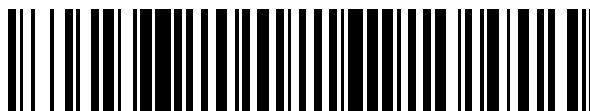


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 528 028**

51 Int. Cl.:

C12Q 1/68 (2006.01)

G06F 19/18 (2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **18.05.2009** **E 09747015 (7)**

97 Fecha y número de publicación de la concesión europea: **12.11.2014** **EP 2288724**

54 Título: **Métodos y sistemas para la selección de portadores universales**

30 Prioridad:

16.05.2008 US 53926 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
03.02.2015

73 Titular/es:

COUNSYL, INC. (100.0%)
180 Kimball Way
South San Francisco, CA 94080, US

72 Inventor/es:

SRINIVASAN, BALAJI S.;
SRINIVASAN, BALAJI K.;
EVANS, ERIC;
SRINIVASAN, RAMJI y
BALAKRISHNAN, KUMARANAYAGAM

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 528 028 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 y sistemas para la selección de portadores universales

5 Antecedentes de la invención

Los análisis genéticos de padres potenciales se pueden usar para predecir la probabilidad de que la descendencia de una pareja tenga determinadas enfermedades genéticas. Personas para las cuales este ensayo es atractivo incluyen aquellos cuyas enfermedades genéticas se llevan en la familia o los pertenecientes a grupos étnicos que tienen una alta incidencia de enfermedades genéticas. Los resultados de estos análisis pueden proporcionar a una pareja información que pueden usar para tomar decisiones acerca de convertirse en padres. Por ejemplo, una pareja puede decidir hacer preparaciones para el cuidado especial que podría ser necesario para criar a un niño con necesidades especiales, o la pareja puede explorar opciones para la reproducción asistida. El documento WO2007/053752 describe métodos basados en ordenador y sistemas para facilitar la evaluación de resultados de análisis de selección genética.

Muchas enfermedades genéticas son raras en una población, por ejemplo, con frecuencias de menos de 1 en 1000. Sin embargo, si una enfermedad genética está causada por un único alelo recesivo de un gen y ambos padres potenciales son portadores del alelo recesivo, entonces la probabilidad de que un niño de la pareja tenga la enfermedad es probablemente del 25 %.

La incidencia de alelos para ciertas enfermedades genéticas es diferente entre diferentes poblaciones. Por ejemplo, en personas con ascendencia judía asquenazí, la frecuencia portadora de la enfermedad de Tay-Sachs es de 1:30; para la enfermedad de Canavan es 1:40; para la enfermedad de Niemann-Pick tipo A es 1:90; para la anemia de Fanconi tipo C es 1:89; para el síndrome de Bloom es 1:100; para la enfermedad de Gaucher tipo 1 es 1:12. En las personas de ascendencia del sudeste asiático, la frecuencia portadora de alfa-talasemia es tan alta como 1:20. En los caucásicos, la frecuencia portadora de fibrosis quística es de 1:25.

Aunque estas enfermedades pueden ser raras individualmente, hay un número suficiente de ellas por lo que la probabilidad de que cualquier individuo sea portador de al menos una es significativamente mayor. Por ejemplo, se ha estimado que el 70 % de la población judía asquenazí tiene al menos un alelo causante de enfermedad. Las estimaciones de carga genética indican que todo ser humano porta de aproximadamente 8 a 30 alelos recesivos deletéreos.

La selección de portadores puede disminuir la incidencia de enfermedades genéticas en una población. Por ejemplo, como resultado de selección de portadores, la incidencia de la enfermedad de Tay Sachs entre la población judía asquenazí ha disminuido en los últimos años. Otro ejemplo de una enfermedad genética mendeliana es la fibrosis quística (CF). La fibrosis quística es a menudo fatal, debilitante e incurable, y costosa de tratar. CF se puede caracterizar por herencia autosómica recesiva y portada por individuos asintomáticos causada por cientos de mutaciones diferentes, cada una de las cuales varía en frecuencia entre los grupos étnicos. El Colegio Americano de Genética Médica (ACMG) recomienda la selección de portadores por análisis genético para todos los padres potenciales para varias enfermedades mendelianas. Desde 1998, se ha recomendado por el ACMG el cribado de toda la población panétnica para el estado portador de fibrosis quística. La lista de enfermedades que el ACMG recomienda ensayar se expande continuamente como una función de los nuevos descubrimientos relacionados con enfermedades genéticas. La selección de portadores se ha adoptado porque los beneficios para la salud son mayores que los costes financieros de ensayar por un margen definitivo.

Actualmente, la selección de los padres potenciales para el estado portador es selectiva. Está indicado para parejas pertenecientes a las poblaciones en riesgo aumentado de afecciones particulares. Hay una necesidad en la técnica de ensayo universal para una amplia diversidad de afecciones genéticas para individuos de cualquier ascendencia.

Sumario de la invención

El alcance de la presente invención se define por las reivindicaciones adjuntas 1 a 19. En un aspecto, se describe un método en este documento que comprende: ensayar un subconjunto de parientes biológicos de un niño o posible niño para determinar la presencia de una pluralidad de variantes genéticas causales correspondientes a al menos una enfermedad genética rara y la presencia de al menos un marcador informativo de ascendencia (AIM); y predecir una probabilidad de un fenotipo del niño o posible niño del subconjunto de parientes biológicos con respecto a la al menos una enfermedad genética rara basándose al menos en parte en la presencia de la pluralidad de variantes genéticas causales y la presencia del al menos un AIM. En algunos casos, la predicción comprende realizar un análisis totalmente probabilístico en los datos recogidos sobre dichas variantes genéticas causales. En algunos casos, la predicción se basa además en información fenotípica sobre al menos un miembro del subconjunto de parientes biológicos. El método puede comprender además la prestación de servicios de asesoramiento genético a al menos un miembro del subconjunto de parientes biológicos. El método puede comprender además la entrega de la probabilidad del fenotipo del niño a un servicio médico de referencia.

En un aspecto, un medio legible por ordenador comprende: lógica configurada para predecir una probabilidad de un fenotipo de un niño o posible niño a partir de un subconjunto de parientes biológicos con respecto a al menos una enfermedad genética rara basándose al menos en parte en los resultados de un ensayo del subconjunto de parientes biológicos para la presencia de una pluralidad de variantes genéticas causales y al menos un marcador informativo de ascendencia (AIM). En algunos casos, la lógica realiza un análisis totalmente probabilístico. En algunos casos, el medio legible por ordenador proporciona una salida en forma de un informe que detalla la presencia del al menos un AIM y la presencia de cualquiera de la pluralidad de variantes genéticas causales en cualquier miembro del subconjunto de parientes biológicos.

En un aspecto, un medio legible por ordenador comprende: lógica configurada para realizar un análisis totalmente probabilístico en los datos correspondientes a una pluralidad de variantes genéticas causales de un hombre y una mujer para predecir una probabilidad de un fenotipo de un niño, donde el hombre y la mujer son padres potenciales del niño. En algunos casos, las variantes genéticas causales de cada uno del hombre y la mujer comprenden uno o más marcadores informativos de ascendencia (AIM), una o más variantes genéticas causales correspondientes a una enfermedad genética rara, una o más variantes genéticas causales correspondientes a un rasgo de la personalidad, o ambos. En algunos casos, el análisis totalmente probabilístico incorpora una pluralidad de fuentes de incertidumbre estadística en la probabilidad. Un medio legible por ordenador en este documento puede comprender además lógica para recibir la entrada de una batería de fenotipos y la asignación de una función de ponderación a la pluralidad de variantes genéticas causales basad en dicha entrada. En algunos casos, la entrada de la batería fenotipos comprende: altura, peso, y la historia de la enfermedad familiar. En algunos casos, el medio legible por ordenador proporciona una salida en forma de un informe que detalla una distribución de probabilidad sobre los riesgos o fenotipos del niño.

En un aspecto, se describe un sistema en el presente documento para predecir un fenotipo de un niño que comprende: un dispositivo de detección de ácido nucleico configurado para detectar una pluralidad de variantes genéticas causales correspondientes a al menos una enfermedad genética rara y al menos un marcador informativo de ascendencia (AIM), donde el dispositivo está en contacto con una muestra de un pariente biológico de un niño o posible niño; un lector configurado para leer los datos de los dispositivos; e instrucciones legibles por ordenador, donde las instrucciones cuando se ejecutan utilizan los datos desde el lector correspondientes a la pluralidad de variantes genéticas causales y los datos del al menos un AIM para predecir una probabilidad de un fenotipo del niño con respecto a la al menos una enfermedad genética rara. En algunos casos, el pariente biológico es una madre potencial o padre potencial del niño o posible niño. En algunos casos, el dispositivo de detección de ácido nucleico comprende una pluralidad de sondas de ácido nucleico que se unen selectivamente a la pluralidad de variantes genéticas causales y el al menos un AIM.

En un aspecto, un sistema para predecir un fenotipo del niño comprende: un dispositivo de detección de ácido nucleico configurado para detectar una pluralidad de variantes genéticas causales correspondientes a más de 85 enfermedades genéticas raras, donde el dispositivo está en contacto con una muestra de una pariente biológico del niño o posible niño; un lector configurado para leer los datos de los dispositivos; e instrucciones legibles por ordenador, donde las instrucciones cuando se ejecutan utilizan los datos desde el lector correspondientes a la pluralidad de variantes genéticas causales para predecir una probabilidad de un fenotipo del niño o posible niño con respecto a las más de 85 enfermedades genéticas raras. En algunos casos, el pariente biológico es una madre potencial o padre potencial del niño o posible niño. En algunos casos, el dispositivo de detección de ácido nucleico comprende una pluralidad de sondas de ácido nucleico que se unen selectivamente a la pluralidad de variantes genéticas causales correspondientes a más de 85 enfermedades genéticas raras. En algunos casos, el dispositivo de detección de ácido nucleico comprende además una pluralidad de sondas de ácido nucleico que se unen selectivamente a al menos un marcador informativo de ascendencia (AIM). En algunos casos, las instrucciones legibles por ordenador cuando se ejecutan utilizan los datos desde el lector correspondientes a al menos un AIM para predecir la probabilidad de un fenotipo del niño.

En un aspecto, un sistema para indicar si un sujeto es portador de una enfermedad genética rara comprende: un lector configurado para leer datos desde un dispositivo de detección de ácido nucleico configurado para detectar una pluralidad de variantes genéticas causales correspondientes a al menos una enfermedad genética rara y al menos un marcador informativo de ascendencia (AIM); e instrucciones legibles por ordenador, donde las instrucciones cuando se ejecutan utilizan los datos desde el lector correspondientes a la pluralidad de variantes genéticas causales y el al menos un marcador informativo de ascendencia para predecir una pluralidad de probabilidades de que el sujeto sea un portador para cada una de la pluralidad de variantes genéticas causales.

En un aspecto, un método comprende: recibir una muestra de un usuario; ensayar la muestra con un dispositivo de detección de ácido nucleico configurado para ensayar una pluralidad de variantes genéticas causales de enfermedades genéticas raras y al menos un marcador informativo de ascendencia (AIM); calcular una pluralidad de probabilidades para los posibles genotipos del usuario en cada localización de la pluralidad de variantes genéticas causales basadas en los resultados de la etapa de ensayo respecto a la pluralidad de variantes genéticas causales y el al menos un AIM; y entregar al usuario la pluralidad de probabilidades correspondientes al usuario que es portador. El método puede comprender además: recibir una muestra de un segundo usuario; ensayar la muestra del segundo usuario con un dispositivo configurado para ensayar una pluralidad de variantes genéticas causales de

enfermedades genéticas raras y al menos un marcador informativo de ascendencia (AIM); calcular una probabilidad de un fenotipo del niño correspondiente a las enfermedades genéticas raras basada en los resultados del ensayo del usuario y el segundo usuario; y entregar la probabilidad del fenotipo del niño a al menos uno del usuario y el segundo usuario. El método puede comprender además proporcionar servicios de asesoramiento genético para al menos a uno del usuario y el segundo usuario. El método puede realizarse como parte de un servicio de predicción de fenotipo del niño. El método puede comprender además la obtención de información fenotípica del usuario; y el uso de la información fenotípica del usuario en las etapas de cálculo. El método puede comprender además la obtención de la historia familiar del usuario; y el uso de la historia familiar del usuario en las etapas de cálculo.

En un aspecto, un dispositivo de detección de ácido nucleico está configurado para ensayar una muestra para una pluralidad de variantes genéticas causales correspondiente a al menos una enfermedad genética rara y uno o más marcadores informativos de ascendencia (AIM). En algunos casos, el dispositivo comprende una pluralidad de sondas de ácido nucleico que se unen selectivamente a la pluralidad de variantes genéticas causales y el uno o más AIM. En algunos casos, el dispositivo comprende una serie de perlas que se unen selectivamente a la pluralidad de variantes genéticas causales y el uno o más AIM.

En otro aspecto, un dispositivo de detección de ácido nucleico está configurado para ensayar una muestra para una pluralidad de variantes genéticas causales correspondientes a más de 85 enfermedades genéticas raras. El dispositivo puede estar configurado adicionalmente para ensayar una muestra para al menos un marcador informativo de ascendencia (AIM). En algunos casos, el dispositivo comprende una pluralidad de sondas de ácido nucleico que se unen selectivamente a una pluralidad de variantes genéticas causales correspondientes a más de 85 enfermedades genéticas raras. El dispositivo puede comprender además una pluralidad de sondas de ácido nucleico que se unen selectivamente a al menos un marcador informativo de ascendencia (AIM). El dispositivo puede comprender una serie de perlas que se unen selectivamente a una pluralidad de variantes genéticas causales correspondientes a más de 85 enfermedades genéticas raras. El dispositivo puede comprender además una serie de perlas que se unen selectivamente a al menos un marcador informativo de ascendencia (AIM). En algunos casos, el dispositivo comprende un ensayo de resecuenciación para detectar la pluralidad de variantes genéticas causales correspondientes a las más de 85 enfermedades genéticas raras. En algunos casos, el dispositivo comprende además un ensayo de resecuenciación para detectar al menos un marcador informativo de ascendencia (AIM).

En algunos casos, el al menos un AIM no es una variante genética causal. En algunos casos, al menos dos de las enfermedades genéticas raras ocurren a frecuencias que difieren en al menos 10 veces en al menos dos poblaciones distintas, donde las al menos dos poblaciones distintas se diferencian por el al menos un AIM.

En un aspecto, un método comprende: la comercialización de un servicio de análisis genético que comprende predecir una probabilidad de un fenotipo del niño a partir de un subconjunto de los parientes biológicos del niño o posible niño, donde la predicción se basa al menos en parte en la presencia de una pluralidad de variantes genéticas causales en cada uno del subconjunto de parientes biológicos y se basa al menos en parte en los ascendientes inferidos de cada uno del subconjunto de parientes biológicos; y la entrega de una probabilidad del fenotipo del niño por una tarifa. La comercialización puede realizarse en relación con un servicio de citas o matrimonio. El método puede comprender, además, remitir a al menos un miembro del subconjunto de parientes biológicos a un médico. En algunos casos, los ascendientes inferidos se infieren por un ensayo para al menos un marcador informativo de ascendencia (AIM).

En un aspecto, se describe un conjunto de combinaciones de ácido nucleico para validar un dispositivo de detección de secuencia de ácido nucleico que comprende un conjunto de sondas de variantes genéticas causales, donde cada combinación de ácido nucleico comprende una pluralidad de segmentos de ácido nucleico que se unen selectivamente a un subconjunto diferente del conjunto de sondas variantes de genéticas causales. En algunos casos, una primera combinación del conjunto comprende un primer segmento de ácido nucleico que interfiere durante la detección con un segundo segmento de ácido nucleico de una segunda combinación del conjunto, y donde la primera combinación no comprende el segundo segmento de ácido nucleico y la segunda combinación no comprende el primer segmento de ácido nucleico. En algunos casos, los segmentos de ácido nucleico de cada combinación son moléculas de ácido nucleico monocatenarias. En algunos casos, los segmentos de ácido nucleico comprenden uno o más plásmidos.

En un aspecto, un método para validar un lote de dispositivos de detección de secuencia de ácido nucleico fabricados comprende: poner en contacto cada uno de una pluralidad de dispositivos de detección de secuencia de ácido nucleico del lote con una combinación de ácido nucleico diferente, donde cada combinación de ácido nucleico comprende una pluralidad de segmentos de ácido nucleico que se unen selectivamente a una pluralidad de sondas de variantes genéticas causales en dichos dispositivos de detección, y donde cada combinación de ácido nucleico se une a un conjunto diferente de la pluralidad de sondas de variantes genéticas causales; y detectar la presencia o ausencia de la pluralidad de sondas de variantes genéticas causales en la pluralidad de dispositivos de detección de ácido nucleico, donde el lote de dispositivos fabricados se valida si toda la pluralidad de sondas de variantes genéticas causales está presente en la pluralidad de dispositivos de detección de ácido nucleico.

En algunos casos, el método comprende además la entrega del lote de dispositivos fabricados cuando los

dispositivos están validados. En algunos casos, el lote de dispositivos fabricados se rechaza si no todas de la pluralidad de sondas de variantes genéticas causales están presentes. En algunos casos, el lote de dispositivos fabricados se modifica y el método se repite si no todas de la pluralidad de sondas de variantes genéticas causales están presentes.

5

Breve descripción de los dibujos

Muchas de las características de la invención se exponen con particularidad en las reivindicaciones adjuntas. Una mejor comprensión de las características y ventajas de este documento se obtendrá por referencia a la siguiente descripción detallada que expone realizaciones ilustrativas, en que se utilizan muchos principios, y los dibujos adjuntos de los cuales:

10

La Figura 1 proporciona una tabla de variantes genéticas causales.

La Figura 2 proporciona una lista de poblaciones a modo de ejemplo.

15

La Figura 3 proporciona varios AIM que distinguen diferentes poblaciones. Las entradas se refieren a los elementos de la base de datos dbSNP, una base de datos de variantes genéticas mantenida por el gobierno de Estados Unidos: <http://www.ncbi.nlm.nih.gov/projects/SNP/>. Los registros conservados en dbSNP contienen información que describe la secuencia y localización de las variantes genéticas, y cuando está disponible, la frecuencia de los alelos de esas variantes en diferentes poblaciones. Los números rs (por ejemplo, rs332, RS25, etc.) son los números de identificación usados para indexar la parte de la base de datos dbSNP.

20

La Figura 4 ilustra una población 1 como un grupo que tiene un hipotético SNP A/G con la secuencia flanqueante dada.

La Figura 5 ilustra la población 2 que tiene exactamente la misma variante con la misma secuencia flanqueante que la Figura 4, pero diferentes proporciones de cada genotipo (0,16, 0,48, 0,36 en lugar de 0,25, 0,50, 0,25).

25

La Figura 6 ilustra la población 3 que tiene una secuencia flanqueante diferente a las Figuras 4 y 5 con una G adicional.

La Figura 7 considera poblaciones por género, donde las Figuras 4-6 consideran las poblaciones en términos de ascendencia geográfica.

30

La Figura 8 muestra un diagrama de dispersión de intensidad para el alelo de delección de beta talasemia similar a los diagramas de dispersión de las Figuras 4-7.

La Figura 9 muestra que los puntos de designación cerca de la frontera están asistidos por el uso de la información de ascendencia.

La Figura 10 muestra que una medición de la intensidad en bruto en un locus causal no es el único valor que importa.

35

La Figura 11 ilustra un ejemplo de código fuente que es útil en la designación.

La Figura 12 muestra una modificación sencilla de la expresión para la probabilidad posterior de si un individuo está en un grupo determinado.

La Figura 13 muestra que un genotipo medido puede tener error de medición representado por una tasa de falso positivo y falso negativo.

40

La Figura 14 muestra que el genotipo inferido está desfasado.

La Figura 15 muestra cómo generar una distribución más posibles recombinantes y, por tanto, posibles gametos a partir de los posibles haplotipos.

La figura 16 muestra cómo repetir este proceso tanto para la madre como para el padre para obtener una distribución de probabilidad sobre las uniones gaméticas, correspondiente a los genotipos organizados del niño (haplotipos aka).

45

La Figura 17 muestra varios tipos diferentes de mapas genotipo-a-fenotipo que pueden usarse con la distribución sobre los haplotipos del niño para producir una distribución sobre posibles fenotipos del niño.

La Figura 18 representa ecuaciones para obtener la distribución sobre los fenotipos estimados del niño dados los genotipos de los padres.

50

La Figura 19 muestra ejemplos de segmentos de ácido nucleico y combinaciones de ácidos nucleicos.

Las Figuras 20-23 muestran procesos a modo de ejemplo de la entrega de una probabilidad de que un usuario es portador de la enfermedad genética rara.

La Figura 24 ilustra etapas a modo de ejemplo de entrada y salida para la generación de informes para los dos padres hipotéticos: mama gallina y papá gallina.

55

Descripción detallada de la invención

Se describen en este documento procedimientos, instrucciones legibles por ordenador, software, sistemas y dispositivos que se utilizan para detectar genotipos de individuos. En este documento, los genotipos se refieren a detectar variantes genéticas causales específicas que causan o pueden causar enfermedades genéticas raras. En muchos casos, las enfermedades genéticas raras son enfermedades mendelianas, donde un individuo es portador de un rasgo (recesivo o dominante) correspondiente a principalmente relacionados con un solo gen. Los métodos, instrucciones legibles por ordenador, software, sistemas y dispositivos pueden usarse para genómica de familia, como generación de distribuciones de probabilidad para muchos miembros de la familia en una familia incluyendo, sin limitaciones: abuelos, padres, hijos, tías y tíos. Los miembros de la familia pueden estar vivos, muertos, embrionarios, o aún no concebidos. El genotipo y/o fenotipo de cualquier miembro de la familia se pueden utilizar

60

65

para obtener información sobre otros miembros de la familia. Los métodos, instrucciones legibles por ordenador, software, sistemas y dispositivos pueden utilizarse como una criba para predecir la probabilidad del fenotipo de un niño o posible niño a partir de dos individuos. En muchos casos, el fenotipo del niño se predice antes de la concepción en base a los genotipos (y algunas veces fenotipos) del padre potencial y la madre potencial. La predicción del fenotipo del niño puede ser útil en diversas circunstancias, como se describe en este documento y proporciona una oportunidad para que un individuo tome una decisión o tome acciones en base a su genotipo personal.

I. Métodos de análisis y predicción

En un aspecto, un método comprende: analizar cada miembro de un conjunto de padres o padres potenciales para determinar la presencia de una pluralidad de variantes genéticas causales correspondientes a al menos una enfermedad genética rara y la presencia de al menos un marcador informativo de ascendencia (AIM); y predecir una probabilidad de un fenotipo de un niño a partir del conjunto de padres o padres potenciales con respecto a la al menos una enfermedad genética rara basada al menos en parte en los resultados de la presencia de la pluralidad de variantes genéticas causales y el al menos un AIM. En algunos casos la etapa de predicción se basa además en información fenotípica sobre al menos un miembro del conjunto de padres o padres potenciales, o información genotípica y/o fenotípica sobre otros parientes biológicos del niño. En algunos casos el método comprende además la prestación de servicios de asesoramiento genético. En algunos casos, una probabilidad sobre el genotipo de una persona, fenotipo, o fenotipo potencial puede entregarse a un servicio médico de referencia. En algunos casos el método comprende además la entrega de la probabilidad del fenotipo del niño a un servicio médico de referencia.

La predicción de la probabilidad de un fenotipo en este documento puede ser una distribución de probabilidad sobre diversos riesgos para los rasgos. La distribución de probabilidad puede ser categórica o continua para cualquier predicción de probabilidad en este documento. Como un ejemplo de una distribución de probabilidad sobre los riesgos del niño, la predicción podría ser $P(D = 1) = 0,80$ y $P(D = 0) = 0,20$, lo que representa una probabilidad del 80 % de que un niño esté afectado ($D = 1$) por una enfermedad mendeliana. Como un ejemplo de una distribución de probabilidad sobre fenotipos del niño, la predicción podría ser una distribución normal con media y varianza especificadas por la posible altura del niño como un adulto.

Una enfermedad genética puede ser cualquier enfermedad que esté influida por una variante genética causal conocida. En algunos casos, una enfermedad genética rara es una enfermedad que está presente a una tasa de 1 por 100 en la población humana. En otros casos, una enfermedad genética rara es una enfermedad que está presente a una tasa de 1 por 1000 en la población humana. En otros casos más, una enfermedad genética rara es una enfermedad que está presente a una tasa de 1 por 10 000, 1 por 100.000, 1 por 1.000.000, o 1 por 10.000.000 en la población humana. En algunos casos, la enfermedad genética rara puede estar presente a una tasa mucho más elevada que 1 por 100 en una población humana étnica o ancestral específica, por ejemplo, 1 por 50, 1 por 20, 1 por 10, o 1 por 5 en una población específica.

A. Enfermedad genética rara

Los métodos, sistemas, software y dispositivos en este documento pueden ensayar o estar configurados para detectar rasgos fenotípicos (también mencionados simplemente como rasgos), es decir, una forma distinta de una característica de un organismo. Por ejemplo, el color de ojos es una característica física; marrón, verde y azul son rasgos fenotípicos. Algunas características físicas relacionadas con la salud tienen, como rasgos asociados, anormales y normales, tal como enfermedad de Huntington (anormal) y no enfermedad de Huntington (normal). Los rasgos pueden ser morfológicos, de desarrollo, bioquímicos, fisiológicos o de comportamiento. La colección de una pluralidad de rasgos fenotípicos exhibidos por un individuo normalmente se menciona como fenotipo del individuo.

Los rasgos mendelianos son rasgos que se heredan mediante un solo gen o están influenciados principalmente por un solo gen. Las enfermedades mendelianas son enfermedades heredadas como un rasgo mendeliano. Los rasgos mendelianos típicos pueden ser autosómicos dominantes, autosómicos recesivos o ligados al sexo (ligados al X o ligados al Y). Los rasgos mendelianos también pueden ser atípicos, incluyendo rasgos con patrones de herencia atípicos (por ejemplo, herencia citoplasmática o penetrancia incompleta) e incluyendo rasgos que surgen como resultado de mutaciones espontáneas pero predecibles (por ejemplo, expansiones de polimorfismos de repetición o puntos calientes de mutación). Los rasgos mendelianos incluyen tanto rasgos mendelianos típicos como rasgos mendelianos atípicos. Normalmente, el porcentaje de varianza en un rasgo mendeliano que se explica por el genotipo es muy elevado, por ejemplo, al menos el 99 %. Sin embargo, el porcentaje de varianza en el fenotipo en un rasgo mendeliano atípico que se explica por el genotipo puede ser menor.

En ciertos casos, dos genes diferentes causan rasgos que parecen similares o idénticos. En este caso, es típico definir subtipos del rasgo de modo que un solo gen se asocie con uno único sub-tipo de rasgo. Por ejemplo, la enfermedad mucopolisacaridosis normalmente se describe además por ser de un tipo particular, tal como mucopolisacaridosis Tipo I o mucopolisacaridosis Tipo VII.

Los rasgos no mendelianos, también llamados rasgos complejos, se heredan mediante más de un gen.

Normalmente, para los rasgos no mendelianos el porcentaje de varianza en el fenotipo explicado por marcadores genotípicos explícitos es bajo, a menudo menos del 50 %. Un rasgo no mendeliano difiere de un rasgo mendeliano atípico en que las diferencias entre los individuos en un rasgo complejo se deben a diferencias en más de un gen mientras que las diferencias entre los individuos para un rasgo mendeliano atípico se deben principalmente a diferencias en un gen. No todos los rasgos no mendelianos tienen varianza explicada genotípicamente de menos del 99 % (Friedman, Naomi P.; Miyake, Akira; Young, Susan E.; DeFries, John C.; Corley, Robin P.; Hewitt, John K. Individual differences in executive functions are almost entirely genetic in origin.. *Journal of Experimental Psychology: General*. 2008 May Vol. 137(2) 201-225). Ejemplos de rasgos no mendelianos son altura, peso y color de piel.

El genotipo de un individuo para un rasgo mendeliano es la identidad de los genes (o gen en el caso de los rasgos ligados al sexo) responsables del rasgo que el individuo porta, por ejemplo, homocigótico para el gen responsable del rasgo, homocigótico para un gen responsable de un rasgo diferente, o heterocigótico.

Las enfermedades genéticas raras, como se describe en este documento son un tipo de rasgo. Una enfermedad genética rara para los propósitos de este documento es una enfermedad que es un rasgo que puede heredarse genéticamente por un niño a partir de un conjunto de padres. La enfermedad o rasgo puede entonces manifestarse en sí mismo en el fenotipo del niño.

Las enfermedades genéticas raras que pueden ensayarse de acuerdo con esta invención incluyen, aunque sin limitación: deficiencia en 21-hidroxilasa, hiperinsulinismo relacionado con ABCC8, ARSACS, acondroplasia, acromatopsia, adenosina monofosfato desaminasa 1, agenesia de cuerpo calloso con neuropatía, alcaptonuria, deficiencia en alfa-1-antitripsina, alfa-manosidosis, alfa-sarcoglicanopatía, alfa-talasemia, enfermedad de Alzheimer, receptor de angiotensina II, Tipo 1, genotipado de apolipoproteína E, argininosuccinaciduria, aspartilglicosaminuria, ataxia con deficiencia de vitamina E, ataxia-telangiectasia, síndrome de poliendocrinopatía autoinmune tipo 1, cáncer hereditario BRCA1 de mama/ovario, cáncer hereditario BRCA2 de mama/ovario, síndrome de Bardet-Biedl, distrofia macular viteliforme de Best, beta-sarcoglicanopatía, beta-talasemia, deficiencia en biotinidasa, síndrome de Blau, síndrome de Bloom, trastornos Relacionados con CFTR, lipofuscinosis ceroides neuronal relacionada con CLN3, lipofuscinosis ceroides neuronal relacionada con CLN5, lipofuscinosis ceroides neuronal relacionada con CLN8, enfermedad de Canavan, deficiencia en carnitina palmitoiltransferasa IA, deficiencia en carnitina palmitoiltransferasa II, hipoplasia de cartílago piloso, malformación cavernosa cerebral, Coroideremia, síndrome de Cohen, cataratas congénitas, dismorfia facial y neuropatía, enfermedad congénita de glicosilación Ia, trastorno congénito de glicosilación Ib, nefrosis congénita finlandesa, enfermedad de Crohn, cistinosis, DFNA 9 (COCH), diabetes y pérdida de audición, distonía primaria de aparición temprana (DYT1), epidermolísis bullosa juntural tipo Herlitz-Pearson, anemia de Fanconi relacionada con FANCC, craneosinostosis relacionada con FGFR1, craneosinostosis relacionada con FGFR2, craneosinostosis relacionada con FGFR3, trombofilia de Leiden de Factor V, trombofilia de mutación R2 de Factor V, deficiencia en Factor XI, deficiencia en Factor XIII, poliposis adenomatosa familiar, disautonomía familiar, hipercolesterolemia familiar Tipo B, fiebre mediterránea familiar, trastornos de almacenamiento de ácido siálico libre, demencia frontotemporal con Parkinsonismo-17, deficiencia en fumarasa, pérdida de audición y sordera no sindrómica DFNA3 relacionada con GJB2, pérdida de audición y sordera no sindrómica DFNB1 relacionada con GJB2, miopatías relacionadas con GNE, galactosemia, enfermedad de Gaucher, deficiencia en glucosa-6-fosfato deshidrogenasa, glutaricacidemia Tipo 1, enfermedad de almacenamiento de glucógeno Tipo 1a, enfermedad de almacenamiento de glucógeno Tipo 1b, enfermedad de almacenamiento de glucógeno Tipo II, enfermedad de almacenamiento de glucógeno Tipo III, enfermedad de almacenamiento de glucógeno Tipo V, síndrome de Gracile, hemocromatosis hereditaria asociada a HFE, AIM Halder, beta-talasemia de Hemoglobina S, intolerancia hereditaria a la fructosa, pancreatitis hereditaria, timina-uraciluria hereditaria, deficiencia en hexosaminidasa, displasia ectodérmica hidrótica 2, homocistinuria causada por deficiencia en cistationina beta-sintasa, parálisis periódica hipercalémica Tipo 1, síndrome de hiperornitinemia-hiperamoniemia-homocitrulinuria, hiperoxaluria primaria Tipo 1, hiperoxaluria primaria Tipo 2, hipocondroplasia, parálisis periódica hipocalémica Tipo 1, parálisis periódica hipocalémica Tipo 2, hipofosfatasa, miopatía infantil y acidosis láctica (formas fatal y no fatal), academias isovaléricas, enfermedad de Krabbe, LGMD21, neuropatía óptica hereditaria de Leber, síndrome de Leigh, tipo franco-canadiense, deficiencia en 3-hidroxiacil de cadena larga-CoA deshidrogenasa, MELAS, MERRF, deficiencia en MTHFR, variante termolábil de MTHFR, pérdida de audición y sordera relacionada con MTRNR1, pérdida de audición y sordera relacionada con MTS1, poliposis asociada a MYH, enfermedad urinaria de jarabe de arce Tipo 1A, enfermedad urinaria de jarabe de arce Tipo 1B, síndrome de McCune-Albright, deficiencia en acil de cadena media-Coenzima A deshidrogenasa, leucoencefalopatía megalencefálica con quistes subcorticales, leucodistrofia metacromática, cardiomiopatía mitocondrial, síndrome de Leigh asociada a ADN mitocondrial y NARP, mucopolisodis IV, mucopolisacaridosis tipo I, tipo mucopolisacaridosis Tipo IIIA, mucopolisacaridosis Tipo VII, neoplasia endocrina múltiple Tipo 2, enfermedad de músculo-ojo-cerebro, miopatía nemalínica, fenotipo neurológico, enfermedad Niemann-Pick debida a deficiencia en esfingomielinasa, enfermedad de Niemann-Pick Tipo C1, síndrome de rotura de Nijmegen, lipofuscinosis ceroides neuronal relacionada con PPT1, deficiencia en hormona de la pituitaria relacionada con PROP1, Síndrome de Pallister-Hall, paramiotonía congénita, síndrome de Pendred, deficiencia en la enzima bifuncional peroxisomal, trastornos de desarrollo dominante, deficiencia en fenilalanina hidroxilasa, inhibidor I del activador del plasminógeno, enfermedad renal poliquística, autosómica recesiva, trombofilia de protrombina G20210A, raquitismo por deficiencia en pseudovitamina D, picnodisostosis, retinitis pigmentosa, autosómica recesiva, Tipo Botnia, síndrome de Rett, condrodisplasia punteada rizomélica Tipo 1, deficiencia en acil de cadena

corta-CoA deshidrogenasa, síndrome de Shwachman-Diamond, síndrome de Sjogren-Larsson, síndrome de Smith-Lemli-Opitz, paraplejia espástica 13, osteocondrodisplasia relacionada con transportador de sulfato, hemocromatosis hereditaria relacionada con TFR2, lipofuscinosis ceroides neuronal relacionada con TPP1, displasia tanatofórica, amiloidosis de transtiretina, deficiencia en proteína trifuncional, DRD deficiente en tirosina hidroxilasa, tirosinemia Tipo I, enfermedad de Wilson, retinoschisis juvenil ligada a X y espectro de síndrome de Zellweger.

