



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 329 364**

51 Int. Cl.:
C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **04703049 .9**

96 Fecha de presentación : **16.01.2004**

97 Número de publicación de la solicitud: **1587946**

97 Fecha de publicación de la solicitud: **26.10.2005**

54 Título: **Análisis de haplotipos.**

30 Prioridad: **17.01.2003 US 441046 P**

45 Fecha de publicación de la mención BOPI:
25.11.2009

45 Fecha de la publicación del folleto de la patente:
25.11.2009

73 Titular/es: **The Trustees of Boston University
One Silber Way
Boston, Massachusetts 02215, US**

72 Inventor/es: **Cantor, Charles, R. y
Ding, Chunming**

74 Agente: **Arias Sanz, Juan**

ES 2 329 364 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

Análisis de haplotipos.

5 **Referencia a solicitudes relacionadas**

La presente solicitud reivindica beneficio de la solicitud provisional de EE.UU. No. de serie 60/441046, presentada el 17 de enero, 2003, que se incorpora a la presente mediante referencia en su totalidad.

10 **Antecedentes de la invención**

Los polimorfismos genéticos son mecanismos bien reconocidos subyacentes a las diferencias interindividuales en el riesgo a enfermedades así como en la respuesta a tratamiento en seres humanos (Evans y Relling (1999) Science 286:487-491; Shields y Harris (2000) J. Clin. Onc. 18:2309-2316). El análisis de polimorfismo de un solo nucleótido (SNP) ha atraído mucha atención con la esperanza de identificar marcadores genéticos para y genes implicados en enfermedades comunes debido a la frecuencia de los SNP. Además, para muchos genes, la detección de los SNP que se sabe confieren pérdida de función proporciona un diagnóstico molecular sencillo para seleccionar medicaciones y dosis óptimas para pacientes individuales (Evans y Relling (1999) Science 286:487-491). Es normal que los genes contengan múltiples SNP, siendo la estructura del haplotipo el determinante principal de las consecuencias fenotípicas (Collins *et al.* (1997) Science 278, 1580-81; Drysdale *et al.* (2000) Proc. Natl. Acad. Sci. 97:10483-8; Krynetski y Evans (1998) Am. J. Hum. Gen. 63:11-16). Por lo tanto, para asociar de forma más precisa riesgos de enfermedad y rasgos farmacogenómicos con polimorfismos genéticos, se necesitan métodos fiables para determinar de forma inequívoca la estructura del haplotipo para múltiples SNP u otros polimorfismos o mutaciones de ácido nucleico en genes así como en secuencias genómicas no codificantes.

Sin embargo, las actuales tecnologías de genotipado sólo son capaces de determinar cada polimorfismo, incluyendo los SNP, por separado. En otras palabras, hay una falta de información sobre cómo se asocian varios polimorfismos entre sí físicamente en un cromosoma. Un haplotipo de ADN, la asociación de fase determinada de varios marcadores polimórficos (por ejemplo, SNP) es un método estadísticamente mucho más poderoso para estudios de asociación de enfermedad. Con todo desgraciadamente, es también mucho más difícil determinar un haplotipo. Los planteamientos experimentales actuales incluyen una separación física de cromosomas homólogos mediante híbrido de línea celular de ratón, clonación en un plásmido y PCR específica de alelo. Ninguno de ellos es un método lo suficientemente sencillo para análisis de alto rendimiento de rutina. Siempre hay modos de determinar haplotipos de forma computacional, pero la precisión de tales análisis computacionales es incierta.

Los planteamientos que se pueden usar para determinar el haplotipo de los SNP u otros polimorfismos de ácido nucleico, modificaciones y/o mutaciones que están en proximidad relativamente cercana incluyen, pero no están limitados a, análisis de polimorfismo conformacional de cadena sencilla (SSCP) (Orita *et al.* (1989) Proc. Natl. Acad. Sci. USA 86:2766-2770), análisis de heterodúplex (Prior *et al.* (1995) Hum. Mutat. 5:263-268), ligación de oligonucleótidos (Nickerson *et al.* (1990) Proc. Natl. Acad. Sci. USA 87:8923-8927) y ensayos de hibridación (Conner *et al.* (1983) Proc. Natl. Acad. Sci. USA 80:278-282). Una desventaja principal de estos procedimientos es que están limitados a la detección de polimorfismos a lo largo de segmentos cortos de ADN y típicamente requieren condiciones de reacción y/o marcaje rigurosas. Las estrategias tradicionales basadas en PCR con Taq polimerasa, tales como PCR-RFLP, amplificación específica de alelo (ASA) (Ruano y Kidd (1989) Nucleic Acids Res. 17:8392), dilución de molécula única (SMD) (Ruano *et al.* (1990) Proc. Natl. Acad. Sci. USA 87:6296-6300), y amplificación y secuenciación acopladas (CAS) (Ruano y Kidd (1991) Nucleic Acids Res. 19:6877-6882), se realizan fácilmente y son muy sensibles, pero estos métodos también están limitados a determinar el haplotipo de SNP a lo largo de segmentos cortos de ADN (<1 kb) (Michalatos-Beloin *et al.* (1996) Nucleic Acids Res. 24:4841-4843; Barnes (1994) Proc. Natl. Acad. Sci. USA 91:5695-5699; Ruano y Kidd (1991) Nucleic Acids Res. 19:6877-6882).

La PCR de largo alcance (LR-PCR) ofrece el potencial de determinar el haplotipo de los SNP que están separados por longitudes de kilobases de ADN genómico. Los productos de LR-PCR se genotipan comúnmente para los SNP, y los haplotipos se infieren usando planteamientos matemáticos (por ejemplo, algoritmo de Clark (Clark (1990) Mol. Biol. Evol. 7:111-122). Sin embargo, inferir los haplotipos de esta manera no produce asignaciones inequívocas de haplotipos cuando los individuos son heterocigotos en dos o más loci (Hodge *et al.* (1999) Nature Genet. 21:360-361). Separar físicamente los alelos mediante clonación, seguido por secuenciación, elimina cualquier ambigüedad, pero este método es laborioso y caro. La amplificación de largo alcance específica de alelo anula ambos problemas, pero está limitada a alelos que contienen SNP que tienen anclajes de inserción/delección heterocigotos para los cebadores de PCR (Michalatos-Beloin *et al.* (1996) Nucleic Acids Res. 24:4841-4843). También se han usado tecnologías más complejas, tales como el análisis de mutaciones monoalélicas (MAMA) (Papadopoulos *et al.* (1995) Nature Genet. 11:99-102) y sondas de nanotubos de carbono (Woolley *et al.* (2000) Nature Biotech. 18:760-763), pero éstas o llevan mucho tiempo (MAMA), o requieren tecnología que no está ampliamente disponible (nanotubos). La solicitud de patente de EE.UU. No. 2002/0081598 divulga un método de determinar el haplotipo que implica el uso de amplificación por PCR y ligación de ADN para acercar mucho los sitios polimórficos de ácido nucleico en un alelo particular para facilitar la determinación de haplotipos que abarcan distancias de kilobases. Sin embargo, este método se basa en al menos dos pasos enzimáticos para crear fragmentos de ADN que se pueden ligar con otros fragmentos de ADN, y posteriormente ligasas para combinar los fragmentos de ADN para formar un fragmento grande con varios sitios

polimórficos en una distancia más corta. Estos pasos adicionales de preparación de la muestra hacen el uso a gran escala y la automatización de esta técnica engorroso y propenso a errores.

Los haplotipos, combinaciones de varios marcadores polimórficos determinados por la fase en un cromosoma, son extremadamente valiosos para estudios como la asociación de enfermedades^{1,2} y evolución de cromosomas. La determinación directa de haplotipo molecular se ha basado firmemente en datos familiares, pero está limitada a regiones genómicas cortas (unas pocas kilobases). La estimación estadística de frecuencias de haplotipo puede ser no definitiva e inexacta³.

Con el rápido descubrimiento y validación de varios millones de polimorfismos de un solo nucleótido (SNP), ahora es cada vez más práctico usar barrido a lo largo del genoma para encontrar genes asociados con enfermedades comunes^{1,2}. Sin embargo, los SNP individuales tienen un poder estadístico limitado para localizar genes de susceptibilidad a enfermedades. Los haplotipos pueden proporcionar poder estadístico adicional en el mapeo de genes de enfermedades⁴⁻⁷.

La determinación del haplotipo de varios marcadores para una célula diploide es complicada ya que las técnicas convencionales de genotipado no pueden determinar las fases de varios marcadores diferentes. Por ejemplo, una región genómica con tres marcadores heterocigotos puede producir 8 posibles haplotipos. Esta ambigüedad puede, en algunos casos, resolverse si están disponibles genotipos de linaje. Sin embargo, incluso para un haplotipo de sólo 3 marcadores, los genotipos de los tríos padre-madre-descendencia pueden fracasar para producir haplotipos de descendientes hasta el 24% de las veces. Los algoritmos computacionales tales como expectación-maximización (EM), sustracción y PHASE se usan para la estimación estadística de haplotipos^{4,8,9}. Sin embargo, estos métodos computacionales tienen serias limitaciones en precisión, número de marcadores y longitud del ADN genómico. Por ejemplo, para un haplotipo de sólo 3 marcadores de individuos doblemente heterocigotos, las tasas de error de los métodos EM y PHASE para la reconstrucción del haplotipo pueden ser tan altas como del 27% y el 19%, respectivamente³. De forma alternativa, se puede usar la determinación molecular directa del haplotipo basada en la separación física de dos ADN genómicos homólogos antes del genotipado. Se han usado clonación de ADN, construcción de híbridos de células somáticas, PCR específica de alelo y PCR de molécula única¹⁰⁻¹², y estos planteamientos son en gran parte independientes de información genealógica. Estos métodos están limitados a regiones genómicas cortas (PCR específica de alelo y PCR de molécula individual) y son propensos a errores.

Por lo tanto, se necesita un método sencillo y más fiable, que también sea adecuado para la determinación de haplotipo a gran escala y automatizada de varios alelos polimórficos separados por varias kilobases de distancia para facilitar el análisis de la estructura del haplotipo en farmacogenómica, patogénesis de enfermedades y estudios epidemiológicos moleculares.

