intervalo, o viceversa. Por ejemplo, en el caso en que el intervalo de puntuaciones completo se encuentre entre 1 y 97, los valores de límite de riesgo se eligen en cada 10ª puntuación entre 20 y 80, y se añaden valores de límite de riesgo adicionales de 55 y 65 para evaluar mejor la parte central de este intervalo (véase la Tabla 1).

En una etapa posterior, cada uno de los valores de límite de riesgo se compara con las puntuaciones calculadas para los individuos en los grupos de casos y control. Específicamente, las puntuaciones para los individuos de grupo de casos (“afectados”) y control (“no afectados”) se utilizan para determinar cuáles de los valores de límite de riesgo proporcionan la mejor sensibilidad, especificidad, valor predictivo positivo (VPP), valor predictivo negativo (VPN), precisión o una combinación de los mismos para diferenciar a los individuos que probablemente muestren el rasgo multifactorial de aquellos que probablemente no muestren el rasgo multifactorial, identificando así un valor de límite de riesgo que sería un buen valor umbral para una prueba de evaluación de riesgos utilizando los SNP asociados. Además, la identificación de un valor umbral apropiado puede implicar adicionalmente el uso de información clínica, demográfica y/o epidemiológica (por ejemplo, con respecto al rasgo multifactorial, población de estudio o individuos a analizar) y/o la interacción del facultativo de la presente invención con una agencia externa (por ejemplo, la FDA, *Food and Drug Administration*, administración de alimentos y fármacos de Estados Unidos)). Este valor umbral puede desarrollarse en una prueba de evaluación de riesgos, por ejemplo, un diagnóstico, con la sensibilidad, la especificidad, el VPP, el VPN y la precisión calculados en función del valor umbral y las puntuaciones para los individuos de los grupos de casos y de control. Los métodos para realizar lo anterior serán muy obvios para los expertos en la técnica a partir de las descripciones de este documento.

Una prueba de evaluación de riesgos de dos clases tiene dos posibles resultados. Un resultado de prueba positivo indica que un individuo presenta o es probable que presente un rasgo de interés, y un resultado de prueba negativo indica que un individuo no presenta y no es probable que presente el rasgo de interés. Como tal, la fiabilidad de una prueba de evaluación de riesgos está relacionada con la frecuencia con la que el resultado de la prueba identifica correctamente a un individuo como “positivo” o “negativo” para el rasgo. Los positivos verdaderos (PV) y los negativos verdaderos (NV) son resultados de pruebas que identifican con precisión a los individuos como positivos (por ejemplo, “afectados”) y negativos (por ejemplo, “no afectados”) respectivamente. Un positivo falso (PF) es un resultado de prueba que clasifica incorrectamente a un individuo como positivo cuando de hecho es negativo para el rasgo. Del mismo modo, un negativo falso (NF) es un resultado de prueba que clasifica incorrectamente a un individuo como negativo cuando de hecho es positivo para el rasgo. Las medidas PV, NV, PF y NF se utilizan para calcular la sensibilidad, la especificidad, el VPP y el VPN para una prueba de evaluación de riesgos.

La “sensibilidad” de una prueba es una medida de la capacidad de la prueba para identificar correctamente a un individuo afectado, o a uno que desarrollará el rasgo de interés. Cuanto más cerca esté la sensibilidad de uno, más precisa será la prueba para identificar a los individuos afectados. Específicamente, la sensibilidad se refiere a la proporción de individuos afectados que son diagnosticados correctamente como tales por la prueba, y se calcula como número de individuos identificados correctamente como afectados (PV) dividido entre el número total de individuos afectados (PV + NF). Se prefiere una alta sensibilidad para que la mayoría de los afectados se identifiquen como tales en la prueba de evaluación de riesgos. La “especificidad” de una prueba es una medida de la capacidad de la prueba para identificar correctamente a un individuo no afectado, o a uno que no desarrollará el rasgo interés. Cuanto más cerca esté la especificidad de uno, más precisa será la prueba para identificar a los individuos no afectados. Específicamente, la especificidad se refiere a la proporción de individuos no afectados que

la prueba identifica correctamente como tales y se calcula como el número de individuos identificados correctamente como no afectados (NV) dividido entre el número total de individuos no afectados (NV + PF). Se prefiere una alta especificidad para minimizar el número de individuos que se identifica incorrectamente como afectados. Por lo tanto, para un valor de límite de riesgo dado, la sensibilidad se calcula como la proporción de casos individuales con una puntuación más alta que el valor de límite de riesgo y la especificidad se calcula como la proporción de individuos de control con una puntuación más baja o igual al valor de límite de riesgo (o, uno menos la proporción de individuos de control con una puntuación más alta que el valor de límite de riesgo).

El “valor predictivo positivo” (VPP) de una prueba de evaluación de riesgos evalúa la fiabilidad de un efecto/resultado de prueba positivo y se calcula como la proporción de individuos con un resultado de prueba positivo que realmente tiene el rasgo de interés. En otras palabras, es la probabilidad de que un resultado de prueba positivo identifique con precisión a un individuo que tiene el rasgo y se calcula como el número de individuos identificados correctamente como afectados (PV) dividido entre el número total de individuos identificados como afectados por la prueba de evaluación de riesgos (PV + PF). En muchos casos, se prefiere un VPP elevado, de modo que la mayoría de los individuos que se identifican como afectados se ven realmente afectados. Por ejemplo, un VPP de 0,98 significa que un individuo con un resultado de prueba positivo tiene un 98% de probabilidad de tener o desarrollar el rasgo. El “valor predictivo negativo” (VPN) de una prueba de evaluación de riesgos evalúa la fiabilidad de un efecto/resultado de prueba negativo y se calcula como la proporción de individuos con un resultado de prueba negativo que no tienen el rasgo de interés. Dicho de otro modo, es la probabilidad de que un resultado de prueba negativo identifique con precisión a un individuo que no tiene el rasgo y se calcula como el número de individuos identificados correctamente como no afectados (NV) dividido entre el número total de individuos identificados como no afectados (NV + NF). En ocasiones, se prefiere un VPN elevado para que la mayoría de los individuos que se identifican como no afectados no se vean realmente afectados (por ejemplo, al excluir sujetos en riesgo de presentar acontecimientos adversos asociados a la administración de un fármaco específico). Por ejemplo, un VPN de 0,999 significa que un individuo con un resultado de prueba negativo tiene solo un 0,1 % de probabilidad de tener o desarrollar el rasgo (por ejemplo, de experimentar el acontecimiento adverso en respuesta al fármaco). Por lo tanto, para un valor de límite de riesgo dado, el VPP puede calcularse como la proporción de todos los individuos con una puntuación más alta que el valor de límite de riesgo que realmente se encuentra en el grupo de casos y el VPN se calcula como la proporción de todos los individuos con una puntuación más baja o igual al valor de límite de riesgo que realmente hay en un grupo de control.

La prevalencia de un rasgo es la frecuencia del rasgo entre la población que se somete a ensayo y se calcula como el número de casos existentes dividido entre la población total en un punto temporal determinado. Aunque la sensibilidad y especificidad de una prueba no están influenciadas por la prevalencia del rasgo en cuestión, tanto el VPP como el VPN están muy influenciados por la prevalencia del rasgo en la población que se somete a ensayo; una menor prevalencia de la enfermedad da como resultado un VPP más bajo y un VPN más alto. Tanto el VPP como en VPN también pueden calcularse como una función de la sensibilidad (sens), especificidad (esp) y prevalencia (prev):

$$\text{VPP} = (\text{sens})(\text{prev}) / [(\text{sens})(\text{prev}) + (1 - \text{espec})(1 - \text{prev})]$$

$$\text{VPN} = (\text{espec})(1 - \text{prev}) / [(\text{espec})(1 - \text{prev}) + (1 - \text{sens})(\text{prev})]$$

Los valores de umbral también pueden seleccionarse utilizando cocientes de verosimilitud para la prueba de evaluación de riesgos. Un cociente de verosimilitud (CV) es una forma de incorporar la sensibilidad y la especificidad de una prueba en una medida y proporciona una indicación de cuánto cambian las probabilidades de tener o desarrollar un rasgo dado basándose en un resultado de prueba positivo o negativo. Dado que la sensibilidad y la especificidad son características fijas de la propia prueba, el CV es independiente de la prevalencia del rasgo en la población, a diferencia del VPP y VPN. Un CV es la probabilidad de que se espere un resultado de prueba determinado en un individuo con el rasgo en comparación con la probabilidad de que se espere el mismo resultado en un individuo sin el rasgo. Un CV para un resultado de prueba positivo (CV+) proporciona una medida de cuánto aumentan las probabilidades de que un individuo tenga o desarrolle el rasgo cuando la prueba es positiva y se calcula como la sensibilidad dividida entre (1 - especificidad). La mejor prueba a utilizar para “determinar” un rasgo es la que tiene el CV+ más elevado. Un CV para un resultado de prueba negativo (CV-) proporciona una medida de cuánto disminuyen las probabilidades de que un individuo tenga o desarrolle el rasgo cuando la prueba es negativa y se calcula como (1 - sensibilidad) dividido entre la especificidad. La mejor prueba a utilizar para “descartar” un rasgo es la que tiene el CV- más bajo. Los CV mayores de diez o menores de 0,1 normalmente se consideran de un alto valor diagnóstico. Los CV se combinan con la “probabilidades previas a la prueba” para determinar las “probabilidades posteriores a la prueba” que el individuo sometido a la prueba ha desarrollado o desarrollará el rasgo de interés (probabilidades posteriores a la prueba = probabilidades previas a la prueba x CV). Las probabilidades previas a la prueba se calculan como información sobre la prevalencia del rasgo, las características de la población y la información sobre el individuo particular que se somete a la prueba y representan la probabilidad de que el individuo tenga o desarrolle el rasgo antes de la prueba. Las probabilidades posteriores a la prueba representan la probabilidad de que el individuo tenga o desarrolle el rasgo dados los resultados de la prueba. En una realización de la presente invención, se selecciona un valor umbral que maximiza el CV para una prueba de evaluación de riesgos.

Otra medida más del valor o de la utilidad de una prueba de evaluación de riesgos es la precisión, que mide la concordancia global entre los resultados de la prueba y la patología real. La precisión se calcula como la suma de los positivos verdaderos y los negativos verdaderos dividida entre el número total de resultados de la muestra  $((PV + NV)/(PV + NV + PF + NF))$ . La precisión de una prueba de evaluación de riesgos puede utilizarse para determinar cuál de un conjunto de valores de límite de riesgo puede ser un valor umbral útil para la prueba.

La sensibilidad, la especificidad, el VPP, el VPN y la precisión se calculan para cada valor de límite de riesgo y la tabla 1 a continuación enumera estos valores para un ejemplo en el que se analizan 102 casos y 405 controles. En la primera columna se muestran los valores de límite elegidos de todo el intervalo de puntuaciones. El número de casos individuales con una puntuación más alta que el valor de límite correspondiente se muestra en la segunda columna. La tercera columna enumera el número de individuos de control con una puntuación más alta que el valor de límite correspondiente. En la cuarta columna se muestra la sensibilidad para una prueba que utiliza cada uno de los valores de límite correspondientes como valores de umbral. En la quinta columna se muestra la especificidad para una prueba que utiliza cada uno de los valores de límite correspondientes como valores de umbral. El VPP y el VPN de una prueba que utiliza cada uno de los valores de límite de riesgo correspondientes como valores de umbral se muestran en la sexta y séptima columna, respectivamente. Finalmente, la precisión de una prueba que utiliza cada uno de los valores de límite correspondientes como valores de umbral se muestra en la octava columna.

Tabla 1

Valores de límite de riesgo	Número de casos (de 102)	Número de controles (de 405)	Sensibilidad	Especificidad	VPP	VPN	Precisión
80	25	0	0,25	1	1	0,84	0,85
70	51	2	0,50	0,995	0,96	0,89	0,90
65	65	8	0,64	0,98	0,89	0,91	0,91
60	79	34	0,77	0,92	0,70	0,94	0,89
55	93	81	0,91	0,80	0,53	0,97	0,82
50	99	154	0,97	0,62	0,39	0,99	0,69
40	102	318	1	0,21	0,24	1	0,37
30	102	394	1	0,03	0,21	1	0,22
20	102	405	1	0	0,20	1	0,20

En condiciones óptimas, una prueba de evaluación de riesgos es muy sensible y muy específica con alto VPP, VPN y precisión, de modo que todos los individuos evaluados se identifican correctamente como que tienen o no el rasgo de interés. Sin embargo, en circunstancias típicas, la selección de un valor de límite de riesgo óptimo puede basarse, por ejemplo, en la mejor combinación de especificidad, sensibilidad, VPP, VPN y precisión, o en un subconjunto de los mismos. Como se muestra en la tabla 1, el uso de un valor de límite de riesgo alto aumenta la especificidad y el VPP de la prueba al tiempo que reduce la sensibilidad y el VPN. Por lo tanto, si una prueba de evaluación de riesgos para determinar la predisposición de un individuo a desarrollar una enfermedad se basa en un valor de límite de riesgo alto, muy pocos individuos serán diagnosticados erróneamente con un riesgo alto de desarrollar la enfermedad, aunque una gran proporción de los que tienen un riesgo alto no sería identificada. Por otro lado, la utilización de un valor de límite de riesgo bajo, aumenta la sensibilidad y el VPN al tiempo que disminuye la especificidad y el VPP, por lo que, aunque la mayoría o todos los individuos con riesgo alto serían identificados como tales, también se identificaría erróneamente un número significativo de individuos con riesgo bajo como de riesgo alto. Por lo tanto, es obvio que ninguno de estos extremos es útil, sino que se puede determinar un equilibrio de sensibilidad, especificidad, VPP y VPN para el rasgo, la población y el individuo particular en cuestión.