B. Variantes genéticas causales

Las variantes genéticas causales son variantes genéticas para las cuales existe evidencia estadística, biológica, y/o funcional de asociación con una enfermedad o rasgo de tal manera que la variante probablemente desempeñe un papel en la etiología. Una sola variante genética causal puede estar asociada con más de un rasgo. Una variante genética causal puede estar asociada con un rasgo mendeliano o un rasgo no mendeliano o ambos.

Aunque es típico referirse a variantes genéticas que ocurren con una frecuencia mayor que o igual al 1 % como polimorfismos y referirse a variantes menos comunes como mutaciones, estos términos se utilizan indistintamente y sin la intención de referirse a una frecuencia particular, salvo que se especifique. Por ejemplo, un SNP es un polimorfismo de un solo nucleótido, y puede ocurrir a cualquier frecuencia. Un ejemplo de una variante genética causal que es un SNP es la variante Hb S de la hemoglobina que produce anemia de células falciformes. Un DIP es un polimorfismo de delección/inserción, y puede ocurrir a cualquier frecuencia. Los DIP también pueden mencionarse como indeles. Un ejemplo de una variante genética causal que es un DIP es la mutación delta508 del gen CFTR que causa fibrosis quística. Un CNV es una variante del número de copias, y puede ocurrir a cualquier frecuencia. Un ejemplo de una variante genética causal que es un CNV es la trisomía del 21, que causa síndrome de Down. Un STR es una variación de repetición en tándem corta, y puede ocurrir a cualquier frecuencia. Los STR también se llaman polimorfismos de repetición. Un ejemplo de una variante genética causal que es un STR es la repetición en tándem que causa la enfermedad de Huntington.

Las variantes genéticas causales pueden manifestarse como variaciones en un polinucleótido de ADN, tales como resulta de un SNP, DIP, CNV, STR, o modificación epigenética heredable (por ejemplo, metilación del ADN). Múltiples instancias de una variante genética causal pueden reconocerse como tales porque son idénticas por estado y/o idénticas por descendencia. Una variante genética causal puede ser también un conjunto de variantes genéticas causales estrechamente relacionadas. Algunas variantes genéticas causales pueden ejercer influencia como variaciones de secuencia en polinucleótidos de ARN. A este nivel, algunas variantes genéticas causales también están indicadas por la presencia o ausencia de una especie de polinucleótidos de ARN. Además, algunas variantes genéticas causales provocan variaciones de secuencia en polipéptidos de proteínas. A este nivel, algunas variantes genéticas causales también están indicadas por la presencia o ausencia de una especie de polipéptidos de proteínas.

La nomenclatura para la catalogación y descripción de las variantes genéticas causales se desarrolló sobre una base ad hoc durante muchos años. Las variantes genéticas causales pueden describirse por más de un nombre. Por ejemplo, las variantes genéticas causales pueden etiquetarse tanto con un nombre común como con un nombre sistemático. Los nombres sistemáticos se basan en una notación evolutiva desarrollada por la Sociedad de Variación del Genoma Humano.

Las variantes genéticas causales se distinguen de las variantes genéticas asociadas indirectamente, que están asociados con un rasgo o enfermedad como resultado de la herencia conjunta correlacionada con una variante genética causal.

Las variantes genéticas causales pueden ser la causa de rasgos mendelianos y no mendelianos. En muchos casos, las variantes genéticas causales que determinan rasgos mendelianos deletéreos y enfermedades mendelianas pueden ser raras y ocurren a una frecuencia de menos del 1 % de una población. Sin embargo, algunas variantes genéticas causales que determinan rasgos mendelianos deletéreos y enfermedades mendelianas ocurren a frecuencias en el intervalo del 1 % al 10 %. En algunos casos, las variantes genéticas causales que determinan rasgos mendelianos deletéreos y enfermedades mendelianas ocurren a frecuencias superiores al 10 %.

Las variantes genéticas causales pueden descubrirse originalmente por los análisis genéticos estadísticos y moleculares de los genotipos y fenotipos de individuos, familias y poblaciones. Las variantes genéticas causales para rasgos mendelianos se identifican normalmente en un proceso de dos etapas. En la primera fase, se examinan familias en las que múltiples individuos poseen el rasgo para el genotipo y el fenotipo. Los datos del genotipo y fenotipo de estas familias se usan para establecer la asociación estadística entre la presencia del rasgo mendeliano y la presencia de varios marcadores genéticos, que normalmente son variantes genéticas indirectamente asociadas que están físicamente proximales a una variante genética causal en el cromosoma. Esta asociación establece una región candidata en la que es probable mapear la variante genética causal. En una segunda fase, se identifica la propia variante genética causal. La segunda etapa normalmente implica la secuenciación de la región candidata. De forma más sofisticada, son posibles procesos de una fase con tecnologías más avanzadas que permiten la identificación directa de una variante genética causal o la identificación de regiones candidatas más pequeñas. Después de descubrir una variante genética causal para un rasgo, pueden descubrirse variantes adicionales para el

mismo rasgo por métodos simples. Por ejemplo, el gen asociado con el rasgo puede secuenciarse en individuos que poseen el rasgo o sus familiares. La invención de nuevos métodos para el descubrimiento de variantes genéticas causales es un área activa de investigación. Se espera que la aplicación de los métodos existentes y la incorporación de nuevos métodos consigan el descubrimiento de variantes genéticas causales adicionales que pueda usarse o ensayarse por los dispositivos, sistemas y métodos de este documento. Muchas variantes genéticas causales se catalogan en bases de datos, incluyendo la base de datos de Herencia Mendeliana en el Hombre Online (MIM) y la Mutación de Genes Humanos (HGMD). Las variantes genéticas causales también se presentan en la bibliografía académica, en conferencias y en comunicaciones personales entre los estudiosos.

En organismos diploides, incluyendo seres humanos y otros mamíferos, una variante genética causal puede estar presente en cero, una, o dos copias. En algunos casos, la variante genética causal puede estar presente en más de dos copias. Los métodos para detectar la presencia de una variante genética causal normalmente también determinan el número de copias de la variante que está presente.

En una realización, al menos una de las variantes genéticas causales causa un rasgo que tiene una incidencia de no más del 1 % en una primera de las poblaciones y al menos otra de las variantes genéticas causales causa un rasgo que tiene una incidencia de no más del 1 % en una segunda de las poblaciones. En otra realización, al menos una de las variantes genéticas causales causa un rasgo que tiene una incidencia de no más de 1/10.000 en una primera de las poblaciones y al menos otra de las variantes genéticas causales causa un rasgo que tiene una incidencia de no más de 1/10.000 en una segunda de las poblaciones.

La Figura 1 proporciona una tabla de variantes genéticas causales a modo de ejemplo para enfermedades genéticas raras.

C. Poblaciones

En algunos casos, los dispositivos, métodos y sistemas de este documento están configurados para ensayar variantes genéticas causales para rasgos, donde los rasgos difieren en frecuencia entre poblaciones. Las poblaciones son grupos de individuos de la misma especie, tal como los seres humanos u otros mamíferos. A las poblaciones humanas se les da a menudo nombres particulares que pueden formar la base para las identidades sociales.

Una población humana puede ser un grupo de personas que comparten una herencia genética común, tal como un grupo étnico (por ejemplo, caucásicos). Una población humana puede ser una población haplotipo o grupo de poblaciones haplotipo (por ejemplo, haplotipo H1, M52). Una población humana puede ser un grupo nacional (por ejemplo, americanos, ingleses, irlandeses). Una población humana puede ser una población demográfica tal como aquellos delineados por edad, sexo y factores socioeconómicos. Las poblaciones humanas pueden ser poblaciones históricas.

Una población puede consistir en individuos distribuidos sobre una gran área geográfica tal que los individuos en los extremos de la distribución pueden no encontrarse nunca entre sí. Los individuos de una población pueden estar dispersados geográficamente en áreas discontinuas. Las poblaciones pueden ser de carácter informativo sobre la ascendencia biogeográfica. Las etiquetas usadas para identificar las poblaciones no tienen que contener una referencia geográfica o de ascendencia explícita para las poblaciones para ser informativas sobre la ascendencia, por ejemplo, una identificación personal de raza.

Las poblaciones también pueden definirse por ascendencia. Los estudios genéticos pueden definir poblaciones. Según la definición de la ascendencia y la genética, las principales poblaciones humanas corresponden a agrupaciones de escala continental, que incluyen Eurasia occidental, África subsahariana, Asia oriental, y nativos americanos. La mayoría de los seres humanos pueden asignarse a al menos una de estas poblaciones sobre la base de la ascendencia. Varias poblaciones más pequeñas también se distinguen como grupos continentales, incluidos indígenas de australianos, de Oceanía y bosquimanos.

Muy a menudo, las poblaciones se pueden descomponer en sub-poblaciones. La relación entre las poblaciones y las subpoblaciones puede ser jerárquica. Por ejemplo, la población de Oceanía puede subdividirse adicionalmente en subpoblaciones que incluyen polinesios, melanesios y micronesios. La población de Eurasia occidental puede subdividirse adicionalmente en subpoblaciones que incluyen europea, asiática occidental/central, asiática del sur y africana del norte. La población europea puede subdividirse adicionalmente en subpoblaciones que incluyen europea noroccidental, europea del sur y poblaciones judías de asquenazíes. La población de Europa noroccidental puede subdividirse adicionalmente en poblaciones nacionales, incluyendo inglesa, irlandesa, alemana, finlandesa, y similares. La población de Asia oriental puede subdividirse adicionalmente en subpoblaciones japonesa china, y coreana. La población del sur de Asia puede subdividirse adicionalmente en poblaciones de India y Pakistán. La población de la India puede subdividirse adicionalmente en personas dravídicas, personas de Brahui, Kannadigas, Malayalis, los tamiles, Telugus, Tuluvas y Gonds.

Un individuo puede identificarse a sí mismo con una o más poblaciones de referencia (por ejemplo, con un grupo

étnico) sobre la base de su ascendencia. Una persona puede identificar su ascendencia en base a la identidad de sus cuatro abuelos biológicos. La ascendencia de algunos individuos (a veces llamada su ascendencia individual o ascendencia biogeográfica individual) puede describirse como procedente de más de una población distinta. La mezcla se define en relación a un conjunto de poblaciones de referencia. La ascendencia de un individuo puede clasificarse en términos de mezcla de una serie de poblaciones de referencia. Estas poblaciones de referencia pueden variar dependiendo de la aplicación. Es común referirse a la mezcla en términos de los cuatro principales grupos continentales porque estos grupos pueden ser fácilmente distinguibles genética y fenonormalmente.

La Figura 2 proporciona una lista de poblaciones a modo de ejemplo.

D. Marcadores Informativos de ascendencia

Los marcadores informativos de ascendencia (AIM) son variantes genéticas que difieren en frecuencia entre poblaciones. Casi todas las variantes genéticas comunes difieren tanto dentro como entre las poblaciones. Sin embargo, las variantes individuales se encuentran a través de un espectro de distribuciones de varianza. En un caso, las variantes que difieren dentro pero no entre poblaciones normalmente no son AIM. En otro caso, las variantes que difieren entre, pero no dentro de poblaciones son especialmente informativas.

Diversos tipos de variantes genéticas pueden ser AIM, incluyendo SNP, DIP, CNV, y STR. Los AIM también pueden ser variaciones de secuencia en polinucleótidos de ARN. Algunos AIM también pueden estar indicados por la presencia o la concentración de una especie de polinucleótidos de ARN. Algunos AIM también pueden ser variaciones de secuencia en polipéptidos de proteínas. Algunos AIM también pueden estar indicados por la presencia o ausencia de una especie de polipéptidos de proteínas.

Varios marcadores informativos de ascendencia se identifican en la Figura 3. Otros también se describen en el documento US 2007/0037182 (Gaskin et al.).

Los marcadores informativos de ascendencia se pueden descubrir determinando la frecuencia de variantes genéticas en una pluralidad de poblaciones. Esto se puede conseguir determinando la frecuencia de variantes ya conocidas en individuos de diversas poblaciones. También se puede conseguir intrínsecamente durante el proceso de descubrimiento de variantes. Ambas tareas se emprendieron por el Proyecto Internacional HapMap, que catalogó polimorfismos SNP.

Los marcadores informativos de ascendencia pueden clasificarse por diversas mediciones que juzgan su poder predictivo. Una medición es la estadística F de Wright, llamada F_{ST} o FST. Esta variable se conoce por otros nombres, incluyendo índice de fijación. Otro parámetro para clasificar AIM es la capacidad informativa. Otro método de clasificación de AIM es el método de SNP correlacionado con PCA de Paschou et al. (Paschou et al. PCA-correlated SNPs for structure identification in worldwide human populations. PLoS Genet (2007) vol. 3 (9) pág. 1672-86).

Para conseguir un grado preseleccionado de confianza en la inferencia de ascendencia sobre la base de marcadores informativos de ascendencia, y para conseguir la inferencia de ascendencia para una pluralidad de poblaciones, normalmente es necesario examinar más de un marcador informativo de ascendencia. Puede usarse un panel suficientemente grande de variantes genéticas seleccionadas aleatoriamente para inferir la ascendencia. Se puede construir un conjunto diana de AIM especialmente adecuados. Muchos investigadores han publicado listas de marcadores informativos de ascendencia sugeridos (por ejemplo: Seldin et al. Application of ancestry informative markers to association studies in European Americans. PLoS Genet (2008) vol. 4 (1) pág. e5; Halder et al. A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. Hum Mutat (2008) vol. 29 (5) pág. 648-58; Tian et al. Analysis and application of European genetic substructure using 300 K SNP information. PLoS Genet (2008) vol. 4 (1) pág. e4; Price et al. Discerning the ancestry of European Americans in genetic association studies. PLoS Genet (2008) vol. 4 (1) pág. e236; Paschou et al. PCA-correlated SNPs for structure identification in worldwide human populations. PLoS Genet (2007) vol. 3 (9) pág. 1672-86; y Bauchet et al. Measuring European population stratification with microarray genotype data. Am J Hum Genet (2007) vol. 80 (5) pp. 948-56.). Pueden usarse estas listas y otras para construir un panel de AIM para los cuales puede configurarse un dispositivo o método de este documento para ensayo.

Algunos AIM son también variantes genéticas causales. Por ejemplo, la variante genética Duffy Null (FY * 0) causa ausencia de un antígeno de la sangre. Esta variante está en casi el 100 % de frecuencia en las poblaciones del África subsahariana y en casi 0 % de frecuencia en poblaciones fuera del África subsahariana. Muchas variantes genéticas causales asociados con la pigmentación son también AIM. Por el contrario, los AIM que no son variantes genéticas causales pueden estar indirectamente asociados con rasgos causados por otros AIM. Cuando se selecciona un conjunto de variantes genéticas causales y marcadores informativos de ascendencia, es útil que los AIM sean diferentes de las variantes genéticas causales.

II. Sistemas

En un aspecto, un medio legible por ordenador comprende lógica que predice una probabilidad de un fenotipo de un niño a partir de un conjunto de padres o padres potenciales con respecto a al menos una enfermedad genética rara basándose al menos en parte, en la presencia o ausencia de una pluralidad de variantes genéticas causales en un primer progenitor, la presencia o ausencia de una pluralidad de variantes genéticas causales en un segundo progenitor, al menos una AIM del primer progenitor y al menos un AIM del segundo progenitor. En otro aspecto, un medio legible por ordenador comprende: lógica para realizar un análisis totalmente probabilístico sobre datos correspondientes a una pluralidad de variantes genéticas causales de un hombre y una mujer para predecir una probabilidad de un fenotipo de un niño, donde el hombre y la mujer son padres potenciales del niño. En algunos casos, la realización de un análisis totalmente probabilístico comprende portar ruido estadístico durante todo el proceso. En algunos casos, las variantes genéticas causales de cada uno del hombre y la mujer comprenden uno o más marcadores informativos de ascendencia (AIM), una o más variantes genéticas causales correspondientes a una enfermedad genética rara, una o más variantes genéticas causales correspondientes a un rasgo de personalidad, o ambos. En algunos casos, el medio legible por ordenador comprende además lógica para recibir la entrada de un cuestionario completado por dicho hombre y/o mujer y la asignación de una función de ponderación a la pluralidad de variantes genéticas causales basada en dicha entrada. En otro caso más, la información del cuestionario comprende uno o más de los siguientes tipos de información respecto al hombre o la mujer: altura, peso, e historia de enfermedad familiar. El medio legible por ordenador puede proporcionar una salida en forma de un informe que detalla la probabilidad del fenotipo del niño. La lógica puede ser instrucciones legibles por ordenador como se describe en este documento.

En un aspecto, se proporcionan instrucciones legibles por ordenador para la predicción de un fenotipo de un niño que cuando se ejecutan: realizan un análisis totalmente probabilístico sobre datos de un hombre y una mujer correspondientes a una pluralidad de variantes genéticas causales para predecir una distribución de probabilidad sobre fenotipos del niño, donde el hombre y la mujer son padres potenciales del niño. En algunos casos, un análisis totalmente probabilístico es un análisis que no hace una determinación basada en una probabilidad en una etapa, sino que porta la probabilidad a lo largo de todo el análisis. Por ejemplo, si el genotipo AA tiene una probabilidad del 91 % y el genotipo AT tiene una probabilidad del 9 % entonces, cuando se calcula el riesgo del niño en etapas posteriores, las probabilidades del 91 % y el 9 % se portan a través de esa etapa de cálculo, en lugar de usar el genotipo más probable AA. En algunos casos, esto puede ser un modelo importante para alelos que tienen baja probabilidad pero riesgo relativo muy alto. Las variantes de cada uno del hombre y la mujer pueden comprender uno o más marcadores informativos de ascendencia (AIM), una o más variantes genéticas causales correspondientes a una enfermedad genética rara, o ambos. Las instrucciones pueden recibir la entrada de información de una batería de fenotipos de dicho hombre y mujer y asignar una función de ponderación a la pluralidad de variantes genéticas causales para predecir la probabilidad del fenotipo del niño. La batería de fenotipos puede recogerse mediante diversos artículos o medios incluyendo, sin limitación: cuestionarios, encuestas, pruebas (tales como pruebas de IQ), registros médicos, y juegos. La información de la batería de fenotipos usada cuando se ejecuta por las instrucciones puede ser la altura, el peso, y la historia de enfermedad familiar. Los elementos de la batería de fenotipos pueden incluir, aunque sin limitación: edad de la primera menopausia, amabilidad, sensibilidad a histamina en las vías respiratorias, dependencia del alcohol, peso al nacer, presión arterial, índice de masa corporal, dependencia de cannabis, dolor pélvico crónico, consumo de café, recuento de sangre completa, responsabilidad, trastorno depresivo mayor DSM-IV, profundidad del sueño, endometriosis, participación en ejercicios, extraversión, desmayos, antecedentes familiares de enfermedad, fatiga, recuento de crestas dactilares, estado físico, volumen espiratorio forzado, evitación del daño, fiebre del heno, altura, niveles de inmunoglobulina E, geometría incisiva, frecuencia alfa individual (EEG), tiempo de inspección, concentración de insulina, IQ, propensión a apendicectomía, propensión a asma, propensión a amigdalectomía, calvicie masculina, tamaño y recuento de lunares, úlceras en la boca, neuroticismo, búsqueda de la novedad, franqueza, osteoartritis en mujeres, latencia P300 parietal, amplitud de ondas lentas parietales, IQ de rendimiento, persistencia, latencia P300 prefrontal, amplitud de ondas lentas prefrontales, parto prematuro para cualquier nacimiento, presencia de trastorno de la alimentación, propensión a la psicosis, calidad del sueño, dependencia de recompensas, iniciación al tabaquismo, tartamudez, susceptibilidad a migrañas, consumo de té, colesterol total, triglicéridos, peso, IQ verbal, comportamiento electoral y VO2max. Las instrucciones pueden proporcionar una salida en forma de un informe que detalla la probabilidad del fenotipo del niño.

En un aspecto, un sistema para predecir un fenotipo del niño comprende: al menos dos dispositivos de detección de ácido nucleico configurados para detectar una pluralidad de variantes genéticas causales correspondientes a al menos una enfermedad genética rara y al menos un marcador informativo de ascendencia (AIM), donde un primer dispositivo está en contacto con una muestra de una mujer y donde un segundo dispositivo está en contacto con una muestra de un hombre, donde el hombre y la mujer son padres o padres potenciales de un niño; un lector configurado para leer los datos de los dispositivos; e instrucciones legibles por ordenador, donde las instrucciones cuando se ejecutan utilizan los datos del lector correspondientes a la pluralidad de variantes genéticas causales y el al menos un marcador informativo de ascendencia para predecir una probabilidad de un fenotipo del niño con respecto a la al menos una enfermedad genética rara. Los al menos dos dispositivos de detección de ácido nucleico pueden comprender una pluralidad de sondas de ácido nucleico que se unen selectivamente a la pluralidad de variantes genéticas causales y el al menos un AIM.

En un aspecto, un sistema para predecir un fenotipo del niño comprende: al menos dos dispositivos de detección de

ácido nucleico configurados para detectar una pluralidad de variantes genéticas causales correspondientes a más de 85 enfermedades genéticas raras, donde un primer dispositivo está en contacto con una muestra de una mujer y donde un segundo dispositivo está en contacto con una muestra de un hombre, donde el hombre y la mujer son padres o padres potenciales de un niño; un lector configurado para leer los datos de los dispositivos; e instrucciones legibles por ordenador, donde las instrucciones cuando se ejecutan utilizan los datos del lector correspondientes a la pluralidad de variantes genéticas causales para predecir una probabilidad de un fenotipo del niño con respecto a las más de 85 enfermedades genéticas raras. Los al menos dos dispositivos de detección de ácido nucleico pueden comprender una pluralidad de sondas de ácido nucleico que se unen selectivamente a la pluralidad de variantes genéticas causales correspondientes a más de 85 enfermedades genéticas raras. Los al menos dos dispositivos de detección de ácido nucleico pueden comprender además una pluralidad de sondas de ácido nucleico que se unen selectivamente a al menos un marcador informativo de ascendencia (AIM). Las instrucciones legibles por ordenador cuando se ejecutan pueden utilizar los datos del lector correspondientes a al menos un AIM para predecir la probabilidad de un fenotipo del niño.

En un aspecto, un sistema para indicar si un sujeto es portador de una enfermedad genética rara comprende: un lector configurado para leer datos desde un dispositivo de detección de ácido nucleico configurado para detectar una pluralidad de variantes genéticas causales correspondientes a al menos una enfermedad genética rara y al menos un marcador informativo de ascendencia (AIM); e instrucciones legibles por ordenador, donde las instrucciones cuando se ejecutan utilizan los datos del lector correspondientes a la pluralidad de variantes genéticas causales y el al menos un marcador informativo de ascendencia para predecir una pluralidad de probabilidades de que el sujeto sea portador para cada una de la pluralidad de variantes genéticas causales.

A. Genotipado de individuos

Hay muchas maneras diferentes de adquirir información sobre una secuencia genética de un individuo. En este documento se describen métodos a modo de ejemplo para el analizar los datos de genotipo producidos por chips de SNP. Algunos de los métodos se refieren directamente a la resolución de algunos de los problemas con la identificación de variantes genéticas causales de enfermedades genéticas raras, por ejemplo: minimizar la tasa de falsos negativos (por ejemplo, la probabilidad de que el ensayo sea negativo cuando la pareja está en riesgo de tener un niño con la enfermedad) manteniendo al mismo tiempo la tasa de falsos positivos por debajo de un umbral especificado. Sin embargo, los mismos conceptos se pueden generalizar a otros métodos para producir datos de secuencia (tal como resecuenciación por captura de secuencia y secuenciación por síntesis) o para averiguar información de secuencia indirectamente (por ejemplo, haciendo mediciones sobre ARN o representantes de proteína).

La designación de genotipos para SNP bialélicos se hace tradicionalmente por agrupación. Dadas los alelos X e Y, puede crearse un diagrama de dispersión de las señales normalizadas para el alelo X e Y. Los genotipos XX, XY, e YY se manifiestan en sí mismos como grupos distintos de puntos en este diagrama de dispersión. Otros tipos de variantes, tales como inserciones, deleciones, y SNP trialélicos se pueden acomodar mediante el uso de múltiples sondas bialélicas. La designación de genotipos puede proceder por agrupación en un espacio dimensional superior, en que están presentes más de dos ejes.

Una designación de falso negativo puede corresponder a informar a una pareja de que no se encuentra en riesgo, cuando en realidad lo están. Se trata de un fallo de un ensayo de selección y puede dar lugar a un niño con una enfermedad. En términos de costes, la medida convencional para el precio del tratamiento de un niño con un trastorno mendeliano raro es de aproximadamente 1 millón de dólares.

Supongamos que el alelo de tipo silvestre es X y el alelo de la enfermedad es Y en un locus dado. En este locus, la mayoría de las parejas serán XX (normales) o XY (portadores), aunque una penetrancia/expresividad incompleta puede conducir a unos pocos individuos YY. La determinación del grupo ancestral de un individuo a partir de AIM y etnicidad auto-informada (como se describe en la reivindicación B a continuación) proporciona probabilidades previas para las frecuencias de genotipo XX, XY, e YY en el locus de la enfermedad. En muchas realizaciones, la determinación de la ascendencia es importante porque muchos loci de enfermedades mendelianas son mucho más frecuentes en una población que en otra.

La mayoría de los software de designación de genotipo usa las probabilidades previas para frecuencias genotípicas como grupos previos, ya que el algoritmo de designación está optimizado para minimizar la probabilidad de clasificación errónea. Por ejemplo, para un SNP bialélico, si A es el grupo ancestral que se está ensayando, MX es la señal normalizada medida en el canal X y MY es la señal normalizada medida en el canal Y, el procedimiento de agrupamiento para designaciones de genotipado es equivalente a la estimación de $P(G|MX,MY,A)$. La mejor manera de hacerlo es con la regla de Bayes, como se muestra en las Figuras 8-12.

En muchos casos, los dispositivos y sistemas de este documento pueden diseñarse err en un lado de identificación de personas como portadores o sin designación si hay alguna duda acerca de que sean de tipo silvestre. Geométricamente, esto corresponde a límites de grupo más flexibles donde se sugeriría ser más agresivos en la designación de XY que las probabilidades basales de portador. Esto puede ser deseable porque los falsos positivos

pueden requerir solamente ensayos de seguimiento, mientras que los falsos negativos pueden representar un fallo de un ensayo de detección.

Matemáticamente, la geometría exacta se determina eligiendo la frontera que minimiza una medida de costes (por ejemplo coste esperado o medio) en lugar de una que minimice la probabilidad de clasificación errónea. Computacionalmente, esto puede requerir software especializado (tal como las instrucciones legibles por ordenador proporcionadas en este documento) para procesar la lectura del chip en bruto.

Para obtener una tasa deseada de falsos positivos de 1/100 o menos, y una tasa de falsos negativos de 1/10000 o menos para un ensayo combinado, puede realizarse el siguiente protocolo: a) inferir la ascendencia para obtener parámetros de grupo específicos de población como se muestra en las Figuras 4-7, b) ensayar cada variante varias veces en el mismo chip, y c) usar una muestra de control con cada lote de chips para confirmar la validez de fabricación.

El uso de mediciones repetidas permite conseguir que estas tasas de falsos positivos y negativos. Como ejemplo, considérese un ensayo para una variante con una precisión media del 99 %, donde la precisión es la probabilidad de que el genotipo real sea el mismo que el genotipo designado. Con error simétrico (por ejemplo, con designación de genotipo por defecto) esto corresponde a una tasa de falsos positivos y falsos negativos del 1 %. Si C es el genotipo designado y G el genotipo real, entonces formalmente tenemos: la tasa de falso positivo: $P(C = XX|G = XY) = 0,01$, y la tasa de falsos negativos: $P(C = XY|G = XX) = 0,01$.

Si se usan tres de estos análisis en paralelo y tienen tasas de fracaso independientes, un voto combinado al mejor de tres para designaciones de genotipo puede conseguir tasas mucho mejores de falsos positivos y negativos. Supóngase que C1, C2, C3 son las designaciones de genotipo de estos tres análisis. Entonces, la probabilidad de que los tres análisis produzcan falsos negativos al mismo tiempo es:

$$P(C1 = XX, C2 = XX, C3 = XX|G = XY) = P(C1 = XX|G = XY) P(C2 = XX|G = XY) P(C3 = XX|G = XY) = (0,01)^3 = 1 \cdot 10^{-6}.$$

Esto es ahora uno en un millón, que es mucho menos de uno en 100, y consigue las tasas de falsos negativos necesarias para un ensayo combinado factible.

Es posible conseguir tasas de fracaso aproximadamente independientes mediante la localización de sondas en diferentes regiones del chip o mediante el uso de múltiples chips o tipos de ensayo. Sin embargo, la suposición de fracaso independiente puede verse comprometida si hay error sistemático, ya sea en el diseño de sondas poco fiables o la presencia de muestras con niveles de ADN extremadamente bajos.

En la práctica, la fiabilidad de la sonda se puede garantizar mediante el uso de un enfoque de embudo, comenzando con muchas sondas redundantes y clasificación de aquellas sondas que funcionan mal en un conjunto representativo de adiestramiento. Específicamente, puede usarse una combinación de muestras aleatorias, plásmidos modificados por ingeniería, y muestras de depósitos de ADN tipo Coriell para determinar empíricamente la sensibilidad y la especificidad para cada variante del ensayo. No se usan variantes con malas estadísticas de designación para el cálculo del riesgo. También puede prepararse una muestra o muestras de control y procesarse junto con ADN del cliente para asegurar que cada lote de análisis/chips se ha fabricado apropiadamente.

B. Inferencia de la ascendencia

Dado un conjunto de mediciones de N marcadores informativos de ascendencia X_1 a de X_N , podemos inferir la pertenencia a un grupo de ascendencia dado a través de técnicas convencionales de clasificación. Por ejemplo, podemos usar la maximización de expectativa basada en agrupación o algoritmos similares para conseguir la clasificación probabilística en categorías, usando variantes ensayadas como características y ascendencia auto-presentada como datos de adiestramiento. A menudo, se prefieren individuos que tienen cuatro abuelos provenientes de la misma área geográfica o grupo étnico como puntos adiestramiento. Aquellos cuyos abuelos provienen de diferentes regiones o grupos étnicos se consideran entonces mezclados. Este análisis se puede ampliar, naturalmente, al caso en que las categorías ancestrales son jerárquicas y/o solapantes. Por ejemplo, es posible definir múltiples tipos de ascendencia: escala continental (Europa, Asia, África), nacional (inglés, francés, alemán), lingüística (francés canadiense), y así sucesivamente. Se pueden usar diferentes clasificadores que emplean diferentes combinaciones de marcadores informativos de ascendencia para distinguir entre prácticamente cualquier colección de tales categorías.

Para los propósitos de este documento, las categorías que son más relevantes son las que corresponden a poblaciones con tasas significativamente aumentadas de endogamia o frecuencias de alelos particulares causantes de enfermedades. Pueden incluirse AIM para identificar todas estas poblaciones en el chip.

Los AIM, en general, son útiles porque el auto-informe de ascendencia, aunque es un punto de partida útil, puede ser engañoso o incompleto. En América del norte, por ejemplo, muchos caucásicos no conocen sus proporciones de

mezcla. Dado que poblaciones como los italianos e ingleses difieren en sus frecuencias de alelos de enfermedad como la variante mediterránea de la beta talasemia (mayor en los italianos: <http://www.genome.gov/10001221>), los AIM pueden proporcionar información importante y útil más allá de la proporcionada por el autoinforme. En particular, para individuos mezclados también es posible asignar la ascendencia sobre una base más granular a regiones particulares de ADN (<http://www.genome.org/cgi/content/abstract/gr.072850.107>).

Finalmente, los AIM son útiles para el chip combinado para rasgos mendelianos poco comunes porque el conocimiento de la ascendencia proporciona una distribución previa sobre las frecuencias de genotipo, lo que reduce la probabilidad de clasificación errónea. Por ejemplo, alguien que no es judío asquenazí probablemente no sea portador de disautonomía familiar (http://en.wikipedia.org/wiki/Familial_dysautonomia), y la información adicional sobre la ascendencia proporciona una verificación independiente, además de los resultados directos del ensayo.

C. Determinación del riesgo genético

El cálculo del riesgo en el contexto de un ensayo para enfermedades genéticas raras puede ser que las tasas de falsos negativos y falsos positivos se mantendrán dentro de parámetros aceptables. La discusión sobre el genotipado anterior es el principal medio para conseguirlo. Dadas las estrategias fiables de genotipado con tasas de falsos positivos del orden del 1 % o menos y tasas de falsos negativos del orden del 0,001 % o menos, pueden aplicarse métodos convencionales de cálculo del riesgo para determinar el riesgo de que una pareja tenga un hijo con una enfermedad genética dada. La aplicación de tales métodos requiere un conocimiento mínimo del modo de herencia de la enfermedad. Para muchos rasgos, los cálculos del riesgo pueden implicar evaluaciones más complejas que implican la posibilidad de penetrancia o expresividad incompleta, mutaciones de novo o recurrentes, anticipación genética, y factores de complicación similares.

Pueden ser importantes preguntas de encuesta, incluyendo una historia familiar en línea, para aumentar los cálculos del riesgo en el caso de que un resultado del ensayo por sí solo no sea determinante. Sin embargo, para muchas de las enfermedades mendelianas ensayadas en el chip, una consulta tradicional puede no proporcionar mucha información adicional porque los portadores a menudo son asintomáticos y la mayoría de los portadores no tienen familiares directos con una enfermedad mendeliana rara. Cualquier resultado positivo de un dispositivo de este documento puede estar sujeto a un ensayo de seguimiento especializado en el caso de que la tasa de falsos positivos sea no despreciable.

D. Predicción del fenotipo a partir del genotipo y el fenotipo de los padres

Los dispositivos y sistemas de este documento proporcionan datos de genotipo que se pueden usar para predecir el riesgo de tener un hijo con una enfermedad determinada. Conceptualmente, esta previsión es posible debido al modo predecible de la herencia de un rasgo mendeliano y la correlación generalmente alta entre tener una variante y tener la afección, a pesar de los problemas implicados con la expresividad y penetrancia variables. Para rasgos complejos, aún es difícil predecir sus valores a partir de mediciones de genotipo solamente. Puede ser posible hacer esto en el futuro mediante midiendo los CNV, a partir de datos de resecuenciación, o combinando múltiples modalidades de ensayo (por ejemplo, metilación, expresión, etcétera), pero los resultados de los chips SNP por sí solos no han explicado gran parte de la varianza. Por ejemplo, un reciente estudio de asociación de genoma completo de IQ encontró que sólo el 1 % de la varianza en el IQ podía predecirse mediante una medición de 500000 SNP (<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1601-183X.2007.00368.x>). Se han encontrado valores igualmente bajos para otros rasgos como la altura, donde grandes estudios con 30000 individuos más han encontrado SNP que explican sólo el 3,7 % de la varianza en la altura.