Compendio de la invención

La presente invención proporciona un modo eficaz para el análisis de haplotipos de alto rendimiento. Se pueden determinar varios marcadores polimórficos de ácido nucleico, tales como SNP, de forma simultánea y fiable mediante PCR multiplex de moléculas individuales de ácido nucleico en varias diluciones paralelas de moléculas individuales y el análisis estadístico consiguiente de los resultados de estas reacciones paralelas de PCR multiplex en moléculas individuales produce la determinación fiable de haplotipos presentes en el sujeto. Los marcadores de ácido nucleico pueden estar a cualquier distancia entre sí en el cromosoma. Además, se puede usar un planteamiento en donde se analizan marcadores solapantes de ADN para unir haplotipos más pequeños en haplotipos mayores. Por consiguiente, la invención proporciona una nueva herramienta poderosa para la determinación diagnóstica de haplotipos y la identificación de nuevos haplotipos.

La presente invención permite la determinación molecular directa del haplotipo de varios marcadores polimórficos separados por varias kilobases incluso que abarquen un cromosoma entero. Distancias de alrededor de 1, 2, 3, 4, 5-10, 15-20 kilobases (kb) o tan lejanas como alrededor de al menos 25, 30, 35, 40, 45 ó 50 kb o más son preferidas.

Los ácidos nucleicos polimórficos útiles según la presente invención incluyen cualquier ácido nucleico polimórfico en cualquier región determinada de ácido nucleico incluyendo, pero no limitado a, sustituciones de un solo nucleótido (polimorfismos de un solo nucleótido o SNP), sustituciones de múltiples nucleótidos, deleciones, inserciones, inversiones, repeticiones cortas en tándem incluyendo, por ejemplo, repeticiones de di-, tri- y tetra-nucleótidos, y metilación y otras diferencias de modificación de ácido nucleico polimórfico. Preferiblemente los nucleótidos polimórficos son SNP.

Primero se diluye una muestra de ácido nucleico, preferiblemente una muestra de ácido nucleico genómico de un sujeto a una dilución de una sola copia. La frase "dilución de una sola copia" se refiere a una dilución en donde sustancialmente sólo está presente una molécula de ácido nucleico o en donde están presentes una o más copias del mismo alelo. Cuando la masa molecular del ácido nucleico es conocida, el experto en la materia puede calcular fácilmente una dilución que produzca una dilución de una sola molécula. Por ejemplo, para ADN genómico humano, alrededor de 3 pg de ADN representa alrededor de una molécula. Debido a la fluctuación estocástica en soluciones de ADN muy diluidas, la muestra diluida puede no tener moléculas molde de ácido nucleico o puede tener dos o más moléculas. Si no hay moléculas presentes en la muestra, no se alcanzará amplificación por PCR y el resultado será "sin genotipo". Si están presentes dos o más moléculas en la muestra, los productos de amplificación resultantes pueden

ES 2 329 364 T3

ser una mezcla de dos alelos diferentes o representar un alelo y por consiguiente se obtiene un genotipo mixto o un genotipo de alelo único, respectivamente.

5 Para obtener peso estadístico para determinar de forma precisa el haplotipo que comprende al menos dos marcadores, se necesitará más de una réplica de diluciones. Por ejemplo, un replicado de cuatro ensayos independientes de genotipado multiplex usando alrededor de 3-4,5 pg de ADN genómico humano, que incluye los pasos de diluir la muestra de ácido nucleico, amplificar la muestra diluida, y genotipado de la muestra amplificada, permite alrededor del 90% de eficacia de determinación directa del haplotipo. Por lo tanto, se pueden incluir preferiblemente al menos alrededor de 4-25, más preferiblemente al menos alrededor de 6-20, 8-20, 10-18, 12-18 y lo más preferiblemente
10 alrededor de 10-12 replicados de la misma muestra en el análisis según la presente invención, incluyendo una réplica los pasos de diluir la muestra de ácido nucleico aislado de un organismo sujeto, amplificación multiplex de la muestra diluida y genotipado de los sitios polimórficos de ácido nucleico de la muestra amplificada.

15 Después del paso de diluir la muestra de ácido nucleico en una dilución de sustancialmente una única molécula de ácido nucleico, las regiones que contienen los sitios polimórficos de interés en el ácido nucleico se amplifican, usando, por ejemplo reacción en cadena de la polimerasa (PCR) y al menos dos, preferiblemente más de dos pares de cebadores que flanquean al menos dos sitios polimórficos diferentes de ácido nucleico en la molécula diana. Los cebadores se pueden seleccionar de modo que amplifiquen un fragmento de al menos alrededor de 50 pares de bases (pb), más preferiblemente al menos alrededor de 100, 200, 300, 400, 500, 600-1000 pb y hasta alrededor de 10000
20 pb, en donde el fragmento contiene al menos un sitio de nucleótido polimórfico. Lo más preferiblemente, los pares de cebadores se diseñan de modo que los productos de amplificación tengan alrededor de 90-350 pb de longitud, aún más preferiblemente alrededor de 100-250 pb de longitud. Es preferible maximizar la eficiencia de amplificación del molde de molécula única y por lo tanto, al menos con la tecnología actual, los fragmentos más cortos son preferidos. Sin embargo, será evidente para el experto en la materia que las técnicas de amplificación de ácidos nucleicos se
25 desarrollan de forma constante y la eficiencia de amplificación de ácidos nucleicos más largos usando cantidades muy pequeñas de molde se pueden perfeccionar y por consiguiente, también se pueden usar cebadores que amplifican fragmentos largos, incluso más largos de los indicados anteriormente, según la presente invención.

Después de la amplificación del molde de molécula única con al menos dos pares diferentes de cebadores, preferiblemente se usan al menos 3, 4, 5, 6, 7, 8, 9, 10 pares de cebadores en una reacción de amplificación multiplex, el producto de amplificación se somete a genotipado. El uso de hasta al menos 15, 20, 30, 40, 50 o más pares de cebadores en una reacción multiplex es preferido.

35 Se puede realizar el genotipado por cualquier medio conocido para el experto en la materia incluyendo, por ejemplo, análisis de polimorfismo de longitud del fragmento de restricción (RFLP) usando enzimas de restricción, análisis de polimorfismo conformacional de cadena sencilla (SSCP), análisis de heteroduplex, análisis por corte químico, ensayos de ligación e hibridación de nucleótidos, amplificación específica de alelo, microsecuenciación en fase sólida, o el sistema MASSARRAY™.

40 El haplotipo se determina posteriormente analizando las réplicas de al menos cuatro reacciones de dilución/amplificación/genotipado de modo que se permite la determinación estadísticamente precisa del haplotipo correcto en el sujeto. Los pasos que incluyen dilución, amplificación y genotipado de la misma muestra del organismo sujeto se repiten varias veces hasta obtener un conjunto de datos que se pueden analizar estadísticamente para revelar el haplotipo correcto en la muestra del organismo sujeto. El planteamiento no se basa en los datos de linaje y no requiere amplificación previa de la región genómica que contiene los marcadores seleccionados simplificando de esta manera el
45 análisis y permitiendo la determinación rápida y automatizada del haplotipo.

La invención permite métodos para determinar un haplotipo nuevo de segmentos de ácido nucleico, particularmente de genes u otros segmentos contiguos de ácido nucleico que comprenden al menos dos, preferiblemente al menos 3,
50 4, 5, 6, 7, 8, 9, 10-15, 20, 30, 40, 50-100 o incluso polimorfismos de ácido nucleico mucho más separados.

Los métodos de la presente invención son útiles en medicina en la determinación de las diferencias en riesgo o susceptibilidad a enfermedad y determinación de la respuesta a tratamiento entre pacientes individuales. Los métodos, sin embargo, no están limitados a aplicaciones en medicina y se pueden usar para determinar la estructura de haplotipo
55 de un gen particular, u otro segmento de ADN contiguo, en un organismo que tiene al menos dos polimorfismos de nucleótido muy separados. De esta manera, los métodos de la invención encuentran un uso adicional en el campo de la agricultura, particularmente en la cría de ganado y plantas cultivadas mejoradas.

En un aspecto, la invención proporciona un método de determinar un haplotipo en una muestra obtenida de un organismo y compararla con haplotipos conocidos para diagnosticar una enfermedad o susceptibilidad a enfermedad de un organismo que comprende los pasos de identificar al menos dos marcadores polimórficos en una región genómica; aislar una muestra de ácido nucleico del organismo sujeto y preferiblemente purificar el ácido nucleico aislado; diluir la muestra de ácido nucleico en sustancialmente una dilución de molécula única; amplificar la muestra de ácido nucleico diluido con al menos dos pares de cebadores, cada uno capaz de amplificar una región diferente que flanquea
65 cada uno de los sitios polimórficos en una reacción de PCR multiplex; genotipado de los sitios polimórficos de la muestra amplificada; producir al menos tres réplicas adicionales de genotipo de la muestra de ácido nucleico del organismo sujeto como se describe anteriormente para permitir la determinación estadísticamente precisa del haplotipo en la muestra del organismo sujeto. En un método preferido el genotipado se realiza usando extensión de cebadores,

nucleótidos terminadores y análisis de espectrometría de masas por desorción/ionización de laser asistida por matriz tiempo de vuelo MALDI-TOF MS. El haplotipo se compara después con una colección de haplotipos existentes tal como una base de datos de haplotipos que comprende haplotipos asociados a enfermedad o a susceptibilidad a enfermedad, o haplotipos asociados con capacidad de respuesta a tratamiento o falta de respuesta de los marcadores polimórficos específicos. Un ejemplo no limitante de una base de datos de haplotipos existente es una base de datos de referencia de haplotipos Y-STR que se puede encontrar en http://ystr.charite.de/index_gr.html.

Por ejemplo, la mutación R117H en el gen del receptor transmembrana de la fibrosis quística (CFTR) muestra un efecto moderado sin la mutación 5T, y un efecto grave cuando la mutación 5T está presente en el mismo cromosoma. De esta manera, un haplotipo de R117H-5T es importante para aplicación clínica para determinar la gravedad del pronóstico de este tipo de fibrosis quística. El método de la presente invención permite la determinación directa de los haplotipos sin que requiera información del genotipo de linaje del paciente, es decir, información de los genotipos de los miembros de la familia del paciente. El mismo planteamiento se puede aplicar en otras enfermedades genéticas donde, por ejemplo, una segunda mutación en el mismo cromosoma puede cambiar la manifestación de la enfermedad a partir de la primera mutación.