La determinación de un valor umbral depende de muchos factores. Por ejemplo, se puede requerir conocimiento clínico, demográfico y/o epidemiológico de la enfermedad para hacer esta determinación. Además, un organismo regulador (por ejemplo, la FDA) puede regular un valor umbral para una prueba de evaluación de riesgos, o un médico puede modificarlo dependiendo, por ejemplo, de información sobre los posibles tratamientos, las

características de la prueba de evaluación de riesgos o los datos específicos de un paciente particular. Además, puede utilizarse, o no, un valor umbral de forma dicotómica. Por ejemplo, el tratamiento de un individuo puede variar dependiendo de si la puntuación del individuo es mayor que el valor umbral (por ejemplo, administrar un fármaco) o menor o igual al valor umbral (por ejemplo, no administrar el fármaco o modificar la administración del mismo (por ejemplo, dosificación, frecuencia, etc.)). Como alternativa, los individuos con puntuaciones más próximas al umbral pueden tratarse de manera diferente de aquellos que tienen puntuaciones más alejadas al umbral. Por ejemplo, un médico puede decidir administrar el fármaco a un individuo con una puntuación que está ligeramente por debajo del umbral en función de factores adicionales, tales como el conocimiento clínico, los datos demográficos, la información epidemiológica y/o comentarios del individuo. Además, el uso de “mayor que” frente a “menor que o igual a” con respecto a la comparación de una puntuación con un valor umbral, es simplemente una cuestión convencional, y en realizaciones alternativas de la presente invención el uso de “mayor que o igual a” frente a “menor que”, se puede utilizar en su lugar, como será obvio para un experto en la técnica.

En un aspecto, la determinación de un valor umbral depende de la gravedad de la enfermedad, que será fácilmente discernible por un experto en la técnica. Por ejemplo, si el rasgo se relaciona con el desarrollo de una enfermedad grave, entonces se preferiría tener una sensibilidad muy alta con independencia de una mayor especificidad, ya que la identificación de los individuos con riesgo alto es crítica para ellos. Por ejemplo, las neoplasias tratables (cánceres *in situ* o enfermedad de Hodgkin) se deben encontrar precozmente, por lo que en la evaluación diagnóstica se deben utilizar pruebas sensibles. De manera similar, se prefiere una prueba con un VPN alto para una enfermedad grave para garantizar que sea baja la cantidad de negativos falsos. Dado que el número de positivos falsos puede ser significativo debido a un VPP menor al ideal, se pueden realizar pruebas adicionales para confirmar el estado de los individuos que dieron positivo/afectados, utilizando, por ejemplo, una prueba de “estándar de oro” muy precisa. Como tal, puede ser más aceptable tener un VPP más bajo cuando se dispone fácilmente de otros diagnósticos confirmatorios. Por ejemplo, la tasa de células cervicales atípicas en la población general es de aproximadamente 1/1000 y la sensibilidad y especificidad de una prueba de Papanicoláu es de 0,70 y 0,90 respectivamente. Según estos valores, el VPP y el VPN de la prueba de Papanicoláu es de 0,00696 y 0,999 respectivamente, lo que significa que una persona con una prueba de Papanicoláu positiva tiene muy pocas probabilidades de tener atipia, mientras que una persona con una prueba de Papanicoláu negativa es casi seguro que no padezca la enfermedad.

En determinados aspectos, se prefiere una alta especificidad y un VPP alto para una prueba de evaluación de riesgos, por ejemplo, cuando hay repercusiones muy poco deseables de resultados de pruebas positivas falsas. Por ejemplo, si la prueba se utiliza para decidir si un individuo recibirá un régimen de tratamiento peligroso (cirugía de trasplante, quimioterapia, radiación, administración de un fármaco con acontecimientos adversos graves, mastectomía, etc.), entonces es importante que los individuos identificados por la prueba que necesiten el tratamiento realmente necesiten el mismo. Por ejemplo, se puede desarrollar una prueba de evaluación de riesgos para identificar a individuos que están en riesgo alto de muerte en ausencia de un procedimiento de trasplante de corazón. Por lo tanto, los individuos que tienen una puntuación más alta que un valor umbral se identifican como propensos a morir a menos que reciban un nuevo corazón. Se preferiría que dicha prueba tuviese un VPP muy alto (~ 1,0) de modo que para un trasplante de corazón solo se tuviese en cuenta a los individuos con una alta probabilidad de muerte. Aunque esto significaría que un número significativo de individuos que morirá sin un trasplante de corazón será excluido del tratamiento (menor VPN) de manera óptima, a ningún individuo se le realizará un trasplante de corazón que no lo necesite en absoluto.

Otro factor fácilmente discernible por un experto en la técnica para determinar un valor umbral apropiado para una prueba de evaluación de riesgos es la prevalencia de la enfermedad en la población como un todo. Por ejemplo, tomando un rasgo que sea extremadamente poco habitual en la población. Una especificidad de 0,95 puede parecer aceptablemente alta, pero significa que el cinco por ciento de los individuos que no tiene un riesgo alto se diagnosticará erróneamente como que tiene un riesgo alto de desarrollar el rasgo. Por lo tanto, para un rasgo que tiene una frecuencia de 1/10000 en la población, aproximadamente 500 individuos serán diagnosticados erróneamente como de “riesgo alto” (positivos falsos) para cada individuo que se identifica correctamente como en riesgo de desarrollar el rasgo. De acuerdo con esto, es más adecuado utilizar un límite con una especificidad más alta para rasgos poco frecuentes y no graves y un límite con una sensibilidad más alta para rasgos comunes y graves. Además, como se ha descrito anteriormente, el VPP y el VPN son muy dependientes de la prevalencia del rasgo de interés. Por ejemplo, el VPP de una prueba de evaluación de riesgos utilizada para identificar individuos en riesgo de desarrollar una enfermedad de una población que tiene una baja prevalencia de la enfermedad será menor que el VPP de la misma prueba de evaluación de riesgos utilizada para identificar a individuos en riesgo de desarrollar la enfermedad de una población que tiene una alta prevalencia de la enfermedad. De manera similar, el VPN de una prueba de evaluación de riesgos utilizada para identificar individuos en riesgo de desarrollar una enfermedad de una población que tiene una baja prevalencia de la enfermedad será mayor que el VPN de la misma prueba de evaluación de riesgos utilizada para identificar individuos en riesgo de desarrollar la enfermedad de una población que tiene una alta prevalencia de la enfermedad. Como tal, aunque una prueba de evaluación de riesgos puede tener un VPP muy alto (o un VPN muy alto) cuando se usa para evaluar individuos en una población, puede no ser útil en otras poblaciones donde la prevalencia del rasgo de interés es diferente, y por lo tanto se puede elegir un valor umbral diferente para diferentes poblaciones según la prevalencia del rasgo de interés. En resumen, un experto en la materia puede seleccionar valores umbral para obtener uno o más parámetros clínicamente útiles,

tales como sensibilidad, especificidad, VPP, VPN, precisión y similares, para una población de pacientes que tiene una prevalencia particular para un rasgo dado que no utiliza solo los métodos presentados en este documento, sino también la intuición y el conocimiento clínico, así como, por ejemplo, interacciones con agencias reguladoras tales como la FDA.

- 5 En una realización de la presente invención, un valor umbral para una prueba de evaluación de riesgos que utiliza los SNP asociados se determina usando una curva ROC (característica operativa del receptor) (Hanley et al. (1982) Radiology 143:29-36; and Beck, et al. (1986) Arch. Pathol. Lab. Med. 110:13-20) basándose en las sensibilidades y especificidades calculadas para los valores de límite de riesgo. Una curva ROC se relaciona con la interrelación intrínseca entre la sensibilidad y la especificidad de una prueba de evaluación de riesgo y se genera representando gráficamente la sensibilidad como una función de uno menos la especificidad para cada valor de límite de riesgo, como se muestra en la figura 1, que ilustra una curva ROC generada utilizando los datos de la tabla 1. Por tanto, cada valor de límite de riesgo corresponde a un “punto de datos” en la curva ROC. El área bajo la curva proporciona una medida de la fiabilidad de la prueba de evaluación de riesgos. Para una prueba de evaluación de riesgos que pueda diferenciar perfectamente entre individuos afectados y no afectados (teniendo la sensibilidad y la especificidad un valor de 1), el área bajo de la curva es de 1. Para una prueba de evaluación de riesgos que no diferencie entre individuos afectados y no afectados, el área bajo la curva es de 0,5. En general, cuanto más cerca esté la curva de los bordes izquierdo y superior del gráfico, más precisa será la prueba de evaluación de riesgos y cuanto más cerca se encuentre la curva de un ángulo de 45 grados del espacio ROC, menos precisa será la prueba. Los programas informáticos normalmente utilizados para analizar las curvas ROC están a disposición del público e incluyen los programas ROCKIT, CORROC2, LABROC4, ROCFIT, CLABROC, ROCPWR, LABMRMC, y PROPROC, pudiendo todos ellos descargarse de los laboratorios Kurt Rossman de Radiological Image Research en el siguiente sitio web: [www.radiology.uchicago.edu/krl/KRL\\_ROC/software\\_index.htm#ROC%20calculations%20Auxiliary%20software](http://www.radiology.uchicago.edu/krl/KRL_ROC/software_index.htm#ROC%20calculations%20Auxiliary%20software). En determinadas realizaciones, se elige un valor umbral de los valores de límite de riesgo, cuyos puntos de datos se encuentran en una parte (por ejemplo, porcentaje) de la curva ROC que está más cerca de la esquina superior izquierda del gráfico. Por ejemplo, si se eligieran puntos de datos del 20 % de la curva ROC más cercana a la esquina superior izquierda del gráfico mostrado en la figura 1 (entre las flechas A y D), se seleccionaría un valor umbral de los puntos de datos correspondientes a valores de límite de riesgo de 55 y 60, indicados como D y E respectivamente. En otras realizaciones, se determina que un valor umbral es el valor de límite de riesgo cuya sensibilidad y especificidad está representada por el punto de datos más cercano a la esquina superior izquierda del gráfico. En la figura 1, este punto de datos (D) corresponde a un valor de límite de riesgo de 55. En otras realizaciones más, se determina un valor umbral a partir de la posición en la curva ROC que está más cerca de la esquina superior izquierda del gráfico. En la figura 1, esta posición se indica como C y corresponde a una sensibilidad de aproximadamente 0,87 y a una especificidad de aproximadamente 0,84. En esta realización, se determina un valor de límite de riesgo que corresponde a la sensibilidad y especificidad representada por esta posición en la curva y ese valor de límite de riesgo se utiliza como valor umbral para una prueba de evaluación de riesgos utilizando los SNP asociados. Por ejemplo, dado que la posición C está entre los puntos de datos D y E, el valor de límite de riesgo óptimo para su uso como un valor umbral debe estar entre 55 y 60. Para determinar el valor de límite de riesgo óptimo, la sensibilidad y especificidad se determinan para todos los valores de límite de riesgo en ese intervalo que se basa en las puntuaciones de los grupos de casos y controles (véase la tabla 2). Se elige el valor de límite de riesgo cuya sensibilidad y especificidad son más próximas a 0,87 y 0,84, respectivamente, y en este ejemplo, ese valor de límite de riesgo es de 56, con una sensibilidad de 0,88 y una especificidad de 0,84. Por lo tanto, como valor umbral para una prueba de evaluación de riesgos que utiliza los alelos asociados, se elige un valor de 56.
- 45 En otra realización de la presente invención, se puede elegir un valor umbral basándose en un resultado clínico deseado específico. Por ejemplo, se puede desarrollar una prueba de evaluación de riesgos para estratificar la población de pacientes como un medio para reducir la incidencia de acontecimientos adversos en individuos que reciben un tratamiento terapéutico en particular. Por ejemplo, un fármaco puede aprobarse para un uso limitado debido a una incidencia del 4 % de acontecimientos adversos, pero podría aprobarse para un uso más amplio si la incidencia de acontecimientos adversos se redujera a al menos un 50 %. En este ejemplo, los “casos” son individuos que tendrían el acontecimiento adverso en respuesta al fármaco y los “controles” serían los que no tendrían el acontecimiento adverso al exponerse al fármaco. El riesgo de que un individuo presente el acontecimiento adverso se determina calculando una puntuación para el individuo en función de sus genotipos en un conjunto de locus asociados y después, por ejemplo, comparando la puntuación con un valor umbral para una prueba de evaluación de riesgos, donde el umbral se determinó mediante el análisis de VPP, VPN, sensibilidad, especificidad, etc., o alguna combinación de los mismos para una prueba de evaluación de riesgos basada en las puntuaciones de un grupo de casos y un grupo de control. Por ejemplo, pueden identificarse individuos con un número de alelos asociados más alto que un valor umbral como en riesgo alto de tener el acontecimiento adverso. El uso de un valor umbral de 60 para los valores de ejemplo ilustrativos que se muestran en la tabla 1, eliminaría el 77 % de los casos y el 8 % de los controles. Dado que se sabe que la incidencia de acontecimientos adversos es del 4 %, una población de pacientes de 1000 tendría ~ 40 casos, de los cuales aproximadamente 31 (77 %) tendrían >60 alelos asociados y aproximadamente 9 de los cuales tendrían ≤60 alelos asociados. La misma población de pacientes tendría ~ 960 controles, aproximadamente 77 (8 %) de los cuales tendrían >60 alelos asociados y aproximadamente 883 de los