De forma importante, aunque es difícil predecir los valores de los rasgos complejos de un individuo a partir de la información del genotipo, a menudo es fácil predecir los valores de los rasgos complejos de un niño a partir de la información de fenotipo de los padres. Es decir, para fenotipos con alta heredabilidad, una medición de los valores de rasgos de la madre y el padre es suficiente para proporcionar una excelente capacidad de predicción para los rasgos del niño. Por ejemplo, la heredabilidad en sentido estricto de la altura es del orden de 0,70 a 0,80, dependiendo de la población y el fondo ambiental (<http://www.sciam.com/article.cfm?id=how-much-of-human-height>). Esto significa que la proporción de la varianza del niño explicada por dos mediciones (la altura de la madre y la altura del padre) varía del 49 % al 64 %, una cifra que es muy superior a lo que se ha conseguido con las predicciones basadas en SNP de la altura individual desde mayo de 2008. Por tanto, es posible explotar este fenómeno con fines de predicción del niño. Para rasgos con heredabilidad no nula, los valores de los rasgos del niño se pueden predecir a partir de mediciones de los padres con bajas tasas de error.

Se puede usar una encuesta de riesgos o fenotipos de los padres y los niños a partir de una muestra aleatoria de 1000 personas de una población diana para proporcionar datos empírica de adiestramiento para las ecuaciones de regresión que relacionan las características de los padres y del niño. También pueden recogerse AIM para facilitar la identificación de miembros de esta población. También es posible incluir datos de otros miembros de la familia, tales como abuelos, tíos y tías, como características adicionales de predicción para predecir rasgos del niño. Estas ecuaciones de regresión pueden ser más sofisticadas que los modelos lineales simples empleados en genética cuantitativa. Puede emplearse cualquier tipo de algoritmo de regresión para predecir el fenotipo niño a partir de las

mediciones de los fenotipos de los padres. Por ejemplo, esto es de utilidad demostrable en el caso de la longevidad, que tiene un patrón de herencia altamente no lineal (<http://longevity-science.org/Genetics-Australia.html>). De forma importante, la no linealidad de la herencia de la longevidad provoca un coeficiente de heredabilidad engañosamente bajo de 0,10.

Estimaciones convencionales de heredabilidad por tanto pueden subestimar la capacidad de predecir los rasgos de la descendencia a partir de los valores paternos, lo que indica que puede ser útil ensayar cualquier rasgo de relevancia comercial potencial en una gran cantidad de tríos padres/hijo de una población dada y determinar empíricamente si la predicción es viable o inviable. Como antes, puede ser útil ensayar otros parientes además de los padres inmediatos. También puede ser útil reportar los intervalos de confianza o cuantiles además de estimaciones puntuales. Las ecuaciones de regresión para predecir riesgos complejos o fenotipos del niño a partir de los fenotipos de los padres y los algoritmos de cálculo del riesgo para predecir el riesgo de enfermedad del niño a partir del genotipo de los padres forman un conjunto unificado de algoritmos para la predicción del niño. Desde una perspectiva de entrada/salida, los padres proporcionan datos de encuestas y muestras de de saliva, y los algoritmos reportan predicciones del niño tanto de riesgos de enfermedades mendelianas como de rasgos complejos.

Como se describe en este documento, la información de la población puede determinar parámetros de grupo. Las Figuras 4-7 ilustran una base teórica mediante la cual la información de la población (según lo revelado por los AUM y/o el auto-informe) puede informar de los parámetros de grupo y la designación de genotipo, y de este modo influyen en la precisión del ensayo de portadores y la predicción de fenotipo del niño. La población 1 es un grupo que tiene un hipotético SNP A/G con la secuencia flanqueante dada en la Figura 4. Si se genotipan 1000 individuos de esta población, se obtienen 1000 mediciones de las intensidades normalizadas de las sondas de canal A y G en ese SNP. Estos valores de intensidad se pueden representar en un diagrama de dispersión mostrado, donde cada punto representa un individuo y los ejes X e Y corresponden a los canales A y G. En este ejemplo, surgen tres grupos, correspondientes a los genotipos AA, AG, y GG. Pueden ajustarse modelos paramétricos de diversos tipos a estos grupos. Uno de los más populares es el modelo de maximización de expectativa (EM), en que se ajusta una distribución gaussiana de múltiples variables con parámetros de μ (el vector media) y σ (la matriz de covarianza) a cada grupo. En este ejemplo, hay tres vectores μ de dimensión 2, y tres matrices σ de dimensión 2×2 (en este caso las matrices de identidad). El área relativa bajo cada gaussiana de múltiples variables se da por el parámetro π_i , correspondiente a la proporción de puntos que se encuentran en ese grupo. En la Figura 5, la población 2 tiene exactamente la misma variante con la misma secuencia flanqueante, pero diferentes proporciones de cada genotipo (0,16, 0,48, 0,36 en lugar de 0,25, 0,50, 0,25). Debido a la biología molecular de la variante y que la secuencia flanqueante es la misma, los grupos están en la misma posición y sólo difiere π_i . En la Figura 6, la población 3 tiene una secuencia flanqueante diferente con una G adicional. Las diferencias de biología molecular en la secuencia flanqueante pueden causar desplazamientos en las posiciones y orientaciones de los grupos, por ejemplo trasladando los grupos hacia arriba en el eje G o introduciendo más varianza a lo largo del eje G. En tal caso, los parámetros μ y σ también pueden diferir junto con los parámetros π_i . Las Figuras 4-6 previas han estado considerando las poblaciones en términos de ascendencia geográfica, sin embargo la Figura 7 considera las poblaciones por género. Específicamente, supóngase que la población está compuesta por 4 varones, que el locus en consideración se encuentra en el cromosoma X, y que todos los miembros de las poblaciones previas eran mujeres. En este caso el diagrama de intensidad para los hombres sería muy diferente. En particular, sólo dos grupos son posibles (A/- y G/-) en lugar de tres (AA, AG, GG). Por tanto, los parámetros de grupo μ , σ y π_i son diferentes de los de la población 1. Estos ejemplos ilustran la importancia de tener en cuenta información sobre la población (por ejemplo, informada por género o etnia) al adiestrar modelos de designación.

En muchos casos, la información de población ayuda a la designación de genotipo. Las Figuras 8-12 muestran un ejemplo de cómo la información de la población influye en la designación de genotipo y el cálculo del riesgo en el contexto de la beta-talasemia, que difiere considerablemente en frecuencia entre dos poblaciones fenonormalmente similares: italianos y caucásicos de la población general. La frecuencia de portador para la variante de beta-talasemia deletérea difiere marcadamente entre las poblaciones europea del norte e italiana, que por lo demás son fenonormalmente similares y muy mezcladas en países como los Estados Unidos. Por ejemplo, la frecuencia de portador en los europeos del norte es menor de 1 por 1.000, mientras que la frecuencia de portador en italianos es mayor de 1 por 100. El diagrama de la Figura 8 muestra un diagrama de dispersión de intensidad para el alelo de delección de beta-talasemia similar a los diagramas de dispersión de las Figuras 4-7. Por simplicidad sólo se muestran dos grupos, correspondientes a los genotipos de tipo silvestre y portador y excluyendo el homocigoto. Una pregunta clave es cómo clasificar puntos con intensidades que no están muy cerca de las medias de sus respectivos grupos, como "A" en la Figura 8. En la Figura 9, se muestra que la designación de puntos cerca de la frontera está asistida por el uso de información de ascendencia. El ejemplo numérico en la Figura 10 muestra que una medición de intensidad en bruto en un locus causal no es el único valor que importa. El acoplamiento de esta medición (X) con el conocimiento de la ascendencia (A) puede cambiar la probabilidad de ser portador en casi 60 veces. Se proporciona un ejemplo de código fuente en la Figura 11. En la práctica no se tendrá un conocimiento completo de la ascendencia. Los AIM ayudan en esta situación. Se puede usar la señal del AIM para producir una distribución sobre las poblaciones, obteniendo la probabilidad de que un locus dado se herede a partir de un antepasado de un grupo de población dado. Como se muestra en la Figura 12, esto requiere una modificación sencilla a la expresión para la probabilidad posterior de si un individuo está en un grupo dado. En lugar de asumir que el valor del parámetro de ascendencia se conoce con probabilidad 1, puede usarse una distribución sobre la ascendencia. Esta técnica de

usar AIM para mejorar la designación en loci causales proporciona una ganancia sustancial en el poder de clasificación y es particularmente útil para ensayos tales como los descritos en este documento, como se indica para muchos grupos de diferente ascendencia.

Hay varias posibles fuentes de ruido que deben justificarse cuando se realizan predicciones de fenotipo del niño usando información de genotipo. Las Figuras 13-18 ilustran el proceso de predicción totalmente probabilística de una distribución sobre los riesgos y fenotipos del niño. El proceso da cuenta de varias fuentes posibles de incertidumbre, incluyendo el error de designación de genotipo, la incertidumbre de organización, la recombinación, la unión gamética, y un mapa de ruido genotipo-a-fenotipo. En primer lugar, como se muestra en la Figura 13, un genotipo medido puede tener error de medición representado por una tasa de falsos positivos y falsos negativos, pero más generalmente se podría especificar una matriz de confusión que dio la probabilidad de predecir el genotipo i dado un genotipo real j . Dado un genotipo medido, el resultado es una distribución sobre genotipos posibles reales. A continuación, como se muestra en la Figura 14, el genotipo inferido está desfasado. Es decir, los loci heterocigóticos adyacentes pueden estar en *cis* o *trans* y esto es desconocido. Uno puede modelar la incertidumbre de fase mediante: el uso de una base de datos adecuada de frecuencias de haplotipos de población en el locus dado; o el uso de uno de varios algoritmos de inferencia de haplotipo, tal como el algoritmo de Clark o la inferencia de haplotipo basada en modelo. La Figura 15 muestra cómo generar una distribución sobre los posibles recombinantes y por tanto posibles gametos procedentes de los posibles haplotipos. La Figura 16 entonces muestra cómo repetir este proceso tanto para la madre como para el padre para obtener una distribución de probabilidad sobre las uniones gaméticas, correspondientes genotipos organizados del niño (haplotipos *aka*). Esto produce una distribución sobre los haplotipos del niño. Finalmente, la Figura 17 revisa varios tipos diferentes de mapas genotipo-a-fenotipo que pueden usarse con la distribución sobre los haplotipos del niño para producir la cantidad deseada: una distribución sobre los posibles riesgos o fenotipos del niño.

La Figura 18 representa ecuaciones para obtener la distribución sobre los riesgos o fenotipos estimados del niño dados los genotipos de los padres, formalizando los diagramas de flujo de las Figuras 13-17. Esta expresión es útil porque es exacta y se puede aproximar para cualquier locus dado. Esto puede hacerse mediante simulación o apreciando que ciertas etapas fracasan; por ejemplo, la predicción del fenotipo del niño basándose en un rasgo de un solo locus no necesita afrontar la recombinación. Es notable que cada distribución condicional puede expresarse como una matriz y la serie de sumas puede expresarse en términos de multiplicaciones de matrices repetidas, una formulación que permite una rápida implementación en ordenador.

En el futuro, según bajen los costes de resecuenciación, es importante señalar que los fenotipos y genotipos de los padres determinados de esta manera pueden entonces retornarse unos sobre los otros. Es decir, la disposición a gran escala voluntaria de la información del genotipo y fenotipo de los padres con el fin de predecir rasgos del niño puede usarse como una fuente de datos para estudios de asociación a escala extremadamente grande.

E. Genómica de los parientes biológicos

Los métodos de este documento pueden utilizarse para genómica familiar. En algunos aspectos, los métodos pueden utilizarse para determinar probabilidades de fenotipos o probabilidades de que un miembro de la familia tenga o haya tenido una enfermedad genética rara en base a selección genética de otros miembros de la familia. Por ejemplo, un subconjunto de parientes biológicos puede seleccionarse como portadores para una pluralidad de variantes genéticas causales de al menos una enfermedad genética rara. Los datos del subconjunto después pueden utilizarse para predecir la probabilidad del fenotipo de otro pariente biológico que no era un miembro del subconjunto ensayado. Por ejemplo, pueden ensayarse tres niños y puede calcularse la probabilidad de que uno de los padres, vivo o muerto, tenga un fenotipo particular o enfermedad genética rara como se describe en este documento. Estos métodos pueden ser útiles para la determinación postmortem de un pariente que puede haber tenido un diagnóstico incierto de una enfermedad en el momento de la muerte o no haber tenido los análisis disponibles para determinar si tenía una enfermedad genética rara.

En algunos casos, los dispositivos y sistemas descritos en este documento que pueden usarse para predecir las probabilidades de fenotipo del niño pueden utilizarse para predecir probabilidades de parientes vivos que tienen ciertos rasgos o enfermedades o fenotipos. Los parientes pueden incluir abuelos, bisabuelos, tíos, tías, madres, padres, hermanos, hijos y cualquier otro miembro de la familia biológica.

En algunos casos, un pariente que no estaba incluido en el subconjunto de parientes biológicos que se genotipó en lo que respecta a variantes genéticas causales se puede remitir a un asesor y/o médico genético. En un caso, un pariente que es miembro o no miembro del subconjunto de parientes biológicos puede seguir un curso de acción médica basado en los resultados de los análisis. Por ejemplo, puede prescribirse aspirina para enfermedades cardíacas a un miembro de la familia en base al riesgo calculado por un método o sistema de este documento. También puede recogerse información a partir de una batería de fenotipos, tal como se discute en este documento, para algunos o todos los del subconjunto de parientes biológicos y utilizarse en los cálculos de las probabilidades para otros parientes. El riesgo de desarrollar una enfermedad para cualquier pariente biológico puede calcularse y ajustada en base a la genómica familiar, tal como se describe en este documento. Los resultados pueden entregarse al pariente y pueden realizarse métodos prácticos. Por ejemplo, puede crearse un árbol familiar genómico y utilizarse para integrar una probabilidad de riesgo de enfermedad para uno, algunos o todos los miembros de su árbol familiar.

III. Dispositivos

Se describe en este documento un dispositivo de selección de portadores universales y métodos para utilizar el dispositivo. También se describe el software informático y el código para la traducción de datos desde los dispositivos y los métodos en probabilidades de que un usuario sea portador o que un conjunto de padres tenga una probabilidad de tener un hijo con una enfermedad genética. Un dispositivo de este documento se puede configurar para ensayar una pluralidad de enfermedades genéticas raras o al menos 5, 10, 85, 100, 200, 500, o 1000 enfermedades genéticas raras.

En un aspecto, un dispositivo de detección de ácido nucleico está configurado para ensayar una muestra para una pluralidad de variantes genéticas causales correspondientes a más de 85 enfermedades genéticas raras. El dispositivo puede estar configurado, además, para ensayar una muestra para al menos un marcador informativo de ascendencia (AIM). En algunos casos, el dispositivo comprende una pluralidad de sondas de ácido nucleico que se unen selectivamente a una pluralidad de variantes genéticas causales correspondientes a más de 85 enfermedades genéticas raras. El dispositivo puede comprender además una pluralidad de sondas de ácido nucleico que se unen selectivamente a al menos un marcador informativo de ascendencia (AIM). El dispositivo puede comprender una serie de perlas que se unen selectivamente a una pluralidad de variantes genéticas causales correspondientes a más de 85 enfermedades genéticas raras. El dispositivo puede comprender además una serie de perlas que se unen selectivamente a al menos un marcador informativo de ascendencia (AIM). En algunos casos, el al menos un AIM no es una variante genética causal. En algunos casos, al menos dos de las enfermedades genéticas raras ocurren a frecuencias que difieren en al menos 10 veces en al menos dos poblaciones distintas, donde las al menos dos poblaciones distintas se diferencian por la al menos un AIM. En algunos casos, el dispositivo comprende un ensayo de captura de secuencia para detectar una pluralidad de variantes genéticas causales correspondientes a más de 85 enfermedades genéticas raras.

En algunos casos, el dispositivo detecta simultáneamente 85, 100, 150, 200, 500, 1000 o más variantes genéticas causales raras. En muchos casos, las variantes genéticas causales indican una enfermedad mendeliana en que un único gen es responsable de una enfermedad. Muchas veces, las enfermedades mendelianas tienen un patrón de herencia recesiva como se demuestra en las Figuras 4-7 y una persona tiene dos alelos recesivos para desarrollar la enfermedad. Los individuos con un alelo mutante y uno no mutante pueden describirse como portadores, ya que portan la variante genética causal, pero no tienen la enfermedad. Si los padres son ambos portadores de la misma enfermedad, sus hijos estarán en riesgo de heredar dos copias del gen mutante y por tanto de desarrollar la enfermedad. Para una enfermedad mendeliana ejemplar, se espera que una cuarta parte de los niños de dos portadores herede la enfermedad. Debido a que el riesgo de tener un hijo afectado es una función determinista de los genes de los padres, se usan comúnmente análisis genéticos para identificar preventivamente los padres que son portadores. Estos análisis se consideran ensayos de selección, ya que se dan a individuos por lo demás sanos. Si se encuentra que una pareja está alto riesgo de tener un niño con una enfermedad mendeliana, en un ejemplo, un asesor genético o un médico pueden ayudarles a considerar sus opciones reproductivas. Muchas enfermedades mendelianas que se entienden a un nivel molecular son graves, provocando muerte temprana o debilidad grave.

Un dispositivo de este documento puede ser un único ensayo destinado para la selección de población panétnica para un panel de enfermedades mendelianas.

En este documento se describe un dispositivo configurado para ensayar variantes genéticas causales, por ejemplo, variantes genéticas que causan un fenotipo particular. El dispositivo también puede configurarse para ensayar la presencia de marcadores informativos de ascendencia (AIM), es decir, marcadores genéticos que se correlacionan con las poblaciones, en particular poblaciones humanas, procedentes de localizaciones geográficas específicas. Las variantes genéticas causales contrastan con los marcadores genéticos en que ciertos marcadores pueden estar altamente correlacionados con la expresión de un rasgo sin causar realmente el rasgo. En algunos casos, pueden seleccionarse variantes genéticas causales donde algunas poblaciones (por ejemplo, poblaciones que se pueden distinguir por AIM) tienen diferentes frecuencias para rasgos o enfermedades causadas por las variantes genéticas causales. En este documento, un dispositivo puede configurarse para ensayar uno o más AIM y los AIM pueden proporcionar un conocimiento previo sobre el usuario del dispositivo y para la población o poblaciones a las que pertenece. El conocimiento previo del fondo de la población de un usuario puede entonces utilizarse por el software y los métodos descritos en este documento para ajustar la probabilidad de que el usuario sea un portador para un rasgo específico o una pluralidad de rasgos. Por ejemplo, si una persona da un resultado positivo en ensayo para un alelo que causa la enfermedad de Tay Sachs y también da un resultado positivo en ensayo para un AIM asociado con los judíos asquenazíes, entonces se puede calcular la probabilidad de que la persona sea un portador para el gen de Tay Sachs. Por otro lado, si una persona da un resultado positivo en ensayo para un alelo que causa la enfermedad de Tay Sachs y da un resultado negativo en ensayo para un AIM asociado con judíos asquenazíes, entonces puede haber razones para creer que el resultado de Tay Sachs es un falso positivo y que puede reflejarse en el cálculo de probabilidad como se describe en este documento.

En un ejemplo, dos usuarios (un hombre, una mujer) reciben su probabilidad de ser portadores para una pluralidad de variantes genéticas causales como se describe en este documento y la probabilidad de que el hijo de los dos

usuarios desarrolle rasgos de las variantes genéticas causales se puede calcular usando los métodos y sistemas de este documento.

En algunos casos, un dispositivo comprende análisis para variantes genéticas causales para rasgos para los cuales diferentes poblaciones están en riesgo aumentado, así como análisis de AIM para estas poblaciones. Los rasgos pueden ser rasgos mendelianos que tienen baja incidencia en una población.

Los dispositivos en este documento están configurados para ensayar variantes genéticas y/o marcadores informativos de ascendencia. Se conocen muchos de estos dispositivos para estos propósitos. Incluyen, sin limitación, series de ácido nucleico (incluyendo series de oligonucleótidos, series de aplicación puntual, series generadas fotolitográficamente, series aleatorias y series completas de hibridación genómica) y secuenciadores de ácidos nucleicos. Estos dispositivos detectan variantes genéticas o marcadores informativos de ascendencia en diversas maneras. Dichas maneras incluyen secuenciación por hibridación, secuenciación por ligamiento, secuenciación por extensión, secuenciación de Sanger, secuenciación de Maxam Gilbert y pirosecuenciación. Estos dispositivos y metodologías son bien conocidos en la técnica. El conocimiento de la secuencia o variante específica que se desea detectar puede proporcionar la información necesaria para configurar el dispositivo.

Un dispositivo de este documento puede crearse a partir de, por ejemplo, series de ácido nucleico (por ejemplo, microseries de ADN Affymetrix o series de perlas Illumina), única secuencia de molécula por métodos de síntesis (por ejemplo, Helicos BioSciences Corporation), amplificación de moléculas de ácido nucleico sobre una perla (por ejemplo, 454 Lifesciences), tecnología de series de una sola molécula clonal (por ejemplo, Solexa, Inc.), polimerización de una sola base usando fluorescencia potenciada de nucleótidos (por ejemplo, Genovoxx GmbH) y series de ácidos nucleicos de una sola molécula (por ejemplo, Drmanac, solicitud de patente de Estados Unidos N° 2007/0099208).

Se describen métodos y dispositivos en muchas patentes incluyendo, por ejemplo, la patente de Estados Unidos 7368265 (Brenner et al. Selective genome amplification), patente de Estados Unidos 7361468 (Liu et al. Methods for genotyping polymorphisms), patente de Estados Unidos 7332277 (Dhallan, Methods for detection of genetic disorders), patente de Estados Unidos 6576424 (Fodor, nucleic acid arrays) y patente de Estados Unidos 7323305 (Leamon et al., Methods of amplifying and sequencing nucleic acids).

A. Tipos de dispositivos y metodología

Con la excepción de las modificaciones epigenéticas estables, las variantes genéticas y marcadores a nivel de ADN normalmente se detectan como variaciones en la secuencia en nucleótidos de ADN. El procesamiento para detectar variaciones en la secuencia en nucleótidos de ADN también se llama genotipado. El genotipado de ADN en una muestra se realiza poniendo en contacto la muestra con un dispositivo o dispositivos.

Es posible detectar variaciones en la secuencia en nucleótidos de ADN mediante secuenciación aleatoria hasta un punto tal que se consiga cobertura suficiente. La secuenciación aleatoria normalmente no es verdaderamente aleatoria sino pseudoaleatoria, y puede estar sesgada. Varias tecnologías permiten la secuenciación aleatoria. Es preferible que la tecnología de secuenciación supere varios estándares de rendimiento, incluyendo la precisión. Puede usarse una tecnología comercial de secuenciación para facilitar el cumplimiento de los estándares de alto rendimiento. Sin embargo, también puede usarse una plataforma no comercial.

En otra realización, las variantes se ensayan por análisis de SNP, hibridación genómica complementaria, secuenciación génica, captura de secuencia, análisis de DIP, análisis de STR, y análisis de CNV. En otra realización, los sub-dispositivos comprenden una serie de ácido nucleico, una serie de perlas, una serie CGH, una colección de sondas inversión molecular, una colección de sondas de oligonucleótido específico de alelo (ASO), pirosecuenciación, secuenciación de Sanger, secuenciación por síntesis y secuenciación por hibridación. En otra realización, la variante genética causal es un SNP (por ejemplo, de sentido erróneo, sin sentido, de corte y ajuste o regulador), DIP (por ejemplo, delección pequeña, inserción pequeña, delección grande, o inserción grande), CNV, polimorfismo de repetición, o reordenamientos complejos.

La secuenciación se puede realizar mediante pirosecuenciación. Esta técnica se basa en la detección de pirofosfato liberado (Hyman, 1988. A new method of sequencing DNA. Anal Biochem. 174:423-36; y Ronaghi, 2001. Pyrosequencing sheds light on DNA sequencing. Genome Res. 11:3-11). La secuenciación se puede realizar mediante secuenciación por síntesis. La secuenciación por síntesis también se llama secuenciación por incorporación (patente de Estados Unidos 7344865). La secuenciación se puede realizar mediante secuenciación por ligamiento. El sistema de secuenciación SOLiD de Applied Biosystems realiza la secuenciación por ligamiento. La secuenciación se puede realizar mediante secuenciación de Sanger. La secuenciación de Sanger también se llama secuenciación de terminación de cadena didesoxi. La secuenciación se puede realizar mediante secuenciación de una sola molécula, que es una clase amplia de tecnologías de secuenciación en diversas fases de desarrollo.

La secuenciación aleatoria puede ser un método ineficaz para detectar variaciones particulares en la secuencia de

ADN. Para aumentar la cobertura eficaz de variantes particulares, la secuenciación aleatoria puede aumentarse por amplificación selectiva de las regiones de ADN que contienen las variantes de interés. Varias tecnologías permiten amplificación selectiva. Es preferible que la tecnología de amplificación selectiva supere varios estándares de rendimiento, incluyendo la fidelidad. Puede usarse una tecnología de amplificación selectiva comercial para facilitar el cumplimiento de los estándares de alto rendimiento. Con la validación y optimización apropiadas, puede usarse una tecnología no comercial. El objetivo de la amplificación selectiva es enriquecer las secuencias particulares dentro de una mezcla compleja.

La amplificación selectiva se puede realizar por reacción en cadena de la polimerasa (PCR). La PCR puede ser combinada. Se pueden realizar múltiples reacciones en paralelo. La PCR requiere la optimización de las condiciones de reacción, una práctica convencional de laboratorio.

La amplificación selectiva se puede realizar por captura de secuencia. La captura de secuencia implica el uso de hibridación para enriquecer secuencias de interés. La captura de secuencia puede realizarse usando una microserie (Albert TJ, et al. Direct selection of human genomic loci by microarray hybridization. *Nature Methods*. 2007 Nov; 4(11):903-5; Okou DT, et al. Microarray-based genomic selection for high-throughput resequencing. *Nature Methods* 2007 Nov; 4(11):907-9; y Hodges E, et al. Genome-wide in situ exon capture for selective resequencing. *Nature Genetics* 2007 Dec; 39(12): 1522-7). Por ejemplo, puede usarse la tecnología de captura de secuencia NimbleGen para amplificación selectiva. La amplificación selectiva se puede realizar mediante amplificación por círculo rodante. La amplificación selectiva se puede realizar por clonación. La clonación puede implicar el uso de una endonucleasa de restricción para escindir una mezcla compleja de ADN en fragmentos de longitudes predecibles.

También es posible detectar variaciones particulares en la secuencia de nucleótidos de ADN por genotipado dirigido. Varias tecnologías permiten el genotipado dirigido. Varias tecnologías permiten la amplificación selectiva. Es preferible que la tecnología de amplificación selectiva supere varios estándares de rendimiento, incluyendo la fidelidad. Puede usarse una tecnología de amplificación selectiva comercial para facilitar el cumplimiento de los estándares de alto rendimiento. Con la validación y optimización apropiadas, puede usarse una tecnología no comercial. La eficacia de las técnicas de genotipado dirigido también puede aumentar la amplificación selectiva de las regiones de ADN que contienen las variantes de interés, incluyendo PCR, captura de secuencia, amplificación por círculo rodante, y clonación.

El genotipado dirigido se puede realizar mediante secuenciación por hibridación. La secuenciación por hibridación se puede realizar utilizando una microserie. El genotipado dirigido se puede realizar mediante microseries. Un ejemplo de una tecnología que permite el genotipado dirigido de microseries es el Affymetrix Affymetrix Genome-Wide Human SNP Array 6.0. El genotipado dirigido se puede realizar mediante series de perlas. Un ejemplo de una tecnología que permite el genotipado dirigido por series de perlas es el Illumina Human1M BeadChip. El genotipado dirigido se puede realizar por series de hibridación genómica comparativa (CGH). Las series CGH implican la co-hibridación de una muestra de referencia y una muestra diana a la misma serie. El genotipado dirigido se puede realizar mediante sondas de inversión molecular. Un ejemplo de una tecnología que permite el genotipado dirigido por sondas de inversión molecular es la plataforma de genotipado dirigido Affymetrix GeneChip. El genotipado dirigido se puede realizar mediante sondas de oligonucleótido específico de alelo (ASO). El diseño de las sondas ASO permite una gran cantidad de permutaciones. Las sondas ASO pueden ser de 15 a 21 nucleótidos de longitud. Las sondas ASO pueden contener cero, uno, dos, tres, o más desapareamientos en comparación con las variantes que se están detectando. Las sondas ASO pueden contener nucleótidos de tramos continuos en un extremo con cero o un desapareamiento con las variantes que se están detectando. El genotipado dirigido se puede realizar por PCR específica de alelo. El genotipado dirigido se puede realizar mediante análisis TaqMan. El genotipado dirigido se puede realizar por extensión de una sola base. El genotipado dirigido se puede realizar por tipado de polimorfismos de fragmentos de longitud de restricción (RFLP). El genotipado dirigido se puede realizar mediante PCR de diagnóstico. Se ha desarrollado diversas técnicas de PCR de diagnóstico. En un ejemplo, la amplificación por PCR de una región que flanquea un DIP producirá fragmentos de diferentes tamaños de acuerdo con la presencia o ausencia del DIP. En otro ejemplo, la amplificación por PCR de una región que flanquea un STR producirá fragmentos de diferentes tamaños de acuerdo con los alelos del STR que estén presentes.

Para aumentar la eficacia, puede usarse una técnica de genotipado combinado. La mayoría de las tecnologías de genotipado permiten de forma intrínseca el genotipado combinado. Aquellas tecnologías que no lo hacen combinado de forma intrínseca, pueden hacerse para funcionar como genotipado combinado mediante la realización de múltiples procedimientos de genotipado singulares en paralelo o en serie. De este modo el genotipado también puede describirse como de alto rendimiento.

La existencia de nucleótidos de un SNP en una muestra puede determinarse por diversas tecnologías, incluyendo una cantidad que no enumerada expresamente aquí. Las series CGH generalmente no son adecuadas para el genotipado de SNP. La existencia de nucleótidos de un DIP en una muestra puede determinarse por diversas tecnologías, dependiendo del tamaño de la delección/inserción. En general, delecciones/inserciones pequeñas, por ejemplo aquellas de menos de 20 nucleótidos de tamaño, pueden genotiparse con las tecnologías usadas para genotipar SNP. DIP más grandes pueden tratarse como CNV. La existencia de nucleótidos de un CNV en una muestra puede determinarse por varias tecnologías. Las tecnologías de genotipado de SNP se han adaptado para

su uso en genotipado de CNV. Las series CGH también son a menudo adecuadas para el genotipado de CNV. La existencia de nucleótidos de un STR en una muestra puede determinarse por varias tecnologías. Los STR son el genotipado más común por PCR de diagnóstico.

Pueden detectarse modificaciones epigenéticas estables, tales como metilación, por varias tecnologías por ejemplo, como se describe en la patente de Estados Unidos 7364855. Por ejemplo, el panel de metilación Illumina GoldenGate y el Illumina HumanMethylation27 BeadChip pueden usarse para detectar diferencias de metilación. Puede usarse secuenciación con bisulfito para determinar la presencia de metilación. La metilación puede detectarse mediante pirosecuenciación, análisis de conformación monocatenaria sensible a metilación (MS-SSCA), análisis de fusión de alta resolución (HRM), extensión del cebador de un único nucleótido sensible a metilación (MS-SnuPE), escisión específica de base/MADLI-TOF, PCR específica de metilación (MSP), y otros.

Podría genotiparse un panel de variantes genéticas y marcadores combinando varias tecnologías diferentes en un único sistema.

Algunas variantes genéticas y marcadores pueden detectarse a nivel de polinucleótido de ARN. Puede usarse reacción en cadena de la polimerasa con transcripción inversa (RT-PCR) en lugar de PCR o además de PCR para detectar variantes genéticas a nivel de polinucleótidos de ARN. Las variantes genéticas pueden alterar la secuencia de un polinucleótido de ARN. La detección de variantes genéticas que afectan a la secuencia de un polinucleótido de ARN es esencialmente equivalente a la detección de variantes genéticas que afectan a la secuencia de un polinucleótido de ADN. Las variantes genéticas se pueden detectar como la presencia o ausencia de un polinucleótido de ARN, o más generalmente como la concentración de un polinucleótido de ARN. Pueden usarse varias técnicas para determinar la presencia, ausencia, o concentración de un polinucleótido de ARN. Puede usarse RT-PCR para determinar la presencia, ausencia, o concentración de un polinucleótido de ARN. Pueden usarse microseries de aplicación puntual, microseries de oligo, y series de perlas para determinar la presencia, ausencia, o concentración de un polinucleótido de ARN.

Puede construirse un dispositivo que detecte variantes genéticas y marcadores, tanto a nivel de ADN como a nivel de ARN.

Algunas variantes genéticas y marcadores pueden detectarse al nivel de polipéptidos de proteínas. Las variantes genéticas pueden alterar la secuencia de un polipéptido de proteína. Las variantes genéticas se pueden detectar como la presencia o ausencia de un polipéptido de proteína, o más generalmente como la concentración de un polipéptido de proteína.

B. Variantes genéticas causales

Los dispositivos de este documento están configurados para ensayar variantes genéticas causales y/o marcadores informativos de ascendencia. La detección de la presencia de una variante o marcador incluye, por implicación, la detección de su ausencia. En algunos casos, los dispositivos están configurados para realizar el genotipado sobre el rasgo, no solamente la presencia de una forma polimórfica de la variante o gen. Los análisis seleccionados pueden incluir variantes para una pluralidad de diferentes poblaciones en riesgo; por ejemplo, al menos 2 poblaciones diferentes, al menos 5 poblaciones diferentes, al menos 10 poblaciones diferentes, al menos 25 poblaciones diferentes, al menos 50 poblaciones diferentes o al menos 100 poblaciones diferentes. Por ejemplo, el dispositivo podría configurarse para ensayar una o más variantes genéticas causales para las cuales la población de Eurasia occidental está en riesgo aumentado, una o más causales variantes genéticas para las cuales la población de África subsahariana está en riesgo aumentado, uno o más variantes genéticas causales para las cuales la población del este de Asia está en riesgo aumentado y una o más variantes genéticas causales para las cuales la población nativa americana se encuentra en riesgo aumentado.