La invención permite un método en donde dos haplotipos que comprenden varios marcadores polimórficos diferentes se pueden combinar para formar un haplotipo mayor que cubre una región genómica mayor. Esto se puede alcanzar usando uno o más pares de cebadores para amplificar un marcador polimórfico común en dos reacciones de amplificación multiplex paralelas después de diluir primero la muestra como se ha descrito anteriormente. El genotipado se realiza como se describe anteriormente y el/los marcador/es solapante/s proporcionan un medio para combinar los dos haplotipos más pequeños en un haplotipo mayor que comprende todos los marcadores analizados en ambas de las dos reacciones de amplificación multiplex diferentes.

La presente invención permite un método para construir una base de datos de haplotipos asociados con una o más enfermedades o rasgos biológicos usando los métodos descritos anteriormente. Tales bases de datos de haplotipos son útiles para aplicaciones de diagnóstico y pronóstico. Se puede comparar un haplotipo obtenido de un organismo sujeto presunto contra la base de datos de haplotipos y permite el diagnóstico y/o pronóstico de una afección de interés. Una afección puede ser una enfermedad o un rasgo bioquímico u otro rasgo biológico que se asocia, por ejemplo, en la capacidad de respuesta a un tratamiento particular o producto farmacéutico y es determinante para elegir una pauta de tratamiento al que, por ejemplo, un paciente humano sería sensible.

En una forma de realización, el polimorfismo es una modificación de ácido nucleico, tal como una diferencia de metilación. Por ejemplo, en una forma de realización, la presente invención proporciona un método para determinar haplotipos que comprenden marcadores incluyendo diferencias de metilación. La muestra de ADN se puede tratar con cualquier composición, por ejemplo, compuestos inorgánicos u orgánicos, enzimas, etc., que afectan diferencialmente al nucleótido modificado, por ejemplo metilado, para crear de forma eficaz polimorfismos basados en estados de metilación. Por ejemplo, la muestra de ADN se trata con bisulfito (Frommer, M., L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, y C. L. Paul. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci U.S.A.* 89:1827-1831) de modo que los residuos de citosina no metilada se convierten en uracilo mientras que las citosinas no metiladas permanecen igual, creando así eficazmente polimorfismos basados en los estados de metilación. Los haplotipos que consisten en polimorfismos en la región de ADN próxima a la región de metilación y la región de metilación misma se pueden determinar en una manera similar a la descrita anteriormente. El ADN tratado con bisulfito se diluye a aproximadamente una única copia, se amplifica mediante PCR multiplex (cada PCR específica para cada polimorfismo), y se genotipa mediante el sistema MassARRAY.

El procedimiento de detección de metilación como se ha descrito anteriormente se repite al menos 3, 4, 5, 6, 7, 8, 9, 10-15, 15-20, 30, 40, 50 o más veces, preferiblemente alrededor de 12-18 veces de modo que permita análisis estadístico del haplotipo de metilación correcto en el organismo sujeto.

En una forma de realización preferida, los métodos de la presente invención usan espectrometría de masas, por ejemplo, el sistema MASSARRAY™, para el genotipado de las muestras.

Por lo tanto, en un aspecto, la presente invención proporciona un método para determinar un haplotipo de un sujeto que comprende los pasos de diluir una muestra de ácido nucleico del sujeto en una dilución de molécula única; amplificar la dilución de molécula única diluida con al menos dos pares de cebadores diseñados para amplificar una región que comprende al menos dos sitios polimórficos en el molde de ácido nucleico; genotipado de los sitios polimórficos en la molécula única de ácido nucleico; y determinar el haplotipo a partir de los genotipos de al menos los dos sitios polimórficos para obtener un haplotipo del sujeto.

Los pasos de dilución, amplificación y genotipado de la muestra de ácido nucleico del sujeto se repiten al menos tres veces a partir de la misma muestra de ácido nucleico para obtener al menos cuatro réplicas de genotipo del mismo sujeto y posteriormente comparar las al menos cuatro réplicas del genotipo para determinar el haplotipo. Preferiblemente, se obtienen al menos 4, 5, 6, 7, 8-10, 10-15, 15-20, 30, 50, 50-100 o más réplicas de genotipo. En una forma de realización se obtienen al menos 12-18 réplicas y los resultados se analizan estadísticamente, usando por ejemplo un método de distribución de Poisson.

ES 2 329 364 T3

En una forma de realización, el método comprende además comparar el haplotipo con un haplotipo de un control o una base de datos de haplotipos controles para determinar la asociación del haplotipo con un rasgo biológico, que puede ser cualquier rasgo biológico incluyendo pero no limitado a varias enfermedades.

5 Los polimorfismos útiles según la presente invención incluyen, pero no están limitados a polimorfismos de un solo nucleótido (SNP), deleciones, inserciones, sustituciones o inversiones. Los polimorfismos también pueden ser una combinación de uno o más marcadores seleccionados del grupo que consiste en un polimorfismo de un solo nucleótido, deleción, una inserción, una sustitución o una inversión u otros tipos de polimorfismos de ácidos nucleicos.

10 En una forma de realización, el paso de genotipado del método descrito anteriormente se realiza usando extensión de cebadores, preferiblemente tecnología MASSARRAY™ y detección por espectrometría de masas, preferiblemente espectrometría de masas MALDI-TOF.

15 En otro aspecto, la invención proporciona un método de diagnosticar una enfermedad o una susceptibilidad a enfermedad determinando un haplotipo relacionado con la enfermedad en un sujeto que comprende los pasos de diluir una muestra de ácido nucleico del sujeto en una dilución de molécula única; amplificar la dilución de molécula única diluida con al menos dos pares de cebadores diseñados para amplificar una región que comprende al menos dos sitios polimórficos en el molde de ácido nucleico; genotipado de los sitios polimórficos en la molécula única de ácido nucleico; determinar el haplotipo a partir del genotipo de al menos los dos sitios polimórficos para obtener un haplotipo para el sujeto; y comparar el haplotipo del sujeto con haplotipos conocidos asociados a enfermedad en donde una coincidencia en el haplotipo muestra con el haplotipo asociado a enfermedad indica que el sujeto tiene la enfermedad o que el sujeto es susceptible a la enfermedad.

20 El método comprende además repetir los pasos de dilución, amplificación y genotipado al menos tres veces a partir de la misma muestra de ácido nucleico para obtener al menos cuatro réplicas de genotipo del mismo sujeto y posteriormente comparar las al menos cuatro réplicas del genotipo para determinar el haplotipo. Preferiblemente, se producen al menos 4, 5, 6, 7, 8, 9, 10-15, 15-20, 25, 30, 40, 50-100 o más réplicas de genotipo. En una forma de realización, se producen alrededor de 12-18 réplicas.

30 La invención también proporciona un método para determinar un haplotipo de un sujeto que comprende los pasos de tratar una muestra de ácido nucleico del sujeto con una composición que afecta diferencialmente a un nucleótido epigenéticamente modificado en la muestra de ácido nucleico para crear eficazmente polimorfismos basados en la modificación epigenética; diluir la muestra de ácido nucleico tratado en una dilución de copia única; amplificar la muestra de ácido nucleico diluido usando al menos dos pares de cebadores; genotipado de la muestra amplificada; y determinar el haplotipo del sujeto de la muestra genotipada. Los términos modificación “epigenética” o nucleótidos modificados “epigenéticamente” como se describen aquí significan ácidos nucleicos que se modifican mediante metilación, acetilación, o de otra manera epigenética, es decir, mediante adición o deleción de una estructura química o molecular en el ácido nucleico cuya adición o deleción tiene un efecto en el fenotipo del sujeto cambiando la función del ácido nucleico modificado.

40 El método comprende además repetir los pasos de dilución, amplificación y genotipado al menos tres veces para obtener al menos cuatro réplicas de genotipo del mismo sujeto y posteriormente determinar el haplotipo del sujeto basado en las réplicas del genotipo. Se pueden producir al menos 4, 5, 6, 7, 8, 9, 10-15, 15-20, 25, 30, 40, 50-100 o más réplicas de genotipo. En una forma de realización preferida, se producen alrededor de 12-18 réplicas.

45 En una forma de realización, la modificación epigenética es metilación.

50 En aún otra forma de realización, la modificación epigenética es metilación y la composición que usa para tratar el ácido nucleico es bisulfito.

55 La invención también proporciona un método para determinar un haplotipo en un sujeto que comprende los pasos de: digerir una muestra de ácido nucleico del sujeto con una enzima de restricción sensible a metilación de modo que bien el ADN sin metilar bien el ADN metilado se deja intacto, dependiendo de la enzima que se use; diluir la muestra de ácido nucleico digerido a una concentración de copia única; amplificar la muestra de ácido nucleico diluido con al menos dos pares de cebadores diferentes; genotipado de la muestra amplificada; y determinación de un haplotipo de un ácido nucleico metilado en donde al menos dos marcadores polimórficos al lado del sitio de metilación, junto con el sitio de metilación, constituyen un haplotipo.

60 La enzima sensible a metilación puede ser HpaII.

65 El método comprende además repetir los pasos de dilución, amplificación y genotipado al menos tres veces para obtener al menos cuatro réplicas de genotipo del mismo sujeto y posteriormente determinar el haplotipo del sujeto basado en las réplicas del genotipo. Preferiblemente, se producen al menos 4, 5, 6, 7, 8, 9, 10-15, 4, 5, 6, 7, 8, 9, 10-15, 15-20, 25, 30, 40, 50-100 o más réplicas de genotipo. En una forma de realización preferida, se producen alrededor de 12-18 réplicas.