5 cuales tendrían  $\leq 60$  alelos asociados. Después de excluir a los 108 individuos con  $\leq 60$  alelos asociados, se puede calcular la incidencia de un acontecimiento adverso en los 892 individuos que no fueron excluidos:  $(9/892) \times 100 = 1\%$ . La incidencia de un acontecimiento adverso en los individuos que se excluyeron se puede calcular de manera similar:  $(31/108) \times 100 = 29\%$ . Utilizando los mismos métodos de cálculo, los valores de límite de riesgo de 59 y 61 también se evaluaron como valores umbral para la prueba de diagnóstico. Un valor de límite de riesgo de 59 dio como resultado una incidencia pronosticada del acontecimiento adverso en los individuos que no se excluyeron del tratamiento del 1 %, pero se excluyeron más individuos en el grupo de control (92), lo que significa que más individuos no corren el riesgo de acontecimiento adverso que se le negaría el tratamiento con el fármaco si ese valor de límite de riesgo se utilizó como un valor umbral en una prueba de diagnóstico. Un valor de límite de riesgo de 61 dio como resultado una incidencia pronosticada del acontecimiento adverso en los individuos que no se excluyeron del tratamiento del 1,2 %, que es mayor que la de un valor de límite de riesgo de 60, sin embargo, se excluyeron menos individuos de control (69), lo que significa que un mayor número de individuos que no corre el riesgo de sufrir el acontecimiento adverso podría beneficiarse del tratamiento con fármaco si este valor de límite de riesgo se utilizase como un valor umbral en una prueba de diagnóstico. Además, si un médico desea maximizar el número de controles que se trata mientras se mantiene el riesgo de acontecimientos adversos en o por debajo del 2 % en la población tratada, un valor umbral de 69 solo excluiría a 10 de los individuos de control y proporcionaría una población tratada con un riesgo del acontecimiento adverso al 2 %. Además, como se muestra en la tabla 2, la sensibilidad solo necesita ser de 0,53 para que la prueba identifique suficientes casos para la eliminación del grupo de pacientes a tratar para reducir el riesgo del acontecimiento adverso al 2 %. Por lo tanto, el uso del valor de límite de riesgo de 69 como un valor umbral en dicha prueba de diagnóstico, disminuirá la incidencia de acontecimientos adversos en la población de individuos tratados con el tratamiento terapéutico particular, mejorando así su perfil de riesgo/beneficio y permitiéndola ampliar su marcador, mientras se maximiza el número total de individuos que no están en riesgo de presentar el acontecimiento adverso que se incluirá en el tratamiento. Claramente, la elección de un riesgo de acontecimientos adversos particular en la población tratada es un factor importante en la determinación de un valor umbral para dicha prueba de diagnóstico, y la determinación de ese nivel de riesgo debe determinarla el médico junto con cualquier agencia reguladora (por ejemplo, la FDA) que estaría implicada en la aprobación de dicho diagnóstico. Por ejemplo, si se desease un riesgo del 1 % de presentar un acontecimiento adverso, se podría elegir un valor umbral de 60, lo que aumentaría el VPN de la prueba (reduciendo así el número real de acontecimientos adversos en la población tratada), mientras que se sacrificaba el VPP (se excluirían más individuos que podrían beneficiarse (controles)). Los pacientes que se excluyen pueden recibir un tratamiento diferente, por ejemplo, con un fármaco diferente, o pueden recibir el fármaco junto con un control cercano del acontecimiento adverso, o con otro tratamiento o agente que contrarreste el acontecimiento adverso.

Tabla 2

Valores de límite de riesgo	Número de casos (de 102)	Número de controles (de 405)	Sensibilidad	Especificidad	VPP	VPN	Precisión
69	54	4	0,53	0,99	0,93	0,89	0,90
61	74	29	0,73	0,93	0,72	0,93	0,89
60	79	34	0,77	0,92	0,70	0,94	0,89
59	80	39	0,78	0,90	0,67	0,94	0,88
58	81	44	0,79	0,89	0,65	0,95	0,87
57	86	53	0,84	0,87	0,62	0,96	0,86
56	90	64	0,88	0,84	0,58	0,97	0,85
55	93	81	0,91	0,80	0,53	0,97	0,82

35 Los conceptos de sensibilidad, especificidad, VPP, VPN, precisión, cocientes de verosimilitud y curvas ROC, y los métodos para elegir un valor umbral apropiado para una prueba de diagnóstico, son muy utilizados y son muy conocidos por los expertos en la materia (véase, por ejemplo, Janssens, et al. (2004) Am. J. Hum. Genet. 74:585-588; [www.bamc.amedd.army.mil/DCI/articles/dci10972.htm](http://www.bamc.amedd.army.mil/DCI/articles/dci10972.htm); Baum M. (1995) Lancet 346:436-437; Forrest P. (1990) "Breast Cancer: the decision to screen"; Nuffield Provincial Hospitals Trust; Morrison, A.S. (1985) "Screening in Chronic Disease" Oxford University Press Inc. USA; [www.genome.gov/10002404;med.usd.edu/som/genetics/curriculum/IIEST7.htm](http://www.genome.gov/10002404;med.usd.edu/som/genetics/curriculum/IIEST7.htm); Bauman A. (1990)

Australian Prescriber 13:62-64; Walker et al. (1986) Med. J. Aust. 145:185-187; Gilbert R. (2001) Western J. Med. 174:405-409; Frohna, J.G. (2001) "Fostering the Efficient, Effective Use of Evidence-based Medicine in the Clinic", 2ª edición, University of Michigan; Raglans, R.A. (2000) "Studying a Study and Testing a Test", 4ª edición, Lippincott Williams & Wilkins; [www.cebm.net/likelihood\\_ratios.asp](http://www.cebm.net/likelihood_ratios.asp); y [www1.elsevier.com/gej-ng/10/22/71/52/140/article.html](http://www1.elsevier.com/gej-ng/10/22/71/52/140/article.html)).

5 Por ejemplo, en un estudio, se evaluó el mejor valor umbral para la alfafetoproteína sérica para diferenciar entre cirrosis hepática y carcinoma hepatocelular, basándose en el área bajo la curva ROC, cocientes de verosimilitud, sensibilidad, especificidad, VPP y VPN (Soresi et al. (2003) Anticancer Res. 23(2C):1747-1753). En otro estudio se compararon la mamografía, la ecografía y la mamografía con RM para determinar si una o una combinación de 2 o  
10 más de estas técnicas proporcionaría los mejores resultados para la detección de cáncer invasivo y enfermedad multifocal, utilizando las medidas de sensibilidad, especificidad, VPP, VPN y precisión (Malur et al. (2001) Breast Cancer Res. 3:55-60). La combinación de las tres técnicas de obtención de imagen, condujo a los mejores resultados con una sensibilidad de 0,994, una especificidad de 0,953, un VPP de 0,939, un VPN de 0,996 y una precisión de 0,97. En otro estudio más, el área bajo las curvas ROC para dos pruebas clínicas, se comparó para determinar si una de las pruebas o una combinación de ambas pruebas era más precisa para identificar la clase de una lesión mamaria (Buscombe et al. (2001) J. Nuc. Med. 42(1):3-8). En otro estudio, se encontró que la prueba del  
15 antígeno prostático específico (PSA, *prostatic-specific antigen*) para detectar cáncer de próstata, tenía una sensibilidad de 0,86 y una especificidad de 0,33 para un límite de 4 ng/ml de PSA, pero eso reducía el límite a 2 ng/ml de PSA aumentando la sensibilidad a 0,95 y redujo la especificidad a 0,20 (Hoffman, et al. (2002) BMC Fam. Pract. 3(1): 19). Una vez que se examinaron todos los valores de límite de riesgo y se calcularon sus respectivas  
20 especificidades, sensibilidades, valores VPP, VPN, CV+ y CV-, y precisiones (o algún subconjunto de los mismos), se puede utilizar un equilibrio óptimo entre estos parámetros, o algún subconjunto de los mismos, en la determinación de un valor umbral. Un experto en la técnica puede elegir un valor umbral que optimice cualquiera de estas medidas, o una combinación de las mismas, para obtener un medio clínicamente útil para estratificar a una población de pacientes, por ejemplo, diagnóstico, pronóstico, farmacogenómica, desarrollo de fármacos, tratánóstico  
25 y similares.

En determinadas realizaciones de la presente invención, puede determinarse y utilizarse más de un valor umbral para clasificar el riesgo de un individuo de presentar un rasgo multifactorial y/o para determinar el tratamiento médico apropiado para el individuo. En una de dichas realizaciones, un primer valor umbral elegido puede basarse en la optimización de la sensibilidad, lo que reduciría el número de individuos que están en riesgo alto, pero que no  
30 son identificados por la prueba (negativos falsos). Los individuos que dan un valor "positivo" en una prueba de evaluación de riesgos utilizando el primer valor umbral se someten después a la misma prueba de evaluación de riesgos utilizando un segundo valor umbral que puede basarse en la optimización de la especificidad. Este segundo valor umbral reducirá el número de individuos que de un valor positivo pero que no están realmente en riesgo alto (positivos falsos). El uso secuencial de dos valores umbral de este tipo, puede servir para aumentar la precisión del  
35 método.

Otra realización de la presente invención en la que se determina y utiliza más de un valor umbral para clasificar el riesgo de un individuo de presentar un rasgo multifactorial, es aquella en la que se utiliza una pluralidad de valores de umbral simultáneamente en la misma prueba de evaluación de riesgos. En una prueba de este tipo, el riesgo de un individuo y/o el tratamiento médico posterior, se determinan en función de los valores umbral en los que la  
40 puntuación fue mayor que, menor que o igual a. En una realización se utiliza al menos aproximadamente dos umbrales, o se utiliza al menos aproximadamente cinco umbrales o al menos aproximadamente 10 umbrales. En determinadas realizaciones, cada puntuación posible para una prueba de evaluación de riesgos dada, se utiliza como umbral; en otras realizaciones, se utiliza un subconjunto de puntuaciones posibles, en el que dicho subconjunto puede abarcar un intervalo específico de puntuaciones o puede incluir puntuaciones elegidas de todo el  
45 intervalo de puntuaciones. Por ejemplo, se puede elegir un primer umbral de tal manera que los individuos que tienen una puntuación más alta que el primer umbral se clasifiquen como muy propensos a desarrollar una enfermedad, y, por lo tanto, se les trate con una intervención médica apropiada (por ejemplo, un fármaco, cirugía, etc.) para prevenir su aparición. Se puede elegir un segundo umbral de tal modo que los individuos que tienen una puntuación inferior al segundo umbral se clasifiquen como que tienen una probabilidad muy alta de desarrollar la  
50 enfermedad, y, por lo tanto, no se les trate para prevenir su aparición. Aquellos individuos con una puntuación que está entre el primer y el segundo umbral pueden clasificarse como que tienen una probabilidad intermedia de desarrollar la enfermedad y, por lo tanto, pueden tratarse de una manera diferente a la de los individuos con una puntuación mayor que el primer umbral o menor que el segundo umbral, por ejemplo, es posible que no se les administre el fármaco, pero se les puede controlar más de cerca para detectar la aparición de la enfermedad en caso  
55 de que aparezca o se les puede administrar el fármaco de una manera diferente (por ejemplo, diferentes dosis, frecuencia, etc.), o pueden tener una cirugía menos radical (por ejemplo, una lumpectomía en lugar de una mastectomía). El tratamiento de individuos con riesgo intermedio puede depender más de otra información, tal como información clínica, demográfica y/o epidemiológica sobre la enfermedad, prueba de evaluación de riesgos, fármaco, paciente, etc., que el tratamiento de individuos que no tienen un riesgo intermedio (es decir, están en riesgo "alto" o  
60 "bajo").

Aunque mediante un estudio de asociación puede identificarse un conjunto de locus asociados, no es necesario utilizar todos los locus asociados en una sola prueba de evaluación de riesgos. Una vez que se identifica un conjunto

de locus asociados, se puede ajustar el número de locus asociados que se utilizarán en una prueba de evaluación de riesgos y analizar el valor de la prueba, por ejemplo, con respecto a su sensibilidad, especificidad, riesgo relativo, cociente de verosimilitud, VPP, VPN, precisión o una combinación de los mismos. Por ejemplo, en determinadas realizaciones, se prefiere un riesgo relativo alto en combinación con una alta sensibilidad. En un aspecto, los métodos de la presente invención pueden utilizarse para determinar un subconjunto (por ejemplo, de al menos aproximadamente 5, 10, 15, 20, 30, 50) de locus asociados que se utilizarán en una prueba de evaluación de riesgos. Por ejemplo, se pueden seleccionar los locus asociados con las mayores diferencias de frecuencia alélica entre el grupo de casos y el de control. En algunas realizaciones, solo se eligen aquellos locus con diferencias de frecuencia alélica de al menos aproximadamente 8 % (0,08), 10 % (0,1), 15 % (0,15) o 25 % (0,25) para su uso en una prueba de evaluación de riesgos. En algunas realizaciones, el subconjunto de locus asociados a utilizar en una prueba de evaluación de riesgos se determina analizando determinadas características de la prueba de evaluación de riesgos resultante utilizando los datos de genotipificación de los grupos de casos y control. Por ejemplo, se puede determinar la sensibilidad, especificidad, riesgo relativo, cociente de verosimilitud, VPP, VPN, precisión, o una combinación de los mismos, para una prueba de evaluación de riesgos hipotética utilizando un subconjunto determinado de locus asociados. De esta manera, se puede analizar una pluralidad de dichas pruebas de evaluación de riesgos hipotéticas y se puede elegir el subconjunto de locus asociados que, en combinación, den como resultado la prueba de evaluación de riesgos con la mejor combinación de estas características. Al igual que ocurre en la determinación de un valor umbral apropiado, como se indicó anteriormente, la mejor combinación de sensibilidad, especificidad, riesgo relativo, cociente de verosimilitud, VPP, VPN, precisión, o un subconjunto de los mismos, para una prueba de evaluación de riesgos, depende de muchos factores clínicos que incluyen, por ejemplo, la gravedad del fenotipo, la prevalencia del fenotipo y de otra información clínica que es específica de la población o del paciente. En determinadas realizaciones, el subconjunto de locus asociados que se utilizará en una prueba de evaluación de riesgos se determina basándose en una combinación de las diferencias de frecuencias alélicas para los locus asociados y de las características de la prueba de evaluación de riesgos resultante. Por lo tanto, utilizando los métodos de la presente invención, se pueden predecir las características de una prueba de evaluación de riesgos utilizando un subconjunto de locus asociados sin tener que realizar un estudio de casos y controles utilizando solo ese subconjunto para medir dichas características.