En un aspecto, se describe un dispositivo que ensaya una pluralidad (por ejemplo, al menos 100, al menos 500, al menos 1000) rasgos. Existen enfermedades en poblaciones con diferente prevalencia. Los rasgos o enfermedades pueden clasificarse de acuerdo a la prevalencia, muchos se agruparán entre los rasgos poco comunes, por ejemplo, rasgos que tienen una prevalencia de 1/1000 o menos, o 1/10.000 o menos. Se dice que dichos rasgos habitan en la larga cola de dicha clasificación. El concepto de una larga cola se visualiza fácilmente por un gráfico de barras en que el eje Y indica la prevalencia y los rasgos están ordenados a lo largo del eje X de más a menos prevalente. En ciertas realizaciones, el número y los rasgos y la incidencia de cada rasgo se pueden elegir de manera que la incidencia colectiva de las enfermedades raras ensayadas ascienda a al menos el 1 %, al menos el 2 %, al menos el 4 %, o al menos el 5 %.

En ciertos dispositivos, las variantes genéticas causales ensayadas tienen diferente incidencia en diferentes poblaciones o subpoblaciones. La diferencia en la incidencia es al menos estadísticamente significativa ($p < 0,05$) y puede aumentarse en al menos un 10 %, al menos un 100 %, al menos 5 veces, al menos 10 veces, al menos 100 veces, al menos 1000 veces o al menos 10.000 en dos poblaciones diferentes. Una variante puede causar un rasgo que tiene la misma incidencia en dos poblaciones diferentes, pero normalmente tendrá incidencia diferente en al menos dos poblaciones diferentes representadas en el dispositivo.

C. Marcadores informativos de ascendencia

Los marcadores informativos de ascendencia que ensaya un dispositivo normalmente se seleccionan con referencia a las variantes que se están ensayando. En ciertas realizaciones, el dispositivo ensaya variantes genéticas causales que tienen diferente incidencia en diferentes poblaciones o subpoblaciones. Pueden distinguirse AIM entre dos poblaciones para las cuales una de las variantes genéticas causales muestra una diferencia en la incidencia. El AIM también puede clasificar a una persona como perteneciente o no perteneciente a la población que está en riesgo aumentado de una de las variantes genéticas causales, por ejemplo, el AIM puede ser de diagnóstico para la población en la que el rasgo está en prevalencia aumentada. En ciertos casos, el AIM puede distinguir entre poblaciones con granularidad más fina, por ejemplo, entre grupos sub-continentales o grupos étnicos relacionados. En este ejemplo, un dispositivo comprende un ensayo para una variante genética causal que tiene diferente incidencia entre estas subpoblaciones.

Debido a que las variantes genéticas causales pueden tener diferentes frecuencias en diferentes poblaciones, tienen algún nivel de valor predictivo sobre la población de la cual proviene el individuo. En ciertas realizaciones, los AIM usados en el dispositivo no son, en sí mismos, variantes genéticas causales. En otras realizaciones se selecciona el AIM o colección AIM usada para predecir la población de origen para que tenga un valor predictivo de al menos el 80 %, 90 %, 95 %, 98 %, o 99 %.

En otra realización se seleccionan los AIM para clasificar al menos tres poblaciones diferentes. En otra realización se seleccionan los AIM para clasificar al menos cuatro diferentes agrupaciones de escala continental. Los AIM pueden seleccionarse para clasificar al menos tres poblaciones diferentes del mismo continente. En otra realización, se seleccionan los AIM para clasificar al menos diez agrupaciones sub-continentales. En otra realización, los AIM se seleccionan para clasificar tres poblaciones seleccionadas entre europeos, asiáticos orientales, asiáticos del sur, africanos subsaharianos y nativos americanos. En otra realización, los AIM están localizados en al menos 11 cromosomas humanos diferentes.

D. Validación de dispositivos

En un aspecto de este documento, los métodos y conjuntos de combinaciones de ácidos nucleicos se proporcionan para validar los dispositivos descritos en este documento. Las muestras de ADN genómico no están disponibles al público para muchas enfermedades genéticas raras. Por lo tanto, puede ser difícil validar o controlar la calidad de un dispositivo configurado para detectar más de 85 enfermedades genéticas raras simultáneamente.

En un aspecto, se describe un conjunto de combinaciones de ácidos nucleicos para validar un dispositivo de detección de secuencias de ácido nucleico, donde cada combinación de ácidos nucleicos comprende una pluralidad de segmentos de ácido nucleico que se unen selectivamente a una pluralidad de sondas de variantes genéticas causales, y donde cada combinación se une a una pluralidad diferente de dichas sondas. En algunos casos, una primera combinación del conjunto comprende un primer segmento de ácido nucleico que interfiere durante la detección con un segundo segmento de ácido nucleico de una segunda combinación del conjunto, y donde la primera combinación no comprende el segundo segmento de ácido nucleico y la segunda combinación no comprende el primer segmento de ácido nucleico. En algunos casos, los segmentos de ácido nucleico de cada combinación son una sola molécula de ácido nucleico. En algunos casos, los segmentos de ácido nucleico comprenden uno o más plásmidos.

En un aspecto, un método para validar un lote de dispositivos de detección de secuencia de ácido nucleico fabricados comprende: poner en contacto una pluralidad de dispositivos de detección de secuencia de ácido nucleico del lote con un conjunto de combinaciones de ácidos nucleicos para validar de un dispositivo de detección de secuencia de ácido nucleico, donde cada combinación de ácidos nucleicos comprende una pluralidad de segmentos de ácido nucleico que se unen selectivamente a una pluralidad de sondas de variantes genéticas causales, y donde cada combinación se une a una pluralidad diferente de dichas sondas; y detectar una presencia de la pluralidad de sondas de variantes genéticas causales de los dispositivos de detección de ácidos nucleicos, donde el lote de dispositivos fabricados se valida si todos de la pluralidad de sondas de variantes genéticas causales están presentes. En algunos casos, el método comprende además la entrega del lote de dispositivos fabricados cuando los dispositivos están validados. En algunos casos, el lote de dispositivos fabricados se rechaza si no todas de la pluralidad de sondas de variantes genéticas causales están presentes. En algunos casos, el lote de dispositivos fabricados se modifica y se repite el método si no todas de la pluralidad de sondas de variantes genéticas causales están presentes.

Como se describe en este documento, las muestras de ADN que representan estas enfermedades genéticas raras pueden generarse por síntesis. Puede sintetizarse ADN bicatenario de novo y clonarse en vectores de plásmidos convencionales de alta cantidad de copias. Este enfoque también se ha reconocido por el documento de orientación de la FDA "Pharmacogenetic Tests and Genetic Tests for Heritable Markers". Los plásmidos proporcionan un recurso renovable de ADN para su uso en la validación y para su uso como controles el genotipado del paciente. Los plásmidos se combinarán, creando una muestra artificial del paciente. Los alelos mutantes para cada variante

genética causal de este documento pueden sintetizarse.

En un caso, todos los plásmidos que comprenden todas las variantes genéticas causales a ensayar por un dispositivo de este documento pueden colocarse en una única combinación o muestra grande. En muchos casos, muchas de las variantes genéticas causales están estrechamente espaciadas entre sí. En estos casos, los plásmidos generados para ensayar una variante son también moldes para otras variantes cercanas y los moldes generarán una señal de tipo silvestre. Esto puede hacer que sea difícil para simular apropiadamente la señal de variante genética causal en las sondas fuera de diana con que interaccionan. En un caso, se dividen secuencias solapantes en diferentes combinaciones de plásmidos que componen un conjunto de segmentos de ácido nucleico. Por ejemplo, se colocan variantes del mismo bloque en diferentes combinaciones. En otro ejemplo, la cantidad total de combinaciones necesaria en un conjunto de combinaciones es igual al tamaño del bloque más grande.

Ejemplos de segmentos de ácidos nucleicos y combinaciones de ácidos nucleicos se muestran en la Figura 19. La Figura 19 ilustra la arquitectura de secuencia de las combinaciones de muestra de validación. La cuestión básica es que la validación de un ensayo combinado de ácidos nucleicos se beneficia del uso de moldes que no interfieren para formar los límites de designación de genotipado.

Para la detección de variantes genéticas causales para validar un dispositivo descrito en este documento, se generan combinaciones de plásmidos mediante la combinación de un grupo de cada tamaño de bloque (por ejemplo, un grupo cada uno de tamaño 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, y 34) en la concentración apropiada. Por tanto, cada combinación contiene 455 variantes sintéticas. Por ejemplo:

- Combinación 1 comprende: Tamaño 1 Grupo 0 + Tamaño 2 Grupo 0 + Tamaño 3 Grupo 0 + Tamaño 4 Grupo 0 + Tamaño 5 Grupo 0 + Tamaño 6 Grupo 0 + tamaño 7 Grupo 0 + Tamaño 8 Grupo 0 + Tamaño 10 Grupo 0 + Tamaño 12 Grupo 0 + Tamaño 34 Grupo 0.
- Combinación 2 comprende: Tamaño 1 Grupo 0 + Tamaño 2 Grupo 1 + Tamaño 3 Grupo 1 + Tamaño 4 Grupo 1 + Tamaño 5 Grupo 1 + Tamaño 6 Grupo 1 + Tamaño 7 Grupo 1 + Tamaño 8 Grupo 1 + Tamaño 10 Grupo 1 + Tamaño 12 Grupo 1 + Tamaño 34 Grupo 1.
- Combinación 3 comprende: Tamaño 1 Grupo 0 + Tamaño 2 Grupo 0 + Tamaño 3 Grupo 2 + Tamaño 4 Grupo 2 + Tamaño 5 Grupo 2 + Tamaño 6 Grupo 2 + Tamaño 7 Grupo 2 + Tamaño 8 Grupo 2 + Tamaño 10 Grupo 2 + Tamaño 12 Grupo 2 + 34 Tamaño del grupo 2.

IV. Métodos comerciales

En un aspecto, un método comprende: recibir una muestra de un usuario; ensayar la muestra con un dispositivo de detección de ácido nucleico configurado para ensayar una pluralidad de variantes genéticas causales de enfermedades genéticas raras y al menos un marcador informativo de ascendencia (AIM); calcular una pluralidad de probabilidades correspondientes al usuario de ser un portador para cada una de la pluralidad de variantes genéticas causales basadas en los resultados de la etapa de ensayo con respecto a la pluralidad de variantes genéticas causales y el al menos un AIM; y entregar al usuario la pluralidad de probabilidades correspondientes al usuario de ser un portador. El método puede comprender además: recibir una muestra de un segundo usuario; ensayar la muestra del segundo usuario con un dispositivo configurado para ensayar pluralidad de variantes genéticas causales de enfermedades genéticas raras y al menos un marcador informativo de ascendencia (AIM); calcular una probabilidad de fenotipo del niño correspondiente a las enfermedades genéticas raras basado en los resultados del ensayo del usuario y el segundo usuario; y entregar la probabilidad del fenotipo del niño a al menos uno del usuario y el segundo usuario. El método puede comprender además proporcionar servicios de asesoramiento genético a al menos uno del usuario y el segundo usuario. El método puede realizarse como parte de un servicio de predicción de fenotipos de niños. El método puede comprender además la obtención de información fenotípica del usuario; y el uso de la información fenotípica del usuario en las etapas de cálculo. El método puede comprender además la obtención de la historia familiar del usuario; y el uso de la historia familiar del usuario en las etapas de cálculo.

También se describe en este documento un método que comprende: comercializar un servicio de análisis genético que comprende predecir una probabilidad de un fenotipo del niño a partir de un conjunto de padres o padres potenciales, donde la predicción se basa al menos en parte, en la presencia de una pluralidad de variantes genéticas causales de más de 85 enfermedades genéticas raras de cada uno de los miembros del conjunto de padres o padres potenciales y se basa al menos en parte, en las ascendencias inferidas de cada uno de los miembros del conjunto de padres o padres potenciales; y entregar una probabilidad del fenotipo del niño del conjunto de padres o conjunto de padres potenciales con respecto a las más de 85 enfermedades genéticas raras por una tarifa. La comercialización puede realizarse en relación con un servicio de citas o matrimonio. El método puede comprender, además, remitir al menos a un miembro del conjunto de padres o conjunto de padres potenciales a un médico. En algunos casos, las ascendencias inferidas se infieren por un ensayo de al menos un marcador informativo de ascendencia (AIM).

A. Análisis genético y Sistema de asesoramiento genético de referencia

Un método comercial ejemplar permite la prestación de un servicio de análisis genético que puede predecir la probabilidad de que un descendiente de una pareja tenga cada uno de una pluralidad de rasgos mendelianos.

Además de los resultados del análisis genético, pueden proporcionarse referencias a asesores genéticos y/u otros profesionales médicos pertinentes para proporcionar ensayos de seguimiento y consulta.

En una realización de un método comercial, el análisis genético de un cliente comienza con un pedido del cliente, donde el cliente paga una cuota a cambio de, por ejemplo, materiales de análisis genético y referencias a asesores genéticos y/u otros profesionales médicos pertinentes. Un cliente puede ser, por ejemplo y sin limitación, un médico, un asesor genético, un centro médico, un individuo, una compañía de seguros, un sitio web, un servicio de citas, un servicio de intermediación, una compañía farmacéutica, un proveedor de servicios de ensayo de laboratorio, o un fabricante de plataformas de diagnóstico. Por ejemplo, un cliente puede ser una pareja de padres potenciales que tratan de saber si su descendencia estará en riesgo de desarrollar algún tipo de enfermedad mendeliana o enfermedad mendeliana rara. Después de que un cliente haga un pedido, puede enviarse un kit de recogida de ADN al cliente. En una realización de un método comercial de la invención, un cliente puede depositar una muestra en el kit de recogida. Cualquier muestra que fuera obvia para un especialista en la técnica puede depositarse en o sobre una muestra de recogida. Una muestra puede ser cualquier compuesto químico que sería obvio para un especialista en la técnica, tal como un fluido corporal, como saliva o una muestra de aliento que contiene compuestos químicos correlacionados (véase, por ejemplo, http://www.popularmechanics.com/blogs/science_news/4220196.html), a partir de la cual es posible identificar y extraer la información genotípica sin procesar requerida (como se discute en este documento). El kit de recogida puede devolverse a la empresa para enviar a un laboratorio de genotipado o puede devolverse directamente al laboratorio de genotipado para su procesamiento. Un laboratorio de genotipado, ya sea interno dentro de la empresa, contratado para trabajar con la empresa, o externo de la empresa, puede aislar el ADN del cliente a partir de la muestra proporcionada. Después de haber aislado el ADN de la muestra, un dispositivo ejemplar de la invención puede ensayar el ADN para la presencia de (i) marcadores informativos de ascendencia y (ii) variantes genéticas causales (la combinación de (i) y (ii) también se menciona en este documento como información genotípica sin procesar). En una realización, el ADN no tiene que aislarse de la muestra para ensayar el ADN para la presencia de información genotípica sin procesar.

La información genotípica sin procesar puede procesarse para inferir la ascendencia del cliente y para confirmar o negar la presencia de variantes genéticas causales (también mencionado en este documento como información genotípica procesada). En una realización, la información genotípica procesada se puede transmitir a un sistema de la invención u otro sistema de cálculo que sería obvio para un especialista en la técnica para predecir la probabilidad de que un descendiente del cliente tenga cada uno de una pluralidad de rasgos causados por variantes genéticas causales, encontradas presentes en el ADN del cliente.

Tras el análisis del ensayo genotípico, los resultados del ensayo se pueden proporcionar al cliente. Los resultados proporcionados a un cliente pueden informar al cliente del estado de portador de un individuo para una enfermedad mendeliana particular y/o las posibilidades de que la descendencia futura del individuo desarrolle la enfermedad mendeliana. En una realización, el cliente también recibe, por ejemplo, a) consulta telefónica directa con un asesor genético empleado por la empresa, o b) información de contacto de asesores genéticos y/u otros profesionales médicos que pueden ofrecer al cliente ensayo de seguimiento y consulta.

También se describe en este documento un método que comprende la comercialización de un servicio de análisis genético en relación con un servicio de citas o matrimonio, donde el servicio de análisis genético comprende: (a) predecir la probabilidad de que un descendiente de la pareja tenga cada uno de una pluralidad de rasgos causados por variantes genéticas causales, donde la predicción se determina en base a los respectivos genotipos de los dos individuos de la pareja y sus respectivas ascendencias inferidas genéticamente; y (b) remitir a la pareja a un asesor genético y/u otro profesional médico (por ejemplo, médico genetista o obstetra/ginecólogo) en base a los resultados del análisis genético. En otra realización, el método comprende además la co-comercialización de un servicio que comprende: c) predecir el fenotipo probable de una descendencia de una pareja para al menos un rasgo, donde la probabilidad se determina basándose en los respectivos fenotipos de los individuos en la pareja. En otra realización, el servicio es un servicio en línea.

En otro componente de un método comercial, puede almacenarse una lista completa de todos los asesores genéticos y/u otros profesionales médicos para ensayos de seguimiento para todas las afecciones patológicas en una base de datos. Las personas en la base de datos pueden ser aquellos que previamente han consentido dar sus nombres y la información de contacto usada en relación con el servicio de análisis genético. En una realización, un conjunto particular de referencias proporcionado a cualquier cliente particular puede variar y dependerá de los resultados del análisis genético. Por ejemplo, es relevante mostrar ciertos asesores genéticos y/u otros profesionales médicos a un cliente sólo en relación con ciertas enfermedades genéticas mendelianas; tales como aquellas para las que tienen una especialidad, por ejemplo. Así ciertas referencias sólo se mostrarán en relación con ciertos resultados del ensayo.

B. Análisis genético y servicios de citas/matrimonio

Los servicios de intermediación en última instancia, tienen como objetivo conectar a dos personas con un propósito de cita y/o matrimonio. Se puede pensar sobre cualquier usuario de un servicio de intermediación como la evaluación de otros diversos usuarios del servicio de intermediación para determinar la probabilidad de un

emparejamiento positivo, por ejemplo, la probabilidad de que dos usuarios cualesquiera sean compañeros compatibles a largo plazo. Como en cualquier proceso de selección de este tipo, se evalúan numerosos factores para determinar si un emparejamiento será un éxito, por ejemplo, el atractivo, los orígenes socioeconómicos, los intereses compartidos, el alma máter compartida son unos pocos seleccionados. Como se describe en este documento, el análisis genético pueden permitir la evaluación de posibles candidatos en relación con un servicio de intermediación de acuerdo con otro factor importante: la salud potencial de la descendencia futura como resultado de un emparejamiento dado.

En un aspecto de un método comercial de la invención, puede ofrecerse un servicio de análisis genético a un cliente en relación con un servicio de intermediación, por ejemplo a través de una sola empresa o una relación de co-comercialización o sociedad. Un usuario de un servicio de intermediación puede solicitar el análisis genético para determinar la probabilidad de que una descendencia resultante del emparejamiento entre el usuario y un posible candidato tenga cada uno de una pluralidad de rasgos mendelianos causados por variantes genéticas causales encontradas presentes en el respectivo ADN del usuario y el candidato. El usuario puede entonces usar esta información para ayudar en la evaluación de los candidatos para un posible emparejamiento. El servicio de intermediación puede ser un servicio en línea, como Shaadi.com, eHarmony.com y Match.com.

En una realización, el análisis genético comienza con un pedido del cliente, donde el cliente paga una cuota a cambio del análisis genético. Por ejemplo, un cliente puede ser un usuario de un servicio de intermediación que esté interesado en la evaluación de otro usuario para un emparejamiento adecuado. Dicho cliente puede usar el análisis genético para saber si la descendencia potencial de un emparejamiento entre dicho cliente y un candidato potencial estará en riesgo de desarrollar una enfermedad mendeliana. Después de seleccionar un candidato a evaluar, un cliente puede pagar tanto para el análisis genético del cliente como del candidato con el consentimiento del candidato. El cliente y el candidato también pueden pagar por separado para el análisis genético. Después de que un cliente haga un pedido, se puede un kit de recogida de ADN enviar al cliente. En una realización de un método comercial de la invención, un cliente puede depositar una muestra en el kit de recogida. Cualquier muestra que fuera obvia para un especialista en la técnica puede depositarse en o sobre un kit de recogida. Una muestra puede ser cualquier compuesto químico que sería obvio para un especialista en la técnica - tal como un fluido corporal, tal como saliva o una muestra de aliento que contiene compuestos químicos correlacionados (véase, por ejemplo, http://www.popularmechanics.com/blogs/science_news/4220196.html) - a partir de la cual es posible identificar y extraer la información genotípica sin procesar requerida. El kit de recogida puede devolverse a la empresa para enviar a un laboratorio de genotipado o puede devolverse directamente al laboratorio de genotipado para su procesamiento. Un laboratorio de genotipado, ya sea interno dentro de la empresa, contratado para trabajar con la empresa, o externo de la empresa, puede aislar el ADN del cliente de la muestra proporcionada. Después de haber aislado el ADN de la muestra, puede usarse un dispositivo ejemplar de la invención u puede usarse otro dispositivo de genotipado para ensayar el ADN para la presencia de información genotípica sin procesar. En una realización, el ADN no tiene que aislarse de la muestra para ensayar el ADN para la presencia de información genotípica sin procesar.

La información genotípica sin procesar puede procesarse para inferir la ascendencia del cliente y para confirmar o negar la presencia de variantes genéticas causales (también mencionado en este documento como información genotípica procesada). En una realización, la información genotípica procesada se puede transmitir a un sistema de la invención u otro sistema de cálculo que sería obvio para un especialista en la técnica para predecir la probabilidad de que un descendiente del cliente tenga cada uno de una pluralidad de rasgos causados por variantes genéticas causales, encontrados presentes en el ADN del cliente. Los resultados del ensayo se proporcionarán entonces al cliente y el candidato. En este punto, el cliente y el candidato pueden saber la probabilidad de que la futura descendencia resultante de un emparejamiento entre el cliente y el candidato desarrolle la enfermedad mendeliana.

También se describe en este documento un método que comprende la comercialización de un servicio análisis genético en relación con un servicio de citas o matrimonio, donde el servicio de análisis genético comprende: (a) predecir la probabilidad de que un descendiente de la pareja tenga cada uno de una pluralidad de rasgos, donde la predicción se determina en base a los respectivos fenotipos y/o antecedentes familiares de los dos individuos de la pareja. En otra realización el método comprende además (b) remitir a la pareja a un asesor genético y/u otro profesional médico (por ejemplo, médico genetista u obstetra/ginecólogo) en base a los resultados del análisis genético.

C. Predicción de fenotipos de la descendencia a partir de los fenotipos de los padres

El análisis genético puede permitir a los padres o parejas potenciales determinar la probabilidad de que la descendencia futura resultante de su unión desarrolle una enfermedad genética mendeliana. Conceptualmente, esta previsión es posible debido al modo predecible de herencia de un rasgo mendeliano y de la correlación generalmente alta entre tener una variante y tener la afección, a pesar de los problemas implicados con la expresividad y penetrancia variables. Para rasgos complejos y/o no mendelianos, todavía es difícil predecir sus valores a partir de mediciones de genotipo solamente. Aunque es difícil predecir los valores de los rasgos complejos de un individuo a partir de la información del genotipo, a menudo es fácil predecir los valores de los rasgos complejos de un niño a partir de la información del fenotipo de los padres. Es decir, para fenotipos con alta

heredabilidad, una medición de los valores de rasgos de la madre y el padre es suficiente para proporcionar una excelente capacidad de predicción para los rasgos del niño (<http://www.amazon.com/Genetics-Analysis-Quantitative-Traits-Michael/dp/0878934812>).

En otro aspecto, esta descripción proporciona un método que comprende predecir el fenotipo probable de una descendencia de una pareja para al menos 10 rasgos no mendelianos, donde la probabilidad se determina basándose en los respectivos fenotipos de los individuos en la pareja. En una realización, los rasgos fenotípicos incluyen un miembro seleccionado del grupo que consiste en la altura, el peso, la capacidad cognitiva, el rendimiento académico, la extraversión, el neuroticismo, la amabilidad, la responsabilidad, la franqueza, la religiosidad y el conservadurismo.

En una realización, las predicciones fenotipo-fenotipo se pueden ofrecer de forma gratuita en relación con el análisis genético como un medio para permitir a los padres o parejas posibles aprender acerca de los rasgos no mendelianos que su futura descendencia pueda tener.

Un cliente puede proporcionar información que describe su fenotipo, así como información que describe el fenotipo su pareja (fenotipos de los padres). En una realización, estos fenotipos de los padres pueden transmitirse a un sistema de la invención u otro sistema de cálculo que sería obvio para un especialista en la técnica para predecir el fenotipo probable de una descendencia del cliente y su pareja basado en la entrada fenotipos de los padres (fenotipo de la descendencia). El fenotipo de la descendencia puede entregarse al cliente de forma gratuita o por una tarifa.

D. Sistemas Informáticos

En otro aspecto, esta descripción proporciona un método que comprende ofrecer un primer y segundo servicio, donde: a) el primer servicio comprende predecir la probabilidad de que un descendiente de la pareja tenga cada uno de una pluralidad de rasgos causados por variantes genéticas causales, donde la predicción se basa en los respectivos genotipos de los dos individuos en la pareja; y b) el segundo servicio comprende predecir el fenotipo probable de un descendiente de la pareja para una pluralidad de rasgos, donde la probabilidad se determina en base a los respectivos fenotipos y/o la historia familiar de los individuos en la pareja. En una realización, al menos una predicción se basa además en los respectivos ascendientes inferidos genéticamente de los individuos. En otra realización, el primer servicio se ofrece como un servicio por una tarifa y el segundo servicio se ofrece como un servicio gratuito.

En otro aspecto, esta descripción proporciona un método que comprende: a) tener en una memoria de ordenador las respuestas a un cuestionario sobre antecedentes familiares de cada individuo de una pareja; b) tener en la memoria del ordenador la información genética de cada individuo de la pareja; c) mostrar: i) el estado de portador de al menos un individuo para al menos un rasgo determinable por la información genética o la historia familiar o ii) características de la descendencia de la pareja determinables por los antecedentes familiares y/o la información genética. En una realización el estado de portador se muestra en un sitio web.

En otro aspecto, la descripción proporciona un sistema que comprende: a) un medio legible por ordenador configurado para almacenar información de historia familiar de cada miembro de una pareja; b) un medio legible por ordenador configurado para almacenar datos que comprenden la información genética de cada miembro de la pareja; c) un medio legible por ordenador que comprende código informático que, cuando se ejecuta: i) predice el estado de portador de cada individuo con respecto a rasgos causados por alelos identificados en la información genética; o ii) predice rasgos probables de la descendencia de la pareja determinables por los antecedentes familiares y/o la información genética; y d) una pantalla que muestra: i) el estado de portador de al menos un miembro de la pareja o ii) los rasgos probables de la descendencia. En una realización, el sistema comprende además e) una página web configurada para aceptar una oferta de compra de un kit de ensayo de ADN. En otra realización, la pantalla es electrónica, por ejemplo, una página web. En otra realización, el sistema comprende además e) una pantalla que muestra referencias a un asesor genético y/u otro profesional médico (por ejemplo, médico genetista u obstetra/ginecólogo) en base a la información genética.

El Internet y la World Wide Web ofrecen acceso a y distribución de información. Por lo tanto, en una realización, un sitio web puede ser particularmente adecuado para proporcionar de forma eficaz diversas funcionalidades para permitir que los clientes compren análisis genéticos y reciban los resultados de los análisis genéticos. El sistema incluirá normalmente un servidor en el que reside la página web. Los usuarios usan una interfaz conectada al servidor, tal como un monitor de ordenador o una pantalla de teléfono, para interactuar con el sitio web haciendo clic o rodando sobre enlaces que despliegan información o dirigen al usuario a otra página web. Los sitios web normalmente son interactivos, lo que permite al usuario introducir información o consultar y obtener una respuesta en la interfaz.

En un aspecto de un sistema método comercial, un sitio web puede permitir a un cliente adquirir, administrar y ver los resultados del análisis genético, así como para aprender de manera más general acerca de la probabilidad de que la potencial descendencia desarrolle una enfermedad mendeliana. Por ejemplo, un cliente puede ser una pareja de padres posibles que buscan para saber si su descendencia estará en riesgo de desarrollar una enfermedad mendeliana.

Después de que un cliente haya creado una cuenta con la empresa en el sitio web, el cliente en primer lugar puede optar por disfrutar de un servicio gratuito que predice la probabilidad de que la descendencia del cliente tenga un rasgo causado por variantes genéticas causales. En otra realización, se proporciona el servicio mencionado anteriormente por una tarifa. El servicio puede ofrecerse al cliente en forma de un cuestionario en línea que se promueve y es accesible a través del sitio web. El cuestionario también puede ser un cuestionario que se solicita y se envía por correo físico o por correo electrónico a un cliente. Un cuestionario puede hacer al cliente una serie de preguntas relacionadas con la historia familiar y la etnia que un cliente puede contestar. Una vez dadas las respuestas un cliente, las respuestas se pueden enviar a un servidor para su almacenamiento y procesamiento. En un ejemplo, el código informático se ejecuta en el servidor para calcular (i) el estado de portador del cliente, si lo hay; y (ii) la probabilidad de que un descendiente del cliente desarrolle cada uno de una pluralidad de rasgos mendelianos, basada en las respuestas proporcionadas por el cliente al cuestionario. Los resultados de esta operación entonces se envían a un servidor para su almacenamiento. El servidor puede transmitir electrónicamente esta información a una interfaz de usuario, tal como el sitio web o un correo electrónico, para su presentación al cliente. En una realización particular, los resultados del cuestionario se presentan al cliente en la sección de resultados del sitio web. Los resultados del cuestionario comprenden (i) el estado de portador del cliente, si lo hay; y (ii) una lista de la probabilidad de que un descendiente del cliente desarrolle cada uno de una pluralidad de rasgos mendelianos.

Al cliente entonces se le puede presentar la oferta de compra del análisis genético para determinar (i) el estado de portador del cliente; y (ii) la probabilidad de que un descendiente del cliente desarrolle cada uno de una pluralidad de rasgos mendelianos, en base a las variantes genéticas causales que se encuentran presentes en el ADN del cliente.

Si el cliente opta por comprar el análisis genético, entonces el cliente puede pagar una cuota, por ejemplo a través de una transacción de tarjeta de crédito en línea, a cambio del análisis genético, consulta telefónica directa con un asesor genético en el personal de la empresa y/o referencias a los asesores genéticos y/u otros profesionales médicos pertinentes. El análisis genético y las referencias pueden pagarse por una cuota en el momento de la compra o pueden incluirse en una cuota inicial de registro del usuario. En una realización, los servicios son gratuitos y los ingresos de la empresa se generan por la publicidad de otros productos en combinación con un producto particular. Por ejemplo, después de que un cliente haya un pedido en línea, la orden se envía a un servidor para su procesamiento. Una vez se ha verificado el pago, el servidor de procesamiento de pedidos puede enviar una notificación electrónica a un vendedor de envíos para enviar por correo un kit de recogida de ADN al cliente. En una realización, el kit de recogida de ADN está separado del servicio de análisis genético o el usuario o cliente ya tiene u obtiene el kit de recogida de ADN de otra fuente. También pueden enviarse electrónicamente notificaciones periódicamente al cliente que comprende confirmación del pedido y actualizaciones de la orden y estado del envío. Una vez que el cliente recibe el kit de recogida, el cliente puede usarlo para obtener información genotípica sin procesar (como se define a continuación). En una realización de un método comercial de la invención, un cliente puede depositar una muestra en el kit de recogida. Cualquier muestra que fuera obvia para un especialista en la técnica puede depositarse en o sobre un kit de recogida. Una muestra puede ser cualquier compuesto químico que sería obvio para un especialista en la técnica, tal como un fluido corporal, como saliva o una muestra de aliento que contiene compuestos químicos correlacionados (véase, por ejemplo, http://www.popularmechanics.com/blogs/science_news/4220196.html), a partir de la cual es posible identificar y extraer la información genotípica sin procesar requerida (como se define a continuación). El kit de recogida puede devolverse a la empresa para enviar a un laboratorio de genotipado o puede devolverse directamente al laboratorio de genotipado para su procesamiento. Un laboratorio de genotipado, ya sea interno dentro de la empresa, contratado para trabajar con la empresa, o externo de la empresa, puede aislar el ADN del cliente de la muestra proporcionada. Después de haber aislado el ADN de la muestra, se puede usar un dispositivo ejemplar de la invención u se puede usar otro dispositivo de genotipado para ensayar el ADN para la presencia de (i) marcadores informativos de ascendencia y (ii) variantes genéticas causales (la combinación de (i) y (ii) también se menciona en este documento como información genotípica sin procesar). En una realización, el ADN no tiene que aislarse de la muestra para ensayar el ADN para la presencia de información genotípica sin procesar.

La información genotípica sin procesar puede enviarse electrónicamente a un servidor para su almacenamiento y procesamiento. El código informático en el servidor se puede ejecutar sobre la información genotípica sin procesar para inferir la ascendencia del cliente y para confirmar la presencia de variantes genéticas causales, si las hay. La información genotípica procesada puede enviarse electrónicamente a un servidor, donde el código informático en el servidor puede ejecutarse sobre la información genotípica procesada para predecir la probabilidad de que un descendiente del cliente tenga cada uno de una pluralidad de rasgos causados por variantes genéticas causales encontradas presentes en la información genotípica procesada del cliente. Los resultados pueden transmitirse electrónicamente a un servidor para su almacenamiento.

En un ejemplo, puede enviarse una notificación al cliente para avisar al cliente de la disponibilidad de los resultados. La notificación puede ser electrónica, tal como por ejemplo, sin limitación, un mensaje de texto, un correo electrónico, o cualquier otro paquete de datos; o la notificación puede ser no electrónica, tal como por ejemplo, sin limitación, una llamada telefónica de un asesor genético o comunicación impresa, tal como un informe enviado por correo. Los resultados proporcionados a un cliente pueden informar al cliente del estado de portador del cliente para

una enfermedad mendeliana particular y/o las posibilidades de que los futuros descendientes del cliente desarrollen enfermedad mendeliana. Después de que el cliente haya recibido los resultados y las referencias, la orden del cliente puede considerarse cumplida, y los resultados y las referencias pueden permanecer accesibles para el cliente a través de una cuenta del sitio web en línea. El cliente puede elegir la continuación proseguir una referencia fuera de línea si el cliente así lo desea, pero fuera de la competencia del sitio web.

Aunque se han mostrado y descrito realizaciones preferidas de la presente invención en este documento, será obvio para los especialistas en la técnica que dichas realizaciones se proporcionan a modo de ejemplo solamente. Se les ocurrirán numerosas variaciones, cambios y sustituciones a los especialistas en la técnica sin apartarse de la invención. Se debe entender que se pueden emplear diversas alternativas a las realizaciones de la invención descrita en este documento en la práctica de la invención. Se pretende que las siguientes reivindicaciones definan el alcance de la invención y que los métodos y estructuras dentro del alcance de estas reivindicaciones y sus equivalentes están cubiertos por el mismo.