Breve descripción de las figuras

Las figuras 1A-1B muestran un diagrama de flujo del genotipado multiplex de moléculas únicas de ADN para el análisis de haplotipo usando polimorfismos de un solo nucleótido (SNP) como marcadores. Los métodos tradicionales de genotipado que usan unos pocos nanogramos (ng) de ADN genómico (alrededor de 1600 copias de moldes genómicos) producen sólo los genotipos de cada marcador SNP individual, pero las fases de estos SNP no se determinan (mostrado en la parte superior derecha en el espectro de masas en la Fig. 1A). El genotipado simultáneo de varios marcadores usando ensayos multiplex con moléculas únicas de ADN (Fig. 1B) permite el análisis de haplotipo ya que los dos alelos se pueden separar físicamente con concentraciones de ADN muy diluidas, mostrado en la parte inferior derecha en el espectro de masas en la Fig. 1B. Al contrario que otros métodos moleculares de determinación del haplotipo, el bloque del haplotipo entero no se tiene que amplificar en este planteamiento. Por el contrario, sólo se amplifican alrededor de 100 pb alrededor de cada SNP individual para el genotipado, produciendo una eficacia muy alta de la amplificación por PCR de las moléculas únicas de ADN. Los marcadores SNP pueden estar tan separados como se desee, mientras que no haya rupturas significativas entre ellos.

La figura 2 muestra los efectos de la concentración de ADN genómico en la eficacia de la determinación del haplotipo. Se usaron alrededor de 3 pg, 5 pg y 9 pg (ó 1, 1,6 y 3 copias de moldes genómicos humanos, respectivamente) para la determinación del haplotipo de tres marcadores SNP en la región CETP. El número de copias de ADN en una reacción específica se estimó mediante la distribución de Poisson. Los resultados de la determinación del haplotipo pueden ser un ensayo fracasado, determinación con éxito del haplotipo, ambos alelos presentes (sin determinación de fase para los marcadores), o un multiplex incompleto. Excepto para multiplex incompletos, los valores son porcentajes de 54 a 144 ensayos multiplex individuales (ver la especificación y el ejemplo para los detalles sobre el cálculo), seguido por valores predichos usando la distribución de Poisson.

La figura 3 muestra ensayos de genotipado multiplex con solapamiento con moléculas únicas de ADN. Se eligieron siete marcadores SNP (A: rs289744, B: rs2228667, C: rs5882, D: rs5880, E: rs5881, F: rs291044, G: 2033254) de una región genómica de 8 kb del locus CETP (se proporcionan detalles de estos SNP, su posición en el cromosoma y los nucleótidos usados para el genotipado en la Tabla 2). Se diseñaron dos ensayos de genotipado 5-plex para estos 7 marcadores y se usaron los SNP heterocigotos solapantes para obtener el haplotipo entero de 7 marcadores SNP. Se usaron ensayos en el individuo 6 para demostrar cómo se lleva esto a cabo. El ensayo multiplex 1 determinó el haplotipo de 5 SNP como AGAGT y CGGGC. El ensayo multiplex 2 determinó el otro haplotipo de 5 SNP como GGGCT y AGGTT. Después, se usaron los genotipos de los SNP solapantes (SNP C, E, F) para combinar los dos haplotipos de 5 SNP en un haplotipo de 7 SNP que cubre la región entera que se investiga.

Descripción detallada de la invención

La presente invención proporciona un planteamiento de determinación directa de haplotipo de molécula que se basa en un descubrimiento sorprendente que se puede usar una dilución de molécula única de ADN genómico para la separación de dos ADN genómicos homólogos y que usando diluciones repetidas de los mismos organismos sujeto como material de partida para la amplificación multiplex de diferentes marcadores de ácido nucleico, se pueden determinar los haplotipos de cualquier organismo sujeto y son estadísticamente precisos. La muestra diluida, amplificada se genotipa después usando, por ejemplo, el sistema MASSARRAY™ (Fig. 1). El genotipado paralelo de varias diluciones diferentes del mismo sujeto produce la determinación del haplotipo estadísticamente precisa en el organismo sujeto.

El planteamiento de la presente invención difiere significativamente del método de PCR de molécula única previo en que el método de la presente invención no requiere la amplificación de la región genómica completa que contiene los marcadores de interés; de esta manera no está limitado a sólo unas pocas kb de ADN. El método de la presente invención alcanza casi el 100% en tasa de éxito en genotipos y haplotipos para moléculas únicas de ADN. Además, el planteamiento del ensayo de genotipado multiplex permite la determinación directa del haplotipo sin información de genotipo de linaje. Se puede alcanzar fácilmente la determinación de haplotipo de alto rendimiento incorporando el método de la presente invención con cualquier sistema de genotipado comercialmente disponible, tal como el sistema MASSARRAY™.

En una forma de realización, la invención proporciona un método de determinar el haplotipo de un sujeto que comprende los pasos de obtener un ácido nucleico, preferiblemente una muestra de ADN genómico, diluir la muestra de ácido nucleico en sustancialmente una dilución de una única molécula, amplificar la muestra de ácido nucleico con al menos dos pares de cebadores diseñados para amplificar una región genómica que contiene un polimorfismo de ácido nucleico en un cromosoma y genotipado del ADN amplificado. Repetir los pasos desde la dilución de la muestra de ácido nucleico, al menos 3 ó más veces y analizar estadísticamente los resultados, determinando de esta manera el haplotipo de los organismos sujeto.

El "sujeto" como se usa en la especificación se refiere a cualquier organismo con al menos genoma diploide incluyendo, pero no limitado a gusanos, peces, insectos, plantas, murinos y otros mamíferos incluyendo animales domésticos tales como vacas, caballos, perros, gatos, y, lo más preferiblemente, seres humanos.

Los métodos de la presente invención son útiles, por ejemplo, en diagnosticar o determinar un pronóstico en una enfermedad que se sabe está asociada con un haplotipo(s) específico(s), para mapear una enfermedad u otro rasgo

ES 2 329 364 T3

biológico cuya causa es actualmente desconocida a una región cromosómica determinada usando haplotipos en el análisis de ligamiento, para determinar nuevos haplotipos, para detectar asociaciones de haplotipos con capacidad de respuesta a productos farmacéuticos.

5 El ADN genómico se puede obtener o aislar de un sujeto usando cualquier método de aislamiento de ADN conocido para los expertos en la materia. Ejemplos de métodos de aislamiento de ADN se pueden encontrar en manuales generales de laboratorio, tal como Sambrook y Russel, MOLECULAR CLONING: A LABORATORY MANUAL, 3ª Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2001), la totalidad del cual se incorpora aquí mediante referencia.

10 Marcadores polimórficos y oligonucleótidos. El número de ácidos nucleicos polimórficos útiles según la presente invención está en continuo aumento. Actualmente, tales marcadores están fácilmente disponibles de una variedad de bases de datos públicamente disponibles y continuamente se añaden nuevos al conjunto de marcadores disponibles. Se pueden usar como marcadores polimórficos según la presente invención marcadores que incluyen polimorfismos de longitud de restricción, repeticiones cortas en tándem tales como repeticiones de di-, tri- y tetra-nucleótidos así como el estado de metilación. Tales marcadores son bien conocidos para el experto en la materia y se pueden encontrar en varias publicaciones y bases de datos incluyendo, por ejemplo la base de datos de repeticiones cortas en tándem (STR) de la ATCC en <http://www.atcc.org/Cultures/str.cfm>.

20 Marcadores particularmente útiles según la presente invención son polimorfismos de un solo nucleótido (SNP). Ejemplos de bases de datos útiles de SNP incluyen, pero no están limitados a la base de datos de SNP humanos en <http://www-genome.wi.mit.edu/snp/human>, página principal de NCBI dbSNP en <http://www.ncbi.nlm.nih.gov/SNP>, <http://lifesciences.perkinelmer.com/SNPDatabase/welcome.asp>, base de datos de SNP humanos de Celera en <http://www.celera.com/genomics/academic/home.cfm?ppage=cds&cpage=snp>, la base de datos de SNP del grupo de análisis del genoma, (GAN) en <http://www-gan.iarc.fr/SNPdatabase.html>.

30 Un número de cebadores de ácidos nucleicos están ya disponibles para amplificar fragmentos de ADN que contienen los polimorfismos y se pueden obtener sus secuencias, por ejemplo, de las bases de datos identificadas anteriormente. También se pueden diseñar cebadores adicionales, por ejemplo, usando método similar al publicado por Vieux, E.F., Kwok, P-Y y Miller, R. D. in BioTechniques (Junio 2002) Vol. 32. Suplemento: "SNPs: Discovery of Marker Disease", pp. 28-32. También se pueden identificar nuevos SNP usando un método del sistema de MASSARRAY™ Discovery-RT (SNP-Discovery) por SEQUENOM Inc. (San Diego, CA).

35 El experto en la materia conoce un número de diferentes métodos de genotipado de polimorfismos de nucleótidos útiles según la presente invención. Métodos tales como análisis de polimorfismo de longitud de restricción (RFLP), análisis de polimorfismo de conformación de cadena sencilla (SSCP), electroforesis en gel en gradiente desnaturante (DGGE), electroforesis en gel en gradiente de temperatura (TGGE), análisis de corte químico, secuenciación directa de ácidos nucleicos usando marcadores incluyendo pero no limitados a marcadores fluorescentes y radioactivos. Todos estos métodos han estado disponibles durante al menos una década y son bien conocidos para el experto en la materia.

40 El genotipado de SNP se puede realizar usando un número de técnicas diferentes que conoce el experto en la materia. Por ejemplo, el genotipado de SNP mediante espectrometría de masas MALDI-TOF se puede realizar usando, por ejemplo, el sistema de espectrometría de masas de Sequenom MASSARRAY™. En este método, después de que se haya realizado la PCR multiplex usando más de un par de cebadores, cada uno flanqueando diferentes SNP, se realiza una reacción de minisequenciación por extensión de cebador en un único pocillo usando nucleótidos terminadores de cadena. El tamaño de los productos de reacción se determina directamente mediante espectrometría de masas MALDI-TOF, dando la información del genotipo. Debería ser posible basado en esta enseñanza. El multiplexado permite la determinación de, por ejemplo, al menos 2, 3, 4 y 5 SNP en un único pocillo de una, por ejemplo, placa de 384 pocillos. Por ejemplo, se puede realizar genotipado de al menos 6, 7, 8, 9, 10-12-plex usando el sistema MASSARRAY™. El sistema MASSARRAY™, por ejemplo, se puede usar para aumentar el nivel de multiplexidad de las reacciones de genotipado hasta incluso más alto, por ejemplo, al menos 12-15, 20, 30, 40 y 50-100 e incluso más alto.