Este aspecto de la presente invención tiene implicaciones prácticas importantes. Por ejemplo, si en un segundo estudio de asociación de validación, determinados locus asociados no se duplican, estos pueden eliminarse del conjunto de locus asociados que se utilizarán en una prueba de evaluación de riesgos, y las características de la prueba de evaluación de riesgos sin los locus “no duplicados” puede determinarse sin realizar otro estudio de asociación. Además, una prueba de evaluación de riesgos que requiere genotipificar una gran cantidad de locus es más costosa de realizar que una prueba de evaluación de riesgos que requiere genotipificar una pequeña cantidad de locus. Por lo tanto, la capacidad de reducir la cantidad de locus asociados en una prueba de evaluación de riesgos mientras que se conservan características específicas deseadas (por ejemplo, sensibilidad, riesgo relativo, etc.) tiene implicaciones directas para la asequibilidad de realizar dicha prueba y, por lo tanto, en la aplicabilidad práctica de dicha prueba.

Además, las pruebas de evaluación de riesgos que suponen que la contribución de cada alelo SNP asociado es aditiva y de magnitudes similares o diferentes del efecto fenotípico y que cada alelo SNP asociado es independiente de (es decir, se segrega independientemente de) todos los demás, se pueden utilizar para predecir el riesgo de desarrollar un rasgo multifactorial para una población a través de generaciones. En una de dichas realizaciones, los genotipos se determinan para un conjunto de SNP en individuos en una población de casos y una población de control, y las frecuencias alélicas de cada SNP en cada población se determinan. Suponiendo que los SNP estén en equilibrio de Hardy-Weinberg entre sí, las frecuencias de los genotipos en cada población se determinan en función de las frecuencias alélicas en cada población. Se puede demostrar una ilustración sencilla de este cálculo en un ejemplo en el que solo se está genotipificando un único SNP. Si el SNP tiene una frecuencia alélica minoritaria (C) de 0,3, la frecuencia del genotipo “CC” es de  $(0,3)^2$ , la frecuencia del genotipo “GG” sería de  $(0,7)^2$  y la frecuencia del genotipo “GC” sería de  $2(0,3)(0,7)$ . Por lo tanto, si uno de los alelos aumenta el riesgo de que un individuo desarrolle un rasgo fenotípico de interés, el riesgo para la población podría determinarse en función de las frecuencias de los genotipos CC, GG, CG en las poblaciones de casos y controles. Dicho riesgo se aplicaría no solo a las poblaciones inmediatas, sino también a las poblaciones posteriores, siempre y cuando las frecuencias alélicas fueran constantes entre generaciones, como es generalmente el caso de grandes poblaciones. En una prueba de evaluación de riesgos, se genotipifica una pluralidad de SNP en poblaciones de casos y controles y se generan dos conjuntos de genotipos. Si los SNP son independientes entre sí, entonces el conjunto de frecuencias alélicas para la pluralidad de SNP puede utilizarse para calcular un conjunto de frecuencias de genotipo para las dos poblaciones. Dado que cada genotipo tiene un rasgo particular asociado basado en el número de alelos asociados que contiene, se puede determinar un riesgo general para los individuos en la población en función de los riesgos que ofrece el conjunto de genotipos, y este riesgo es el mismo para los individuos en otras poblaciones (por ejemplo, generaciones posteriores) siempre que las frecuencias alélicas permanezcan relativamente constantes.

60 Identificación de individuos en riesgo de desarrollar una enfermedad multifactorial.

Una vez que se han determinado uno o más valores umbral, para determinar el riesgo de que el individuo desarrolle

o presente el rasgo multifactorial de interés se puede examinar a un individuo (“individuo de prueba”) que no sea un miembro de los grupos de casos o controles. En determinadas realizaciones de la presente invención, el individuo de prueba es de la misma especie que la de los individuos en los grupos de casos y controles. El individuo de prueba se genotipifica en cada uno de los locus de SNP asociados (o un subconjunto de los mismos, como se ha descrito anteriormente). Se calcula una puntuación para el individuo de prueba que se basa al menos en su genotipo en cada uno de los locus de SNP de la misma manera que se calcularon las puntuaciones para los individuos en los grupos originales de casos y controles (como se ha indicado anteriormente). En una realización de la presente invención, la puntuación calculada para el individuo de prueba se compara con uno o más valores umbral para determinar si es probable que ese individuo muestre o no el rasgo multifactorial (por ejemplo, una enfermedad). Por ejemplo, si un individuo de prueba tiene una puntuación mayor que un primer valor umbral, se puede considerar probable que un individuo de prueba desarrolle o presente la enfermedad, y si la puntuación del individuo de prueba es igual o menor que un segundo valor umbral se puede considerar que un individuo de prueba tiene un riesgo bajo de desarrollar la enfermedad. El primer y segundo valor umbral puede ser igual o diferente. Por ejemplo, en una realización en la que se elige 55 como primer y segundo umbral, entonces un individuo de prueba que tenga una puntuación mayor de 55 puede diagnosticarse como propenso a desarrollar la enfermedad y un individuo de prueba que tenga una puntuación de 55 o menor puede diagnosticarse como poco propenso a desarrollar la enfermedad. Además, basándose en la prevalencia de la enfermedad y en la sensibilidad y especificidad de la prueba de evaluación de riesgos, se puede calcular la probabilidad o posibilidad de que una persona identificada como de riesgo alto en la prueba realmente tenga o desarrolle la enfermedad (por ejemplo, probabilidades después de una prueba, como se analiza a continuación). Del mismo modo, se puede calcular la probabilidad o posibilidad de que una persona identificada como de riesgo bajo por la prueba en realidad no tenga ni desarrolle la enfermedad.

En otra realización de la presente invención, se calcula un riesgo relativo para un individuo de prueba para analizar adicionalmente la posibilidad de que el individuo desarrolle o muestre la enfermedad. El riesgo relativo es una medida de cuánto influye un factor de riesgo particular en el riesgo de un resultado específico. Por ejemplo, un riesgo relativo de 2 asociado a un factor de riesgo significa que las personas con ese factor de riesgo tienen un riesgo dos veces mayor de tener un resultado específico que las personas sin ese factor de riesgo. En un aspecto, un riesgo relativo para una enfermedad es un aumento de veces del riesgo en relación con el riesgo del rasgo (por ejemplo, enfermedad) en la población general. Un riesgo relativo se determina calculando la relación entre el porcentaje de individuos en el grupo de casos y el porcentaje de individuos en el grupo de control que cumplen o exceden una puntuación dada basándose en al menos sus genotipos en el conjunto de SNP que están asociados con el rasgo. Utilizando los datos presentados en la tabla 1, por ejemplo, el riesgo relativo de un individuo con una puntuación de al menos 65 es  $(0,64)/(0,02)=32$ , lo que significa que el individuo tiene un riesgo 32 veces mayor de desarrollar el rasgo. Para comparar, el riesgo relativo de un individuo con una puntuación de al menos 70 es  $(0,5)(0,005)=100$ , lo que significa que el individuo tiene un riesgo 100 veces mayor de desarrollar el rasgo. En una realización de la presente invención, se calcula una puntuación para un individuo de prueba y se analizan los grupos de casos y controles para determinar qué porcentaje de individuos del grupo de casos y qué porcentaje de individuos del grupo de control tienen una puntuación que sea al menos tan grande como la del individuo de prueba. A continuación, el porcentaje de casos individuales con una puntuación al menos tan grande como la del individuo de prueba se divide entre el porcentaje de individuos del grupo de control con una puntuación al menos tan grande como la del individuo de prueba para calcular el riesgo relativo para el individuo de prueba.

Como se indicó anteriormente, el riesgo relativo proporciona un aumento de veces del riesgo en relación con el riesgo de la enfermedad en la población general. Por lo tanto, para determinar el riesgo del individuo de prueba de desarrollar la enfermedad, el riesgo relativo para el individuo debe combinarse con información clínica con respecto a la prevalencia de la enfermedad. Por ejemplo, si la enfermedad tiene una prevalencia de 1:100, entonces un individuo con un riesgo relativo de 32 tiene una probabilidad de desarrollar la enfermedad de 32:100 o de 0,32. Sin embargo, para una enfermedad que tiene una prevalencia de 1:1.000.000 un individuo con un riesgo relativo de 32 tiene una probabilidad de desarrollar la enfermedad de 32:1.000.000 o 0,000032. Por lo tanto, aunque los riesgos relativos fueron los mismos en estos dos ejemplos, la probabilidad real de desarrollar la enfermedad fue muy frecuente para estas dos enfermedades. En determinadas realizaciones de la presente invención, el riesgo de un individuo de prueba de desarrollar un factor multifactorial de interés se calcula multiplicando el riesgo relativo determinado para el individuo por la prevalencia del rasgo multifactorial en la población general. La determinación del riesgo relativo es muy conocida y los expertos en la materia la realizan de manera habitual (véase Sackett, et al. (1991) *Clinical Epidemiology: a basic science for clinical medicine* (segunda edición) Little Brown, Boston).

Además, el VPP y el PNV de una prueba de evaluación de riesgos pueden proporcionar información sobre el riesgo de que un individuo tenga o desarrolle una enfermedad en función del resultado de la prueba. Por ejemplo, si un individuo da “positivo” en la prueba para la enfermedad utilizando una prueba con un VPP de 0,87 y un VPN de 0,99, entonces el individuo tiene un 87 % de probabilidad de tener o desarrollar la enfermedad. Del mismo modo, si otro individuo da “negativo” en la prueba para la enfermedad utilizando la misma prueba, entonces ese individuo tiene solo un 1 % de probabilidad de tener o desarrollar la enfermedad.

Los cocientes de verosimilitud utilizan la sensibilidad y especificidad de una prueba para proporcionar una media de cuánto cambia un resultado de prueba particular la probabilidad de que un paciente tenga o no un rasgo

multifactorial de interés, como se indicó anteriormente. El cociente de verosimilitud (CV) de un resultado de prueba positivo (CV+) se calcula como la sensibilidad dividida entre (1- especificidad) y el (CV) de un resultado de prueba negativo (CV-) se calcula como (1- sensibilidad) dividido entre la especificidad. Estos valores de CV se multiplican por las probabilidades previas a la prueba para calcular las probabilidades posteriores a la prueba, lo que representa las posibilidades de que el individuo tenga o desarrolle el rasgo multifactorial al incorporar información sobre la prevalencia de la enfermedad, el grupo de pacientes y los factores de riesgo específicos del paciente (probabilidades previas a la prueba) e información sobre la propia prueba de diagnóstico (CV). Las probabilidades posteriores a la prueba se pueden utilizar para calcular la probabilidad posterior a la prueba dividiendo las probabilidades posteriores a la prueba entre (1 + probabilidades posteriores a la prueba). Por ejemplo, si un individuo que da positivo tiene una probabilidad previa a la prueba de 1 a 66 basada en una prevalencia de 1,5 % y la prueba tiene un CV+ de 6,6, entonces las probabilidades después de la prueba serán de 0,1 y la probabilidad posterior a la prueba será de 0,09 lo que significa que el individuo tiene un 9 % de probabilidad de tener la enfermedad. Del mismo modo, si un individuo que da negativo tiene una probabilidad previa a la prueba de 1 a 3 y la prueba tiene un CV de 0,09, entonces las probabilidades posteriores a la prueba serán de 0,03, lo que corresponde a una probabilidad posterior a la prueba del 3 % de que el individuo tenga la enfermedad. De esta forma, los cocientes de verosimilitud y la prevalencia del rasgo multifactorial se pueden utilizar para calcular la probabilidad de que un individuo tenga o desarrolle un rasgo multifactorial de interés basándose en un resultado de prueba dado.

#### Usos de pronóstico y diagnóstico

Las medidas preventivas son satisfactorias para prevenir muchas enfermedades diferentes, pero estas medidas solo son satisfactorias si se puede identificar a los individuos que corren un riesgo de desarrollar la enfermedad antes de que aparezca la misma. La aparición de enfermedades multifactoriales es especialmente difícil de predecir debido al complejo conjunto de factores que influyen en su desarrollo. Como tal, los individuos a menudo no saben que corren el riesgo de desarrollar una enfermedad multifactorial hasta que es demasiado tarde para prevenirla. Para un experto en la técnica será obvio que los métodos presentados pueden servir como herramientas valiosas para que los médicos tomen decisiones médicas con respecto a la atención de sus pacientes. La determinación del riesgo es un aspecto importante del análisis clínico de un individuo utilizado para determinar si las intervenciones médicas están justificadas o no y cuáles son las intervenciones más apropiadas para un individuo determinado (Bucher, et al. (1994) BMJ 309(6957):761-764; Forrow, et al. (1992) Am JMed 92(2)121-124).

En determinadas realizaciones, la presente invención proporciona métodos para identificar individuos con riesgo de desarrollar una enfermedad (pronóstico), permitiendo así la implementación de medidas para prevenir o retrasar la aparición de la enfermedad. En una realización, el riesgo del individuo de desarrollar una enfermedad dada se puede determinar comparando una puntuación basada al menos en el genotipo del individuo en un conjunto de SNP asociados a la enfermedad con al menos un valor umbral. Si la puntuación del individuo supera un valor umbral, la institución de medidas preventivas (por ejemplo, radioterapia o farmacoterapia) puede estar justificada. En otra realización, el riesgo de un individuo de desarrollar una enfermedad puede determinarse calculando un riesgo relativo para el individuo y multiplicando el riesgo relativo por la prevalencia de la enfermedad. En otra realización, la sensibilidad, especificidad, VPP, VPN y/o precisión de una prueba de evaluación de riesgos se utiliza para calcular el riesgo de un individuo de desarrollar la enfermedad. En otra realización más, el CV para la prueba se utiliza para calcular las probabilidades/posibilidades posteriores a la prueba de que el individuo desarrolle la enfermedad. En otra realización, se utiliza una combinación de los métodos descritos anteriormente para determinar el riesgo de un individuo de tener o desarrollar la enfermedad. Esta información puede utilizarla un médico para determinar mejor un régimen de tratamiento apropiado para el individuo. A menudo, esta información se utiliza en combinación con información clínica, demográfica y/o epidemiológica sobre la enfermedad, el paciente, o la población de la que proviene el paciente. En algunos aspectos, los métodos presentados en este documento también se pueden utilizar para identificar individuos que son resistentes a una enfermedad. Por ejemplo, algunos individuos que tienen antecedentes familiares de una enfermedad (por ejemplo, cáncer de mama) nunca desarrollan la enfermedad. Este conocimiento podría evaluar mejor el riesgo de estos individuos de desarrollar la enfermedad en cuestión, proporcionando tranquilidad a quienes no están en riesgo alto y, en algunos casos, impediría tratamientos profilácticos drásticos (por ejemplo, mastectomía electiva). Los métodos presentados en este documento también se pueden utilizar para identificar individuos con un mayor riesgo de desarrollar una afección adversa no relacionada con la enfermedad y por lo tanto motivar cambios en el estilo de vida para prevenir la aparición de la afección. Por ejemplo, la prueba de evaluación de riesgos que comprende un conjunto un SNP asociados con hipertensión, podría proporcionar un fuerte incentivo para aquellas personas que se encuentren en riesgo alto de hacer ejercicio y llevar una dieta saludable.