Ejemplo 1

Se proporciona un método en este documento para validar o controlar la calidad de un lote de dispositivos fabricados configurados para ensayar variantes genéticas causales de más de 85 enfermedades genéticas raras. Las muestras de ADN genómico no están disponibles públicamente para muchos genotipos raros, por tanto, puede ser difícil obtener una cantidad apropiada de muestras para ensayar el dispositivo. También puede ser deseable evitar la dilución del ensayo con alelos de tipo silvestre y para evitar la necesidad una gran cantidad de dispositivos para ensayar el lote fabricado. En este ejemplo, se generaron por síntesis muestras de ADN que representan las variantes genéticas causales. Se sintetizó ADN bicatenario de novo y se clonó en vectores de plásmidos convencionales alto número de copias. Estas muestras representan 662 variantes que causan enfermedades y 139 enfermedades. Este enfoque ha sido descrito en varios informes recientes relacionados con el ensayo de portador de FQ (Berwouts et al., 2008; Christensen et al., 2007). Se verificó en secuencia que todos los plásmidos contenían cero errores por metodología BigDye de paseo con cebador, bicatenario.

Los segmentos de ácido nucleico de las variantes genéticas causales estaban organizados en bloques para evitar interferencias durante la detección de cualquiera de las variantes genéticas causales. Las variantes en el mismo bloque se colocaron en diferentes combinaciones. La cantidad total de combinaciones necesaria es igual al tamaño del bloque más grande. Usando una sonda de solapamiento máxima de un 50 % de similitud de secuencia, el bloque más grande es el tamaño de 34 en este ejemplo.

Los recuentos por tamaños de bloque son: 320 bloques de tamaño 1 (variantes singleton); 84 de tamaño 2; 21 de tamaño 3; 12 de tamaño 4; 2 de tamaño 5; 7 de tamaño 6; 3 de tamaño 7; 1 de tamaño 8; 1 de tamaño 10; 3 de tamaño 12; y 1 de tamaño 34.

Los insertos se dividieron en conjuntos que pueden combinarse en uno o más plásmidos para la síntesis. El primer conjunto es el grupo de 320 insertos sin zonas de solapamiento. Los siguientes 2 conjuntos son los grupos de insertos de bloques de tamaño dos (cada uno con 84 insertos). El siguiente es 3 grupos de 21 insertos cada uno para tamaño de bloque 3. El siguiente es 4 grupos de 12 insertos de cada uno, y así sucesivamente de forma que haya n grupos generados por los bloques de tamaño n . En total, se crearon 92 grupos de insertos para la síntesis de plásmidos.

Ejemplo 2

Los procesos a modo de ejemplo para la entrega de una probabilidad de que un usuario es un portador de una enfermedad genética rara se muestran en las Figuras 20-23. Las Figuras 20-21 ilustran las vías para el cumplimiento de pedidos para clientes de web y médicos respectivamente. Puede realizarse un pedido por un médico o un consumidor. Un pedido se puede realizar para un único ensayo o para una pareja o familia. La orden puede aceptarse a través de un sitio web. El sistema de pedidos puede aceptar información de contacto, detalles demográficos e información de facturación. La información de contacto puede incluir, sin limitación, nombre, dirección, número de teléfono y direcciones de correo electrónico. La información demográfica puede incluir, sin limitación, sexo, fecha de nacimiento, y etnia auto-remitida. Puede enviarse una notificación de confirmación de pedido usando la información de contacto proporcionada. Se añaden órdenes aceptables a una base de datos, y los estados de estas órdenes pueden posteriormente mantenerse por una máquina de estados.

Se envía un kit de recogida de muestra al usuario. Se recoge una muestra que es cualquier tejido o fluido humano. La muestra también puede ser ADN aislado de un ser humano. Ejemplos de muestras útiles para este ejemplo incluyen, aunque sin limitación: saliva, sangre, orina, células bucales, líquido amniótico, raspaduras de células, y cultivo celular. La muestra después se genotipa, usando un dispositivo descrito en este documento. La solicitud del fenotipo, por ejemplo, la recuperación de auto-identificación de rasgos fenotípicos de un usuario, puede realizarse en paralelo con el procesamiento de la muestra.

La recogida de muestras se puede realizar en casa, en la oficina de un médico o en un centro de recogida

especializado. La recogida de muestras y el retorno pueden rastrearse por el avance del estado de la máquina de estado de seguimiento orden. Las muestras recibidas por la instalación de registro se pueden registrar en el sistema de base de datos mediante el avance de su estado en la máquina del estado. Después de la aceptación en la instalación de registro, las muestras pueden entregarse a la instalación de genotipado. La instalación de genotipado puede devolver los datos genómicos sin procesar a un servidor de almacenamiento seguro de datos por protocolos de transferencia segura de archivos. La carga de archivos puede provocar un avance de la máquina de estados. Este avance puede desencadenar que un servidor configurado para realizar la designación de genotipo recupere los datos genómicos sin procesar desde el servidor de almacenamiento de datos, así como cualquier dato de fenotipo asociado con la orden. El algoritmo de genotipado puede producir una designación de genotipo totalmente probabilística.

Las figuras 22-23 ilustran las vías de procesamiento de muestras de alto nivel y las vías computacionales detalladas respectivamente. Los algoritmos descritos en las figuras previas y la etapa de procesamiento mostrada en la Figura 22, se describen en la Figura 23. Se reciben lotes de chips y se miden con fines de control de calidad (lote pasa QC). La información tal como la historia familiar, el género o la ascendencia ancestro auto-remitida se utiliza para servir como verificación independiente en la designación para cada muestra

(datos de fenotipo recuperados de muestras de lote). Paralelamente a este proceso, se actualiza constantemente un informe con las predicciones del niño. Se entregan primero los cálculos de riesgo previos al ensayo, en base al fenotipo (tal como historia familiar y otras respuestas a un cuestionario en línea). Una vez que se recibe y procesa una muestra de genotipo, se dan los cálculos posteriores al ensayo. Después se genera el informe y se envía a las fases finales de la vía, para la aprobación del personal de laboratorio y del médico, como se muestra en la Figura 22.

Pueden generarse parámetros de control de calidad a partir del proceso de designación. Un parámetro ejemplar de control de calidad es el porcentaje de designaciones probabilísticas de genotipo en que al menos un genotipo tiene una probabilidad posterior mayor que un valor umbral. Un lote de muestras se procesa conjuntamente. Cuando se procesan como un lote, las designaciones probabilísticas de genotipo individuales pueden usarse para generar estadísticas de control de calidad a nivel de lote. Las designaciones probabilísticas de genotipo pueden almacenarse en una base de datos. El éxito de la designación de genotipo puede desencadenar un avance del estado de la orden. Para una orden de pareja o familia, la máquina de estados puede retenerla hasta completarse la totalidad de la orden, también pueden pasarse órdenes individuales al siguiente estado. Si se requieren de datos de fenotipo para el cálculo del riesgo, entonces la máquina de estados puede demorar el procesamiento hasta que se recogen todos los datos de fenotipo. La máquina de estado también puede desencadenar una notificación al paciente de que se requieren datos de fenotipo. Si todos los datos de genotipo y fenotipo están listos, entonces la máquina de estados puede avanzar, activar el servidor de cálculo de riesgos para realizar el cálculo del riesgo. Los resultados del cálculo de riesgo se pueden serializar y transferirse a un sistema de presentación de resultados. Este es un formato legible por máquina de los resultados. La máquina de estados puede avanzar la orden cuando se complete la transferencia. Los resultados del servidor de informes pueden combinar el cálculo probabilístico de de riesgos con el texto y formato apropiados para generar un informe legible por seres humanos. Este informe legible por seres humanos puede tener formato adicional para su visualización en un sitio web. Este informe legible por seres humanos puede tener formato adicional para otros medios tales como archivos PDF para su impresión. Los informes de resultados finales se pueden liberar de forma automática mediante un sistema de autoverificación. Un ser humano puede revisar los informes para su liberación. En una instancia los revisores son un científico de laboratorio clínico y un médico. Los resultados son accesibles a través de un portal web que enlaza con una vista de los resultados y un resumen de los parámetros de control de calidad. La aceptación del informe por el científico de laboratorio clínico libera los resultados al médico. El médico puede revisar los resultados en un portal similar y aprobar la versión final de los resultados.

Ejemplo 3

La Figura 24 ilustra etapas a modo de ejemplo de entrada y salida para la generación de informes para dos padres hipotéticos: mama gallina y papá gallina. Se produce una predicción de niño que incorpora genotipos de madre y el padre, fenotipos de madre y padre y genotipos y fenotipos de parientes. Cualquiera o todas estas variables pueden ser valores perdidos, con incumplimientos inicializados a partir de individuos demográficamente similares (y si esto no se sabe, a partir de la población mundial). La predicción resultante del niño puede incluir no solamente el riesgo de la enfermedad, sino también otras variables tales como la altura y el peso. Diferentes variables en la predicción del niño utilizarán diferentes pesos de genotipo y fenotipo.

REIVINDICACIONES

1. Un método que comprende:

- 5 a) usar un análisis genético combinado para analizar a una posible madre o a un posible padre de un niño potencial para determinar la presencia de una pluralidad de variantes genéticas causales correspondientes cada una a al menos una enfermedad genética mendeliana rara y la presencia de al menos un marcador informativo de ascendencia (AIM); y
- 10 b) predecir una probabilidad de un fenotipo del niño potencial con respecto a la al menos una enfermedad genética mendeliana rara basándose al menos en parte, en la presencia de la pluralidad de variantes genéticas causales y la presencia del al menos un AIM, en donde dicha predicción comprende realizar un análisis totalmente probabilístico en los datos recogidos sobre dichas variantes genéticas causales, en donde la información obtenida de la presencia de al menos un AIM de la posible madre o el posible padre se usa para reducir la probabilidad de clasificación errónea de dichas variantes genéticas causales en el análisis genético combinado.
- 15 2. El método de cualquiera de las reivindicaciones anteriores, en el que la predicción en la etapa b) se basa adicionalmente en información fenotípica sobre la posible madre o el posible padre.
- 20 3. El método de cualquiera de las reivindicaciones anteriores que comprende analizar a una posible madre y a un posible padre del niño potencial.
4. El método de cualquiera de las reivindicaciones anteriores, en el que una pluralidad de las enfermedades genéticas tienen, cada una, una frecuencia de menos de 1 en 1000 en seres humanos.
- 25 5. El método de cualquiera de las reivindicaciones anteriores, en el que la pluralidad de enfermedades genéticas raras es de más de 85 enfermedades genéticas mendelianas raras.
- 30 6. El método de la reivindicación 5, en el que la pluralidad de las enfermedades se seleccionan entre fibrosis quística, Tay Sachs, deficiencia en 21-hidroxilasa, hiperinsulinismo relacionado con ABCC8, ARSACS, acondroplasia, acromatopsia, adenosina monofosfato desaminasa 1, agenesia del cuerpo calloso con neuropatía, alcaptonuria, deficiencia en alfa-1-antitripsina, alfa-manosidosis, alfa-sarcoglicanopatía, alfa-talasemia, receptor de angiotensina II, Tipo 1, argininosuccinaciduria, aspartilglicosaminuria, ataxia con deficiencia de vitamina E, ataxia-telangiectasia, síndrome de poliendocrinopatía autoinmune tipo 1, síndrome de Bardet-Biedl, distrofia macular viteliforme de Best, beta-sarcoglicanopatía, beta-talasemia, deficiencia en biotinidasa, síndrome de Blau, síndrome de Bloom, trastornos relacionados con CFTR, lipofuscinosis ceroides neuronal relacionada con CLN3, lipofuscinosis ceroides neuronal relacionada con CLN5, lipofuscinosis ceroides neuronal relacionada con CLN8, enfermedad de Canavan, deficiencia en carnitina palmitoiltransferasa IA, deficiencia en carnitina palmitoiltransferasa II, hipoplasia de cartílago piloso, malformación cavernosa cerebral, coroideremia, síndrome de Cohen, cataratas congénitas, dismorfia facial y neuropatía, enfermedad congénita de glicosilación, trastorno congénito de glicosilación Ib, nefrosis congénita finlandesa, cistinosis, DFNA 9 (COCH), distonía primaria de aparición temprana (DYT1), epidermolisis bullosa juntural tipo Herlitz-Pearson, anemia de Fanconi relacionada con FANCC, craneosinostosis relacionada con FGFR1, craneosinostosis relacionada con FGFR2, craneosinostosis relacionada con FGFR3, trombofilia de Leiden de Factor V, trombofilia de mutación R2 de Factor V, deficiencia en Factor XI, deficiencia en Factor XIII, disautonomía familiar, hipercolesterolemia familiar Tipo B, fiebre mediterránea familiar, trastornos de almacenamiento de ácido siálico libre, demencia frontotemporal con Parkinsonismo-17, deficiencia en fumarasa, pérdida de audición y sordera no sindrómica DFNA3 relacionada con GJB2, pérdida de audición y sordera no sindrómica DFNB1 relacionada con GJB2, miopatías relacionadas con GNE, galactosemia, enfermedad de Gaucher, deficiencia en glucosa-6-fosfato deshidrogenasa, glutaricacidemia Tipo 1, enfermedad de almacenamiento de glucógeno Tipo 1a, enfermedad de almacenamiento de glucógeno Tipo II, enfermedad de almacenamiento de glucógeno Tipo III, enfermedad de almacenamiento de glucógeno Tipo V, síndrome de Gracile, hemocromatosis hereditaria asociada a HFE, beta-talasemia de Hemoglobina S, intolerancia hereditaria a la fructosa, pancreatitis hereditaria, timina-uraciluria hereditaria, deficiencia en hexosaminidasa A, displasia ectodérmica hidrótica 2, homocistinuria causada por deficiencia en cistationina beta-sintasa, parálisis periódica hipercalémica Tipo 1, síndrome de hiperornitinemia-hiperamonemia-homocitrulinuria, hiperoxaluria primaria Tipo 1, hiperoxaluria primaria Tipo 2, hipocondroplasia, parálisis periódica hipocalémica Tipo 1, parálisis periódica hipocalémica Tipo 2, hipofosfatasa, acidemias isovaléricas, enfermedad de Krabbe, LGMD2I, tipo franco-canadiense, deficiencia en 3-hidroxiacil de cadena larga-CoA deshidrogenasa, deficiencia en MTHFR, variante termolábil de MTHFR, pérdida de audición y sordera relacionada con MTT51, poliposis asociada a MYH, enfermedad urinaria de jarabe de arce Tipo 1A, enfermedad urinaria de jarabe de arce Tipo 1B, deficiencia en acil de cadena media-Coenzima A deshidrogenasa, leucoencefalopatía megalencefálica con quistes subcorticales, leucodistrofia metacromática, cardiomiopatía mitocondrial, síndrome de Leigh asociada a ADN mitocondrial y NARP, mucopolisacaridosis IV, mucopolisacaridosis tipo I, mucopolisacaridosis Tipo IIIA, mucopolisacaridosis Tipo VII, neoplasia endocrina múltiple Tipo 2, enfermedad de músculo-ojo-cerebro, miopatía nemalínica, enfermedad Niemann-Pick debida a deficiencia en esfingomielinasa, enfermedad de Niemann-Pick Tipo C1, síndrome de rotura de Nijmegen, lipofuscinosis ceroides neuronal relacionada con PPT1, deficiencia en hormona de la pituitaria relacionada con PROP1, Síndrome de
- 65

Pallister-Hall, paramiotonía congénita, síndrome de Pendred, deficiencia en la enzima bifuncional peroxisomal, deficiencia en fenilalanina hidroxilasa, inhibidor I del activador del plasminógeno, enfermedad renal poliquística, autosómica recesiva, trombofilia de protrombina G20210A, picnodisostosis, retinitis pigmentosa, autosómica recesiva, Tipo Botnia, síndrome de Rett, condrodisplasia punteada rizomélica Tipo 1, deficiencia en acil de cadena corta-CoA deshidrogenasa, síndrome de Shwachman-Diamond, síndrome de Sjogren-Larsson, síndrome de Smith-Lemli-Opitz, paraplejia espástica 13, osteocondrodisplasia relacionada con transportador de sulfato, hemocromatosis hereditaria relacionada con TFR2, lipofuscinosis ceroides neuronal relacionada con TPP1, displasia tanatofórica, amiloidosis de transtiretina, deficiencia en proteína trifuncional, DRD deficiente en tirosina hidroxilasa, tirosinemia Tipo I, enfermedad de Wilson, retinosquiasis juvenil ligada a X y espectro de síndrome de Zellweger.

7. El método de cualquiera de las reivindicaciones anteriores, en el que la pluralidad de variantes genéticas causales corresponden a una o más enfermedades genéticas para las cuales la población de Asia oriental está en riesgo aumentado, una o más enfermedades genéticas para las cuales la población de África subsahariana está en riesgo aumentado y una o más enfermedades genéticas para las cuales la población de Eurasia occidental está en riesgo aumentado.

8. El método de la reivindicación 7 en el que:

- (i) el riesgo aumentado es al menos 10 veces en comparación con otras poblaciones, o
- (ii) la población de Eurasia occidental se selecciona entre las subpoblaciones europea, asiática occidental/central, asiática meridional, y norteafricana, o
- (iii) la población de Asia oriental se selecciona entre las subpoblaciones china, japonesa y coreana.

9. El método de cualquiera de las reivindicaciones 1-6, en el que las variantes genéticas causales corresponden a una o más enfermedades genéticas para las cuales la población de nativos americanos están en riesgo aumentado o una o más enfermedades para las cuales la población de judíos asquenazíes están en riesgo aumentado.

10. El método de cualquiera de las reivindicaciones anteriores, en el que al menos un AIM no es una variante genética causal.

11. El método de cualquiera de las reivindicaciones anteriores, en el que los AIM incluyen al menos un AIM que distingue poblaciones africanas y europeas, al menos un AIM que distingue poblaciones africanas y asiáticas y al menos un AIM que distingue poblaciones europeas y asiáticas.

12. El método de la reivindicación 11, en el que los AIM incluyen al menos un AIM que distingue poblaciones africanas y americanas, poblaciones europeas y americanas, poblaciones asiáticas y americanas, poblaciones noreuropeas y sureuropeas, poblaciones noreuropeas y judías asquenazíes, poblaciones sureuropeas y judías asquenazíes, poblaciones irlandesas e inglesas, poblaciones hispanas y caucásicas, poblaciones chinas y japonesas, o poblaciones surasiáticas, centroasiáticas y asiáticas orientales.

13. El método de cualquiera de las reivindicaciones anteriores, en el que los AIM se seleccionan entre AIM de la Figura 3.

14. El método de la reivindicación 1, en el que el ensayo comprende el uso de una serie de ácidos nucleicos, secuenciación por hibridación, secuenciación por ligamiento, secuenciación por extensión, secuenciación de Sanger, secuenciación de Maxam Gilbert o pirosecuenciación.

15. Un medio legible por ordenador que comprende un código informático que cuando se ejecuta predice una probabilidad de un fenotipo de un niño potencial, a partir de la posible madre o del posible padre, con respecto a al menos una enfermedad genética mendeliana rara, basada al menos en parte en los resultados de un análisis de la posible madre o del posible padre para la presencia de una pluralidad de variantes genéticas causales y al menos un marcador informativo de ascendencia (AIM), en el que la probabilidad de un fenotipo del niño potencial se calcula de acuerdo con el método de una cualquiera de las reivindicaciones 1 a 13.

16. Un sistema para predecir un fenotipo de un niño que comprende:

- un dispositivo de detección de ácido nucleico combinado configurado para detectar una pluralidad de variantes genéticas causales correspondientes a al menos una enfermedad genética mendeliana rara y al menos un marcador informativo de ascendencia (AIM), en donde el dispositivo está en contacto con una muestra de una posible madre o de un posible padre de un niño potencial;
- un lector configurado para leer los datos de los dispositivos; y
- un medio legible por ordenador de acuerdo con la reivindicación 15.

17. Un sistema para indicar si un sujeto es portador de una enfermedad genética mendeliana rara que comprende: un lector configurado para leer los datos de un dispositivo de detección de ácido nucleico combinado configurado

para detectar una pluralidad de variantes genéticas causales correspondientes a al menos una enfermedad genética mendeliana rara y al menos un marcador informativo de ascendencia (AIM), y un medio legible por ordenador de acuerdo con la reivindicación 15.

5 18. Un método que comprende:

recibir una muestra de un usuario;

analizar la muestra con un dispositivo de detección de ácido nucleico combinado configurado para analizar una pluralidad de variantes genéticas causales de enfermedades genéticas mendelianas raras y al menos un
10 marcador informativo de ascendencia (AIM);

calcular una pluralidad de probabilidades para los posibles genotipos del usuario en cada localización de la pluralidad de variantes genéticas causales basándose en los resultados de la etapa de análisis respecto a la pluralidad de variantes genéticas causales y el al menos un AIM,

15 en donde dicho cálculo comprende realizar un análisis totalmente probabilístico en los datos recogidos sobre dichas variantes genéticas causales, en donde la información obtenida del al menos un AIM respecto a la ascendencia del usuario se usa para reducir la probabilidad de clasificación errónea de dichas variantes genéticas causales en el dispositivo de detección de ácido nucleico combinado; y

entregar al usuario la pluralidad de probabilidades correspondientes al usuario que es portador.

20 19. El método de cualquiera de las reivindicaciones 1-14, en el que al menos dos de las enfermedades genéticas mendelianas raras ocurren a frecuencias que difieren en al menos 10 veces en al menos dos poblaciones distintas, en donde las al menos dos poblaciones distintas se diferencian en al menos un AIM.

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Deficiencia en 21-hidroxilasa	CYP21A2	F306+t
Deficiencia en 21-hidroxilasa	CYP21A2	F306+t
Deficiencia en 21-hidroxilasa	CYP21A3	g.655A/C>G
Deficiencia en 21-hidroxilasa	CYP21A4	g.655A/ C>G
Deficiencia en 21-hidroxilasa	CYP21A6	G110del8nt
Deficiencia en 21-hidroxilasa	CYP21A5	G110del8nt
Deficiencia en 21-hidroxilasa	CYP21A7	I172N, rs34607927
Deficiencia en 21-hidroxilasa	CYP21A2	I236N
Deficiencia en 21-hidroxilasa	CYP21A2	M239K, rs6476
Deficiencia en 21-hidroxilasa	CYP21A2	P30L
Deficiencia en 21-hidroxilasa	CYP21A2	P453S
Deficiencia en 21-hidroxilasa	CYP21A2	Q318X
Deficiencia en 21-hidroxilasa	CYP21A2	R356W
Deficiencia en 21-hidroxilasa	CYP21A2	V237E, rs12530380
Deficiencia en 21-hidroxilasa	CYP21A2	V281L, rs6471
Hiperinsulinismo relacionado con ABCC8	ABCC8	3992-9G>A
Hiperinsulinismo relacionado con ABCC8	ABCC8	delF1388
Hiperinsulinismo relacionado con ABCC8	ABCC8	delF1388
Hiperinsulinismo relacionado con ABCC8	ABCC8	V187D
Acondroplasia	FGFR3	G375C
Acondroplasia	FGFR3	G380R, rs28931614
Acromatopsia	CNGB3	c.1148delC
Acromatopsia	CNGB3	c.1148delC
Acromatopsia	CNGB3	c.819-826del8
Acromatopsia	CNGB3	c.819-826del8
Acromatopsia	CNGB3	c.886-896del11 1 insT
Acromatopsia	CNGB3	c.886-896del11 1 insT
Acromatopsia	CNGB3	c.991-3T>G
Acromatopsia	CNGB3	p.Arg403Gln
Acromatopsia	CNGB3	p_Glu336X
Adenosina monofosfato		
Desaminasa 1	AMPD1	P48L
Adenosina monofosfato		
Desaminasa 1	AMPD1	Q12X, rs17602729
Agnesia de cuerpo calloso con neuronopatía	SLC12A6	c.2436delG
Agnesia de cuerpo calloso con neuronopatía	SLC12A6	c.2436delG
Alcaptonuria	HGD	c.174delA
Alcaptonuria	HGD	c.174delA
Alcaptonuria	HGD	c.457_458insG
Alcaptonuria	HGD	c.457_458insG
Alcaptonuria	HGD	G161R
Alcaptonuria	HGD	G270R

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Alcaptonuria	HGD	IVS1-1G>A
Alcaptonuria	HGD	IVS5+1G>A
Alcaptonuria	HGD	Met368Val
Alcaptonuria	HGD	P230S
Alcaptonuria	HGD	S47L
Alcaptonuria	HGD	V300G
Deficiencia en alfa-1-antitripsina	SERPINA1	Arg101His, rs709932
Deficiencia en alfa-1-antitripsina	SERPINA1	Glu264Val
Deficiencia en alfa-1-antitripsina	SERPINA1	Glu342Lys, rs28929474
Deficiencia en alfa-1-antitripsina	SERPINA1	Glu376Asp,rs1303
Alfa-manosidosis	MAN2B1	IVS14+1G>C
Alfa-manosidosis	MAN2B1	p.L809P
Alfa-manosidosis	MAN2B1	P.R750W
Alfa-sarcoglicanopatía	SGCA	R77C, rs28933693
Alfa-talasemia	HBA2	H19D
Alfa-talasemia	HBA1	HbQ
Alfa-talasemia	HBA1,HBA2	3,7 kb (tipo I) delección alfa-2
Alfa-talasemia	HBA1.HBA2	3,7 kb (tipo I) delección alfa-2
		HbVar nº id de base de datos 1086, -(SEA); delección de ~20 kb incluyendo ambos genes de alfa-globina alfa-Thal-1-1
Alfa-talasemia	HBA1,HBA2	HbVar nº id de base de datos 1086, -(SEA); delección de ~ 20 kb incluyendo ambos genes de alfa-globina alfa-Thal-1-1
Alfa-talasemia	HBA1.HBA2	HbVar nº id de base de datos 1087, -(MED-I); delección de ~ 17,5 kb incluyendo ambos genes de alfa-globina alfa-Thal-1
Alfa-talasemia	HBA1.HBA2	1 HbVar nº id de base de datos 1087, -(MED-I); delección de ~ 17,5 kb incluyendo ambos genes de alfa-globina alfa-Thal-1
Alfa-talasemia v	HBA1,HBA2-	1 HbVar nº id de base de datos 1088, -(alfa)20,5; esta delección de 20,5 kb implica alfa2 y el extremo 5' de alfa 1; alfa-Thal-1
Alfa-talasemia	HBA1.HBA2	HbVar nº id de base de datos 1088, -(alfa)20,5; esta delección de 20,5 kb implica alfa2 y el extremo 5' de alfa 1; alfa-Thal-1
Alfa-talasemia	HBA1.HBA2	HbVar nº id de base de datos 1094, --(FIL); una delección de 30-34 kb que implica los genes alfa 1, alfa2, y zeta alfa-Thal-1
Alfa-talasemia	HBA1.HBA2	HbVar nº id de base de datos 1094, --(FIL); una delección de 30-34 kb que implica los genes alfa 1, alfa2, y zeta alfa-Thal-1
Alfa-talasemia	HBA1.HBA2	HbVar nº id de base de datos 1095,

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
		--(THAL-1); una delección de 34-38 kb que implica los genes alfa 1, alfa2, y zeta alfa-Thal-1
Alfa-talasemia	HBA1.HBA2	HbVar nº id de base de datos 1095, --(THAL-1); una delección de 34-38 kb que implica los genes alfa 1, alfa2, y zeta alfa-Thal-1
Alfa-talasemia	HBA1.HBA2	HbVar nº id de base de datos 1097, --(MED-II); una delección de 26,5 kb que implica los dos genes alfa y zeta alfa-Thal-1
Alfa-talasemia	HBA1.HBA2	HbVar nº id de base de datos 1097, --(MED-II); una delección de 26,5 kb que implica los dos genes alfa y zeta alfa-Thal-1
Alfa-talasemia	HBA1.HBA2	
Alfa-talasemia	HBA2	HbVar nº id de base de datos 187
		HbVar nº id de base de datos 2598, IVS I-5 (G>A)
Alfa-talasemia	HBA1.HBA2	HbVar nº id de base de datos 703
Alfa-talasemia	HBA1.HBA2	HbVar nº id de base de datos 704
		HbVar nº id de base de datos 705, Hb Koya
Alfa-talasemia	HBA1.HBA2	Dora
		HbVar nº id de base de datos 707,
Alfa-talasemia	HBA1.HBA2	rs41412046
Alfa-talasemia	HBA1	HbVar nº id de base de datos 87
		HbVar nº id de base de datos 969, Poli A (A->G); AATAAA->AATGAA
Alfa-talasemia	HBA1,HBA2	beta+
		HbVar nº id de base de datos 971, Poli A (A->G); AATAAA->AATAAG
Alfa-talasemia	HBA1.HBA2	beta+
Alfa-talasemia	HBA2	MIT
Alfa-talasemia	HBA1	W14X
Receptor de angiotensina II, Tipo I	AGTR1	A1166C
Genotipado de apolipoproteína E	APOE	p.C112R, rs429358
Genotipado de apolipoproteína E	APOE	p.R158C, rs7412
Argininosuccinicaciduria	ASL	R385C
ARSACS	SACS	5254C>T
ARSACS	SACS	6594delT
ARSACS	SACS	6594delT
Aspartilglicosaminuria	AGA	c.199_200delGA
Aspartilglicosaminuria	AGA	c.199_200delGA
Aspartilglicosaminuria	AGA	C163S
Ataxia con deficiencia de vitamina E	TTPA	744delA
Ataxia con deficiencia de vitamina E	TTPA	744delA
Ataxia-Telangiectasia	ATM	R35X

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Poliendocrinopatía autoinmune		
Síndrome Tipo 1 autoinmune	AIRE	c.1163_1164insA
Síndrome de poliendocrinopatía tipo 1 autoinmune	AIRE	c.1163_1164insA
Síndrome de poliendocrinopatía tipo 1 autoinmune	AIRE	C.769C>T
Síndrome de poliendocrinopatía tipo 1 autoinmune	AIRE	c.967_979del
Síndrome de poliendocrinopatía tipo 1 autoinmune	AIRE	c.967_979del
Poliendocrinopatía síndrome tipo 1	AIRE	Y85C
Síndrome de Bardet-Biedl	BBS1	M390R
Síndrome de Bardet-Biedl	BBS 10	p.C91LfsX4
Síndrome de Bardet-Biedl	BBS10	p.C91LfsX4
Distrofia macular viteliforme de Best	BEST1	c.G383C
Beta-Sarcoglicanopatía	SGCB	S114F
Beta-talasemia	HBB	-28 (A->G) beta+
Beta-talasemia	HBB	-29 (A->G) beta+
Beta-talasemia	HBB	-29A>G
Beta-talasemia	HBB	-30 (T->A) beta+
Beta-talasemia	HBB	-87 (C->G) beta+
Beta-talasemia	HBB	-88C>T
Beta-talasemia	HBB	CAP+1 (A->C) beta+
Beta-talasemia	HBB	Codón 15 (G->A); TGG(Trp)->TAG(codón de parada) beta0, rs34716011
Beta-talasemia	HBB	Codón 15 (G->A); TGG(Trp)->TAG(codón de parada) beta0, rs34716011
Beta-talasemia	HBB	Codón 16 (-C); GGC(Gly)->GG-beta0
Beta-talasemia	HBB	Codón 16 (-C); GGC(Gly)->GG-beta0
Beta-talasemia	HBB	Codón 17 (A->T); AAG(Lys)->TAG(codón de parada) beta0
Beta-talasemia	HBB	Codón 24 (T->A); GGT(Gly)->GGA(Gly) beta+
Beta-talasemia^ --~'	HBB	Codón 39 (C->T); CAG(Gln)->TAG(codón de parada) beta0
Beta-talasemia	HBB	Codón 5 (-CT); CCT(Pro)->C beta0
Beta-talasemia	HBB	Codón 5 (-CT); CCT(Pro)->C beta0
Beta-talasemia	HBB	Codón 6 (-A); GAG(Glu)->G-G beta0

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Beta-talasemia	HBB	Codón 6 (-A); GAG(Glu)->G-G beta0
Beta-talasemia	HBB	Codón 8 (-AA); AAG(Lys)->--G beta0
Beta-talasemia	HBB	Codón 8 (-AA); AAG(Lys)->--G beta0
Beta-talasemia	HBB	Codones 41/42 (-TTCT); TTCTTT(Phe-Phe)->—TT beta0
Beta-talasemia	HBB	Codones 41/42 (-TTCT); TTCTTT(Phe-Phe)->--TT beta0
Beta-talasemia	HBB	Codones 71/72 (+A); TTT AGT(Phe Ser)->TTT A AGT; beta0
Beta-talasemia	HBB	Codones 71/72 (+A); TTT AGT(Phe Ser)->TTT A AGT; beta0
Beta-talasemia	HBB	Codones 8/9 (+G); AAG TCT(Lys;Ser)->AAG G TCT beta0
Beta-talasemia	HBB	Codones 8/9 (+G); AAG TCT(Lys;Ser)->AAG G TCT beta0
Beta-talasemia	HBB	HbVar nº id de base de datos 889, IVS-II- 654 (C->T); AAGGCAATA- >AAG [^] GTAATA beta+(severa)
Beta-talasemia	HBB	HbVar nº id de base de datos 890, IVS-II- 705 (T->G); GATGTAAGA- >GAG [^] GTAAGA beta+
Beta-talasemia	HBB	HbVar nº id de base de datos 891, IVS-II- 745 (C->G); CAGCTACCAT- >AG [^] GTACCAT beta+
Beta-talasemia	HBB	HbVar nº id de base de datos 979, delección de 619 pb de beta0
Beta-talasemia	HBB	delección de beta 0
Beta-talasemia	HBB	delección de 619 pb de beta0
Beta-talasemia	HBB	IVS-I-1 (G->A); AG [^] GTTGGT- >AGATTGGT beta0
Beta-talasemia	HBB	IVS-I-1 (G->T); AG [^] GTTGGT- >AGTTTGGT beta0
Beta-talasemia	HBB	IVS-I-110 (G->A) beta+; mutación es 21 nucleótidos 5' al
Beta-talasemia	HBB	sitio aceptor de corte y ajuste AG [^] GC
Beta-talasemia	HBB	IVS-I-5' (G->C) beta+(severa)
Beta-talasemia	HBB	IVS-II-1 (G->A); beta0
Beta-talasemia	HBB	IVS-II-844 (C->G); beta+
Beta-talasemia	HBB	IVS1+6T>C
Beta-talasemia	HBB	IVS11-849A>C
Beta-talasemia	HBB	IVS11-849A>G
Deficiencia en biotinidasa	BTB	A171T, rs13073139
Deficiencia en biotinidasa	BTB	D252G, rs28934601
Deficiencia en biotinidasa	BTB	D444H, rs13078881
Deficiencia en biotinidasa	BTB	F403V
Deficiencia en biotinidasa	BTB	G98:d7i3
Deficiencia en biotinidasa	BTB	G98:d7i3
Deficiencia en biotinidasa	BTB	Q456H
Deficiencia en biotinidasa	BTB	R157H
Deficiencia en biotinidasa	BTB	R538C