55 De forma alternativa, el análisis de fragmentos para la detección de SNP se puede realizar por lotes de varias muestras en un sistema de electroforesis capilar, por ejemplo un analizador genético ABI PRISM® 3100 (Applied Biosystem, Foster City, CA). Para el análisis de electroforesis capilar, los cebadores se pueden marcar usando colorantes, incluyendo, pero no limitados a FAM, HEX, NED, LIZ, ROX, TAMRA, PET y VIC.

60 La discriminación alélica de SNP individuales se puede llevar a cabo además usando el sistema de detección de secuencia ABI PRISM® 7900HT (Applied Biosystem, Foster City, CA), que permite el análisis de polimorfismos de un solo nucleótido (SNP) usando el ensayo fluorogénico de 5' nucleasa.

65 Aún otro método útil disponible según la presente invención es una extensión de cebador en matriz (APEX) que es un método de resecuenciación para la identificación rápida de polimorfismos que combina la eficacia de un ensayo basado en micromatrices (alternativo a métodos basados en gel, ver, por ejemplo, la patente de EE.UU. No. 6153379 y Shumaker *et al.* Hum. Mutat. 7(4):346-354, 1996) con el método de secuenciación de ácidos nucleicos de Sanger (Sanger *et al.*, Proc. Natl. Acad. Sci. 74:5463-5467 (1977)). En general, las micromatrices son microchips, por ejemplo

ES 2 329 364 T3

portaobjetos de vidrio, que contienen miles de segmentos de ADN en una disposición ordenada, que permite el análisis simultáneo de miles de marcadores genéticos.

5 Aún otro método de genotipado útil según la presente invención es una técnica de mini-secuenciación en fase sólida, que también se basa en una reacción de extensión de cebador y que se puede usar para el genotipado de SNP y que también se puede automatizar fácilmente (Patente de EE.UU. No. 6013431, Suomalainen *et al*; Mol. Biotechnol. Jun; 15(2):123-31, 2000).

10 En general, una reacción de extensión de cebador es una reacción de secuenciación cíclica modificada en la que está presente al menos un didesoxinucleótido (terminador) y no están presentes todos los desoxinucleótidos a concentración significativa. Cuando un terminador se incorpora en una hebra de ADN, no se puede producir extensión adicional en esa hebra. En una reacción de secuenciación cíclica estándar, los terminadores están presentes sólo en concentraciones pequeñas junto con altas concentraciones de nucleótidos típicos. En reacciones de extensión de una sola base para ensayos de SNP, se usan dos o más nucleótidos terminadores marcados fluorescente o radioactivamente
15 (correspondientes a los dos o más alelos presentes en el SNP cuyo tipo se va a determinar).

Los pasos del método de la presente invención incluyen diluir la muestra de ácido en una dilución de nucleótido único, amplificar la muestra diluida, y determinar el genotipo de la muestra amplificada. Estos pasos se repiten al menos 3 veces, preferiblemente al menos 4, 5, 6, 7, 8, 9, 10-15, 15-20, 20-25, o incluso 25-50 veces. Preferiblemente,
20 los pasos se repiten alrededor de 12-18 veces de modo que los resultados se puedan analizar estadísticamente. Se usa el análisis de distribución de Poisson para analizar los resultados usando los métodos conocidos para los expertos en la materia. El análisis se describe en detalle, por ejemplo, en Stephens *et al*. Am J Hum Genet 46: 1149-1155, 1990.

25 Se define haplotipo como una combinación de alelos o polimorfismos de ácido nucleico, tales como SNP de loci estrechamente unidos que se encuentran en un único cromosoma y que se tienden a heredar juntos. Las recombinaciones suceden con diferente frecuencia en diferentes partes del genoma y por lo tanto, la longitud de los haplotipos varía a lo largo de las regiones cromosómicas y cromosomas. Para un segmento génico específico, con frecuencia hay muchas combinaciones teóricamente posibles de SNP, y por lo tanto hay muchos haplotipos teóricamente posibles.

30 Tradicionalmente, se ha usado la información sobre el flujo genético en una genealogía para reconstruir los haplotipos más probables para familias e individuos. Sin embargo, incluso si las muestras de ácido nucleico de todos los miembros de la familia estuvieran disponibles, que es raramente el caso, el análisis de haplotipos basado en estadística con frecuencia no revela la fase correcta, es decir, haplotipo, de los marcadores. Además, la recogida de grandes materiales de muestra de, por ejemplo familias humanas, lleva mucho tiempo y es cara. La presente invención permite un método en donde se pueden determinar nuevos haplotipos usando polimorfismos de ácidos nucleicos establecidos o nuevos. Por ejemplo, los SNP nuevos primero se identifican usando muestras de ácido nucleico aisladas de varios organismos sujeto de la misma especie, se determina después el genotipo de cada marcador SNP polimórfico de un sujeto por separado, por ejemplo usando alrededor de 1-10 ng, preferiblemente alrededor de 5 ng de ADN genómico.
40 La muestra de ADN genómico se diluye entonces en alrededor de 1 copia de molde genómico por dilución. El haplotipo se determina determinando los SNP en una muestra diluida, es decir, una muestra diluida en sustancialmente una dilución de una única molécula. De forma alternativa, se puede determinar primero el genotipo de la muestra o en paralelo para cada marcador usando una solución de ácido nucleico más concentrada. Esto se puede usar para verificar o controlar la determinación de haplotipo usando réplicas de la muestra diluida.

45 La región genómica de la que se va a determinar el haplotipo usando el método de la presente invención es preferiblemente al menos alrededor de 1, 2, 3, 4, 5, 6, 7, 8, ó 9 kb, más preferiblemente al menos alrededor de 10 kb o más, al menos alrededor de 15 kb o más, al menos alrededor de 20 kb o más. El tamaño de la región que contiene los nucleótidos polimórficos puede ser al menos alrededor de 25 kb o más, al menos alrededor de 35 kb o más, al menos
50 alrededor de 40-45 kb, o 45-50 o incluso alrededor de 50-100 kb o más. Lo más preferiblemente la región genómica puede ser alrededor de 25 kb o más.

Al determinar los haplotipos, tanto las reacciones de PCR como de genotipado son preferiblemente “multiplexadas” término que se pretende que incluya combinar al menos dos, preferiblemente más de al menos 3, 4, 5, 6, 7, 8, 9,
55 10-15 ó 20-25 cebadores de extensión en la misma reacción se usan para identificar, preferiblemente más de al menos 3, 4, 5, 6, 7, 8, 9, 10-15 ó 20-25 regiones polimórficas de ácido nucleico en la misma reacción de genotipado. Se pueden usar al menos 30 pares de cebadores o más.

60 En una forma de realización, el polimorfismo es al menos una modificación de ácido nucleico, tal como una diferencia de metilación. En una forma de realización, la presente invención proporciona un método de determinar haplotipos que comprenden marcadores que incluyen diferencias de metilación. El método de determinar haplotipos por diferencias de metilación según la presente invención comprende los pasos de diluir una muestra de ácido nucleico de un organismo sujeto en dos diluciones paralelas de sustancialmente una única molécula. Las dos diluciones se someten posteriormente a un ensayo de detección de metilación, por ejemplo, un ensayo AFLP (ver, por ejemplo, Vos
65 *et al*. Nucleic Acids Res 23: 4407-4414, 1995; Xu *et al*., Plant Molecular Biology Reporter 18: 361-368, 2000). El ensayo descrito por Vos y col. y Xu y col. se modifica para realizar según el método de la presente invención.

ES 2 329 364 T3

Brevemente, se digieren dos diluciones de molécula única en dos reacciones paralelas con una mezcla que comprende una enzima sensible a metilación y otra enzima, preferiblemente una enzima de restricción de corte menos frecuente, en donde la enzima de restricción de corte menos frecuente en ambas reacciones de digestión es la misma y la enzima sensible a metilación añadida a las dos reacciones paralelas difiere en su capacidad de digerir ácidos nucleicos metilados/no metilados. Por ejemplo, una dilución se digiere con una combinación de EcoRI y HpaII y la dilución paralela se digiere por tratamiento con EcoRI y MspI. Las dos muestras digeridas se ligan después usando un solución de ligación adaptador como se describe en Vos y col. y Xu y col., y se amplifica en reacciones paralelas usando al menos dos, preferiblemente más de dos pares de cebadores que son capaces de reconocer los sitios de reconocimiento de las enzimas de restricción en los moldes. En el ejemplo descrito anteriormente, se usan cebadores EcoRI y HpaII-MspI. Uno de los cebadores está marcado para permitir la detección de los fragmentos de las digestiones usando, por ejemplos métodos electroforéticos en gel o detección de espectrometría de masas.

El procedimiento de detección de metilación como se describe anteriormente se repite al menos 3 veces más, preferiblemente alrededor de 6-12 veces para permitir el análisis estadístico del haplotipo correcto de metilación en el organismo sujeto.

A la luz de esta divulgación, se pueden adaptar fácilmente otras tecnologías de detección de modificación de ácidos nucleicos incluyendo técnicas de detección de metilación para ser usadas según los pasos de principio de la presente invención incluyendo dilución a molécula única, digestión, amplificación multiplex y genotipado multiplex. Los métodos de detección de metilación también se pueden combinar para detectar tanto metilación como otros marcadores polimórficos, tales como los SNP. En tal forma de realización, la amplificación después de la digestión con enzimas de restricción se realiza no sólo con cebadores específicos de metilación sino también con cebadores diseñados para amplificar fragmentos que contienen polimorfismos conocidos de ácidos nucleicos, tales como SNP.

La invención permite un método de crear haplotipos de varios nucleótidos polimórficos usando ensayos de genotipado multiplex con solapamiento con moléculas únicas de ADN. Por ejemplo, se eligen marcadores de una región genómica grande y se realizan una o más reacciones de amplificación multiplex a partir de diluciones de un solo nucleótido y se usan marcadores de polinucleótidos heterocigotos y que solapan para obtener el haplotipo entero.