Algunas enfermedades son difíciles de diagnosticar basándose únicamente en los síntomas físicos aparentes de un paciente. El diagnóstico de estas enfermedades a menudo se confunde por la variedad de formas en las que dicha enfermedad puede en sí misma manifestarse en diferentes individuos, y/o por el hecho de que sus síntomas pueden ser similares a los de diversas enfermedades no relacionadas. En un aspecto adicional de la presente invención, se puede utilizar conjunto de SNP asociados con dicha enfermedad para ayudar en el diagnóstico de un individuo que presenta un fenotipo que puede ser indicativo de la enfermedad. Por lo tanto, la genotipificación del individuo para el conjunto de SNP asociados y la determinación del riesgo del individuo de presentar la enfermedad, podría confirmar

o argumentar en contra del diagnóstico sugerido por los síntomas físicos. Si el diagnóstico fue confirmado, un médico podría utilizar esta información para tomar decisiones de tratamiento para el individuo, tal como iniciar un régimen de tratamiento para la enfermedad. Por ejemplo, la enfermedad celiaca es un trastorno autoinmunitario del sistema digestivo que daña al intestino delgado e interfiere con la absorción de nutrientes de los alimentos.

5 Específicamente, la enfermedad celiaca causa una respuesta inflamatoria en el intestino delgado en respuesta al gluten, una proteína que se encuentra en el trigo, centeno y cebada, y el único tratamiento para la enfermedad celiaca es una dieta sin gluten. Es difícil diagnosticar la enfermedad celiaca porque diferentes individuos muestran diferentes síntomas. Por ejemplo, algunos tendrán principalmente síntomas gastrointestinales, tales como distensión abdominal o diarrea, mientras que otros solo tendrán irritabilidad o depresión. Además, la afección puede diagnosticarse fácilmente de manera errónea porque sus síntomas son similares a otros muchos trastornos, tales como síndrome del intestino irritable, enfermedad de Crohn, colitis ulcerosa, diverticulitis, infecciones intestinales, síndrome de fatiga crónica y depresión. Los métodos presentados en este documento pueden utilizarse para identificar un conjunto de locus genéticos asociados a la enfermedad celiaca, y estos locus pueden utilizarse para explorar individuos que muestren síntomas indicativos de enfermedad celiaca. Aquellos individuos que se encuentran en riesgo alto de desarrollar enfermedad celiaca en función de su composición genética pueden ser diagnosticados con enfermedad celiaca y se les puede aplicar una dieta sin gluten.

En otras realizaciones, los métodos presentados en este documento pueden utilizarse para ayudar a determinar si una terapia profiláctica está justificada o no para prevenir el desarrollo de, por ejemplo, una enfermedad en un individuo. Por ejemplo, existen terapias aprobadas para la prevención del cáncer de mama que dependen de información del historial clínico, tal como antecedentes familiares, el inicio del primer periodo menstrual, el número de hijos, etc. Estos factores, aunque son útiles para calcular las probabilidades previas a la prueba, solos son marginalmente predictivos de si una mujer desarrollará o no cáncer de mama. Una prueba de evaluación de riesgos que se utilizará en combinación con las probabilidades previas a la prueba proporcionaría un medio muy superior de decidir si se trata o no profilácticamente a un individuo (por ejemplo, con tamoxifeno) al proporcionar una manera mucho más precisa de identificar y cuantificar su riesgo de desarrollar cáncer de mama.

En el presente documento también se describe un ensayo de pronóstico o diagnóstico que comprende una matriz de ácidos nucleicos que contiene sondas diseñadas para detectar la presencia del conjunto de SNP asociados en una muestra biológica. Los ácidos nucleicos se aíslan de una muestra biológica de un individuo de prueba y se hibridan con las sondas en la matriz de ácidos nucleicos. Las intensidades de la sonda se analizan para proporcionar un genotipo para el individuo de prueba en cada una de las posiciones de SNP asociadas. Los genotipos se utilizan para calcular una puntuación para el individuo de la prueba y el riesgo del individuo de desarrollar la enfermedad se determina de acuerdo con los métodos presentados en este documento.

El conjunto de SNP asociados se puede utilizar además para identificar regiones del genoma que están implicadas en el desarrollo del fenotipo de la enfermedad. Estos SNP pueden estar directamente implicados en la manifestación de la enfermedad, o pueden estar en desequilibrio de ligamiento con locus que están directamente implicados. Por ejemplo, un SNP asociado a una enfermedad puede influir directamente en la expresión o función de una proteína asociada a la enfermedad, o puede estar en desequilibrio de ligamiento con otro locus que influye en la expresión o función de la proteína. Los ejemplos de efectos directos a la expresión o función de una proteína incluyen, pero sin limitación, un polimorfismo que altera la secuencia polipeptídica de la proteína y un polimorfismo que se produce en una región reguladora (es decir, promotor, potenciador, etc.) dando como resultado la expresión aumentada o disminuida de la proteína. En ciertas realizaciones, las regiones genómicas que contienen el conjunto de SNP asociados se analizan para identificar genes que están directamente implicados en la base biológica de la enfermedad ("genes identificados").

Los SNP asociados que se encuentran en región codificante de un gen pueden utilizarse para detectar o cuantificar la expresión de un alelo asociado en una muestra biológica para su uso como un marcador de diagnóstico para la enfermedad. Por ejemplo, los ácidos nucleicos que contienen los SNP asociados se pueden utilizar como sondas oligonucleotídicas para controlar los niveles de ARN o ARNm en el organismo a analizar o una parte del mismo, tal como un tejido u órgano específico, para determinar si el gen que codifica el ARN o ARNm contiene un alelo asociado. En un aspecto, se proporciona un kit de diagnóstico o pronóstico que comprende sondas oligonucleotídicas para su uso en la detección de un alelo asociado en una muestra biológica. Del mismo modo, si el alelo asociado causa un cambio en la secuencia polipeptídica de la proteína codificada, la constitución alélica del gen puede analizarse a nivel de proteína utilizando cualquier técnica habitual, tal como métodos inmunológicos (por ejemplo, transferencias Western, radioinmunoprecipitación y similares) o ensayos basados en actividad que miden una actividad asociada con el producto génico. En un aspecto, se proporciona un kit de diagnóstico o pronóstico que comprende un ensayo para detectar un polipéptido codificado por un alelo asociado en una muestra biológica. La manera en la que se exploran las células para detectar la presencia de secuencias de nucleótidos o polipéptidos particulares, está bien establecida en la bibliografía y no requiere una elaboración adicional en el presente documento, sin embargo, véase, por ejemplo, Sambrook, et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Nueva York) (2001).

60

Terapias

El conjunto de SNP asociados puede ser útil para desarrollar terapias para la prevención de enfermedades. En un aspecto, los genes identificados pueden utilizarse para terapia génica. Por ejemplo, si se encuentra que un gen identificado está regulado negativamente en individuos que presentan la enfermedad, entonces la regulación positiva del gen podría ser una estrategia eficaz para prevenir la aparición de una enfermedad en individuos de prueba. La regulación positiva del gen identificado puede realizarse incorporando un alelo del gen que no esté asociado a la enfermedad en un vector de expresión e introduciendo adicionalmente el vector en un organismo, regulando así la expresión del gen en el organismo. Dichos vectores generalmente tienen sitios de restricción convenientes localizados cerca de la secuencia promotora para proporcionar la inserción de secuencias de ácido nucleico en un genoma receptor. Se pueden preparar casetes de transcripción que comprendan una región de inicio de la transcripción, el gen diana o un fragmento de mismo, y una región de terminación de la transcripción. Los casetes de transcripción se pueden introducir en una variedad de vectores, por ejemplo, plásmido; retrovirus, por ejemplo, lentivirus; adenovirus y similares, donde los vectores pueden mantenerse transitoria o establemente en las células. El gen o producto proteico se puede introducir directamente en los tejidos o células del hospedador por diversas vías, entre las que se incluyen infección vírica, microinyección o fusión de vesículas. La inyección a chorro también puede utilizarse para la administración intramuscular, como se describe en Furth, et al., *Anal. Biochem*, 205: 365-68 (1992). Como alternativa, el ADN puede recubrirse con micropartículas de oro y administrarse por vía intradérmica mediante un dispositivo de bombardeo de partículas o "pistola de genes" como se describe en la bibliografía (véase, por ejemplo, Tang, et al., *Nature*, 356: 152-54 (1992)).

Las proteínas codificadas por los genes identificados pueden ser dianas para la terapia de anticuerpos si hay un cambio de aminoácido en la secuencia de la proteína que está asociada con una predisposición a la enfermedad. Por ejemplo, si un alelo asociado codifica una variante de proteína que es un factor causante de la enfermedad, los anticuerpos específicos para la variante de proteína asociada a la enfermedad pueden administrarse a un paciente como un medio para inhibir el desarrollo de la enfermedad. En determinadas realizaciones, para prevenir la aparición de una enfermedad, a un paciente se le puede administrar una combinación de anticuerpos, cada uno específico de una proteína asociada a una enfermedad diferente.

Las moléculas antisentido pueden utilizarse para regular negativamente la expresión de un alelo asociado de un gen identificado en las células. Una molécula antisentido forma un dúplex con el ARNm codificado por un alelo de un gen, regulando por lo tanto su expresión y bloqueando la traducción de la proteína correspondiente. Por ejemplo, se puede desarrollar un reactivo antisentido basándose en la secuencia del ARNm codificado por un alelo asociado. Este agente antisentido puede después administrarse a un paciente heterocigoto (que posee un alelo asociado y un alelo que no está asociado con la enfermedad) para disminuir la expresión del alelo asociado permitiendo que predomine la expresión del alelo no asociado. El reactivo antisentido puede ser oligonucleótidos antisentido, particularmente oligonucleótidos antisentido sintéticos que tienen modificaciones químicas, o construcciones de ácido nucleico que expresan dichas moléculas antisentido como ARN. Se puede administrar una combinación de moléculas antisentido, donde una combinación puede comprender múltiples secuencias diferentes.

Como una alternativa a los inhibidores antisentido, pueden utilizarse compuestos de ácido nucleico catalíticos, por ejemplo, ribozimas, conjugados antisentido, etc., para inhibir la expresión de alelos asociados. Las ribozimas pueden sintetizarse *in vitro* y se administran al paciente o pueden estar codificadas en un vector de expresión a partir del cual se sintetiza la ribozima en una célula diana (por ejemplo, véase la solicitud de patente internacional WO 9523225, y Beigelman, et al., *Nucl. Acids Res.* 23: 4434-42 (1995)). En el documento WO 9506764 se describen ejemplos de oligonucleótidos con actividad catalítica. En Bashkin, et al., *Appl. Biochem. Biotechnol.* 54: 43-56 (1995) se describen conjugados de oligonucleótidos antisentido con un complejo metálico, por ejemplo, terpiridilCu(II), capaces de mediar la hidrólisis del ARNm.

En ensayos de exploración de fármacos para identificar ligandos o sustratos que se unen, modulan o imitan la acción de ese producto proteico, pueden utilizarse una proteína expresada codificada por un gen identificado y de este modo identificar agentes terapéuticos para proporcionar, por ejemplo, un reemplazo o mejora de la función proteica en las células afectadas, o un agente que module o anule la función de la proteína. Para este propósito puede utilizarse una amplia variedad de ensayos incluidos los ensayos realizados *in vitro* de unión entre proteínas marcadas, ensayos de unión entre proteína y ADN, ensayos de cambio de movilidad electroforética, inmunoensayos de unión a proteínas y ensayos similares. El término "agente", como se usa en el presente documento, describe cualquier molécula, por ejemplo, una proteína o una molécula pequeña, con capacidad de alterar, imitar o enmascarar, directa o indirectamente, la función fisiológica de un gen o producto génico identificado. En general, se realizan varios ensayos en paralelo con diferentes concentraciones del agente para obtener una respuesta diferencial a las diversas concentraciones. Típicamente, una de estas concentraciones sirve como control negativo, por ejemplo, una concentración de cero o por debajo del nivel de detección. Además, todo o un fragmento de una variante de proteína purificada puede utilizarse para la determinación de la estructura cristalina tridimensional, que puede utilizarse para determinar la función biológica de la proteína o una parte de la misma, modelar interacciones intermoleculares, fusión de membrana, etc.

Los agentes candidatos abarcan numerosas clases químicas, aunque típicamente son moléculas orgánicas o complejos, preferentemente compuestos orgánicos pequeños, que tienen peso molecular mayor de 50 y menor que aproximadamente 2.500 daltons. Los agentes candidatos comprenden grupos funcionales necesarios para la interacción estructural con proteínas, particularmente enlaces de hidrógeno, e incluyen típicamente al menos un grupo amina, carbonilo, hidroxilo o carboxilo y frecuentemente al menos dos de los grupos químicos funcionales. Los agentes candidatos a menudo comprenden estructuras cíclicas de carbono o heterocíclicas y/o estructuras aromáticas o poliaromáticas sustituidas con uno o más de los grupos funcionales anteriores. Los agentes candidatos también se encuentran entre las biomoléculas que incluyen, pero sin limitación: péptidos, sacáridos, ácidos grasos, esteroides, purinas, pirimidinas, derivados, análogos estructurales o combinaciones de los mismos.