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Síndrome de Blau	NOD2	E383K
Síndrome de Blau	NOD2	L469F
Síndrome de Blau	NOD2	R334Q
Síndrome de Blau	NOD2	R334W
Síndrome de Bloom	BLM	2407insT
Síndrome de Bloom	BLM	2407insT
		736delATCTGAinsTAGATTC
Síndrome de Bloom	BLM	(2281del6/ins7)
		736delATCTGAinsTAGATTC
Síndrome de Bloom	BLM	(2281del6/ins7)
Cáncer hereditario BRCA1 de mama/ovario	BRCA1	185delAG
Cáncer hereditario BRCA1 de mama/ovario	BRCA1	185delAG
Cáncer hereditario BRCA1 de mama/ovario	BRCA1	5382insC
Cáncer hereditario BRCA1 de mama/ovario	BRCA1	5382insC
Cáncer hereditario BRCA1 de mama/ovario	BRCA1	Tyr978X
Cáncer hereditario BRCA2 de mama/ovario	BRCA2	6174delT
Cáncer hereditario BRCA2 de mama/ovario	BRCA2	6174delT
Cáncer hereditario BRCA2 de mama/ovario	BRCA2	8765delAG
Cáncer hereditario BRCA2 de mama/ovario	BRCA2	8765delAG
Enfermedad de Canavan	ASPA	A305E (9140A), rs28940574
Enfermedad de Canavan	ASPA	E285A (854A>C), rs28940279
Enfermedad de Canavan	ASPA	IVS2-2A>G (433-2A>G)
Enfermedad de Canavan	ASPA	Y231X(693C>A)
Deficiencia de carnitina palmitoiltransferasa IA	CPT1A	G710E
Deficiencia de carnitina palmitoiltransferasa IA	CPT1A	P479L
Deficiencia de carnitina palmitoiltransferasa II	CPT2	G549D
Deficiencia de carnitina palmitoiltransferasa II	CPT2	L178F 534 ins/25 pb del
Deficiencia de carnitina palmitoiltransferasa II	CPT2	L178F 534 ins/25 pb del
Deficiencia de carnitina palmitoiltransferasa II	CPT2	P227L
Deficiencia de carnitina palmitoiltransferasa II	CPT2	P50H

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Deficiencia de carnitina palmitoiltransferasa II	CPT2	P604S
Deficiencia de carnitina palmitoiltransferasa II	CPT2	Q413fs
Deficiencia de carnitina palmitoiltransferasa II	CPT2	Q413fs
Deficiencia de carnitina palmitoiltransferasa II	CPT2	Q550R
Deficiencia de carnitina palmitoiltransferasa II	CPT2	R124X
Deficiencia de carnitina palmitoiltransferasa II	CPT2	R503C
Deficiencia de carnitina palmitoiltransferasa II	CPT2	R631C
Deficiencia de carnitina palmitoiltransferasa II	CPT2	S113L
Deficiencia de carnitina palmitoiltransferasa II	CPT2	s38fs
Deficiencia de carnitina palmitoiltransferasa II	CPT2	s38fs
Deficiencia de carnitina palmitoiltransferasa II	CPT2	Y628S, rs28936673
Hipoplasia de cartílago piloso	RMRP	g.262G>T
Hipoplasia de cartílago piloso	RMPR	g.70A>G
Trastornos relacionados con CFTR	CFTR	1811+1,6kbA->G
Trastornos relacionados con CFTR	CFTR	2183AA>G
Trastornos relacionados con CFTR	CFTR	2183AA>G
Trastornos relacionados con CFTR	CFTR	3849+10kbC>T
Trastornos relacionados con CFTR	CFTR	A455E
Trastornos relacionados con CFTR	CFTR	A559T
Trastornos relacionados con CFTR	CFTR	C524X
Trastornos relacionados con CFTR	CFTR	574delA, 574delA
Trastornos relacionados con CFTR	CFTR	574delA, 574delA
Trastornos relacionados con CFTR	CFTR	2108delA, 2108delA
Trastornos relacionados con CFTR	CFTR	2108delA, 2108delA
Trastornos relacionados con CFTR	CFTR	3171delC, 3171delC
Trastornos relacionados con CFTR	CFTR	3171delC, 3171delC
Trastornos relacionados con CFTR	CFTR	621+1G->T
Trastornos relacionados con CFTR	CFTR	2105-2117del3insAGAAA
Trastornos relacionados con CFTR	CFTR	2105-2117del3insAGAAA
Trastornos relacionados con CFTR	CFTR	711+1G->T

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Trastornos relacionados con CFTR	CFTR	711+5G->A
Trastornos relacionados con CFTR	CFTR	712-1G->T
Trastornos relacionados con CFTR	CFTR	1288insTA, 1288insTA
Trastornos relacionados con CFTR	CFTR	1288insTA, 1288insTA
Trastornos relacionados con CFTR	CFTR	936delTA
Trastornos relacionados con CFTR	CFTR	936delTA
Trastornos relacionados con CFTR	CFTR	[delta]F311
Trastornos relacionados con CFTR	CFTR	[delta]F311
Trastornos relacionados con CFTR	CFTR	1078delT, 1078delT
Trastornos relacionados con CFTR	CFTR	1078delT, 1078delT
Trastornos relacionados con CFTR	CFTR	1161delC, 1161delC
Trastornos relacionados con CFTR	CFTR	1161delC, 1161delC
Trastornos relacionados con CFTR	CFTR	1609delCA, 1609delCA
Trastornos relacionados con CFTR	CFTR	1609delCA, 1609delCA
Trastornos relacionados con CFTR	CFTR	[delta]I507
Trastornos relacionados con CFTR	CFTR	[delta]I507
Trastornos relacionados con CFTR	CFTR	rs332, [delta]F508
Trastornos relacionados con CFTR	CFTR	rs332, [delta]F508
Trastornos relacionados con CFTR	CFTR	1677delTA, 1677delTA
Trastornos relacionados con CFTR	CFTR	1677delTA, 1677delTA
Trastornos relacionados con CFTR	CFTR	1717-1G->A
Trastornos relacionados con CFTR	CFTR	1812-1G->A
Trastornos relacionados con CFTR	CFTR	1898+1G->A
Trastornos relacionados con CFTR	CFTR	1898+1G->T
Trastornos relacionados con CFTR	CFTR	1898+5G->T
Trastornos relacionados con CFTR	CFTR	1949del84, 1949del84
Trastornos relacionados con CFTR	CFTR	1949del84, 1949 del 84
Trastornos relacionados con CFTR	CFTR	2043delG, 2043delG
Trastornos relacionados con CFTR	CFTR	2043delG, 2043delG
Trastornos relacionados con CFTR	CFTR	2055delI9->A
Trastornos relacionados con CFTR	CFTR	2055delI9->A
Trastornos relacionados con CFTR	CFTR	2143delT, 2143delT
Trastornos relacionados con CFTR	CFTR,	2143delT,2143delT
Trastornos relacionados con CFTR	CFTR	2184delA,2184delA
Trastornos relacionados con CFTR	CFTR	2184delA,2184delA
Trastornos relacionados con CFTR	CFTR	2184insA, 2184insA
Trastornos relacionados con CFTR	CFTR	2184insA, 2184insA
Trastornos relacionados con CFTR	CFTR	2307insA, 2307insA
Trastornos relacionados con CFTR	CFTR	2307insA, 2307insA
Trastornos relacionados con CFTR	CFTR	296+12T->C
Trastornos relacionados con CFTR	CFTR	2789+5G->A
Trastornos relacionados con CFTR	CFTR	2869insG, 2869insG
Trastornos relacionados con CFTR	CFTK	2869insG, 2869insG
Trastornos relacionados con CFTR	CFTR	3120G->A
Trastornos relacionados con CFTR	CFTR	3120+1G->A
Trastornos relacionados con CFTR	CFTR	3272-26A->G
Trastornos relacionados con CFTR	CFTR	3659delC, 3659delC
Trastornos relacionados con CFTR	CFTR	3659delC, 3659delC

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Trastornos relacionados con CFTR	CFTR	3667del4, 3667del4
Trastornos relacionados con CFTR	CFTR	3667del4, 3667del4
Trastornos relacionados con CFTR	CFTR	3791delC,3791delC
Trastornos relacionados con CFTR	CFTR	3791delC, 3791delC
Trastornos relacionados con CFTR	CFTR	3821delT,3821delT
Trastornos relacionados con CFTR	CFTR	3821delT,3821delT
Trastornos relacionados con CFTR	CFTR	3905insT, 3905insT
Trastornos relacionados con CFTR	CFTR	3905insT, 3905insT
Trastornos relacionados con CFTR	CFTR	4016insT,4016insT
Trastornos relacionados con CFTR	CFTR	4016insT,4016insT
Trastornos relacionados con CFTR	CFTR	394delTT, 394delTT
Trastornos relacionados con CFTR	CFTR	394delTT, 394delTT
Trastornos relacionados con CFTR	CFTR	405+IG->A
Trastornos relacionados con CFTR	CFTR	405+3A->C
Trastornos relacionados con CFTR	CFTR	444delA
Trastornos relacionados con CFTR	CFTR	444delA
Trastornos relacionados con CFTR	CFTR	3876delA, 3876delA
Trastornos relacionados con CFTR	CFTR	3876delA, 3876delA
Trastornos relacionados con CFTR	CFTR	457TAT->G
Trastornos relacionados con CFTR	CFTR	457TAT->G
Trastornos relacionados con CFTR	CFTR	3199del6,3199del6
Trastornos relacionados con CFTR	CFTR	3199del6,3199del6
Trastornos relacionados con CFTR	CFTR	406-IG->A
Trastornos relacionados con CFTR	CFTR	663delT, 663delT
Trastornos relacionados con CFTR	CFTR	663delT, 663delT
Trastornos relacionados con CFTR	CFTR	935delA, 935delA
Trastornos relacionados con CFTR	CFTR	935delA, 935delA
Trastornos relacionados con CFTR	CFTR	CFTRdele2,3(21kb)
Trastornos relacionados con CFTR	CFTR	CFTRdele2,3(21kb)
Trastornos relacionados con CFTR	CFTR	D1152H
Trastornos relacionados con CFTR	CFTR	E60X
Trastornos relacionados con CFTR	CFTR	E92X
Trastornos relacionados con CFTR	CFTR	F508C, rs 1800093
Trastornos relacionados con CFTR	CFTR	G178R
Trastornos relacionados con CFTR	CFTR	G330X
Trastornos relacionados con CFTR	CFTR	G480C
Trastornos relacionados con CFTR	CFTR	G542X
Trastornos relacionados con CFTR	CFTR	G551D
Trastornos relacionados con CFTR	CFTR	G622D
Trastornos relacionados con CFTR	CFTR	G85E
Trastornos relacionados con CFTR	CFTR	G91R
Trastornos relacionados con CFTR	CFTR	I148T, rs35516286
Trastornos relacionados con CFTR	CFTR	I506V
Trastornos relacionados con CFTR	CFTR	IVS8-5T
Trastornos relacionados con CFTR	CFTR	IVS8-7T
Trastornos relacionados con CFTR	CFTR	IVS8-9T
Trastornos relacionados con CFTR	CFTR	K710X
Trastornos relacionados con CFTR	CFTR	L206W

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Trastornos relacionados con CFTR	CFTR	M1101K, rs36210737
Trastornos relacionados con CFTR	CFTR	N1303K
Trastornos relacionados con CFTR	CFTR	P574H
Trastornos relacionados con CFTR	CFTR	Q1238X
Trastornos relacionados con CFTR	CFTR	Q359K/T360K_wt
Trastornos relacionados con CFTR	CFTR	Q493X
Trastornos relacionados con CFTR	CFTR	Q552X
Trastornos relacionados con CFTR	CFTR	Q890X
Trastornos relacionados con CFTR	CFTR	R1066C
Trastornos relacionados con CFTR	CFTR	R1070Q
Trastornos relacionados con CFTR	CFTR	R1158X
Trastornos relacionados con CFTR	CFTR	R1162X
Trastornos relacionados con CFTR	CFTR	R117C
Trastornos relacionados con CFTR	CFTR	R117H
Trastornos relacionados con CFTR	CFTR	R1283M
Trastornos relacionados con CFTR	CFTR	R334W
Trastornos relacionados con CFTR	CFTR	R347H
Trastornos relacionados con CFTR	CFTR	R347P
Trastornos relacionados con CFTR	CFTR	R352Q
Trastornos relacionados con CFTR	CFTR	R553X
Trastornos relacionados con CFTR	CFTR	R560T
Trastornos relacionados con CFTR	CFTR	R709X
Trastornos relacionados con CFTR	CFTR	R75X
Trastornos relacionados con CFTR	CFTR	R764X
Trastornos relacionados con CFTR	CFTR	S1196X
Trastornos relacionados con CFTR	CFTR	S1235R, rs34911792
Trastornos relacionados con CFTR	CFTR	S1251N
Trastornos relacionados con CFTR	CFTR	S1255X
Trastornos relacionados con CFTR	CFTR	S364P
Trastornos relacionados con CFTR	CFTR	S549I
Trastornos relacionados con CFTR	CFTR	S549N
Trastornos relacionados con CFTR	CFTR	S549R
Trastornos relacionados con CFTR	CFTR	S549R
Trastornos relacionados con CFTR	CFTR	T338I
Trastornos relacionados con CFTR	CFTR	V520F
Trastornos relacionados con CFTR	CFTR	W1089X
Trastornos relacionados con CFTR	CFTR	W1204X
Trastornos relacionados con CFTR	CFTR	W1204X
Trastornos relacionados con CFTR	CFTR	W1282X
Trastornos relacionados con CFTR	CFTR	Y1092X
Trastornos relacionados con CFTR	CFTR	Y122X
Coroideremia	CHM	c.1609+2dupT
Coroideremia	CHM	c.1609+2dupT
Lipofuscinosis ceroides neuronal relacionada con CLN3	CLN3	c.461_677del
Lipofuscinosis ceroides neuronal relacionada con CLN3	CLN3	c.461_677del
Lipofuscinosis ceroides neuronal relacionada con CLN3	CLN3	c.791_1056del

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Lipofuscinosis ceroides neuronal relacionada con CLN3	CLN3	c.791_1056del
Lipofuscinosis ceroides neuronal relacionada con CLN5	CLN5	c.1175_1176delAT
Lipofuscinosis ceroides neuronal relacionada con CLN5	CLN5	c.H75_1176delAT
Lipofuscinosis ceroides neuronal relacionada con CLN5	CLN5	c.225G>A
Lipofuscinosis ceroides neuronal relacionada con CLN8	CLN8	c.70C>G
Síndrome de Cohen	VPS13B	c.3348_3349delCT
Síndrome de Cohen	VPS13B	c.3348_3349delCT
Cataratas congénitas		
Dismorfismo facial, y Neuropatía	CTDP1	IVS6+389C>T
Trastorno congénito de Glicosilación Ia	PMM2	P.F119L
Trastorno congénito de Glicosilación Ia	PMM2	P.R141H
Trastorno congénito de Glicosilación Ib	MPI	R295H, rs28928906
Nefrosis congénita finlandesa	NPHS1	c.121_122del
Nefrosis congénita finlandesa	NPHS1	c.121_122del
Nefrosis congénita finlandesa	NPHS1	c.3325C>T
Enfermedad de Crohn	NOD2	3020 ins C
Enfermedad de Crohn	NOD2	3020 ins C
Enfermedad de Crohn	NOD2	G908R, rs2066845
Enfermedad de Crohn	NOD2	R702W, rs2066844
Cistinosis	CTNS	1035insC
Cistinosis	CTNS	1035insC
Cistinosis	CTNS	537del21
Cistinosis	CTNS	537del21
Cistinosis	CTNS	deleción 57kb
Cistinosis	CTNS	deleción 57kb
Cistinosis	CTNS	D205N
Cistinosis	CTNS	L158P
Cistinosis	CTNS	W138X
DFNA 9 (COCH)	COCH	P51S
Diabetes y pérdida auditiva	mtDNA	3234A>G
Diabetes y pérdida auditiva	mtDNA	3271T>C
Diabetes y pérdida auditiva	mtDNA	G8363A
Diabetes y pérdida auditiva	mtDNA	X147Q9C
Distonía primaria de aparición temprana (DYT1)	TOR1A	904_906delGAG
Distonía primaria de aparición temprana (DYT1)	TOR1 A	904_906delGAG
Epidermolisis bullosa juntural tipo Herlitz-Pearson	LAMB3	3024delT

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Epidermólisis bullosa juntural tipo Herlitz-Pearson	LAMB3	3024delT
Epidermólisis bullosa juntural tipo Herlitz-Pearson	LAMB3	p.Q243X
Epidermólisis bullosa juntural tipo Herlitz-Pearson	LAMB3	R144X
Epidermólisis bullosa juntural tipo Herlitz-Pearson	LAMB3	R42X
Epidermólisis bullosa juntural tipo Herlitz-Pearson	LAMB3	R635X
Epidermólisis bullosa juntural tipo Herlitz-Pearson	LAMA3	R650X
Epidermólisis bullosa juntural tipo Herlitz-Pearson	LAMC2	R95X
Tormbofilia de Leiden de Factor V	F5	H1299R
Tormbofilia de Leiden de Factor V	F5	R506Q, rs6025
Tormbofilia de mutación R2 de Factor V	F5	rs6027
Deficiencia de Factor XI	FII	E1 17X (576G>T)
Deficiencia de Factor XI	FII	F283L(1074T>C)
Deficiencia de Factor XI	FII	IVS14+1G>A
Deficiencia de Factor XI	FII	IVS14del4
Deficiencia de Factor XI	FII	IVS14del4
Deficiencia de Factor XIII	F13A1	V34L, rs5985
Poliposis adenomatosa familiar	APC	I1307K, rs1801155
Disautonomía familiar	KBKAP	2507+6T>C
Disautonomía familiar	IKBKAF	P914Lr
Disautonomía familiar	KBKAP	R696P
Hipercolesterolemia familiar Tipo B	APOB	R3500Q, rs5742904
Hipercolesterolemia familiar Tipo B	APOB	R3500W
Hipercolesterolemia familiar Tipo B	APOB	S3531C, rs12713559
Fiebre mediterránea familiar	MEFV	A744S (2230G>T)
Fiebre mediterránea familiar	MEFV	delI692 (del2076_2078)
Fiebre mediterránea familiar	MEFV	delI692 (del2076_2078)

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Fiebre mediterránea familiar	MEFV	E148Q (442 G>C), rs3743930
Fiebre mediterránea familiar	MEFV	E167D (501 G>C)
Fiebre mediterránea familiar	MEFV	F479L (1437 C>G)
Fiebre mediterránea familiar	MEFV	K695R (2084A>G)
Fiebre mediterránea familiar	MEFV	M680I (2040G>C)
Fiebre mediterránea familiar	MEFV	M694I (2082G>A), rs28940578
Fiebre mediterránea familiar	MEFV	M694V (2080A>G)
Fiebre mediterránea familiar	MEFV	P369S (1105 C>T), rsl 1466023
Fiebre mediterránea familiar	MEFV	R408Q (1223G>A), rsl 1466024
Fiebre mediterránea familiar	MEFV	R653H (1958G>A)
Fiebre mediterránea familiar	MEFV	R761H (2282G>A)
Fiebre mediterránea familiar	MEFV	T267I (800 C>T)
Fiebre mediterránea familiar	MEFV	V726A (2177T>C), rs28940579
Anemia de Fanconi relacionada con FANCC	FANCC	322delG
Anemia de Fanconi relacionada con FANCC	FANCC	322delG
Anemia de Fanconi relacionada con FANCC	FANCC	IVS4+4A>T (711 +4A>T)
Anemia de Fanconi relacionada con FANCC	FANCC	Q13X (37C>T)
Anemia de Fanconi relacionada con FANCC	FANCC	R547X
Craneosinostosis relacionada con FGFR1	FGFR1	P252R
Craneosinostosis relacionada con FGFR2	FGFR2	P253R
Craneosinostosis relacionada con FGFR2	FGFR2	S252W
Craneosinostosis relacionada con FGFR3	FGFR3	A391E, rs28931615
Craneosinostosis relacionada con FGFR3	FGFR3	P250R, rs4647924
Trastornos de almacenamiento de ácido siálico libre	SLC17A5	c.1007_1008delTA
Trastornos de almacenamiento de ácido siálico libre	SLC17A5	c.1007_1008delTA
Trastornos de almacenamiento de ácido siálico libre	SLC17A5	c.115C>T
Demencia frontotemporal con parkinsonismo-17	MAPT	IVS10+16

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Demencia frontotemporal con parkinsonismo-17	MAPT	P301L
Demencia frontotemporal con parkinsonismo-17	MAPT	P301S
Demencia frontotemporal con parkinsonismo-17	MAPT	R406W
Deficiencia de fumarasa	FH	c.1431_1433dupAAA
Deficiencia de fumarasa	FH	c.1431_1433dupAAA
Galactosemia	GALT	delección génica de 5,0 Kb
Galactosemia	GALT	delección génica de 5,0 Kb
Galactosemia	GALT	5'UTR-119del
Galactosemia	GALT	5'UTR-119del
Galactosemia	GALT	IVS2-2 A>G
Galactosemia	GALT	K285N
Galactosemia	GALT	L195P T>C
Galactosemia	GALT	L218L
Galactosemia	GALT	N314D, rs2070074
Galactosemia	GALT	Phel71Ser
Galactosemia	GALT	Q169K
Galactosemia	GALT	Q188R
Galactosemia	GALT	S135L
Galactosemia	GALT	T138M C>T
Galactosemia	GALT	X380R
Galactosemia	GALT	Y209C A>G
Enfermedad de Gaucher	GBA	1035insG
Enfermedad de Gaucher	GBA	1035insG
Enfermedad de Gaucher	GBA	84insG
Enfermedad de Gaucher	GBA	84insG
Enfermedad de Gaucher	GBA	D409H, rs1064651
Enfermedad de Gaucher	GBA	D409V
Enfermedad de Gaucher	GBA	IVS2(+1)G>A
Enfermedad de Gaucher	GBA	L444P (1448T>C), rs35095275
Enfermedad de Gaucher	GBA	N370S
Enfermedad de Gaucher,	GBA	R463C
Enfermedad de Gaucher	GBA	R463H
Enfermedad de Gaucher	GBA	R496H(1604G>A)
Enfermedad de Gaucher	GBA	V394L
Pérdida de audición y sordera no sindrómica DFNA 3 relacionada con GJB2	GJB2	167delT
Pérdida de audición y sordera no sindrómica DFNA 3 relacionada con GJB2.	GJB2	167delT
Pérdida de audición y sordera no sindrómica DFNA 3 relacionada con GJB2	GJB2	235delC
Pérdida de audición y sordera no sindrómica DFNA 3 relacionada con GJB2	GJB2	235delC

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Pérdida de audición y sordera no sindrómica DFNA 3 relacionada con GJB2	GJB2	35delG
Pérdida de audición y sordera no sindrómica DFNA 3 relacionada con GJB2	GJB2	35delG
Pérdida de audición y sordera no sindrómica DFNA 3 relacionada con GJB2	GJB2	IVS1+1G>A
Pérdida de audición y sordera no sindrómica DFNB 1 relacionada con GJB2	GJB2	101delAG
Pérdida de audición y sordera no sindrómica DFNB 1 relacionada con GJB2	GJB2	313del14
Pérdida de audición y sordera no sindrómica DFNB 1 relacionada con GJB2	GJB2	313del14
Pérdida de audición y sordera no sindrómica DFNB 1 relacionada con GJB2	GJB2	delE120
Pérdida de audición y sordera no sindrómica DFNB 1 relacionada con GJB2	GJB2	delE120
Pérdida de audición y sordera no sindrómica DFNB 1 relacionada con GJB2	GJB2	M34T, rs35887622
Pérdida de audición y sordera no sindrómica DFNB 1 relacionada con GJB2	GJB2	Q124X
Pérdida de audición y sordera no sindrómica DFNB 1 relacionada con GJB2	GJB2	R184P
Pérdida de audición y sordera no sindrómica DFNB 1 relacionada con GJB2	GJB2	V37I
Pérdida de audición y sordera no sindrómica DFNB 1 relacionada con GJB2	GJB2	W24X
Pérdida de audición y sordera no sindrómica DFNB 1 relacionada con GJB2	GJB2	W77R
Pérdida de audición y sordera no sindrómica DFNB 1 relacionada con GJB2	GJB2	W77X
Deficiencia en glucosa-6-fosfato deshidrogenasa	G6PD	A335V
Deficiencia en glucosa-6-fosfato deshidrogenasa	G6PD	R459L
Deficiencia en glucosa-6-fosfato deshidrogenasa	G6PD	R459P
Deficiencia en glucosa-6-fosfato deshidrogenasa	G6PD	rs1050828, rs1050828

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Deficiencia en glucosa-6-fosfato deshidrogenasa	G6PD	rs1050829, rs1050829
Deficiencia en glucosa-6-fosfato deshidrogenasa	G6PD	rs5030868, rs5030868
Glutaricacidemia Tipo 1	GCDH	A421V
Glutaricacidemia Tipo 1	GCDH	R402W
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	459insTA
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	459insTA
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	727G/T
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	del F327
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	del F327
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	G188R
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	G270V
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	Q242X
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	Q27fsdelC
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	Q27fsdelC
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	Q347X
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	R83C
Enfermedad de almacenamiento de glucógeno Tipo Ia	G6PC	R83H
Enfermedad de almacenamiento de glucógeno Tipo Ib	G6PT1	1211delCT
Enfermedad de almacenamiento de glucógeno Tipo Ib	G6PT1	A367T
Enfermedad de almacenamiento de glucógeno Tipo Ib	G6PT1r	G339C
Enfermedad de almacenamiento de glucógeno Tipo Ib	G6PT1	G339D
Enfermedad de almacenamiento de glucógeno Tipo Ib	G6PT1	W118R
Enfermedad de almacenamiento de glucógeno Tipo II	GAA	Arg854X
Enfermedad de almacenamiento de glucógeno Tipo II	GAA	Asp645Glu, rs28940868
Enfermedad de almacenamiento de glucógeno Tipo II	GAA	IVS1(~13t>g)
Enfermedad de almacenamiento de glucógeno Tipo III	AGL	1484delT
Enfermedad de almacenamiento de glucógeno Tipo III	AGL	1484delT
Enfermedad de almacenamiento de glucógeno Tipo III	AGL	17delAG

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Enfermedad de almacenamiento de glucógeno Tipo III	AGL	17delAG
Enfermedad de almacenamiento de glucógeno Tipo III	AGL	Q6X
Enfermedad de almacenamiento de glucógeno Tipo V	PYGM	G204S
Enfermedad de almacenamiento de glucógeno Tipo V	PYGM	K542T
Enfermedad de almacenamiento de glucógeno Tipo V	PYGM	K542X
Enfermedad de almacenamiento de glucógeno Tipo V	PYGM	R49X
Miopatías relacionadas con GNE	GNE	M712T, rs28937594
Síndrome de Gracile	BCS1L	c.232A>G, rs28937590
Beta-talasemia de hemoglobina S	HBB	c.19G>A
Beta-talasemia de hemoglobina S	HBB	c.20A>T
Beta-talasemia de hemoglobina S	HBB	c.79G>A
Beta-talasemia de hemoglobina S	HBB	Hb CS
Beta-talasemia de hemoglobina S	HBB	Hb D
Beta-talasemia de hemoglobina S	HBB	Hb O
Intolerancia hereditaria a la fructosa	ALDOB	A149P, rs 1800546
Intolerancia hereditaria a la fructosa	ALDOB	A174D
Intolerancia hereditaria a la fructosa	ALDOB	Delta4E4
Intolerancia hereditaria a la fructosa	ALDOB	Delta4E4
Intolerancia hereditaria a la fructosa	ALDOB	N334K
Intolerancia hereditaria a la fructosa	ALDOB	Y203X
Pancreatitis hereditaria	PRSS1	A16V
Pancreatitis hereditaria	SPINK1	MIT
Pancreatitis hereditaria	PRSS1	N29I
Pancreatitis hereditaria	SPINK1	N34S, rs17107315
Pancreatitis hereditaria	PRSS1	R122C
Pancreatitis hereditaria	PRSS1	R122H
Timina-uraciluria hereditaria	DPYD	rs3918290
Deficiencia en hexosaminidasa A	HEXA	1278insTATC
Deficiencia en hexosaminidasa A	HEXA	1278insTATC
Deficiencia en hexosaminidasa A	HEXA	G269S (805G>A)

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Deficiencia en hexosaminidasa A	HEXA	IVS12 + 1G>C
Deficiencia en hexosaminidasa A	HEXA	IVS7 + 1G>A
Deficiencia en hexosaminidasa A	HEXA	rVS9 + 1G>A
Deficiencia en hexosaminidasa A	HEXA	R178C
Deficiencia en hexosaminidasa A	HEXA	R178H
Deficiencia en hexosaminidasa A	HEXA	R247W (739C>T)
Deficiencia en hexosaminidasa A	HEXA	R249W (745C>T)
Hemocromatosis hereditaria asociada a HFE	HFE	E168Q
Hemocromatosis hereditaria asociada a HFE	HFE	E168X
Hemocromatosis hereditaria asociada a HFE	HFE	HM971246, H63H
Hemocromatosis hereditaria asociada a HFE	HFE	P160delC
Hemocromatosis hereditaria asociada a HFE	HFE	P160delC
Hemocromatosis hereditaria asociada a HFE	HFE	Q127H, rs28934595
Hemocromatosis hereditaria asociada a HFE	HFE	Q283P
Hemocromatosis hereditaria asociada a HFE	HFE	rs1799945, rs1799945
Hemocromatosis hereditaria asociada a HFE	HFE	rs1800562, rs1800562
Hemocromatosis hereditaria asociada a HFE	HFE	rs1800730, rs1800730
Hemocromatosis hereditaria asociada a HFE	HFE	V53M
Hemocromatosis hereditaria asociada a HFE	HFE	V59M
Hemocromatosis hereditaria asociada a HFE	HFE	W169X
Displasia ectodérmica hidrótica 2	GJB6	A88V, rs28937872
Displasia ectodérmica hidrótica 2	GJB6	G11R
Displasia ectodérmica hidrótica 2	GJB6	V37E
Homocistinuria causada por deficiencia en cistationina Beta-Sintasa	CBS	G307S 919G->A
Homocistinuria causada por deficiencia en cistationina Beta-Sintasa	CBS	I278T 833T->C, rs5742905
Parálisis periódica hipercalémica Tipo 1	SCN4A	I693T

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Parálisis periódica hipercalémica Tipo 1	SCN4A	L689I
Parálisis periódica hipercalémica Tipo 1	SCN4A	L689V
Parálisis periódica hipercalémica Tipo 1	SCN4A	M1360V
Parálisis periódica hipercalémica Tipo 1	SCN4A	M1592 V
Parálisis periódica hipercalémica Tipo 1	SCN4A	p.A1156T
Parálisis periódica hipercalémica Tipo 1	SCN4A	p.M1370V
Parálisis periódica hipercalémica Tipo 1	SCN4A	p.R1448C
Parálisis periódica hipercalémica Tipo 1	SCN4A	p.T1313M
Parálisis periódica hipercalémica Tipo 1	SCN4A	R675G
Parálisis periódica hipercalémica Tipo 1	SCN4A	R675Q
Parálisis periódica hipercalémica Tipo 1	SCN4A	R675W
Parálisis periódica hipercalémica Tipo 1	SCN4A	T704M
Parálisis periódica hipercalémica Tipo 1	SCN4A	V781I
Síndrome de hiperornitinemia-hiperamonemia-homocitrulinuria	SLC25A15	F188del
Síndrome de hiperornitinemia-hiperamonemia-homocitrulinuria	SLC25A15	F188del
Hiperoxaluria, primaria, Tipo 1	AGXT	33insC
Hiperoxaluria, primaria, Tipo 1	AGXT	33insC
Hiperoxaluria, primaria, Tipo 1	AGXT	F152I
Hiperoxaluria, primaria, Tipo 1	AGXT	G170R
Hiperoxaluria, primaria, Tipo 1	AGXT	1244T
Hiperoxaluria, primaria, Tipo 2	GRHPR	103delG
Hiperoxaluria, primaria, Tipo 2	GRHPR	103delG
Hipocondroplasia	FGFR3	Asn328Ile
Hipocondroplasia	FGFR3	I538V
Hipocondroplasia	FGFR3	K650M
Hipocondroplasia	FGFR3	K650N 1950G>T
Hipocondroplasia	FGFR3	K650Q
Hipocondroplasia	FGFR3	N540K 1620C>A

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Hipocondroplasia	FGFR3	N540S
Hipocondroplasia	FGFR3	N540T
Parálisis periódica hipocalémica Tipo 1	CACNA1S	R528G
Parálisis periódica hipocalémica Tipo 1	CACNA1S	R528H
Parálisis periódica hipocalémica Tipo 1	CACNA1S	rs28930068, rs28930068
Parálisis periódica hipocalémica Tipo 1	CACNA1S	rs28930069, rs28930069
Parálisis periódica hipocalémica Tipo 2	SCN4A	R669H
Parálisis periódica hipocalémica Tipo 2	SCN4A	R672C
Parálisis periódica hipocalémica Tipo 2	SCN4A	R672G
Parálisis periódica hipocalémica Tipo 2	SCN4A	R672H
Parálisis periódica hipocalémica Tipo 2	SCN4A	R672S
Hipocondroplasia	ALPL	Asp361Val
Hipocondroplasia	ALPL	c.1559delT
Hipocondroplasia	ALPL	c.1559delT
Hipocondroplasia	ALPL	E174K
Hipocondroplasia	ALPL	G317D
Hipocondroplasia	ALPL	Phe310Leu
Acidemia isovalérica	IVD	A282V
Acidemia isovalérica	IVD	rs28940889
Enfermedad de Krabbe	GALC	EX11-17DEL
Enfermedad de Krabbe	GALC	EX11-17DEL
Enfermedad de Krabbe	GALC	G270D
Enfermedad de Krabbe	GALC	rs1805078, rs1805078
Enfermedad de Krabbe	GALC	rs398607
Neuropatía óptica hereditaria de Leber	mtDNA	14484T>C
Neuropatía óptica hereditaria de Leber	mtDNA	15257G>A
Neuropatía óptica hereditaria de Leber	mtDNA	G14459A
Neuropatía óptica hereditaria de Leber	mtDNA	G3460A
Neuropatía óptica hereditaria de Leber	mtDNA	m.11778G>A
Neuropatía óptica hereditaria de Leber	mtDNA	m.13708G>A
Neuropatía óptica hereditaria de Leber	mtDNA	m.15812G>A
Neuropatía óptica hereditaria de Leber	mtDNA	m.3394T>C
Neuropatía óptica hereditaria de Leber	mtDNA	m.4216T>C