Por ejemplo, la figura 3 muestra 7 marcadores SNP (A: rs289744, B: rs2228667, C: rs5882, D: rs5880, E: rs5881, F: rs291044, G: 2033254) de una región genómica de 8 kb del locus CETP que se eligieron para determinar un haplotipo. Los detalles de estos SNP, sus posiciones en el cromosoma y los oligonucleótidos usados para el genotipado se proporcionan en la tabla 2. Se diseñaron dos ensayos de genotipado 5-plex para los 7 marcadores y se usaron los SNP heterocigotos solapantes para obtener el haplotipo entero de 7 marcadores SNP. Los ensayos en el individuo No. 6 se usaron para demostrar cómo se lleva esto a cabo. El ensayo multiplex 1 determinó el haplotipo de 5 SNP como AGAGT y CGGGC. El ensayo multiplex 2 determinó el otro haplotipo de 5 SNP como GGGCT y AGGTT. Después, los genotipos de los SNP solapantes (SNP C, E, F) se usaron para combinar los dos haplotipos de 5 SNP en un haplotipo de 7 SNP que cubría la región entera que se investiga.

Ejemplo

Se determinaron los efectos de la concentración de ADN sobre la eficacia de la determinación del haplotipo como sigue. Se usaron 3 picogramos (pg), 5 pg y 9 pg (equivalente de 1, 1,6 y 3 copias de molde genómico) de ADN genómico para la amplificación por PCR y genotipado de 3 SNP en la región CEPT de 12 individuos. Cada ensayo 3-plex se repitió de 12-18 veces para evaluar la eficacia de la PCR y de la determinación del haplotipo. El resultado de un ensayo típico se resume en la Tabla 1. El número de copia de la región de ADN genómico de interés para soluciones de ADN muy diluidas se estima mediante la distribución de Poisson¹³. Los resultados de la determinación del haplotipo se clasificaron en 4 grupos (Tabla 1).

Los ensayos fallidos pueden resultar bien de fallo en la amplificación por PCR de ADN de copia única o simplemente de que no haya molde presente debido a la fluctuación estocástica de soluciones de ADN muy diluidas.

Las determinaciones de genotipado parcialmente fallidas (o multiplex incompletos) son aquellas en las que se han genotipado con éxito sólo 1 ó 2 SNP. Esto es debido lo más probablemente a PCR sin éxito para 1 ó 2 de las regiones de ADN de SNP, ya que en la mayoría de los casos los 3 marcadores SNP están presentes o ausentes al mismo tiempo debido a la cercana proximidad de los marcadores SNP (< 628 pb). La distribución de Poisson también puede producir la presencia de ambos alelos en la solución y por lo tanto la incapacidad de resolver la fase de los SNP.

El análisis con éxito de la determinación de haplotipo se alcanza cuando están presentes una única copia del alelo o copias múltiples del mismo alelo y el genotipado tiene éxito.

El genotipado multiplex incompleto se puede usar para estimar la eficacia del genotipado a partir de moléculas de ADN de copia única. Una determinación de genotipado parcial sugiere la presencia de ADN de SNP pero un fallo para determinar el genotipo de algunos de los SNP. Típicamente se observó un 5-10% de determinaciones de genotipado multiplex incompleto (Fig. 2), lo que sugiere una eficiencia de PCR de alrededor del 90-95% con moléculas únicas de ADN. Este planteamiento puede sobreestimar la eficiencia de PCR, puesto que no se tomaron en consideración los ensayos completamente fallidos. También se llevaron a cabo la comparación detallada entre valores observados

ES 2 329 364 T3

y teóricos de ensayos fallidos, la determinación de haplotipo con éxito y la presencia de ambos alelos (Fig. 2 y ver la sección de métodos para detalles de cálculo). Los valores teóricos se basan en la distribución de Poisson de soluciones de ADN muy diluidas y la asunción de una eficiencia de amplificación por PCR del 100%. El acuerdo cercano entre la estimación teórica y la observación experimental sustancia la estimación anterior de eficiencia de PCR extremadamente alta con moléculas únicas de ADN.

La alta eficiencia de PCR se debe principalmente a la alta eficiencia de amplificación de amplicones muy cortos (típicamente 100 pb) y la alta sensibilidad de la detección por espectrometría de masas MALDI-TOF de oligonucleótidos de ADN. La alta eficiencia de PCR es preferida para el análisis de determinación de haplotipo de alto rendimiento. Por ejemplo, con la actual eficiencia de PCR, se puede alcanzar una eficacia de determinación de haplotipo del 40-45% con una única reacción usando 3-4,5 pg de ADN genómico. Un replicado de 4 ensayos de genotipado multiplex independientes permitirá alrededor del 90% de eficacia directa de determinación de haplotipo.

A continuación se demostró un planteamiento para determinar haplotipos donde hay demasiados marcadores a ser determinados en un ensayo de genotipado multiplex. Se usaron SNP informativos solapantes para combinar haplotipos de varios ensayos multiplex. Se eligieron seis marcadores SNP en una región genómica CETP de 8 kb, y se usaron 2 ensayos de genotipado 4-plex solapantes para el análisis de determinación de haplotipo (Fig. 3). Se pudieron determinar los haplotipos de los 12 individuos para esta región genómica, sin absolutamente ninguna optimización del sistema de ensayo.

El planteamiento presentado aquí proporciona una plataforma de tecnología poderosa y única para el análisis de la determinación molecular directa de haplotipo de regiones genómicas de largo alcance. Este planteamiento es completamente independiente información de genotipo de linaje.

Además se ha incorporado a esta técnica el sistema MASSARRAY™ comercialmente disponible para aplicaciones de alto rendimiento. Esta tecnología es extremadamente útil en la determinación de haplotipo de largo alcance y diagnóstico basado en haplotipo.

Materiales y Métodos

ADN genómicos y oligonucleótidos. Las muestras de ADN genómico humano usadas para la determinación del haplotipo del locus CETP fueron suministradas por SEQUENOM Inc. (San Diego, CA). Estos ADN se aislaron usando el kit de aislamiento de ADN Puregene (Gentra Systems) de muestras de sangre adquiridas del Banco de Sangre (San Bernardino, CA). Los antecedentes personales de los donantes de sangre no son accesibles para estas muestras. Las muestras de ADN genómico humano para la determinación del haplotipo de un segmento de 25 kb en el cromosoma 5q31 fueron ADN de familia de CETP adquiridos de Coriell Cell Repositories (ver Tabla 3). La información sobre los SNP y los oligonucleótidos para el genotipado se proporcionan en la Tabla 2.

Análisis de genotipado y de determinación del haplotipo. Los análisis de genotipado se llevaron a cabo usando el sistema MassArray™ (SEQUENOM Inc.). Cada SNP de cada individuo se genotipó primero por separado usando 5 ng de ADN genómico. Para el análisis de determinación del haplotipo, se llevaron a cabo ensayos de genotipado multiplex usando 3 pg (o aproximadamente 1 copia de molde genómico, a menos que se especifique de otra manera) de ADN genómico.

Análisis de los efectos de la concentración de ADN genómico sobre la determinación del haplotipo. Para calcular el porcentaje de ensayos fallidos, simplemente se contaron todos los ensayos fallidos (sin determinaciones de cada SNP), divididos por el número total de ensayos. Típicamente se hacen de 12 a 18 replicados para cada 6 ó 12 individuos. El porcentaje de ensayos incompletos se calcula de la misma manera. Para calcular el porcentaje de determinaciones con éxito del haplotipo y ambos alelos, se excluyeron los datos de aquellos individuos con haplotipos homocigotos. Las predicciones teóricas se basan en la distribución de Poisson de soluciones de ADN muy diluidas, según un métodos publicado¹³.

ES 2 329 364 T3

TABLA 1

Análisis de haplotipo de muestra con ensayo de genotipado triplex^a

	Repetición	Determinaciones de genotipo
5	1	GGC ^b
	2	GGC
10	3	- ^c
	4	-GC ^d
	5	-
15	6	GGC
	7	-
	8	ACA
20	9	-GC
	10	A/G C/G A/C ^e
	11	ACA
25	12	ACA

^a Se determinaron los genotipos de 3 marcadores SNP con ensayos triplex a partir de 3 pg de ADN genómico.

^b Los 3 SNP son genotipo G, G, C respectivamente

^c Fallo de genotipado para cualquiera de los 3 SNP.

^d Fallo de genotipado del primer SNP, los dos restantes SNP son G y C respectivamente.

^e Fallo para separar los dos alelos, de esta manera los genotipos son A/G, C/A y A/C para los 3 SNP.

Tabla 2. Marcadores de polimorfismo de un solo nucleótido (SNP), sus localizaciones cromosómicas, pares de cebadores para amplificar los marcadores y mezclas terminadoras usadas en la reacción.

SNP ID	Posición crom.	Cebador de PCR 1	Cebador de PCR 2	Cebador de extensión	Mezcla term.
rs289741	47282625	TCTACCAGCTTGGCTCCCTC (SEQ ID NO.: 1)	AAGTCCATCAGCAGCAGCAG (SEQ ID NO.: 2)	GGGAGTCAGCCCAGCTC (SEQ ID NO.: 3)	ACT
rs289742	47282337	ACTGGTGAGACAATCCCTTC (SEQ ID NO.: 4)	CCACTGGCATTAAGTGCTG (SEQ ID NO.: 5)	AGCCACAGAAGAAGGACTCC (SEQ ID NO.: 6)	ACT
rs289744	47281997	TACCAGAAAACCAGACCTCTG (SEQ ID NO.: 7)	ATGCTGGACAGAAAAGTGAG (SEQ ID NO.: 8)	TGAGGATGTTGGGAGGG (SEQ ID NO.: 9)	ACT
rs289744 ^a	47281997	TCTACCAGAAAACCAGACCTC (SEQ ID NO.: 10)	ACTGCTGGACAGAAAAGTGAG (SEQ ID NO.: 11)	ACCTCTAGGGCCCCCTTAC (SEQ ID NO.: 12)	CG
rs2228667 ^a	47282820	CTCGAGTGATAATCTCAGGG (SEQ ID NO.: 13)	AGGTAGTGTTTACAGCCCTC (SEQ ID NO.: 14)	TGATGATGTCTCGAAGAGGCTCATG (SEQ ID NO.: 15)	CG
rs5882 ^a	47284007	TTACGAGACATGACCTCAGG (SEQ ID NO.: 16)	GCATTTGATTTGGCAGAGCAG (SEQ ID NO.: 17)	CTGCAGGAAGCTCTGGATG (SEQ ID NO.: 18)	CG
rs5882 ^b	47284007	GCATTTGATTTGGCAGAGCAG (SEQ ID NO.: 19)	TTACGAGACATGACCTCAGG (SEQ ID NO.: 20)	AGACAGCTCCGAGTCC (SEQ ID NO.: 21)	ACT
rs5880 ^b	47285008	GCAGCACATACTGGAATCC (SEQ ID NO.: 22)	TTTCTCTCCCCAGGATATCG (SEQ ID NO.: 23)	GCCTTTTCTTAGAATAGGAGG (SEQ ID NO.: 24)	ACT
rs5881 ^a	47288087	AGATCTTGGGCACTTTGAGG (SEQ ID NO.: 25)	ACCCCTGTCTTCCACAGGTT (SEQ ID NO.: 26)	TGGGCTGGCTGGGGAAGC (SEQ ID NO.: 27)	CG
rs5881 ^b	47288087	ACCCCTGTCTTCCACAGGTT (SEQ ID NO.: 25)	AGATCTTGGGCACTTTGAGG (SEQ ID NO.: 26)	TGCTTCCACAGGTTCTCGGC (SEQ ID NO.: 27)	ACT