Los agentes candidatos se obtienen a partir de una amplia variedad de fuentes que incluyen bibliotecas de compuestos sintéticos o naturales. Por ejemplo, se dispone de numerosos medios para la síntesis aleatoria y dirigida de una amplia variedad de compuestos orgánicos y biomoléculas, que incluyen la expresión de oligonucleótidos y oligopéptidos aleatorizados. Como alternativa, las bibliotecas de compuestos naturales en forma de extractos bacterianos, fúngicos, vegetales y animales, están disponibles o se producen fácilmente. Adicionalmente, las bibliotecas y compuestos producidos de manera natural o sintética se modifican fácilmente a través de medios químicos, físicos y bioquímicos convencionales y se pueden utilizar para producir bibliotecas combinatorias. Los agentes farmacológicos conocidos pueden someterse a modificaciones químicas dirigidas o aleatorias, tales como acilación, alquilación, esterificación, amidificación, etc., para producir análogos estructurales.

Cuando el ensayo de exploración es un ensayo de unión, una o más de las moléculas se pueden acoplar a un marcador, pudiendo dicho marcador proporcionar directa o indirectamente una señal detectable. Varios marcadores incluyen radioisótopos, fluorescentes, quimiluminiscentes, enzimas, moléculas de unión específica, partículas, por ejemplo, partículas magnéticas y similares. Las moléculas de unión específica incluyen pares, tales como biotina y estreptavidina, digoxina y antidigoxina, etc. Para los miembros de unión específica, el miembro complementario normalmente se marcaría con una molécula que proporciona detección de acuerdo con procedimientos conocidos.

En el ensayo de exploración puede incluirse una variedad de otros reactivos. Estos incluyen reactivos tales como sales, proteínas neutras, por ejemplo, albumina, detergentes, etc., que se utilizan para facilitar la unión óptima entre proteínas y/o reducir las interacciones no específicas o de fondo. Pueden utilizarse reactivos que mejoren la eficacia del ensayo, tales como inhibidores de proteasa, inhibidores de nucleasa, agentes antimicrobianos, etc.

Los agentes se pueden combinar con un vehículo o diluyente farmacéuticamente aceptable, que incluye cualquiera y todos los disolventes, medios de dispersión, recubrimientos, agentes antioxidantes, isotónicos y retardadores de absorción y similares. El agente puede combinarse con aditivos convencionales, tales como lactosa, manitol, almidón de maíz o de patata; con aglutinantes, tales como celulosa cristalina, derivados de celulosa, goma arábiga, almidón de maíz o gelatinas; con desintegrantes, tales como almidón de maíz, almidón de patata o carboximetilcelulosa de sodio; con lubricantes, tales como talco o estearato de magnesio; y si se desea, con agentes taponantes, agentes humectantes, conservantes y agentes aromatizantes. El uso de dichos medios y agentes para sustancias farmacéuticamente activas es muy conocido en la técnica y está fácilmente disponible al público. Además, las sustancias auxiliares farmacéuticamente aceptables, tales como agentes ajustadores y tamponadores de pH, agentes ajustadores de tonicidad, estabilizadores, agentes humectantes y similares, están fácilmente disponibles al público. Excepto en la medida en que cualquier medio o agente convencional sea incompatible con el ingrediente activo, se contempla su uso en las composiciones terapéuticas y métodos descritos en este documento. En las composiciones también pueden incorporarse ingredientes activos complementarios.

Los siguientes métodos y excipientes son meramente ejemplares y de ninguna manera limitativos. Los agentes identificados de la divulgación se pueden incorporar en una variedad de formulaciones para la administración terapéutica. Más particularmente, los complejos pueden formularse en composiciones farmacéuticas por combinación con vehículos o diluyentes farmacéuticamente aceptables, apropiados, tal como se indicó anteriormente y pueden formularse en preparaciones en formas sólidas, semisólidas, líquidas o gaseosas, tales como comprimidos, cápsulas, polvos, gránulos, pomadas, soluciones, geles, microesferas y aerosoles. Adicionalmente, los agentes se pueden formular en preparaciones para inyecciones disolviéndolas, suspendiéndolas o emulsionándolas en un disolvente acuoso o no acuoso, tal como aceites vegetales u otros aceites similares, glicéridos de ácido alifático sintético, ésteres de ácidos alifáticos superiores o propilenglicol; y si se desea, con aditivos convencionales tales como agentes solubilizantes, isotónicos, de suspensión, emulsionantes, estabilizantes y conservantes. Además, los agentes se pueden utilizar en una formulación en aerosol para administrar por inhalación. Los agentes identificados por los métodos presentados en este documento pueden formularse en propulsores presurizados aceptables tales como diclorofluorometano, propano, nitrógeno y similares. Como alternativa, los agentes pueden convertirse en supositorios para administración rectal, mezclando con una variedad de bases tales como bases emulsionantes o solubles en agua y pueden incluir vehículos tales como manteca de cacao, carboceras y polietilenglicoles, que se funden a la temperatura corporal, pero que son sólidos a temperatura ambiente.

Los implantes para formulaciones de liberación sostenida son muy conocidos en la técnica. Los implantes se formulan como microesferas, tabletas, etc., con polímeros biodegradables o no biodegradables. Por ejemplo, los

polímeros de ácido láctico y/o ácido glicólico forman un polímero erosionable que es bien tolerado por el hospedador. El implante que contienen los agentes identificados puede colocarse cerca del sitio de acción, de modo que la concentración local de agente activo aumenta con respecto al resto del cuerpo. Se puede proporcionar formas de dosificación unitarias para administración oral o rectal tales como jarabes, elixires y suspensiones en las que  
 5 cada unidad de dosificación, por ejemplo, cucharadita, cucharada, cápsula de gel, comprimido o supositorio, contiene una cantidad predeterminada de las composiciones de la presente divulgación. De manera similar, las formas de dosificación unitarias para inyección o administración intravenosa pueden comprender el compuesto de la presente divulgación en una composición como una solución en agua estéril, solución salina normal u otro vehículo farmacéuticamente aceptable. Las especificidades para las nuevas formas de dosificación unitarias dependen del  
 10 compuesto particular empleado y del efecto a obtener y la farmacodinámica asociada con cada agente activo en el hospedador.

La administración de los agentes se puede obtener de varias maneras. La formulación puede administrarse por vía oral, por inhalación, o puede inyectarse, por ejemplo, inyección intravascular, intratumoral, subcutánea, intraperitoneal, muscular, etc. Los agentes pueden ser tópicos, sistémicos o pueden estar localizados mediante el uso de un implante que actúa reteniendo la dosis activa en el sitio del implante. La dosificación de la formulación terapéutica variará, dependiendo del agente específico y de la formulación utilizada, de la naturaleza de la enfermedad, la frecuencia de la administración, la forma en la que se administra, la eliminación del agente del hospedador y similares, de tal manera que sea suficiente para tratar la enfermedad o sus síntomas, mientras que se minimizan los efectos secundarios. En algunos casos, la administración oral requerirá una dosis diferente que si se  
 15 administra por vía intravenosa. Los compuestos se administrarán a una dosificación eficaz de tal manera que, durante un periodo de tiempo adecuado, la progresión de la enfermedad pueda detenerse sustancialmente. La dosis inicial puede ser mayor, seguida de una dosis de mantenimiento más pequeña. La dosis puede administrarse ocasionalmente como una vez, semanalmente o cada dos semanas o fraccionarse en dosis más pequeñas y administrarse una vez al día, dos veces por semana, etc., para mantener un nivel de dosis eficaz. El tratamiento puede ser durante cortos periodos de tiempo, por ejemplo, después de la fibrilación ventricular, o durante periodos de tiempo prolongados, por ejemplo, en la prevención de nuevos episodios de fibrilación ventricular. Se contempla que la composición se obtenga y utilice siguiendo las orientaciones de un médico para su uso *in vivo*.

#### Farmacogenómica

En otras realizaciones, el conjunto de SNP asociados, identificados por los métodos de la presente divulgación se usan para la farmacogenómica y desarrollo de fármacos. Debido a la gran cantidad de opciones de tratamiento disponibles para las enfermedades multifactoriales comunes, a menudo es difícil determinar cuál de un grupo de opciones de tratamiento será más eficaz para un paciente determinado. Por lo general, se pueden probar varias opciones diferentes antes de encontrar una que sea segura y eficaz. Mientras tanto, el paciente continuará padeciendo los efectos de la enfermedad, y quizá también experimente acontecimientos adversos en respuesta a una más de las opciones de tratamiento probadas. Los métodos presentados en este documento son útiles para  
 30 estratificar poblaciones de pacientes antes de iniciar un régimen de tratamiento o un ensayo clínico. Se identifican locus polimórficos que están asociados a la respuesta de un paciente o un fármaco u otro tratamiento médico. La respuesta puede ser un acontecimiento adverso o puede estar relacionada con la eficacia del tratamiento. Los locus asociados se utilizan para explorar poblaciones de pacientes para generar perfiles de evaluación de riesgo (por ejemplo, puntuaciones, información fenotípica, información demográfica, información clínica, información epidemiológica, factores ambientales, datos de expresión génica, información sobre la afección médica, información sobre el individuo, información sobre posibles opciones de tratamiento, datos aportados del individuo, datos aportados por una agencia reguladora, o una combinación de los mismos) relacionados con los locus asociados a los pacientes que ayudarán a los médicos a determinar a qué individuos se les debe dar un fármaco o tratamiento  
 35 médico y a cuáles no, cómo se puede modificar dicho tratamiento para obtener una mejor respuesta en el paciente, o si incluir o no a un individuo en un ensayo clínico. Por ejemplo, los individuos que están predispuestos a presentar un acontecimiento adverso y los que es poco probable que muestren una respuesta eficaz a un fármaco pueden excluirse del tratamiento con ese fármaco, y en su lugar pueden tratarse por medios alternativos (un fármaco diferente u otro tratamiento médico) o tratarse alterando la administración del fármaco (por ejemplo, diferente dosificación, frecuencia de administración, etc.).  
 40  
 45  
 50

En una de dichas realizaciones, los individuos se exploran con respecto a un conjunto de SNP que están asociados a una enfermedad que confiere un riesgo conocido de una respuesta adversa a un tratamiento farmacológico particular. Aquellos individuos con riesgo alto de desarrollar la enfermedad están excluidos del régimen de tratamiento. Por ejemplo, los individuos con SQT (síndrome de QT largo), tienen un riesgo alto de fibrilación ventricular cuando reciben administración de fármacos antiarrítmicos. Sería beneficioso explorar una población de pacientes para un conjunto de locus asociados con SQT antes de administrar dicho fármaco y excluir a aquellos individuos con riesgo alto de desarrollar SQT. El conjunto de SNP asociados con la enfermedad se determina realizando un estudio de asociación, y un riesgo del individuo de desarrollar la enfermedad se realiza tal como se describió anteriormente. Un riesgo alto de desarrollar la enfermedad puede considerarse como un factor de riesgo de acontecimientos adversos en respuesta a un fármaco antiarrítmico y esta información puede utilizarla un médico para determinar las opciones de tratamiento apropiadas para el individuo. Por ejemplo, si un individuo tiene un  
 55  
 60

riesgo alto de desarrollar la enfermedad, entonces se puede impedir la administración del fármaco. Si el individuo tiene un riesgo bajo de desarrollar la enfermedad, la administración del fármaco puede ser una opción de tratamiento viable.

5 En otra realización de la presente invención, la eficacia de un régimen de tratamiento farmacológico se predice en un individuo basándose en los genotipos del individuo en un conjunto de SNP asociados con la eficacia del fármaco. Esta información se utiliza para determinar la probabilidad de que el fármaco sea un tratamiento eficaz para el individuo, un régimen de tratamiento apropiado (por ejemplo, dosificación, etc.) para el individuo, o si en su lugar se deben tener en cuenta otros fármacos u opciones de tratamiento. Por ejemplo, se puede realizar un estudio de asociación utilizando un grupo de casos de individuos que no tienen una respuesta eficaz al fármaco (“no respondedores”) y un grupo de control de individuos que tienen una respuesta eficaz (“respondedores”). Los miembros de los grupos de casos y control se genotifican en una pluralidad de posiciones de SNP, las frecuencias alélicas relativas se calculan para cada uno de los SNP y se identifica un conjunto de SNP asociado con una respuesta eficaz como aquellos SNP que tienen diferencias de frecuencia alélica que son significativamente diferentes entre los grupos de casos y controles. Para cada miembro de los grupos de casos y controles se calcula una puntuación según sus genotipos en los SNP asociados, y estas puntuaciones se utilizan para determinar uno o más valores umbral apropiados para una prueba de evaluación de riesgos que predecirá el riesgo de que un individuo no tenga una respuesta eficaz al fármaco. La determinación de un valor umbral apropiado también puede incluir uno o más de lo siguiente: conocimiento clínico, demográfico y/o epidemiológico del fármaco, la indicación que se está tratando y la población del paciente así como el cálculo de la sensibilidad, especificidad, VPP, VPN, precisión, CV+ y CV- de la prueba de evaluación de riesgos. Un individuo que es candidato para recibir el fármaco se genotifica en cada una de las posiciones de SNP asociadas, y se calcula una puntuación para el individuo en función de al menos sus genotipos en el conjunto de SNP asociados. Si el individuo tiene una puntuación que es mayor que un valor umbral, el individuo puede clasificarse como no respondedor y se pueden considerar tratamientos alternativos. Si el individuo tiene una puntuación igual o menor que un valor umbral, el individuo puede clasificarse como probablemente respondedor y se puede recomendar la administración del fármaco. En otra realización, el riesgo de un individuo de ser no respondedor puede determinarse calculando un riesgo relativo para el individuo y multiplicando el riesgo relativo por la prevalencia de los no respondedores basándose en la eficacia conocida del fármaco. En otra realización, la probabilidad del individuo de ser un respondedor se calcula utilizando la precisión, el CV+, CV-, VPP y/o VPN de la prueba de evaluación de riesgos. Después, esta información puede utilizarse un médico para decidir sobre los tratamientos apropiados para el individuo.