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Neuropatía óptica hereditaria de Leber	mtDNA	m.4917A>G
Síndrome de Leigh, tipo francés-canadiense	LRPPRC	A354V
LGMD2I	FKRP	L276I, rs28937900
Deficiencia en 3-hidroxiacil de cadena larga-CoA deshidrogenasa	HADHA	E474Q c.1528G>C
Deficiencia en 3-hidroxiacil de cadena larga-CoA deshidrogenasa	HADHA	Q342X1132C>T
Enfermedad urinaria de jarabe de arce Tipo 1A	BCKDHA	Y438N
Enfermedad urinaria de jarabe de arce Tipo 1B	BCKDHB	E372X
Enfermedad urinaria de jarabe de arce Tipo 1B	BCKDHB	G278S
Enfermedad urinaria de jarabe de arce Tipo 1B	BCKDHB	R183P
Síndrome de McCune-Albright	GNAS	R201C
Síndrome de McCune-Albright	GNAS	R201G
Síndrome de McCune-Albright	GNAS	R201H
Deficiencia en acil de cadena media-coenzima A deshidrogenasa	ACADM	244insT
Deficiencia en acil de cadena media-coenzima A deshidrogenasa	ACADM	244insT
Deficiencia en acil de cadena media-coenzima A deshidrogenasa	ACADM	250C>T
Deficiencia en acil de cadena media-coenzima A deshidrogenasa	ACADM	583G>A
Deficiencia en acil de cadena media-coenzima A deshidrogenasa	ACADM	616C>T
Deficiencia en acil de cadena media-coenzima A deshidrogenasa	ACADM	617G>A
Deficiencia en acil de cadena media-coenzima A deshidrogenasa	ACADM	799G>A
Deficiencia en acil de cadena media-coenzima A deshidrogenasa	ACADM	K304E
Deficiencia en acil de cadena media-coenzima A deshidrogenasa	ACADM	Y42H

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Leucoencefalopatía megaloencefálica con quistes subcorticales	MLC1	135insC
Leucoencefalopatía megaloencefálica con quistes subcorticales	MLC1	135insC
MELAS	mtDNA	3243A>G
MELAS	mtDNA	3250T>C
MELAS	mtDNA	3252A>G
MELAS	mtDNA	A12770G
MELAS	mtDNA	C3256T
MELAS	mtDNA	G13513A
MELAS	mtDNA	T3291C
MELAS	mtDNA	T8356C
MELAS	mtDNA	T9957C
MERRF	mtDNA	8361G>A
MERRF	mtDNA	A8296G
MERRF	mtDNA	m.8344A>G
Leucodistrofia metacromática	ARSA	c.459+1G>A
Leucodistrofia metacromática	ARSA	p.P426L, rs28940893
Leucodistrofia metacromática	ARSA	p.T274M
Leucodistrofia metacromática	ARSA	P377L
Cardiomiopatía mitocondrial	mtDNA	A3260T
Cardiomiopatía mitocondrial	mtDNA	A4300G
Cardiomiopatía mitocondrial	mtDNA	C3303T
Cardiomiopatía mitocondrial	mtDNA	T9997C
Síndrome de Leigh asociado a ADN mitocondrial y NARP	mtDNA	5537insT
Síndrome de Leigh asociado a ADN mitocondrial y NARP	mtDNA	5537insT
Síndrome de Leigh asociado a ADN mitocondrial y NARP	mtDNA	8993T>C
Síndrome de Leigh asociado a ADN mitocondrial y NARP	mtDNA	8993T>G
Síndrome de Leigh asociado a ADN mitocondrial y NARP	mtDNA	C11777A
Síndrome de Leigh asociado a ADN mitocondrial y NARP	mtDNA	T10158C

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Síndrome de Leigh asociado a ADN mitocondrial y NARP	mtDNA	T10191C
Síndrome de Leigh asociado a ADN mitocondrial y NARP	mtDNA	T8851C
Síndrome de Leigh asociado a ADN mitocondrial y NARP	mtDNA	T9176C
Síndrome de Leigh asociado a ADN mitocondrial y NARP	mtDNA	T9176G
Deficiencia en MTHFR	MTHFR	1298A>C
Deficiencia en MTHFR	MTHFR	rs1801133, rs1801133
Pérdida de audición y sordera relacionada con MTRNR1	mtDNA	1095T>C
Pérdida de audición y sordera relacionada con MTRNR1	mtDNA	1494C>T
Pérdida de audición y sordera relacionada con MTRNR1	mtDNA	1555A>G
Pérdida de audición y sordera relacionada con MTRNR1	mtDNA	961T>G
Pérdida de audición y sordera relacionada con MTRNR1	mtDNA	A7445G
Pérdida de audición y sordera relacionada con MTRNR1	mtDNA	7443A>G
Pérdida de audición y sordera relacionada con MTRNR1	mtDNA	7444G>A
Pérdida de audición y sordera relacionada con MTRNR1	mtDNA	7472insC
Pérdida de audición y sordera relacionada con MTRNR1	mtDNA	7472insC
Pérdida de audición y sordera relacionada con MTRNR1	mtDNA	7510T>C
Pérdida de audición y sordera relacionada con MTRNR1	mtDNA	7511T>C
Pérdida de audición y sordera relacionada con MTRNR1	mtDNA	75T2T>C
Mucopolisacaridosis IV	MCOLN1	delEx1 3 Ex7 (511>6944del)
Mucopolisacaridosis IV	MCOLN1	delEx1 3 Ex7 (511 >6944del)
Mucopolisacaridosis IV	MCOLN1	IVS-2A>G
Mucopolisacaridosis Tipo I	IDUA	c.46_57del
Mucopolisacaridosis Tipo I	IDUA	c.46_57del
Mucopolisacaridosis Tipo I	IDUA	p.A327P
Mucopolisacaridosis Tipo I	IDUA	p.P533R
Mucopolisacaridosis Tipo I	IDUA	Q70X
Mucopolisacaridosis Tipo I	IDUA	W402X

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Mucopolisacaridosis Tipo IIIA	SGSH	p.R245H
Mucopolisacaridosis Tipo IIIA	SGSH	p.R74C
Mucopolisacaridosis Tipo IIIA	SGSH	p.S66W
Mucopolisacaridosis Tipo VII	GUSB	p.D152N
Neoplasia endocrina múltiple Tipo 2	RET	2047T>A
Neoplasia endocrina múltiple Tipo 2	RET	2047T>A
Neoplasia endocrina múltiple Tipo 2	RET	2047T>C
Neoplasia endocrina múltiple Tipo 2	RET	2047T>G
Neoplasia endocrina múltiple Tipo 2	RET	2048G>A
Neoplasia endocrina múltiple Tipo 2	RET	A883F 2647 G>T
Neoplasia endocrina múltiple Tipo 2	RET	Glu768Asp G>C
Neoplasia endocrina múltiple Tipo 2	RET	M918T
Enfermedad músculo-ojo-cerebro	POMGNT1	C.1539+1G>A
Poliposis asociada a MYH	MUTYH	C.1376C>A
Poliposis asociada a MYH	MUTYH	c.494A>G, rs34612342
Poliposis asociada a MYH	GEN_SYMBOL_T BD	rs36053993, rs36053993
Enfermedad de Niemann-Pick debida a deficiencia en esfingomielinasa	SMPD1	c.990delC
Enfermedad de Niemann-Pick debida a deficiencia en esfingomielinasa	SMPD1	c.990delC
Enfermedad de Niemann-Pick debida a deficiencia en esfingomielinasa	SMPD1	fsP330
Enfermedad de Niemann-Pick debida a deficiencia en esfingomielinasa	SMPD1	fsP330
Enfermedad de Niemann-Pick debida a deficiencia en esfingomielinasa	SMPD1	L302P
Enfermedad de Niemann-Pick debida a deficiencia en esfingomielinasa	SMPD1	R496L
Enfermedad de Niemann-Pick debida a deficiencia en esfingomielinasa	SMPD1	R608del
Enfermedad de Niemann-Pick Tipo C1	NPC1	I1061T

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Síndrome de rotura de Nijmegen	NBN	657del5
Síndrome de rotura de Nijmegen	NBN	657del5
Síndrome de Pallister-Hall	GLI3	2012delG
Síndrome de Pallister-Hall	GLI3	2012delG
Síndrome de Pallister-Hall	GLI3	2023delG
Síndrome de Pallister-Hall	GLI3	2023delG
Síndrome de Pendred	SLC26A4	1197delT
Síndrome de Pendred	SLC26A4	1197delT
Síndrome de Pendred	SLC26A4	E384G
Síndrome de Pendred	SLC26A4	IV58+1(G->A)
Síndrome de Pendred	SLC26A4	L236P
Síndrome de Pendred	SLC26A4	T416P
Deficiencia de enzima bifuncional peroxisomal	HSD17B4	c.302+IG>C
Deficiencia de enzima bifuncional peroxisomal	HSD17B4	c.303-IG>A
Trastornos de desarrollo dominante	NLGN4X	D396X
Trastornos de desarrollo dominante	NLGN4X	D396X
Trastornos de desarrollo dominante	NLGN4X	NLGN4X:1253delAG
Trastornos de desarrollo dominante	NLGN4X	NLGN4X:1253delAG
Trastornos de desarrollo dominante	NLGN3	R451C
Deficiencia en fenilalanina hidroxilasa	PAH	G272X
Deficiencia en fenilalanina hidroxilasa	PAH	I65T
Deficiencia en fenilalanina hidroxilasa	PAH	IVS12+1G>T
Deficiencia en fenilalanina hidroxilasa	PAH	L48S, rs5030841
Deficiencia en fenilalanina hidroxilasa	PAH	R158Q, rs5030843
Deficiencia en fenilalanina hidroxilasa	PAH	R252W, rs5030847
Deficiencia en fenilalanina hidroxilasa	PAH	R261Q, rs5030849
Deficiencia en fenilalanina hidroxilasa	PAH	R408Q, rs5030859
Deficiencia en fenilalanina hidroxilasa	PAH	R408W, rs5030858
Deficiencia en fenilalanina hidroxilasa	PAH	rs5030855
Deficiencia en fenilalanina hidroxilasa	PAH	rs5030861
Deficiencia en fenilalanina hidroxilasa	PAH	Y414C, rs5030860

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Inhibidor I de activador de plasminógeno	SERPINE1	-844 G>A
Inhibidor I de activador de plasminógeno	SERPINE1	4G/5G
Enfermedad renal poliquística, recesiva autosómica	PKHD1	C.10412T>G
Enfermedad renal poliquística, recesiva autosómica	PKHD1	C.107C>T
Enfermedad renal poliquística, recesiva autosómica	PKHD1	C.1486C>T
Enfermedad renal poliquística, recesiva autosómica	PKHD1	c.5895dupA
Enfermedad renal poliquística, recesiva autosómica	PKHD1	c.5895dupA
Enfermedad renal poliquística, recesiva autosómica	PKHD1	c.9689delA
Enfermedad renal poliquística, recesiva autosómica	PKHD1	c.9689delA
Lipofuscinosis ceroides neuronal relacionada con PPT 1	PPT1	c.364A>T
Lipofuscinosis ceroides neuronal relacionada con PPT 1	PPT1	p.L10X
Lipofuscinosis ceroides neuronal relacionada con PPT 1	PPT1	p.R151X
Lipofuscinosis ceroides neuronal relacionada con PPT 1	PPT1	T75P
Deficiencia de hormona de la pituitaria relacionada con PROP 1	PROP1	301-302delAG
Deficiencia de hormona de la pituitaria relacionada con PROP 1	PROP1	301-302delAG
Trombofilia de protrombina	F2	rs1799963
Trombofilia de protrombina	F2	rs6025, rs6025
Raquitismo por deficiencia en pseudovitamina D	CYP27B1	duplicación de 7 pb en el exón 8
Raquitismo por deficiencia en pseudovitamina D	GYP27B1	duplicación de 7 pb en el exón 8
Raquitismo por deficiencia en pseudovitamina D	CYP27B1	958delG
Raquitismo por deficiencia en pseudovitamina D	CYP27B1	958delG
Síndrome de Rett	MECP2	806delG
Síndrome de Rett	MECP2	806delG
Síndrome de Rett	MECP2	A140V, rs28934908
Síndrome de Rett	MECP2	P152R
Síndrome de Rett	MECP2	P225R
Síndrome de Rett	MECP2	R106W, rs28934907
Síndrome de Rett	MECP2	R133C
Síndrome de Rett	MECP2	R168X
Síndrome de Rett	MECP2	R255X
Síndrome de Rett	MECP2	R270X
Síndrome de Rett	MECP2	R294X

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Síndrome de Rett	MECP2	R306C, rs28935468
Síndrome de Rett	MECP2	S134C
Síndrome de Rett	MECP2	T158M, rs28934906
Condrodisplasia punteada rizomélica Tipo 1	PEX7	p.A218V
Condrodisplasia punteada rizomélica Tipo 1	PEX7	p.G217R
Condrodisplasia punteada rizomélica Tipo 1	PEX7	p.L292X, rs1805137
Deficiencia en acil de cadena corta-CoA deshidrogenasa	ACADS	c.511C>T, rs1800556
Deficiencia en acil de cadena corta-CoA deshidrogenasa	ACADS	c.625G>A
Deficiencia en acil de cadena corta-CoA deshidrogenasa	ACADS	R107C
Síndrome de Shwachman-Diamond	SBDS	183_184TA>CT
Síndrome de Shwachman-Diamond	SBDS	183_184TA>CT
Síndrome de Shwachman-Diamond	SBDS	258+2T>C
Síndrome de Sjogren-Larsson	ALDH3A2	C.943C>T
Síndrome Smith-Lemli-Opitz	DHCR7	C380Y
Síndrome Smith-Lemli-Opitz	DHCR7	IVS8-1G>C
Síndrome Smith-Lemli-Opitz	DHCR7	L109P
Síndrome Smith-Lemli-Opitz	DHCR7	L157P
Síndrome Smith-Lemli-Opitz	DHCR7	R352Q
Síndrome Smith-Lemli-Opitz	DHCR7	R352W
Síndrome Smith-Lemli-Opitz	DHCR7	R404C
Síndrome Smith-Lemli-Opitz	DHCR7	R446Q
Síndrome Smith-Lemli-Opitz	DHCR7	T93M
Síndrome Smith-Lemli-Opitz	DHCR7	V326L
Síndrome Smith-Lemli-Opitz	DHCR7	W151X
Síndrome Smith-Lemli-Opitz	DHCR7	W151X
Paraplejia espástica 13	HSPD1	V72I
Osteocondrodisplasia relacionada con transportador de sulfato	SLC26A2	340delV
Osteocondrodisplasia relacionada con transportador de sulfato	SLC26A2	340delV

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Osteocondrodисplasia relacionada con transportador de sulfato	SLC26A2	C.837C>T
Osteocondrodисplasia relacionada con transportador de sulfato	SLC26A2	C653S
Osteocondrodисplasia relacionada con transportador de sulfato	SLC26A2	IVS1+2T>C
Osteocondrodисplasia relacionada con transportador de sulfato	SLC26A2	R178X
Hemocromatosis hereditaria relacionada con TFR2	TFR2	AVAQ594-597del
Hemocromatosis hereditaria relacionada con TFR2	TFR2	AVAQ594-597del
Hemocromatosis hereditaria relacionada con TFR2	TFR2	E60X
Hemocromatosis hereditaria relacionada con TFR2	TFR2	E60X
Hemocromatosis hereditaria relacionada con TFR2	TFR2	M172K
Hemocromatosis hereditaria relacionada con TFR2	TFR2	Y250X
Displasia tanatofórica	FGFR3	G370C
Displasia tanatofórica	FGFR3	K650E
Displasia tanatofórica	FGFR3	R248C
Displasia tanatofórica	FGFR3	S249C
Displasia tanatofórica	FGFR3	S371C
Displasia tanatofórica	FGFR3	X807C A>T
Displasia tanatofórica	FGFR3	X807G
Displasia tanatofórica	FGFR3	X807L
Displasia tanatofórica	FGFR3	X807R
Displasia tanatofórica	FGFR3	X807S
Displasia tanatofórica	FGFR3	X807W
Displasia tanatofórica	FGFR3	Y373C
Lipofuscinosis ceroides neuronal relacionada con TPP1	TPP1	C.509-1G>A
Lipofuscinosis ceroides neuronal relacionada con TPP1	TPP1	c.509-1G>C
Lipofuscinosis ceroides neuronal relacionada con TPP1	TPP1	G284V
Lipofuscinosis ceroides neuronal relacionada con TPP1	TPP1	p.R208X
Amiloidosis de transtiretina	TTR	c.148G>A
DRD deficiente en tirosina hidroxilasa	TH	L205P
DRD deficiente en tirosina hidroxilasa	TH	R202H
Tirosinemia Tipo I	FAH	E357X
Tirosinemia Tipo I	FAH	IVS12+5 G>A
Tirosinemia Tipo I	FAH	IVS7-6 T>G
Tirosinemia Tipo I	FAH	IVS8-1G>C
Tirosinemia Tipo I	FAH	p.W262X
Tirosinemia Tipo I	FAH	P261L
Tirosinemia Tipo I	FAH	Q64H

FIGURA 1 - Tabla de variantes genéticas causales a modo de ejemplo

Enfermedad	Gen	Nombre variante
Enfermedad de Wilson	ATP7B	1340del4
Enfermedad de Wilson	ATP7B	3402delC
Enfermedad de Wilson	ATP7B	3402delC
Enfermedad de Wilson	ATP7B	H1069Q
Enfermedad de Wilson	ATP7B	R778G
Enfermedad de Wilson	ATP7B	W779X
Enfermedad de Wilson	ATP7B	W779X
Retinoschisis juvenil ligada a X	RSI	E72K
Retinoschisis juvenil ligada a X	RSI	G109R
Retinoschisis juvenil ligada a X	RSI	G74V
Espectro de síndrome de Zellweger	PEX1	c.2097_2098insT
Espectro de síndrome de Zellweger	PEX1	c.2097_2098insT
Espectro de síndrome de Zellweger	PEX1	c.2916delA
Espectro de síndrome de Zellweger	PEX1	c.2916delA
Espectro de síndrome de Zellweger	PEX1	p.G843D

FIGURA 2

Poblaciones que son notables con respecto a tener elevados riesgos de algunas enfermedades mendelianas:

1. Nivel continental:

Europa:

Asia del este

África

América

2. Nivel sub-continental

Arabia

Judíos asquenazíes

Canadá

Central

Chipre

Dinamarca

República dominicana

Inglaterra

Ascendencia europea

Finlandia

Hungría

India

Irán

Israel

Italia

FIGURA 2

Costa de Marfil
Japón
Kenia
Corea
Kuwait
Mediterránea
Mediterránea
Micronesia
Este central
Marruecos
Navajo, occidental
Nueva Zelanda
Nigeria
Noruega
Polonia
Portugal
Rusia
Arabia Saudita
Escandinavia
Eslovaquia
Eslovenia
Sudeste
Suiza
Taiwán

FIGURA 2

Túnez

Turquía

EEUU

Vanuatu

Nivel local:

Bretaña, Francia

Quebec, Canadá

Región Charlevoix-Saguenay de Quebec

Terranova, Canadá

Adra Pradesh, India

Judíos iraníes

Sardianos, Italia

Beduinos en Kuwait

Pingelap, Islas Carolina

Asquenazí in EEUU

Esquimales Yupik de Alaska

FIGURA 3

AIM que distinguen poblaciones africanas y europeas

rs376350, rs675924, rs676290, rs708156, rs717090, rs717091, rs717225, rs718092, rs718387, rs719776, rs720225, rs720888, rs720891, rs720966, rs721684, rs723632, rs723802, rs724729, rs725416, rs725438, rs725472, rs725510, rs725667, rs725908, rs726391, rs726700, rs726777, rs726996, rs727342, rs728606, rs728647, rs736723, rs764051, rs874816, rs951308, rs951412, rs951431, rs951666, rs952109, rs952165, rs952902, rs953035, rs959960, rs1002835, rs1027583, rs1057187, rs1074075, rs1074086, rs1112730, rs1113480, rs1152537, rs1153849, rs1342008, rs1352405, rs1365720, rs1366842, rs1367996, rs1369290, rs1377456, rs1382934, rs1404694, rs1405467, rs1407716, rs1431948, rs1465648, rs1478785, rs1485765, rs1487214, rs1506069, rs1516238, rs1526028, rs1541836, rs1584385, rs1831024, rs1858465, rs1861498, rs1904464, rs1923416, rs1928415, rs1929609, rs1945248, rs1950284, rs1979541, rs2021779, rs2021781, rs2077681, rs2078588, rs2169462, rs2207782, rs2211771, rs2225251, rs2263039, rs2317212, rs2341823

AIM que distinguen poblaciones africanas y asiáticas

rs1063, rs326798, rs376350, rs441728, rs717225, rs717531, rs718387, rs718424, rs719213, rs719776, rs720225, rs720888, rs720891, rs720966, rs721684, rs723632, rs723802, rs724729, rs725472, rs725667, rs725908, rs726391, rs726700, rs726777, rs726996, rs728606, rs736723, rs874816, rs879831, rs952109, rs952902, rs959157, rs967445, rs1073768, rs1074075, rs1074086, rs1074182, rs1113480, rs1152537, rs1318822, rs1366842, rs1369290, rs1371048, rs1377456, rs1382934, rs1391681, rs1404694, rs1405467, rs1407716, rs1431948, rs1465648, rs1478785, rs1485765, rs1487214, rs1506069, rs1526028, rs1807912, rs1820556, rs1831024, rs1851204, rs1858465, rs1861498, rs1904464, rs1923416, rs1928415, rs1929609, rs1934009, rs1945248, rs1961640, rs1979541, rs2017684, rs2021779, rs2021781, rs2077681, rs2078588, rs2115467, rs2169462, rs2341823

AIM que distinguen poblaciones africanas y americanas

1063, rs326798, rs376350, rs441728, rs483109, rs717090, rs717091, rs717171, rs717225, rs717531, rs718424, rs719213, rs719776, rs720225, rs720888, rs720891, rs720966, rs721684, rs722559, rs722869, rs723632, rs723802, rs723822, rs724246, rs724247, rs724729, rs725416, rs725472, rs725667, rs725908, rs726700, rs726777, rs726996, rs728606, rs764419, rs768324, rs874816, rs879831, rs925140, rs938431, rs951554, rs951784, rs952159,

FIGURA 3

rs953899, rs956065, rs959030, rs959157, rs963170, rs963171, rs967445, rs998410, rs1048610, rs1073768, rs1074075, rs1074086, rs1074182, rs1112806, rs1112828, rs1113480, rs1152537, rs1318822, rs1320892, rs1333208, rs1344870, rs1352405, rs1366842, rs1366847, rs1369290, rs1371048, rs1377456, rs1382934, rs1391681, rs1396799, rs1401385, rs1401608, rs1403454, rs1404694, rs1405467, rs1407716, rs1411106, rs1431948, rs1435090, rs1465648, rs1478785, rs1485765, rs1487214, rs1506069, rs1526028, rs1533224, rs1584385, rs1807912, rs1820556, rs1831024, rs1832443, rs1851204, rs1858465, rs1861498, rs1900099, rs1904464, rs1923416, rs1928415, rs1929609, rs1934009, rs1945248, rs1953054, rs1961273, rs1961320, rs1961640, rs1979541, rs1980888, rs1986644, rs2006996, rs2017684, rs2060319, rs2077681, rs2078588, rs2115467, rs2137636, rs2169462, rs2203099, rs2889670

AIM que distinguen poblaciones europeas y asiáticas

rs1063, rs242687, rs326798, rs441728, rs483109, rs708156, rs708726, rs717090, rs717091, rs717373, rs717531, rs718092, rs718686, rs719213, rs720496, rs722559, rs722869, rs723822, rs724246, rs724247, rs724496, rs724806, rs725264, rs725438, rs727342, rs728647, rs754798, rs764051, rs879831, rs925140, rs938431, rs951308, rs951378, rs951412, rs951431, rs951554, rs951666, rs952165, rs953035, rs953899, rs959030, rs959157, rs959960, rs963170, rs963171, rs967445, rs1002835, rs1037796, rs1048610, rs1074182, rs1112730, rs1112806, rs1153849, rs1318822, rs1320892, rs1327805, rs1333208, rs1350462, rs1364184, rs1365720, rs1371048, rs1391681, rs1396799, rs1401385, rs1401608, rs1403454, rs1411106, rs1435090, rs1461227, rs1516238, rs1541836, rs1584385, rs1807912, rs1820556, rs1851204, rs1934009, rs1938684, rs1939546, rs1944872, rs1950284, rs1953054, rs1961273, rs1961640, rs1980888, rs2017684, rs2021215, rs2115467, rs2137636, rs2203099, rs2207782, rs2225251, rs2263039, rs2312211, rs2317212, rs2366882, rs2889670

AIM que distinguen poblaciones europeas y americanas

1063, rs242687, rs441728, rs483109, rs675924, rs676290, rs708156, rs708726, rs717171, rs717373, rs717531, rs718092, rs718686, rs719213, rs720050, rs720051, rs720496, rs722559, rs722869, rs723822, rs724246, rs724247, rs724496, rs724806, rs725264, rs727342, rs754798, rs764051, rs764419, rs768324, rs925140, rs938431, rs951378, rs951412, rs951554, rs951666,

FIGURA 3

rs951784, rs952159, rs952165, rs953035, rs953899, rs956065, rs957862, rs958790, rs959030, rs959157, rs959960, rs963170, rs963171, rs998410, rs1002835, rs1027582, rs1027583, rs1037796, rs1048610, rs1112730, rs1112806, rs1112828, rs1153849, rs1318822, rs1320892, rs1327805, rs1333208, rs1344870, rs1350462, rs1364184, rs1365720, rs1366847, rs1371048, rs1395475, rs1396799, rs1401385, rs1401608, rs1403454, rs1411106, rs1435090, rs1461227, rs1533224, rs1541836, rs1820556, rs1822488, rs1832443, rs1900099, rs1934009, rs1938684, rs1939180, rs1939546, rs1944872, rs1950284, rs1953054, rs1961273, rs1961320, rs1980888, rs1986644, rs2006996, rs2021215, rs2060319, rs2086214, rs2115467, rs2137636, rs2203099, rs2207782, rs2211771, rs2225251, rs2312211, rs2317212, rs2366882, rs2889670

AIM que distinguen poblaciones asiáticas y americanas

1900099, rs768324, rs1344870, rs717171, rs958790, rs1939180, rs956065, rs1366847, rs1822488, rs1980888, rs951784, rs720051, rs720050, rs2203099, rs1533224, rs1403454, rs720496, rs1037796, rs1832443, rs1961320, rs764419, rs2317212, rs2060319, rs1401385, rs959030, rs1435090, rs1112828, rs1320892, rs1027582, rs0, rs2086214, rs1027583, rs1411106, rs957862, rs242687, rs998410, rs723822, rs725264, rs708726, rs1395475, rs2366882, rs0, rs2006996, rs1327805, rs1939546, rs1986644, rs963170, rs952159, rs1350462, rs952109, rs963171, rs676290, rs959157, rs675924, rs2207782, rs2211771, rs0, rs2021215, rs1342008, rs717373, rs722559, rs725510, rs708156, rs483109, rs953899, rs754798, rs719213, rs1401608, rs2137636, rs718092, rs725908, rs2889670, rs1112730, rs952165, rs1112806, rs951666, rs1953054, rs1333208, rs0, rs1365720, rs2341823, rs1938684, rs1961273, rs717531, rs967445, rs1073768, rs2021779, rs441728, rs726391, rs1944872, rs2021781, rs727342, rs2312211, rs736723, rs1153849, rs951554, rs1057187, rs2225251, rs1934009

AIM que distinguen europeos del norte de europeos del sur:

rs1129038, rs3769005, rs2596501, rs1364394, rs969539, rs974020, rs1922286, rs2171209, rs1560569, rs17443616, rs2596834, rs1890131, rs2418844, rs1157492, rs1922086, rs6432110, rs8041327, rs1416467, rs33706, rs379773, rs10496610, rs725974, rs822759, rs1517407, rs2187684, rs7997100, rs959763, rs1003306, rs1854226, rs2367191, rs202546, rs2905347, rs2236876, rs10508372, rs959260, rs920590, rs986642, rs2219248, rs923031, rs495347, rs103294, rs2003092, rs1373557,

FIGURA 3

rs16891982, rs1073321, rs2014303, rs1873195, rs1408794, rs9290675,
 rs7108371, rs1045873, rs523776, rs7163907, rs10853962, rs1777689,
 rs1032143, rs10512122, rs7908825, rs4859259, rs10516982, rs7965049,
 rs153595, rs7277342, rs10486207, rs3809125, rs2847502, rs10509384,
 rs2097884, rs4639533, rs10255965, rs1976033, rs1107820, rs10483853,
 rs4686497, rs3822616, rs4832640, rs2086085, rs2685159, rs12502036,
 rs9861816, rs2251432, rs8040452, rs7552548, rs3007711, rs1879558,
 rs1476162, rs17864053, rs477627, rs2047058, rs1065674, rs9860730,
 rs1448314, rs554788, rs1582398, rs1660964, rs10509954, rs17457687,
 rs4307284, rs11586379, rs12124147, rs7548659, rs760607, rs1981135,
 rs749663, rs1334804, rs10889750, rs1288367, rs213496, rs11207865,
 rs4915691, rs1566246, rs3101336, rs544858, rs10493649, rs12758138,
 rs6660484, rs2968485, rs712886, rs12061503, rs12046602, rs731756,
 rs10776798, rs6587597, rs912572, rs1419074, rs488150, rs1322780,
 rs859362, rs6680701, rs2146060, rs2784101, rs2494302, rs10494878,
 rs1930903, rs2275302, rs4916113, rs587913, rs6723966, rs10929646,
 rs4260216, rs340747, rs1489688, rs4665797, rs1438131, rs13032262,
 rs10205008, rs605832, rs7594173, rs6544718, rs6713506, rs7577894,
 rs2075375, rs6720799, rs41420, rs13032535, rs11164050, rs11123861,
 rs6739285, rs1955394, rs16829231, rs6730157, rs2117742, rs1037266,
 rs6752189, rs1227131, rs6749847, rs7558428, rs9288172, rs3769823,
 rs10189499, rs1955117, rs10189760, rs6436553, rs6436883, rs10510268,
 rs3804989, rs17043611, rs342042, rs2128162, rs3916092, rs11914832,
 rs744751, rs285327, rs11544593, rs6773085, rs11130841, rs6795735,
 rs2054956, rs9310279, rs6551458, rs7630043, rs9289584, rs4535234,
 rs7634702, rs4678297, rs9839394, rs10513729, rs7636818, rs12635682,
 rs4859259, rs6775595, rs6449375, rs13132286, rs6831024, rs6448770,
 rs9884706, rs10008492, rs12186184, rs1389037, rs6832891, rs1109501,
 rs4261956, rs3923243, rs1455313, rs6820697, rs1343921, rs13349989,
 rs7669241, rs2850971, rs2194860, rs1448817, rs1557815, rs4574434,
 rs1541745, rs4128688, rs692157, rs1390009, rs10519410, rs1507500,
 rs10517660, rs1492468, rs1430975, rs11730565, rs1435442, rs4241802,
 rs10032784, rs2736122, rs7718757, rs1533019, rs29460, rs464923,
 rs1001935, rs2279095, rs1373967, rs173686, rs6452788, rs174015,
 rs10478046, rs6863510, rs6866231, rs346650, rs1528961, rs171718,
 rs4868204, rs2135053, rs7727897, rs248327, rs1933652, rs2326106,
 rs4960257, rs2876167, rs1992387, rs2237148, rs7763768, rs6918101,
 rs12660883, rs382259, rs2076173, rs804847, rs12524885, rs4610536,
 rs638473, rs2451688, rs9389124, rs7753036, rs4896663, rs17551120,
 rs11756366, rs10428822, rs10950641, rs2041362, rs2189947, rs10267453,
 rs697518, rs2529015, rs10951140, rs1003549, rs933360, rs17135491,
 rs4020771, rs7791143, rs4730287, rs7781715, rs12705973, rs12706793,
 rs1593306, rs11761774, rs4571660, rs7795021, rs7806048, rs1548353,
 rs4565458, rs12548107, rs13273386, rs1127379, rs10109984, rs7830163,

FIGURA 3

rs6993747, rs7827918, rs2380646, rs4634634, rs11780821, rs11784678, rs1026804, rs1841316, rs4871195, rs4736413, rs396861, rs10758823, rs566820, rs10962589, rs2840790, rs10511715, rs10812520, rs867469, rs2130118, rs10746763, rs10908907, rs884886, rs10739277, rs7860625, rs1009473, rs7090242, rs7921493, rs11255712, rs1324322, rs4948508, rs3858126, rs4746826, rs7072160, rs4691, rs499437, rs1857459, rs7097946, rs11186543, rs701873, rs1023331, rs7071247, rs10509826, rs1325172, rs540609, rs11146457, rs2045272, rs7931276, rs1477569, rs4755844, rs2218868, rs2237997, rs12804561, rs2282504, rs7342241, rs2015747, rs488753, rs7131355, rs4936969, rs740851, rs7976721, rs10770437, rs699039, rs12370505, rs2088170, rs2860493, rs2723891, rs11107018, rs7307510, rs12231308, rs7312155, rs991817, rs11066320, rs1955105, rs2650170, rs1405050, rs3815210, rs11059335, rs2398513, rs377318, rs7323018, rs1979558, rs9539642, rs13379032, rs4885167, rs2175075, rs1678386, rs9556553, rs4982420, rs7152286, rs2038281, rs10498472, rs7492698, rs2180611, rs1531631, rs8003492, rs17126387, rs9671457, rs4906226, rs1947745, rs11634609, rs2611605, rs1153860, rs1906433, rs10519005, rs2292745, rs1320205, rs387727, rs8025028, rs3922394, rs1510058, rs8043261, rs2072986, rs1891325, rs896401, rs949429, rs30237, rs8047148, rs7205880, rs12446160, rs1872678, rs9937047, rs4238802, rs3751834, rs6539986, rs2253820, rs7215135, rs7220080, rs685098, rs2165846, rs11870879, rs196948, rs12939848, rs8071270, rs736632, rs400839, rs1786153, rs11564361, rs16948113, rs4941246, rs17079195, rs11151863, rs4799268, rs4536588, rs6510672, rs1552046, rs715159, rs8111998, rs2082455, rs17496703, rs7976, rs7252868, rs8103620, rs4814697, rs6076623, rs6118234, rs1041200, rs8116153, rs1884783, rs6126462, rs6129532, rs6032343, rs477627, rs4811158, rs2041317, rs6092326, rs3916504, rs6071491, rs977712, rs1735899, rs2832643, rs2836824, rs11702531, rs2235338, rs848728, rs11147509, rs4131364