ES 2 329 364 T3

TABLA 3

Muestras de ADN usadas en el ejemplo

5	Número de depósito	Tipo de muestra	Descripción de la muestra	Relación
10	<u>GM12547</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	padre
	<u>GM12548</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	madre
	<u>GM12549</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	hijo
15	<u>GM12550</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	hija
	<u>GM12551</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	hija
	<u>GM12552</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	hijo
20	<u>GM12553</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	hija
	<u>GM12554</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	hija
	<u>GM12555</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	hijo
25	<u>GM12556</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	abuelo paterno
	<u>GM12557</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	abuela paterna
	<u>GM12558</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	abuelo materno
30	<u>GM12559</u>	Linfoblasto	<u>CEPH/LINAJE FRANCÉS 66</u>	abuela materna
	<u>GM07038</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	padre
35	<u>GM06987</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	madre
	<u>GM07004</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	hijo
	<u>GM07052</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	hijo
40	<u>GM06982</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	hijo
	<u>GM07011</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	hija
	<u>GM07009</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	hijo
45	<u>GM07678</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	hijo
	<u>GM07026</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	hijo
	<u>GM07679</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	hijo
50	<u>GM07049</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	abuelo paterno
	<u>GM07002</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	abuela paterna
	<u>GM07017</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	abuelo materno
55	<u>GM07341</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	abuela materna
	<u>GM11820</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1333</u>	hija
60	<u>GM07029</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	padre
	<u>GM07019</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	madre
	<u>GM07062</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	hija
65	<u>GM07053</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	hija

ES 2 329 364 T3

	<u>GM07008</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	hijo
5	<u>GM07040</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	hijo
	<u>GM07342</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	hijo
	<u>GM07027</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	hijo
10	<u>GM06994</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	abuelo paterno
	<u>GM07000</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	abuela paterna
	<u>GM07022</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	abuelo materno
15	<u>GM07056</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	abuela materna
	<u>GM11821</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1340</u>	hijo
20	<u>GM07349</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	padre
	<u>GM07348</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	madre
	<u>GM07350</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	hija
25	<u>GM07351</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	hijo
	<u>GM07352</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	hijo
	<u>GM07353</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	hijo
30	<u>GM07354</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	hija
	<u>GM07355</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	hijo
	<u>GM07356</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	hijo
35	<u>GM07347</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	abuelo paterno
	<u>GM07346</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	abuela paterna
	<u>GM07357</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	abuelo materno
40	<u>GM07345</u>	Linfoblasto	<u>CEPH/LINAJE DE UTAH 1345</u>	abuela materna

Referencias

- 45 Las referencias citadas aquí y a lo largo de la especificación se incorporan a la presente mediante referencia en su totalidad.
- 50 1. **Grupe, A. et al.** In silico mapping of complex disease-related traits in mice. *Science* 292, 1915-8. (2001).
 - 55 2. **Hirschhorn, J.N., Lohmueller, K., Byrne, E. & Hirschhorn, K.** A comprehensive review of genetic association studies. *Genet Med* 4, 45-61. (2002).
 3. **Zhang, S., Pakstis, A.J., Kidd, K.K. & Zhao, H.** Comparisons of two methods for haplotype reconstruction and haplotype frequency estimation from population data. *Am J Hum Genet* 69, 906-14. (2001).
 - 60 4. **Templeton, A.R., Sing, C.F., Kessling, A. & Humphries, S.** A cladistic analysis of phenotype associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. *Genetics* 120, 1145-54. (1988).
 5. **Kruglyak, L.** Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 22, 139-44. (1999).
 - 65 6. **Judson, R, Stephens, J.C. & Windemuth, A.** The predictive power of haplotypes in clinical response. *Pharmacogenomics* 1, 15-26. (2000).

ES 2 329 364 T3

7. **Martin**, ER. *et al.* Analysis of association at single nucleotide polymorphisms in the APOE region. *Genomics* 63, 7-12. (2000).

5 8. **Clark**, A.G. Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol Biol Evol* 7, 111-22. (1990).

9. **Stephens**, M., **Smith**, N.J. & **Donnelly**, P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68, 978-89. (2001).

10 10. **Ruano**, G. & **Kidd**, K.K. Direct haplotyping of chromosomal segments from multiple heterozygotes via allele-specific PCR amplification. *Nucleic Acids Res* 17, 8392. (1989).

11. **Ruano**, G., **Kidd**, K.K. & **Stephens**, J.C. Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules. *Proc Natl Acad Sci USA* 87, 6296-300. (1990).

15 12. **Douglas**, J.A., **Boehnke**, M., **Gillanders**, E., **Trent**, J.M. & **Gruber**, S.B. Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. *Nat Genet* 28, 361-4. (2001).

20 13. **Stephens**, J.C., **Rogers**, J. & **Ruano**, G. Theoretical underpinning of the single-molecule-dilution (SMD) method of direct haplotype resolution. *Am J Hum Genet* 46, 1149-55. (1990).

14. **Daly**, M.J., **Rioux**, J.D., **Schaffner**, S.F., **Hudson**, T.J. & **Lander**, E.S. High-resolution haplotype structure in the human genome. *Nat Genet* 29, 229-32. (2001).

25 15. **Gabriel**, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* 296, 2225-9. (2002).

30

35

40

45

50

55

60

65

REIVINDICACIONES

1. Un método para determinar un haplotipo en un sujeto que comprende los pasos de:

- (a) diluir una muestra de ácido nucleico del sujeto a una dilución de molécula única;
- (b) amplificar la dilución de molécula única con al menos dos pares de cebadores diferentes diseñados para amplificar al menos dos regiones diferentes de ácido nucleico, comprendiendo cada región al menos dos sitios polimórficos en el molde de ácido nucleico;
- (c) determinar el genotipo de los sitios polimórficos en la molécula única de ácido nucleico; y repetir adicionalmente los pasos a-c al menos tres veces a partir de la misma muestra de ácido nucleico para obtener al menos cuatro réplicas de genotipo del mismo sujeto; y
- (d) determinar el haplotipo a partir de las al menos cuatro réplicas de genotipo de al menos dos sitios polimórficos para obtener un haplotipo para el sujeto.

2. El método de la reivindicación 1, en donde los sitios polimórficos pueden estar a cualquier distancia entre sí en el cromosoma.

3. El método de la reivindicación 1, en donde los sitios polimórficos están separados por una distancia de una kilobase o más.

4. El método de la reivindicación 1, que comprende además comparar el haplotipo con un haplotipo de un control a una base de datos de haplotipos de controles para determinar la asociación del haplotipo con un rasgo biológico.

5. El método de la reivindicación 1, en donde el polimorfismo es un polimorfismo de un solo nucleótido.

6. El método de la reivindicación 1, en donde el polimorfismo es una delección, una inserción, una sustitución o una inversión.

7. El método de la reivindicación 1, en donde el polimorfismo es una combinación de uno o más marcadores seleccionados del grupo que consiste en un polimorfismo de un solo nucleótido, delección, una inserción, una sustitución o una inversión.

8. El método de cualquiera de las reivindicaciones 1 y 3 a 7 en donde el genotipado se realiza usando extensión de cebador y detección por espectrometría de masas.

9. El método de la reivindicación 1, en donde se producen 12-18 réplicas de genotipo.

10. Un método para diagnosticar una enfermedad o susceptibilidad a enfermedad determinando un haplotipo relacionado con enfermedad en un sujeto que comprende los pasos de:

- (a) diluir una muestra de ácido nucleico del sujeto a una dilución de molécula única;
- (b) amplificar la dilución de molécula única con al menos dos pares de cebadores diferentes diseñados para amplificar al menos dos regiones diferentes de ácido nucleico, comprendiendo cada región al menos dos sitios polimórficos en el molde de ácido nucleico;
- (c) determinar el genotipo de los sitios polimórficos en la molécula única de ácido nucleico;
- (d) repetir los pasos a-c al menos tres veces a partir de la misma muestra de ácido nucleico para obtener al menos cuatro réplicas de genotipo del mismo sujeto;
- (e) determinar el haplotipo a partir de las al menos cuatro réplicas de genotipo de al menos dos sitios polimórficos para obtener un haplotipo para el sujeto; y
- (f) comparar el haplotipo del sujeto con haplotipos asociados a enfermedad conocidos, en donde una coincidencia en el haplotipo de la muestra con un haplotipo relacionado con enfermedad indica que el sujeto tiene la enfermedad o que el sujeto es susceptible a la enfermedad.

11. El método de la reivindicación 10, en donde se producen 12-18 réplicas.

12. Un método de determinar un haplotipo de un sujeto que comprende los pasos de:

ES 2 329 364 T3

- (a) tratar una muestra de ácido nucleico de un sujeto con una composición que afecta diferencialmente a un nucleótido epigenéticamente modificado en la muestra de ácido nucleico para crear de forma eficaz polimorfismos basados en la modificación epigenética;
- 5 (b) diluir la muestra de ácido nucleico tratada a una dilución de copia única;
- (c) amplificar la dilución de copia única usando al menos dos pares de cebadores diferentes que amplifican al menos dos regiones diferentes de ácido nucleico;
- 10 (d) determinar el genotipo de la muestra amplificada;
- (e) repetir los pasos b-d al menos tres veces para obtener al menos cuatro réplicas de genotipo del mismo sujeto; y
- 15 (f) determinar el haplotipo del sujeto a partir de las al menos cuatro réplicas de genotipo.