En una realización relacionada, se puede desarrollar un diagnóstico para una zona terapéutica para permitir que un médico individualice mejor el tratamiento de los pacientes. En lugar de centrarse en un solo fármaco, el diagnóstico de la zona terapéutica proporcionaría información sobre la probabilidad de que un paciente responda a una serie de fármacos relacionados con una sola zona terapéutica. Por ejemplo, en el comercio existe una gran cantidad de fármacos para tratar la depresión, incluidos los ISRS (inhibidores selectivos de la recaptación de serotonina), los ATC (antidepresivos cíclicos), los IMAO (inhibidores de monoaminooxidasa) y triazolopiridinas. Para identificar locus polimórficos asociados con la eficacia de cada uno de estos tipos de fármacos se pueden utilizar estudios de asociación, y estos locus podrían utilizarse para explorar poblaciones de pacientes para determinar qué clase de fármacos sería más eficaz para un individuo determinado. Para cada fármaco, un grupo de casos comprende individuos con depresión que tuvieron una respuesta eficaz al fármaco, y un grupo de control que comprende individuos que no tuvieron una respuesta eficaz al fármaco. Los SNP asociados se identifican como aquellos que tienen una frecuencia alélica significativamente diferente en el grupo de casos en comparación con el grupo de controles. Para cada clase de fármaco, se determinan umbrales que identificarán a los individuos con un nivel alto (por ejemplo, >80 %, o >90 %, o >95 % o >98 %) de posibilidad de tener una respuesta eficaz. Un individuo que necesite terapia antidepresiva se explora para los SNP que están asociados con cada uno de los tipos de fármaco y un médico determina una terapia apropiada elegida para el individuo en función de la información del genotipo del individuo y de los umbrales determinados para cada clase de fármaco.

En una realización relacionada adicional, los SNP asociados con la eficacia de un fármaco pueden utilizarse para mejorar la eficacia del fármaco estratificando poblaciones de pacientes para excluir probables no respondedores del tratamiento. En un ejemplo, un porcentaje de ~32 % de los pacientes expuestos a un fármaco se clasifican como respondedores. Se realizó un estudio de asociación con un grupo de casos de pacientes respondedores y con un grupo de control de pacientes no respondedores y se descubrió que 25 SNP estaban asociados al fenotipo respondedor. Basándose en las puntuaciones calculadas para los casos y controles, se descubrió que el 81 % de los respondedores y el 40 % de los no respondedores tenían una puntuación de >19. Por lo tanto, el uso de 19 como un valor umbral para estratificar una población de pacientes antes de administrar el fármaco mejora la eficacia global del fármaco de ~32 % a ~50 %. Al hacerlo, la cantidad de pacientes no respondedores expuestos al fármaco disminuye considerablemente, y los excluidos pueden tratarse en seguida con terapias alternativas. Un cambio en la eficacia de esta magnitud podría ayudar a obtener un nuevo fármaco aprobado, o podría alentar un uso más amplio de un fármaco ya aprobado.

60 En otra realización más, los métodos presentados en este documento pueden utilizarse para evaluar si debería utilizarse un fármaco de marca registrada, o si en su lugar se puede sustituir por un genérico más asequible. Por

ejemplo, se realizaría un estudio de asociación para identificar locus genéticos asociados con una respuesta clínica positiva a la alternativa genérica. Los pacientes que necesitan tratamiento se genotificarían después en estos locus asociados y se calcularía una puntuación. La puntuación del individuo se utilizaría después para predecir la eficacia del fármaco genérico en el individuo, y un médico utilizaría esta información para decidir acerca del tratamiento para el individuo. Esta aplicación de los métodos divulgados también podría utilizarse para decidir acerca de reembolsos de costes médicos. Por ejemplo, si se descubre que era poco probable que el fármaco genérico fuera eficaz en el individuo A, entonces el fármaco de marca comercial se administraría a A y el coste del fármaco de marca comercial podría reembolsarse a A; sin embargo, si el individuo B pudiera tener una respuesta eficaz al genérico, entonces al individuo B no se le daría el fármaco de marca comercial más costoso, y solo sería reembolsable el coste del genérico.

En otra realización de la presente invención, el riesgo de que un individuo experimente un acontecimiento adverso en respuesta a la administración de un fármaco, se determina basándose en los genotipos del individuo en un conjunto de SNP asociados con la aparición de acontecimientos adversos relacionados con el fármaco. Si se descubre que un individuo tiene un riesgo alto de experimentar un acontecimiento adverso en respuesta a un régimen de tratamiento, entonces se puede evitar el régimen de tratamiento y se pueden considerar otras opciones de tratamiento. Por ejemplo, se puede realizar un estudio de asociación utilizando un grupo de casos de individuos que presentó un acontecimiento adverso en respuesta al fármaco y un grupo de control de individuos que no presentó el acontecimiento adverso. Los miembros de los grupos de casos y control se genotificaron en una pluralidad de posiciones de SNP, se calcularon las frecuencias alélicas relativas para cada uno de los SNP y los SNP asociados con el acontecimiento adverso se identifican como aquellos SNP que tenían diferencias de frecuencia alélica significativamente diferentes entre los grupos de casos y control. Para cada miembro de los grupos de casos y control se calculó una puntuación en función de sus genotipos en los SNP asociados, y estas puntuaciones se utilizaron para determinar uno o más valores umbral apropiados para una prueba de evaluación de riesgos que predecirá el riesgo de que un individuo experimente un acontecimiento adverso en respuesta al fármaco con niveles adecuados de sensibilidad, especificidad, VPP, VPN, CV+, CV- y/o precisión. Como se indicó anteriormente, la selección de un valor umbral también puede basarse en factores clínicos, tales como la gravedad del acontecimiento adverso, la enfermedad o trastorno que se trata y el historial médico del individuo que se está tratando. Por ejemplo, si el acontecimiento adverso es la muerte, entonces es esencial una alta sensibilidad para identificar a aquellos individuos que tienen una alta probabilidad de morir si se les administra el fármaco. Antes de recibir el fármaco, un individuo se genotifica en cada una de las posiciones de SNP asociadas y se calcula una puntuación para el individuo basándose al menos en sus genotipos en el conjunto de SNP asociados. Por ejemplo, si el individuo tiene una puntuación que es mayor que un valor umbral determinado a partir de las puntuaciones de los grupos de casos y de control, el individuo puede clasificarse como que probablemente experimente un acontecimiento adverso si se administra el fármaco, y el uso del fármaco puede impedirse. Si el individuo tiene una puntuación igual o inferior a un valor umbral, individuo puede clasificarse como que probablemente no experimente un acontecimiento adverso y se puede recomendar la administración del fármaco. Si el individuo tiene una puntuación menor o igual a un valor umbral y mayor que otro valor umbral, el individuo puede clasificarse como que tiene una probabilidad intermedia de experimentar un acontecimiento adverso y se pueden utilizar terapias farmacológicas alternativas, o el fármaco puede administrarse, por ejemplo, solo con una estrecha supervisión, o utilizando un método de administración diferente (por ejemplo, diferente dosificación, frecuencia, etc.) o en combinación con otro agente terapéutico para contrarrestar el acontecimiento adverso. En otro ejemplo de una prueba de evaluación de riesgos que utiliza múltiples valores umbral, si se determina que un paciente tiene un riesgo bajo de experimentar un acontecimiento adverso en respuesta a un fármaco dado, el médico puede prescribirle una dosis más alta de ese fármaco que la que se prescribiría a un paciente con un riesgo intermedio de experimentar el acontecimiento adverso. La determinación del mejor régimen de tratamiento para un individuo con un riesgo intermedio de experimentar un acontecimiento adverso, puede depender más de otra información (por ejemplo, datos clínicos, información facilitada por la FDA o por el paciente, etc.) que de la determinación del mejor régimen de tratamiento para un individuo con un riesgo muy alto o bajo. En otra realización, el riesgo de un individuo a experimentar un acontecimiento adverso puede determinarse calculando un riesgo relativo para el individuo y multiplicando el riesgo relativo por la prevalencia conocida de individuos que experimentan acontecimientos adversos. Esta información puede utilizarla después un médico para decidir sobre los tratamientos apropiados para el individuo. Los acontecimientos adversos en respuesta a la administración de un fármaco incluyen, pero si limitación, reacciones alérgicas, arritmia cardíaca, ictus, broncoespasmo, trastornos gastrointestinales, desmayos, impotencia, erupciones cutáneas, fiebre, mialgia, cefalea, náuseas, defectos de nacimiento, sofocos, cambios de humor, mareos, agitación, vómitos, trastornos del sueño, somnolencia, insomnio, adicción al fármaco y muerte.

En una realización relacionada, los SNP asociados con la seguridad de un fármaco pueden utilizarse para mejorar la seguridad del fármaco estratificando poblaciones de pacientes para excluir del tratamiento a aquellos individuos que probablemente presenten un acontecimiento adverso en respuesta a la administración del fármaco. En un ejemplo, se descubre que un nuevo fármaco tiene una excelente eficacia, tolerancia y conveniencia, sin embargo, el 4 % de los individuos tratados con el fármaco experimentan un acontecimiento adverso grave y esta incidencia de acontecimientos adversos ha limitado el uso del fármaco, por ejemplo, solo a aquellos individuos para quienes han fracasado otras terapias. Sin embargo, una agencia reguladora ha estipulado que, si la incidencia del acontecimiento adverso se redujera a al menos un 50 %, entonces el fármaco podría aprobarse para un uso más amplio. Esto se

conseguiría si los individuos que probablemente experimentan el acontecimiento adverso, pudiesen identificarse antes del tratamiento, de modo que se realice un estudio de asociación con un grupo de casos de individuos que experimentan el acontecimiento adverso y un grupo de individuos de control que no identificó un conjunto de 20 SNP asociados con el acontecimiento adverso. En la tabla 3 se muestran los resultados del estudio de asociación, mostrando los valores de límite de riesgo en la primera columna, el % de casos con puntuaciones mayores que el valor de límite de riesgo correspondiente en la segunda columna, el % de controles con puntuaciones mayores que el valor de límite de riesgo correspondiente en la tercera columna, el riesgo relativo en la cuarta columna, el porcentaje de sensibilidad en la quinta columna, el porcentaje de especificidad en la sexta columna, el VPP (como un porcentaje) en la séptima columna y el VPN (como un porcentaje) en la octava columna.

10

Tabla 3

Valor de límite de riesgo	% de casos	% de controles	Riesgo relativo	Sensibilidad	Especificidad	VPP	VPN
20	40,0 %	2,8 %	14,2	40,0 %	97,2 %	37,3 %	97,5 %
19	51,6 %	5,6 %	9,2	51,6 %	94,4 %	27,7 %	97,9 %
18	58,0 %	9,9 %	5,4	58,0 %	90,1 %	19,6 %	98,1 %
16	75,0 %	28,5 %	2,5	75,0 %	71,5 %	9,9 %	98,6 %
15	91,2 %	39,8 %	2,3	91,2 %	60,2 %	8,7 %	99,4 %

Utilizando estos valores, se descubrió que el uso de 19 como un valor umbral, eliminaría aproximadamente al 51,6 % de los pacientes en mayor riesgo de experimentar el acontecimiento adverso, mientras que solo eliminaría al 5,6 % de los que podrían beneficiarse del fármaco. Por lo tanto, si se exploraron 1000 sujetos usando 19 como valor umbral y suponiendo que el 4 % de ellos están en riesgo alto de experimentar el acontecimiento adverso, 74 [(1000)(0,04)(0,516) + (1000)(0,96)(0,056)] serían excluidos y los 926 restantes podrían tratarse. El riesgo de experimentar acontecimientos adversos para los tratados sería por tanto [(1000)(0,04)(1-0,516)/926 = 0,02] o 2 %. Por lo tanto, el uso de 19 como valor umbral en un diagnóstico para estratificar poblaciones de pacientes antes de administrar el fármaco, reduciría la incidencia de acontecimientos adversos del 4 % al 2 %, lo que capacitaría al fármaco para un uso más amplio. Del mismo modo, el valor 18 también podría utilizarse como un valor umbral que excluiría a 23/1000 individuos y daría lugar a una incidencia esperada de acontecimientos adversos del 1,9 % en los individuos tratados. Sin embargo, esta disminución en la incidencia de acontecimientos adversos, se combina con una disminución tanto en la especificidad como en el VPP de la prueba. La selección de un valor umbral de diagnóstico de riesgo/beneficio apropiado, puede requerir no solo información acerca de la propia prueba (especificidad, sensibilidad, VPP, VPN, etc.) sino también interacción entre el profesional que lleva a cabo los métodos presentados en este documento y una agencia reguladora (por ejemplo, la FDA) y el criterio basado en la utilidad clínica. El objetivo de dicha prueba farmacogenómica sería maximizar el VPN (reducir la incidencia del acontecimiento adverso en los individuos tratados) mientras que se equilibra el VPP (minimizando la exclusión de pacientes que podrían beneficiarse del fármaco). El uso de los métodos descritos en el presente documento para reducir la frecuencia de acontecimientos adversos podría ayudar a obtener un nuevo fármaco aprobado, o podría alentar un uso más amplio de un fármaco ya aprobado. Por ejemplo, asociando dicho diagnóstico con un fármaco, puede ser posible reducir la frecuencia de acontecimientos adversos a niveles que son aceptables desde un punto de vista comercial, rescatando en efecto un fármaco que de otro modo no habría sido aprobado.

Será obvio para los expertos en la técnica, que un valor umbral apropiado para la aprobación de un diagnóstico que se asocia con un fármaco, depende en gran medida de las negociaciones entre un patrocinador del fármaco (por ejemplo, una compañía farmacéutica) y las autoridades reguladoras (por ejemplo, la FDA). Este sería el caso de si el diagnóstico es para mejorar la eficacia o la seguridad de un fármaco. Por ejemplo, aunque en el ejemplo anterior la frecuencia de acontecimientos adversos se reduce al 2 %, las autoridades reguladoras podrían requerir un nivel de seguridad más estricto y, por lo tanto, un valor umbral más bajo para identificar a individuos que se excluyan del tratamiento con el fármaco, sacrificando de este modo el VPP por un VPN más alto.