13. El método de la reivindicación 12, en donde se producen 12-18 réplicas.

20 14. El método de la reivindicación 12, en donde el nucleótido epigenéticamente modificado es un nucleótido metilado.

15. El método de la reivindicación 14, en donde la muestra de ácido nucleico se trata con bisulfito.

25 16. Un método para determinar un haplotipo en un sujeto que comprende los pasos de:

- (a) digerir una muestra de ácido nucleico del sujeto con una enzima de restricción sensible a metilación de modo que bien el ADN no metilado bien el ADN metilado se deja intacto, dependiendo de la enzima que se use;
- 30 (b) diluir la muestra de ácido nucleico digerida a una dilución de molécula única;
- (c) amplificar la dilución de molécula única usando al menos dos pares de cebadores diferentes;
- 35 (d) determinar el genotipo de la muestra amplificada;
- (e) repetir los pasos b-d al menos tres veces para obtener al menos cuatro réplicas de genotipo del mismo sujeto; y
- 40 (f) determinar un haplotipo de un ácido nucleico metilado a partir de al menos cuatro réplicas de genotipo en donde al menos un marcador polimórfico próximo al sitio de metilación, junto con el sitio de metilación, constituye un haplotipo.

45

50

55

60

65

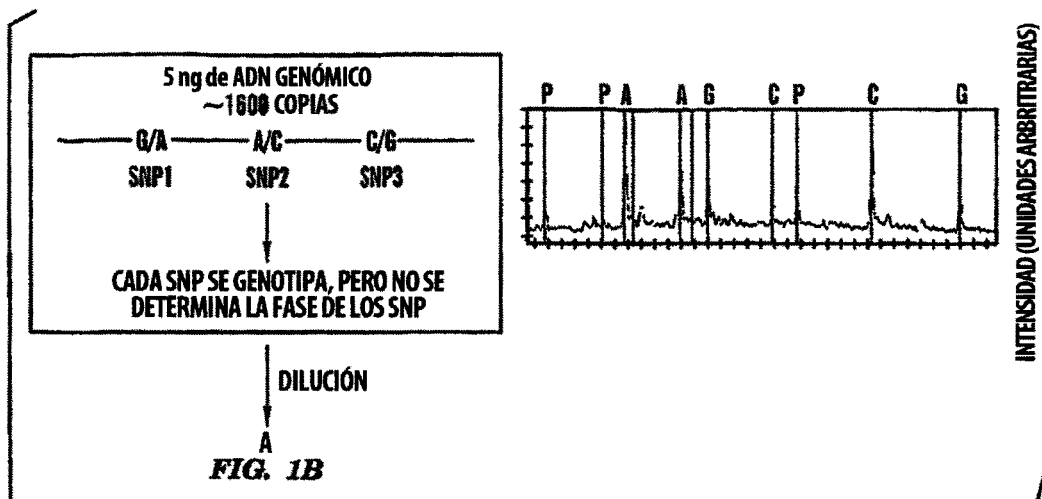


FIG. 1A

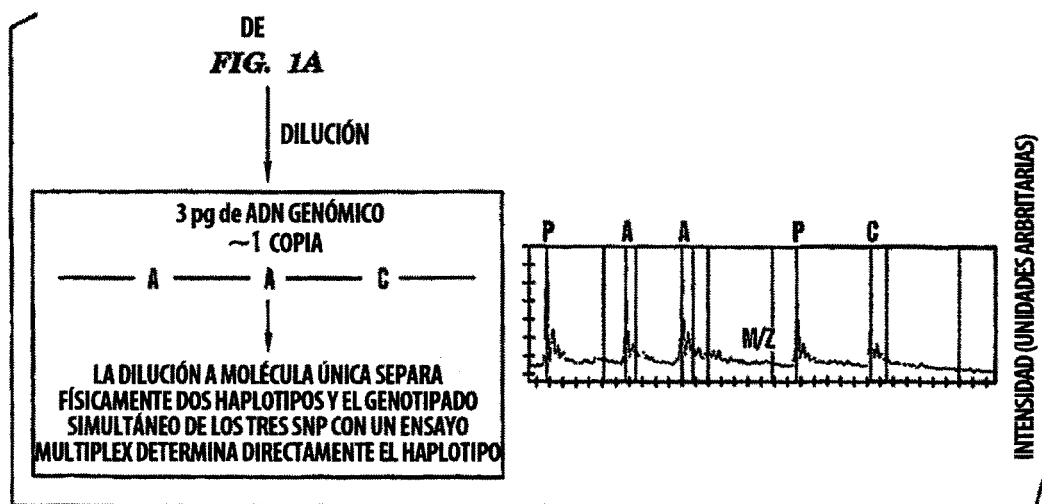


FIG. 1B

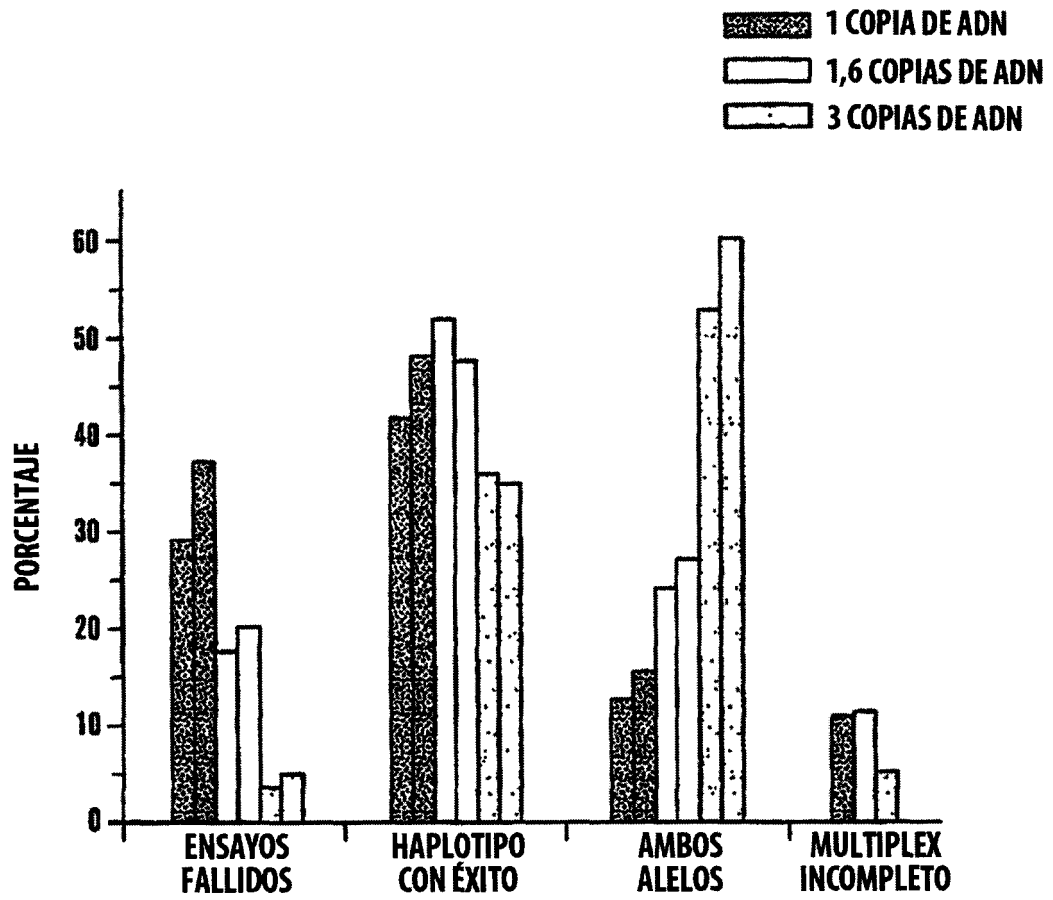


FIG. 2

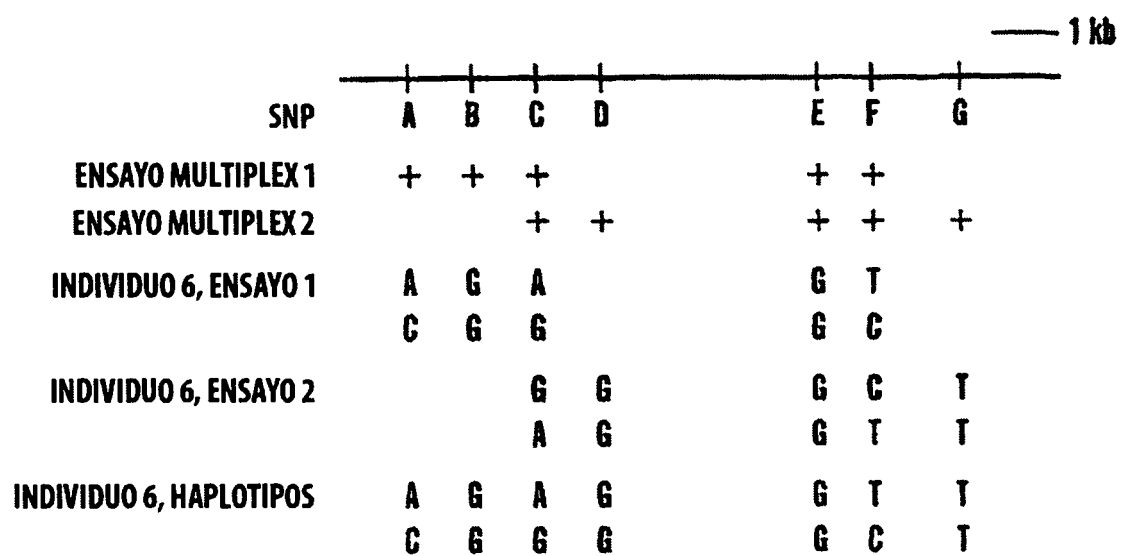


FIG. 3