En algunas realizaciones de la presente invención, se utiliza una prueba de evaluación de riesgos para seleccionar pacientes para su inclusión en un ensayo clínico, por ejemplo, para probar la capacidad de un agente (por ejemplo, un fármaco) para tratar una enfermedad u otra afección médica. Estos métodos pueden ser útiles, por ejemplo, para explorar a aquellos individuos cuyo perfil de evaluación de riesgos indique que es probable que experimenten un acontecimiento adverso; por lo tanto, el ensayo clínico puede llevarse a cabo utilizando poblaciones ricas de individuos con pocas probabilidades de experimentar el acontecimiento adverso. Del mismo modo, los métodos

también pueden ser útiles para identificar una población de individuos con mayor probabilidad de tener la enfermedad o afección médica, de tal modo que solo esos individuos estén incluidos en el ensayo clínico para probar un tratamiento para esa enfermedad o afección médica. Sin dicha prueba, en el ensayo clínico pueden incluirse individuos que realmente padecen una enfermedad diferente con síntomas fenotípicos similares, lo que  
 5 podría confundir los resultados de un ensayo clínico. Anteriormente se han descrito métodos para utilizar pruebas de evaluación de riesgos para diagnosticar a un individuo que presenta un fenotipo que puede ser indicativo de la enfermedad.

Un ensayo clínico se realiza preferentemente en poblaciones compuestas por individuos con perfiles de riesgo similares (que incluyen, por ejemplo, información genotípica y fenotípica). El uso de dichas poblaciones elimina o  
 10 reduce la variación en el resultado del tratamiento debido a factores además del diseño del ensayo clínico (por ejemplo, comparación de grupos que reciben tratamiento frente a grupos tratados con placebo) lo que conduce a una evaluación más precisa del valor del agente en el tratamiento de la enfermedad (por ejemplo, eficacia). Pueden utilizarse algoritmos implementados por ordenador para identificar subpoblaciones más homogéneas en las que el tratamiento o la profilaxis tiene un efecto significativo a pesar de que el tratamiento o la profilaxis no sean eficaces en poblaciones más heterogéneas y más grandes. En dichos métodos, se proporcionan datos para una primera  
 15 población con una enfermedad u otra afección, a tratar con el agente y para una segunda población también con la enfermedad o afección, a tratar con un placebo. Se realiza una prueba de evaluación de riesgos en todos los individuos en las dos poblaciones para determinar el riesgo de cada individuo, por ejemplo, de no tener una respuesta eficaz o de tener un acontecimiento adverso en respuesta al tratamiento. Después, para el ensayo clínico se seleccionan subpoblaciones de cada una de la primera y segunda subpoblaciones de manera que los individuos de las subpoblaciones tengan una mayor similitud de perfiles de evaluación de riesgos entre sí que los individuos de la primera y segunda población original. En algunas realizaciones, los individuos que probablemente tengan un acontecimiento adverso o que probablemente no tengan una respuesta eficaz al tratamiento, se retiran de la primera y segunda población. En algunas enfermedades, los individuos que es poco probable que tengan la enfermedad o el  
 20 trastorno de interés, se retiran de la primera y segunda población. En algunas realizaciones, los perfiles de evaluación de riesgos se realizan en todos los posibles participantes del ensayo clínico, antes de determinar qué incluir en la primera subpoblación y qué incluir en la segunda subpoblación. Se proporcionan otros ejemplos, por ejemplo, en el documento U.S.S.N. 11/344.975 publicado como 2006-0228728 el 12 de octubre de 2006.

En determinados aspectos, la presente invención proporciona métodos muy mejorados para determinar el riesgo de  
 30 un individuo de desarrollar o presentar un rasgo multifactorial. En determinados aspectos, los métodos se utilizan además para desarrollar pronósticos, diagnósticos o terapias para una enfermedad multifactorial. En otros aspectos, los métodos se utilizan además para predecir la respuesta a un fármaco en los individuos antes de la administración de un régimen terapéutico. Los métodos presentados en este documento pueden ayudar además a reducir el coste global de un tratamiento médico, al proporcionar un medio para encontrar rápidamente la intervención médica correcta (más eficaz, más segura, más asequible, etc.) para un individuo, de modo que no se malgaste el valioso  
 35 tiempo y dinero en terapias de valor limitado. Debe entenderse que la descripción anterior pretende ser ilustrativa y no restrictiva. Todas las publicaciones mencionadas en este documento se citan con el fin de describir y desvelar reactivos, metodologías y conceptos que pueden utilizarse junto con la presente invención. Nada de lo que se indica en este documento debe interpretarse como una admisión de que estas referencias son técnicas anteriores en relación con las invenciones descritas en el presente documento. A lo largo de la divulgación, se hace referencia a  
 40 diversas patentes, solicitudes de patente y publicaciones.

**REIVINDICACIONES**

1. Un método de determinación del riesgo de un individuo de desarrollar o presentar una afección médica multifactorial, que comprende:

5 a) determinar una puntuación para el individuo basándose en información genotípica obtenida de una muestra del individuo, en la que la información genotípica es la presencia o la ausencia de alelos asociados a la afección médica en una pluralidad de locus polimórficos, en el que a un alelo asociado se le asigna un valor diferente para determinar la puntuación de otros alelos asociados, y en el que la magnitud del efecto sobre el riesgo de los alelos se utiliza para determinar la puntuación, y en el que la determinación de la puntuación para el individuo se basa además en información no genética; y

10 b) comparar la puntuación con al menos un valor umbral, en el que el resultado de dicha comparación es indicativo del riesgo del individuo de desarrollar o presentar la afección médica multifactorial.

2. El método de la reivindicación 1, en el que la información genotípica comprende genotipos para el individuo en una pluralidad de locus polimórficos bialélicos, en el que cada uno de dicha pluralidad tiene un alelo asociado y un alelo no asociado, y en el que además cada uno de los genotipos se selecciona del grupo que consiste en homocigoto para el alelo asociado, heterocigoto y homocigoto para el alelo no asociado.

3. El método según la reivindicación 2, que comprende además identificar los alelos asociados y los alelos no asociados para dicha pluralidad de locus polimórficos bialélicos realizando un estudio de asociación con un grupo de individuos de casos y un grupo de individuos de control, determinando así un conjunto de alelos de dicho locus polimórfico que son significativamente más abundantes en el grupo de casos que en el grupo de control, en el que dicho conjunto de alelos o un subconjunto de los mismos, son los alelos asociados.

4. El método de la reivindicación 3, en el que dicha realización de un estudio de asociación comprende además

a) genotipificar dicho grupo de casos y dicho grupo de control en un conjunto de locus polimórficos que comprende dicha pluralidad de locus polimórficos bialélicos;

25 b) calcular una frecuencia alélica relativa para cada uno de dicho conjunto de locus polimórficos de cada uno de dicho grupo de casos y dicho grupo de control;

30 c) para cada uno de dicho conjunto de locus polimórficos, comparar la frecuencia alélica relativa calculada para el grupo de casos con la frecuencia alélica relativa calculada para el grupo de control, identificando así un subconjunto de dicho conjunto de locus polimórficos, en el que cada uno de dicho subconjunto tiene una frecuencia alélica relativa que es significativamente diferente para el grupo de casos que para el grupo de control; y

d) determinar un alelo para cada uno de dicho subconjunto que es más abundante en dicho grupo de casos que en dicho grupo de control, en el que dicho alelo es uno de dichos alelos asociados.

35 5. El método de la reivindicación 3, que comprende además validar dichos alelos asociados realizando un segundo estudio de asociación con un segundo grupo de casos y un segundo grupo de control, determinando así cuáles de dichos alelos asociados son significativamente más abundantes en el segundo grupo de casos que en el segundo grupo de control, en el que aquellos de dichos alelos asociados que son significativamente más abundantes en el segundo grupo de casos que en el segundo grupo de control, son los alelos asociados validados.

6. El método de la reivindicación 3, que comprende además determinar uno de dicho al menos un valor umbral mediante un método que comprende

40 a) calcular una puntuación para cada miembro de dicho grupo de casos y dicho grupo de control;

b) seleccionar una serie de valores de límite de riesgo;

c) computar un conjunto de valores para cada una de dicha serie de valores de límite de riesgo en el que dicho conjunto de valores comprende al menos un valor entre una sensibilidad, una especificidad, un VPP, un VPN, una precisión, un riesgo relativo, un CV+, un CV- e información clínica;

45 d) elegir uno de dicha serie de valores de límite de riesgo como dicho uno de dichos al menos un valor umbral basándose en dicho conjunto de valores, determinando así dicho uno de dichos al menos un valor umbral.

7. El método de la reivindicación 6, en el que el cálculo de una puntuación para cada miembro de dicho grupo de casos y dicho grupo de control comprende
- 5 a) determinar un genotipo para cada dicho miembro en dicha pluralidad de locus polimórficos bialélicos, en el que el genotipo se selecciona del grupo que consiste en un homocigoto para un alelo asociado, heterocigoto y homocigoto para un alelo no asociado;
  - b) asignar un primer valor a cada uno de dichos locus polimórficos que tiene un genotipo que es homocigoto para un alelo que no es el alelo asociado;
  - c) asignar un segundo valor a cada uno de dichos locus polimórficos que tiene un genotipo que es heterocigoto;
  - 10 d) asignar un tercer valor a cada uno de dichos locus polimórficos que tiene un genotipo que es homocigoto para el alelo asociado;
  - e) sumar los valores determinados en las etapas b) a d) para todos dichos locus polimórficos, calculando así una puntuación para cada dicho miembro de dicho grupo de casos y dicho grupo de control.
8. El método de la reivindicación 6, en el que dicha selección de una serie de valores de límite de riesgo comprende
- 15 identificar una puntuación más alta de las puntuaciones calculadas para cada miembro de dicho grupo de casos y dicho grupo de control;
  - determinar un intervalo de límite de riesgo, en el que el intervalo va de 1 a dicha puntuación más alta;
  - seleccionar una serie de valores en todo el intervalo de límite de riesgo, seleccionando así dicha serie de valores de límite de riesgo.
9. El método de la reivindicación 8, en el que dicha selección de dichas series de valores en todo el intervalo de límite de riesgo comprende un método seleccionado del grupo que consiste en
- 20 seleccionar cada valor dentro del intervalo de límite de riesgo;
  - seleccionar cada enésimo valor dentro del intervalo de límite de riesgo;
  - dividir el intervalo de límite de riesgo en porcentajes y seleccionar un valor a cada enésimo porcentaje del intervalo de límite de riesgo;
  - 25 seleccionar un número de valores más elevado de una parte central del intervalo de límite de riesgo que de una parte superior o inferior del intervalo de límite de riesgo; y
  - seleccionar un número de valores más elevado de una parte superior o inferior del intervalo de límite de riesgo que de una parte central del intervalo de límite de riesgo.
10. El método en la reivindicación 5, en el que dicha determinación de dicho uno de dicho al menos un valor umbral comprende además utilizar una curva ROC basada en dicha sensibilidad y dicha especificidad calculadas en c), en el que una representación gráfica de dicha curva ROC se denomina gráfica.
- 30
11. El método de la reivindicación 10, que comprende además elegir como dicho uno de dicho al menos un valor umbral de un valor de límite de riesgo correspondiente a un punto de datos de dicha curva ROC que está más cerca de una esquina superior izquierda de dicho gráfico que cualquier otro punto de datos de dicha curva ROC, en el que cada punto de datos en dicha curva ROC corresponde a un valor de límite de riesgo diferente, y/o adicionalmente comprende
- 35 a) determinar una posición en dicha curva ROC que esté más cerca de la esquina superior izquierda de dicho gráfico y determinar una sensibilidad y una especificidad que correspondan a dicha posición;
  - b) analizar dichas puntuaciones para cada miembro de dicho grupo de casos y dicho grupo de control para identificar un valor de límite de riesgo cuya sensibilidad y especificidad son las más cercanas a dicha sensibilidad y especificidad que corresponden a dicha posición, en el que dicho valor de límite de riesgo cuya sensibilidad y especificidad corresponden a dicha posición es dicho uno de dicho al menos un valor umbral.
  - 40

12. El método de la reivindicación 6, en el que, para un valor de límite de riesgo dado, dicho riesgo relativo se calcula mediante un método que comprende
- a) determinar un porcentaje de dichos miembros de dicho grupo de casos que tienen una puntuación que es al menos tan grande como dicho valor de límite de riesgo dado;
- 5        b) determinar un porcentaje de dichos miembros de dicho grupo de control que tienen una puntuación que es al menos tan grande como dicho valor de límite de riesgo dado;
- c) dividir dicho porcentaje determinado en a) entre dicho porcentaje determinado en b) para calcular dicho riesgo relativo.
- 10      13. El método de la reivindicación 1, en el que dicha determinación de una puntuación para el individuo comprende adicionalmente
- a) determinar un genotipo para dicho individuo en dicha pluralidad de locus polimórficos bialélicos, en el que el genotipo se selecciona del grupo que consiste en homocigoto para un alelo asociado, heterocigoto y homocigoto para un alelo no asociado;
- 15        b) asignar un primer valor a cada uno de dichos locus polimórficos que tienen un genotipo que es homocigoto para un alelo que no es el alelo asociado;
- c) asignar un segundo valor a cada uno de dichos locus polimórficos que tiene un genotipo que es heterocigoto;
  - d) asignar un tercer valor a cada uno de dichos locus polimórficos que tiene un genotipo que es homocigoto para el alelo asociado;
- 20        e) sumar los valores determinados en las etapas b) a d) para todos los dichos locus polimórficos, determinando así una puntuación para el individuo.
14. El método de acuerdo con una cualquiera de las reivindicaciones 1 o 13, que es un método implementado por ordenador.
- 25      15. El método según cualquiera de las reivindicaciones 1, 13 o 14, en el que los valores asignados a los alelos asociados en función de la magnitud del efecto sobre el riesgo de los alelos, se basan en la frecuencia de los alelos en una población de casos y en una población de control.
16. El método de acuerdo con la reivindicación 1, que comprende, antes de la etapa (a), genotipificar una muestra obtenida de dicho individuo para obtener dicha información genotípica.

Figura 1