AIM que distinguen europeos del norte de europeos judíos asquenazíes:

rs3769005, rs7551844, rs2596501, rs3007711, rs1129038, rs6996185, rs847851, rs17443616, rs1922086, rs974020, rs7108371, rs1065674, rs554788, rs1364394, rs1890131, rs1922286, rs10516982, rs477627, rs6448770, rs13084044, rs1355170, rs6432110, rs9860730, rs12713956, rs6028505, rs9576338, rs986642, rs4756052, rs7983897, rs10506555, rs822759, rs6751522, rs1991718, rs10508372, rs10493430, rs33706, rs379773, rs202546, rs920590, rs4307284, rs7163907, rs2251432, rs10762340, rs1660964, rs10853962, rs764255, rs4859259, rs923031, rs959260, rs8029021, rs2596834, rs2418844, rs998401, rs10496610, rs10509384, rs2003092, rs1560569, rs8016025, rs6477998, rs1517407, rs224378, rs17079195, rs10460810, rs6687300, rs2252815, rs1032143,

FIGURA 3

rs1476162, rs17457687, rs4642918, rs4832640, rs7637803, rs7997100, rs6432398, rs6490700, rs4771165, rs359291, rs1373557, rs2171209, rs1073321, rs4787923, rs920559, rs90213, rs10497233, rs3822616, rs7819806, rs12430668, rs12441220, rs10486207, rs2377689, rs4336881, rs1598452, rs605832, rs2219248, rs8078782, rs1045873, rs552976, rs10431948, rs2793471, rs2905347, rs7316723, rs1118962, rs1032355, rs1582398, rs6841252, rs1743789, rs1408794, rs1710807, rs914505, rs2154328, rs1870590, rs12706769, rs1854226, rs2014303, rs881433, rs10868791, rs10504275, rs6502048, rs1904406, rs17026435, rs8041327, rs1400174, rs11744977, rs9290675, rs1462368, rs518457, rs1435260, rs2063092, rs1011340, rs718387, rs9285474, rs2236876, rs2847502, rs1777689, rs706397, rs9993173, rs2363127, rs217627, rs10512122, rs725974, rs7699090, rs7038314, rs7301622, rs10495413, rs1381795, rs2047058, rs969539, rs10494870, rs489381, rs2804756, rs7975512, rs3927745, rs17149147, rs11713873, rs1994952, rs523776, rs10520475, rs349187, rs618897, rs6021183, rs4815707, rs313478, rs3844253, rs10509954, rs12502036, rs1531746, rs7561808, rs163018, rs2176246, rs11248060, rs11245545, rs6864793, rs9861816, rs17310125, rs1157492, rs103294, rs7501724, rs3764796, rs3809125, rs284509, rs11586379, rs153595, rs2419063, rs730566, rs2187684, rs838958, rs10825992, rs4782202, rs12500040, rs7965049, rs10484547, rs7920193, rs4686497, rs2241332, rs7689609, rs12251462, rs10458810, rs2097884, rs7598757, rs2287101

AIM que distinguen europeos del sur de europeos judíos asquenazíes:

rs7551844, rs6996185, rs1129038, rs6477998, rs1710807, rs1743789, rs6841252, rs6448770, rs11586379, rs1462368, rs8029021, rs13084044, rs847851, rs764255, rs17079195, rs10431948, rs12706769, rs8016025, rs3007711, rs12713956, rs7983897, rs2793471, rs8060233, rs12441220, rs1355170, rs4756052, rs554788, rs881433, rs6687300, rs1065674, rs10493430, rs1011340, rs10825992, rs8078782, rs40459, rs920559, rs605832, rs9285474, rs4815707, rs838958, rs10460810, rs1976033, rs10506555, rs9576338, rs6432398, rs914505, rs969539, rs477627, rs1531746, rs6028505, rs224378, rs6578195, rs6502048, rs4642918, rs6751522, rs1991718, rs4797909, rs11744977, rs1870590, rs2685159, rs2154328, rs7598757, rs9860730, rs163018, rs6678209, rs7108371, rs1994952, rs8040452, rs10516982, rs262216, rs618897, rs284509, rs6490700, rs17026435, rs11713873, rs6907950, rs11259206, rs10762340, rs9993173, rs17149147, rs4639533, rs1381795, rs7316723, rs359291, rs2446653, rs273747, rs2063092, rs10255965, rs4976606, rs7155872, rs730566, rs4307284, rs2377689, rs7689609, rs2411128, rs10517807, rs1435260, rs3927745, rs10458810, rs17310125, rs1660964, rs7301622,

FIGURA 3

rs706397, rs1416467, rs998401, rs313478, rs11854557, rs4131048, rs13008, rs2241332, rs4243408, rs951412, rs7920193, rs3844253, rs718387, rs777710, rs4771165, rs132549, rs1389560, rs3856744, rs6460989, rs2252815, rs2251432, rs349187, rs2363127, rs10504275, rs7372209, rs495347, rs4770401, rs1410418, rs1118962, rs7975512, rs7038314, rs1904406, rs489381, rs7277342, rs7699090, rs10497233, rs10494870, rs9323490, rs1157492, rs1032355, rs552976, rs10495413, rs12430668, rs440431, rs6063178, rs12645879, rs1598452, rs2285369, rs7679675, rs7561808, rs7900067, rs1024487, rs7637803, rs217627, rs4787923, rs17457687, rs1400174, rs17439723, rs1435090, rs7819806, rs90213, rs2547116, rs3764796, rs9435, rs986642, rs2589654, rs16922018, rs1352411, rs1984282, rs2287101, rs4859259, rs1107820, rs1476162, rs959763, rs6511703, rs12500040, rs6021183, rs10868791, rs7501724, rs7163907, rs10853962, rs4336881, rs959567, rs6864793, rs10520475, rs11245545, rs17799799, rs518457, rs10509384, rs2366687, rs1003306, rs2367191, rs10504924, rs10507632, rs2728945, rs1905135, rs10516096

AIM que distinguen irlandeses de ingleses (y distinguen otras poblaciones del norte de Europa de otras diferentes):

rs354690, rs1200826, rs6133219, rs4937688, rs17595617, rs592229, rs7632151, rs10158939, rs6673923, rs11761305, rs2824976, rs13143572, rs11735755, rs1491610, rs1697143, rs12677633, rs2331295, rs12028395, rs1035218, rs11812285, rs17487804, rs2831507, rs11119274, rs2866904, rs3211663, rs2180046, rs1791648, rs1874777, rs11169282, rs10735745, rs11629910, rs7637171, rs1975944, rs17751406, rs3133719, rs7802273, rs4682664, rs1424341, rs17495345, rs2101869, rs622917, rs11234095, rs11579554, rs2574824, rs985246, rs7172022, rs17043135, rs3934253, rs6796183, rs9400660, rs6542847, rs9537662, rs4391081, rs921656, rs12621435, rs608082, rs969517, rs2279578, rs4741859, rs697387, rs9939450, rs300386, rs201930, rs6569343, rs6978997, rs2742059, rs2616982, rs7834280, rs17057633, rs4521229, rs3132069, rs13322103, rs4568821, rs2241466, rs1546963, rs1039621, rs11152343, rs7615026, rs1124602, rs7998134, rs1168553, rs8027997, rs7585767, rs7015718, rs4854502, rs1916977, rs11759031, rs12282752, rs924247, rs3115769, rs2489772, rs2066284, rs12589358, rs2284759, rs692612, rs1420109, rs17166207, rs6855114, rs2292702, rs150881, rs213759, rs7937428, rs150139, rs2029121, rs3805331, rs9366778, rs2389197, rs4242426, rs2154254, rs1345941, rs2120991, rs384267, rs10982246, rs283310, rs2832146, rs3936367, rs906671, rs11163055, rs725926, rs9353982, rs2251530, rs13032261, rs9981318, rs4561129, rs12193402, rs4854647, rs10045343, rs3937773, rs10016411, rs12618502, rs4857192, rs11706497, rs1361168, rs603089, rs7521746, rs3096700, rs1859218, rs6727787,

FIGURA 3

rs3798315, rs1812458, rs12327639, rs10496354, rs6593652, rs12202888,
rs12427378, rs126280, rs1496292, rs13281284, rs10197026, rs3764622,
rs632853, rs1334811, rs11258614, rs1432770, rs762324, rs10506479,
rs17583414, rs4889126, rs11635677, rs7538393, rs4607068, rs2148379,
rs1544908, rs1604528, rs10148024, rs5761163, rs1495085, rs6727258,
rs8718, rs1409590, rs13071295, rs12132550, rs4693421, rs1861532,
rs10896958, rs6738275, rs9348266, rs1925391, rs42985, rs516925,
rs11162802, rs4771561, rs12706465, rs13215804, rs4818059, rs201739,
rs13059876, rs17745881, rs3823518, rs4903820, rs2214020, rs6992907,
rs9556820, rs10249706, rs3798236, rs7701328, rs12229047, rs10846585,
rs936110, rs11851852, rs9300943, rs1926150, rs7872577, rs4868253,
rs6586395, rs4832680, rs1979689, rs17325754, rs2803353, rs2109302,
rs2840131, rs12645288, rs4241627, rs12065099, rs10511771, rs17264096,
rs3026886, rs569908, rs6749268, rs7591709, rs1450878, rs8015618,
rs17068823, rs4699356, rs6926578, rs10095717, rs11841049, rs17735123,
rs1918098, rs17158154, rs1928027, rs7685881, rs4130719, rs6742202,
rs1294875, rs12319392, rs9788522, rs6061663, rs12407446, rs6705555,
rs9551221, rs4688639, rs494791, rs11748140, rs10885355, rs4395073,
rs6824429, rs11204020, rs2644640, rs6455468, rs10163400, rs2005104,
rs907942, rs1007042, rs6452431, rs2901100, rs6799064, rs16880382,
rs1805313, rs12199314, rs1243974, rs10171287, rs352418, rs10953428,
rs4974389, rs7254735, rs7257916, rs639813, rs6431565, rs1435887,
rs10761660, rs10118378, rs12533837, rs2503848, rs12363937, rs6752254,
rs12906896, rs17537169, rs2268933, rs6939639, rs3775479, rs1691018,
rs17463958, rs6741326, rs7019271, rs4509106, rs4743564, rs13395560,
rs4751890, rs8058014, rs7946208, rs2023952, rs6066892, rs153645,
rs4443343, rs9297395, rs10895115, rs10419661, rs1833208, rs129074,
rs1889085, rs12127377, rs32441, rs10886671, rs1363926, rs13221445,
rs10501445, rs11811998, rs13225753, rs8034124, rs7906816, rs11189831,
rs11147671, rs2809823, rs6538980, rs500906, rs13182055, rs7143468,
rs4636972, rs10519510, rs7521399, rs278729, rs1468329, rs10515441,
rs566821, rs1882396, rs11906231, rs7909464, rs12030971, rs10047997,
rs17866606, rs170644, rs4129526, rs1869092, rs2149973, rs4660438,
rs12404676, rs16928653, rs11138141, rs11756439, rs3925053, rs10843894,
rs4259484, rs11772445, rs1598859, rs6476962, rs6676680, rs2836604,
rs4843426, rs12626954, rs4621050, rs1794275, rs1801212, rs10889272,
rs3847580, rs808868, rs10741808, rs2622892, rs17497293, rs7171366,
rs7685182, rs12280220, rs4944000, rs6030839, rs1437683, rs11000000,
rs1122821, rs6767356, rs10795942, rs10917268, rs605383, rs957881,
rs10780596, rs7720751, rs1928091, rs3760440, rs12129709, rs1259866,
rs17600119, rs17141252, rs10954006, rs137296, rs9623117, rs10517796,
rs1755833, rs17680913, rs4321843, rs10209427, rs7123425, rs17590228,
rs16957064, rs1941603, rs4403789, rs827628, rs6782694, rs2335629,
rs1405134, rs2651458, rs6478230, rs7152796, rs531539, rs1874494,

FIGURA 3

rs4843507, rs17273267, rs7320534, rs10868262, rs2113574, rs4814983,
 rs7587228, rs17624713, rs17508376, rs1997458, rs2176313, rs251867,
 rs1413710, rs7257183, rs1554808, rs1572228, rs7006668, rs338926,
 rs7135506, rs6427832, rs6766822, rs7591141, rs4781927, rs13106298,
 rs12874003, rs13071953, rs1910863, rs10433559, rs6092900, rs10196343,
 rs6448638, rs2110420, rs2871647, rs9873664, rs12634613, rs9465438,
 rs11767724, rs16999738, rs12488465, rs6992802, rs4902059, rs4308943,
 rs11183847, rs2887631, rs4278155, rs10486014, rs10795076, rs7204868,
 rs945704, rs3742673, rs10468473, rs897646, rs11916893, rs4897646,
 rs1523902, rs972124, rs10510254, rs2718324, rs4740358, rs12440104,
 rs6582999, rs4257073, rs2127247, rs17006704, rs11165281, rs11642715,
 rs4968938, rs10235162, rs2218935, rs9353512, rs13096852, rs8037001,
 rs992499, rs7305812, rs12643434, rs7094594, rs9330294, rs7798936,
 rs1008645, rs2272786, rs11757369, rs4886026, rs243099, rs1131620,
 rs10279772, rs7619189, rs10792322, rs11539202, rs2145449, rs11119145,
 rs12570141, rs3851384, rs527507, rs11590681, rs3804639, rs7584576,
 rs201328, rs201430, rs10097270, rs133074, rs1964196, rs2135091, rs16885,
 rs4394088, rs16943989, rs11055697, rs9646474, rs16854813, rs9290227,
 rs3928425, rs6949364, rs7586333, rs12091966, rs6831436, rs10083961,
 rs10932040, rs6798997, rs861079, rs12867481, rs6881628, rs753281,
 rs10496342, rs3117582, rs1463242, rs2408239, rs7744432, rs564968,
 rs7870926, rs609305, rs4887443, rs1931902, rs11692435, rs2048161,
 rs4684146, rs4628026, rs1930146, rs11735972, rs1866997, rs3745509,
 rs12412656, rs295117, rs11686077, rs10282703, rs10869031, rs2047171,
 rs2616612, rs16874127, rs12249377, rs845787, rs4072374, rs2552353,
 rs1377554, rs400688, rs1507599, rs4728490, rs10484867, rs845561,
 rs4334271, rs385773, rs4480424, rs2304572, rs11117791, rs10964808,
 rs9813516, rs2034781, rs11607746, rs626364, rs6536415, rs6090443,
 rs2328492, rs1379411, rs2583692, rs9938659, rs9315681, rs11622517,
 rs13359372, rs10143058, rs7590571, rs12064002, rs10835941, rs17134141,
 rs1343227, rs10077241, rs10132, rs2761291, rs6432594, rs10499748,
 rs6746287, rs7709399, rs12669805, rs11793897, rs17762729, rs12596290,
 rs11153162, rs6940057, rs1316883, rs7186783, rs6982890, rs937816,
 rs7725438, rs10504180, rs8094458, rs241529, rs980274, rs855395,
 rs10948327, rs12149783, rs12275853, rs2155854, rs4816612, rs2669331,
 rs497321, rs669336, rs1333592, rs2270834, rs6076248, rs9692165,
 rs4776567, rs4652192, rs6942609, rs6565261, rs1339287, rs631661,
 rs12345503, rs10491367, rs1348338, rs2876507, rs17069574, rs6416581,
 rs1370031, rs10174187, rs10783028, rs7904364, rs16932027, rs12993717,
 rs12568609, rs550010, rs4896634, rs6574859, rs1558160, rs1772579,
 rs1878424, rs288057, rs12647859, rs130539, rs7934426, rs7278046,
 rs1886690, rs1950221, rs17747739, rs13034680, rs3790116, rs6031215,
 rs10016081, rs966149, rs793096, rs8044334, rs16987794, rs12156286,
 rs2803183, rs7871735, rs1187763, rs858550, rs2301570, rs1010656,

FIGURA 3

rs1815619, rs2184925, rs10521057, rs11184219, rs2726675, rs10485678,
 rs7782814, rs489466, rs1093309, rs12439063, rs2396081, rs4887111,
 rs6979324, rs4233629, rs2189864, rs1267472, rs7006436, rs2839475,
 rs6874663, rs952579, rs10794696, rs11634215, rs1256992, rs6747820,
 rs8024685, rs1477798, rs8109759, rs6076954, rs17034929, rs4915886,
 rs2560301, rs6700849, rs1368406, rs2724031, rs2292719, rs1553664,
 rs3935801, rs4148298, rs4698844, rs4459724, rs1423089, rs12021671,
 rs16839940, rs8021670, rs4948235, rs13200910, rs308793, rs10924440,
 rs7329258, rs4780267, rs282152, rs10824914, rs16907003, rs7153598,
 rs3010888, rs6957465, rs12441839, rs10192315, rs727637, rs11786174,
 rs1488214, rs11635749, rs2729910, rs4869578, rs9459128, rs212434,
 rs1863471, rs12372169, rs472465, rs2908201, rs6010652, rs807704,
 rs6695712, rs1282557, rs10496954, rs2295682, rs9820762, rs12198061,
 rs11698898, rs2139747, rs700278, rs2192883, rs6312, rs1486731, rs2684796,
 rs12795982, rs2839589, rs2184221, rs9640302, rs4746826, rs632547,
 rs7764278, rs6651164, rs16750, rs13020203, rs9905820, rs1951797,
 rs10969030, rs10002315, rs4920104, rs7257017, rs9514044, rs2872338,
 rs3750111, rs1512155, rs4785173, rs2151842, rs2218724, rs1822917,
 rs1688015, rs11221731, rs887106, rs7809872, rs4532570, rs12554199,
 rs2773207, rs12314080, rs11881242, rs7225060, rs2518201, rs1010779,
 rs8048863, rs9299261, rs4586385, rs12642133, rs10777342, rs16901689,
 rs1859569, rs9300717, rs1154278, rs11174549, rs7140955, rs10108974,
 rs7852159, rs9419387, rs10831905, rs10498377, rs269850, rs9454096,
 rs6967514, rs13132569, rs12327342, rs11799558, rs1325156, rs2844651,
 rs2556097, rs2891409, rs12538916, rs8005808, rs2456220, rs6859290,
 rs924884, rs12501016, rs11644921, rs907238, rs2331175, rs2800257,
 rs717966, rs10825066, rs7112734, rs1524600, rs10491030, rs1387329,
 rs7996510, rs17461905, rs11839068, rs9312336, rs7204900, rs1672376,
 rs12677340, rs9575619, rs2765584, rs7205649, rs2839349, rs1454149,
 rs4907404, rs2166488, rs1230313, rs908822, rs2402800, rs913984,
 rs17651643, rs10979066, rs3024536, rs17173637, rs567384, rs9562077,
 rs10942957, rs4658702, rs7785682, rs2392147, rs613479, rs12710777,
 rs2719163, rs12094260, rs483241, rs12083651, rs12279152, rs7971883,
 rs17527048, rs2231895, rs512458, rs7321658, rs9923061, rs2298608,
 rs865151, rs11754258, rs270274, rs17567580, rs8021112, rs6506515,
 rs1553255, rs2492501, rs3781578, rs10780514, rs11695049, rs2022725,
 rs6959225, rs12796185, rs1369256, rs10063082, rs11966566, rs4427879,
 rs2395653, rs12018676, rs10154899, rs135551, rs9538094, rs10518452,
 rs17482565, rs2761384, rs10496557, rs7707491, rs943068, rs4713949,
 rs4233167, rs767249, rs12191488, rs6855792, rs1541335, rs2868146,
 rs6564940, rs1497305, rs1465010, rs11603501, rs10230087, rs899069,
 rs7864459, rs10517677, rs4708459, rs1545255, rs11634465, rs2284944,
 rs894392, rs10021708, rs7748977, rs9385824, rs7116354, rs11642445,
 rs3806300, rs4705295, rs2303802, rs791656, rs4387618, rs16972092,

FIGURA 3

rs4891825, rs11712075, rs1978368, rs16858789, rs1446394, rs2060530,
 rs157237, rs1482363, rs9502971, rs3757791, rs11768706, rs10888101,
 rs16936082, rs6585474, rs3857176, rs9510565, rs7787246, rs10954638,
 rs10968042, rs2700230, rs7652486, rs11602256, rs12492695, rs1197701,
 rs4918844, rs12646895, rs12377016, rs11995144, rs227906, rs4735339,
 rs7673403, rs7855536, rs760560, rs913259, rs9442235, rs11940490,
 rs3742076, rs9407554, rs7321548, rs1523250, rs5751691, rs2691244,
 rs6089457, rs849734, rs897111, rs9613630, rs1432053, rs2616667,
 rs10159191, rs12619236, rs7660552, rs12731666, rs2841455, rs6454198,
 rs1466835, rs637001, rs7594872, rs470490, rs1017104, rs10403583,
 rs7526907, rs10484854, rs1355829, rs12273605, rs691127, rs10930718,
 rs1543603, rs4936578, rs722113, rs3846456, rs3934711, rs11101312,
 rs10068902, rs11988550, rs4331742, rs12923978, rs582962, rs12247354,
 rs10923142, rs9912143, rs2468638, rs11543651, rs16997896, rs2827845,
 rs10112398, rs1412471, rs1012769, rs153785, rs4368537, rs7820789,
 rs3813135, rs1554690, rs4890891, rs2114088, rs2327506, rs445275,
 rs6842443, rs9423942, rs11593576, rs244050, rs1007018, rs2192754,
 rs945249, rs17048973, rs17088268, rs6105221, rs2798950, rs9855729,
 rs4808416, rs10883479, rs6729229, rs875569, rs2475229, rs2047153,
 rs10493770, rs10128640, rs7154324, rs17780310, rs1944362, rs3793504,
 rs12452093, rs4283980, rs2070713, rs11118045, rs12585880, rs7598637,
 rs10887047, rs915538, rs17293152, rs904269, rs2987380, rs717751,
 rs6765166, rs10486461, rs7858808, rs6599127, rs11136596, rs16943315,
 rs7336525, rs9366950, rs6073394, rs1908984, rs2280699, rs2238508,
 rs7460771, rs7502772, rs7038373, rs1020042, rs11791419, rs11780780,
 rs11111689, rs743082, rs17242960, rs10500956, rs10431150, rs1689803,
 rs12743521, rs11778372, rs17271148, rs1360646, rs8066468, rs11831940,
 rs2966487, rs7018906, rs11180273, rs544603, rs4786780, rs7658820,
 rs13068538, rs7963499, rs4902562, rs10515998, rs4331170, rs204538,
 rs666649, rs12691565, rs16876243, rs10504268, rs4397025, rs6793927,
 rs2735202, rs11172147, rs11632449, rs10733358, rs2048816, rs10942889,
 rs1475048, rs7698580, rs12504106, rs716953, rs12111723, rs7075768,
 rs11854173, rs10520656, rs316582, rs13295839, rs824453, rs4596664,
 rs9289539, rs10275488, rs4981674, rs4955713, rs2387698, rs17277113,
 rs250413, rs17590785, rs10462465, rs833010, rs7814885, rs10733333,
 rs6772196, rs1342744, rs10168477, rs10273, rs7221778, rs17187133,
 rs17827858, rs17279118, rs373531, rs12372697, rs10357, rs2822548,
 rs1387910, rs4859864, rs17319408, rs1265203, rs12904370, rs10518627,
 rs10050980, rs3782814, rs17820124, rs2475357, rs1795065, rs4414922,
 rs7487814, rs13194998, rs7807708, rs17660328, rs4471501, rs8077267,
 rs4304313, rs2387673, rs3846677, rs17012485, rs4234474, rs11731647,
 rs6712681, rs2058131, rs2270409, rs6089784, rs17698283, rs6494039,
 rs2937785, rs260445, rs9641312, rs4911442, rs4688104, rs905570, rs99521,
 rs12997609, rs504341, rs1748168, rs11196967, rs1791468, rs17455729,

FIGURA 3

rs1393861, rs4623797, rs4819591, rs1009728, rs17019119, rs4943731,
rs1036295, rs2793471, rs9283119, rs4248914, rs733044,

AIM que distinguen hispanos de caucásicos:

rs118003, rs166673, rs169088, rs261056, rs280181, rs286804, rs347413,
rs374672, rs382515, rs382558, rs455249, rs578363, rs600558, rs640330,
rs688858, rs707077, rs717251, rs717471, rs717614, rs718996, rs719628,
rs719750, rs720030, rs720229, rs721128, rs721246, rs721288, rs721361,
rs721362, rs721363, rs724130, rs724658, rs726512, rs726610, rs727923,
rs728506, rs741933, rs754291, rs831451, rs869538, rs894198, rs950132,
rs951572, rs951784, rs952020, rs952562, rs953111, rs955770, rs955771,
rs956193, rs956377, rs957862, rs958985, rs959763, rs961463, rs965892,
rs965917, rs967294, rs967906, rs979050, rs979322, rs988462, rs1016120,
rs1023534, rs1032143, rs1033042, rs1035701, rs1073354, rs1074149,
rs1074716, rs1074805, rs1075263, rs1075906, rs1108929, rs1113082,
rs1113144, rs1114000, rs1116138, rs1116853, rs1158619, rs1202645,
rs1210110, rs1331819, rs1343809, rs1359247, rs1363234, rs1363235,
rs1366233, rs1368387, rs1368697, rs1373904, rs1375575, rs1375989,
rs1376392, rs1377470, rs1378702, rs1380415, rs1382859, rs1384118,
rs1388175, rs1390476, rs1391961, rs1393736, rs1395162, rs1395604,
rs1399664, rs1400347, rs1402458, rs1404759, rs1405633, rs1410592,
rs1411445, rs1412171, rs1414149, rs1417503, rs1418706, rs1419702,
rs1420117, rs1449050, rs1451371, rs1454027, rs1478810, rs1479285,
rs1485254, rs1498458, rs1501626, rs1537371, rs1537523, rs1546561,
rs1548299, rs1549944, rs1572902, rs1579029, rs1585217, rs1585946,
rs1587120, rs1599710, rs1603681, rs1605727, rs1820072, rs1824683,
rs1830876, rs1854673, rs1900672, rs1917152, rs1917199, rs1927014,
rs1929646, rs1931621, rs1951095, rs1951207, rs1951507, rs1951519,
rs1961491, rs1986508, rs1990542, rs1992539, rs1999333, rs2008163,
rs2008927, rs2010392, rs2011812, rs2014303, rs2014533, rs2015963,
rs2017892, rs2017903, rs2028021, rs2040847, rs2040848, rs2057127,
rs2067084, rs2078426, rs2134897, rs2139246, rs2194127, rs2196302,
rs2210254, rs2216629, rs2226672, rs2351002, rs2355205, rs2358295,
rs2366101, rs2366977, rs2369860, rs2373584, rs2374545, rs2389774,
rs2409523, rs2418400, rs2419925, rs2865116, rs2868932, rs2886937,
rs2887280, rs2899875

AIM que distinguen asiáticos del sur, central y este:

rs32826, rs63319, rs170872, rs187145, rs187861, rs191644, rs205474,
rs219562, rs235946, rs266175, rs266177, rs303450, rs305851, rs353599,

FIGURA 3

rs355191, rs366220, rs409914, rs411301, rs428294, rs498046, rs632159, rs664574, rs714857, rs717205, rs717757, rs718137, rs719006, rs719866, rs719910, rs720376, rs720578, rs720844, rs721002, rs721407, rs721684, rs722559, rs722653, rs722711, rs723195, rs723212, rs723686, rs723698, rs723723, rs723794, rs724246, rs724247, rs724258, rs724259, rs725389, rs726108, rs726217, rs726472, rs726493, rs726692, rs727424, rs727708, rs728497, rs728618, rs728647, rs730991, rs791856, rs864736, rs896019, rs910552, rs923976, rs950198, rs950303, rs950400, rs951062, rs951199, rs951998, rs952503, rs952785, rs952976, rs952993, rs953624, rs953680, rs956070, rs956408, rs958478, rs958976, rs963096, rs963170, rs963171, rs966134, rs966285, rs967990, rs978486, rs978488, rs985933, rs985934, rs988913, rs1000313, rs1004246, rs1011873, rs1013063, rs1018616, rs1029554, rs1035076, rs1073109, rs1073119, rs1073985, rs1114000, rs1178328, rs1318710, rs1336672, rs1340846, rs1343981, rs1352514, rs1356763, rs1359932, rs1362987, rs1366267, rs1371063, rs1374499, rs1381331, rs1381972, rs1383965, rs1383970, rs1391221, rs1391961, rs1392096, rs1404501, rs1405567, rs1406012, rs1406218, rs1407043, rs1408830, rs1413724, rs1416615, rs1417503, rs1418102, rs1422004, rs1422323, rs1422895, rs1427356, rs1428702, rs1434193, rs1446966, rs1448229, rs1470248, rs1480951, rs1509421, rs1532517, rs1549690, rs1563353, rs1565138, rs1565534, rs1570699, rs1578243, rs1583355, rs1586149, rs1586597, rs1812624, rs1817186, rs1822811, rs1822812, rs1822813, rs1823100, rs1826873, rs1876545, rs1892114, rs1903189, rs1903746, rs1924257, rs1928045, rs1934008, rs1940006, rs1944086, rs1945146, rs1950284, rs1954119, rs1985835, rs2007397, rs2008011, rs2008686, rs2013708, rs2035799, rs2035801, rs2050137, rs2052291, rs2065443, rs2077811, rs2101551, rs2134897, rs2138212, rs2145231, rs2192874, rs2327790, rs2369898, rs2370672, rs2382193, rs2392002, rs2406519, rs2421172, rs2877455

AIM que distinguen poblaciones chinas y japonesas:

rs466256, rs1024052, rs6462942, rs6685064, rs244757, rs3924094, rs11087655, rs944664, rs11087655, rs2402820, rs2280100, rs11118295, rs721953, rs2377527, rs296364, rs10894532, rs2842186, rs6462942, rs4820389, rs4632285, rs4838113, rs16966142, rs1500127, rs4925258, rs11118295, rs2280100, rs766144, rs6462942, rs766144, rs2241880, rs1159114, rs4962061, rs5924101, rs7900633, rs11087655, rs715846, rs6817562, rs2185369, rs331332, rs4937787, rs2241880, rs12484697, rs4669614, rs11087655, rs2377527, rs5924101, rs4838113, rs2838665, rs12493799, rs12505641, rs3019711, rs5924101, rs1286041, rs4669614, rs614502, rs2280100, rs4362557, rs5924766, rs6937974, rs782758, rs10488533, rs614502, rs2240800, rs7003901, rs4362557, rs1858234,

FIGURA 3

rs2724863, rs1748944, rs1077834, rs1259807, rs12576800, rs2970869,
rs9308712, rs6855567, rs6457300, rs331332, rs2293072, rs4829981,
rs1077834, rs4947803, rs7016174, rs7991928, rs2271738, rs9397264,
rs4849159, rs1554656, rs1383887, rs12221798, rs1077834, rs2358521,
rs1383887, rs12025272, rs2763002, rs9397264, rs4669614, rs7660417,
rs2114716, rs4749159, rs1383887, rs542172, rs3805466, rs7016174,
rs7991928, rs9397264, rs16993741, rs7016174, rs1259807

Población 1

AA .25, AG .50, GG .25
...ACTG[A/G]GTAG...

Parámetros EM
(AA,AG,GG)

$$\begin{aligned}\pi_1 &= [0,25, 0,50, 0,25] \\ \mu_1 &= [[0,9, 0,1], [0,5, 0,3], [0,1, 0,7]] \\ \Sigma_1 &= [I, I, I]\end{aligned}$$

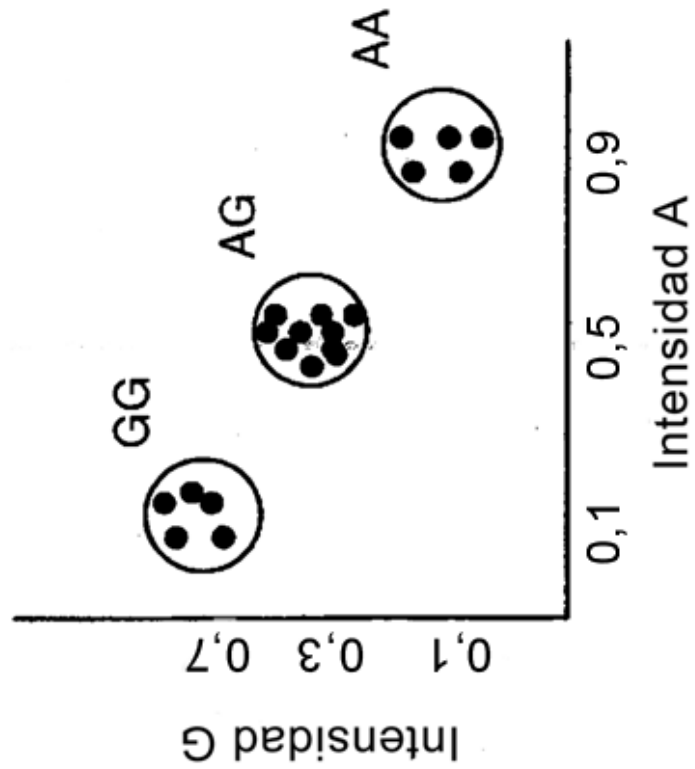


FIGURA 4

Población 2

AA .16, AG .48, GG .36
...ACTG[A/G]GTAG...

Parámetros EM
(AA,AG,GG)

$\pi_2 = [0,16, 0,48, 0,36]$
 $\mu_2 = [[0,9, 0,1], [0,5, 0,3], [0,1, 0,7]]$
 $\Sigma_2 = [I, I, I]$

Comparación con población 1

$\pi_1 \neq \pi_2$
 $\mu_1 = \mu_2$
 $\Sigma_1 = \Sigma_2$

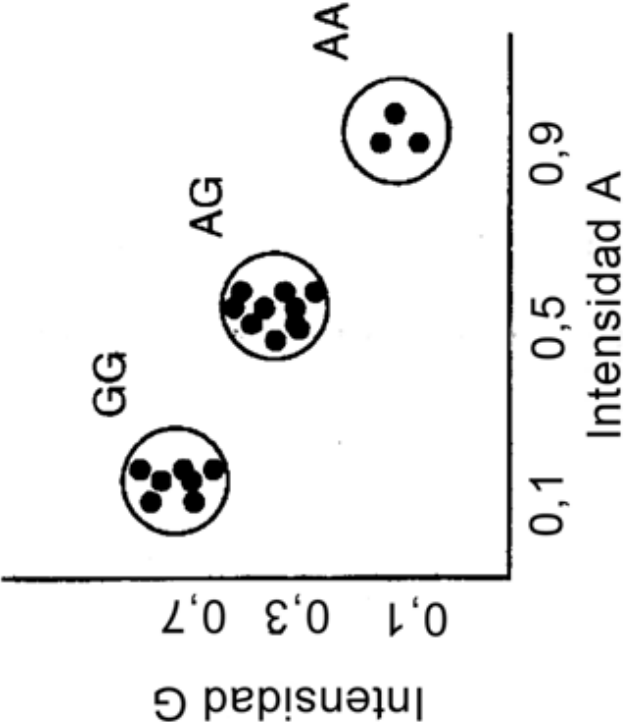


FIGURA 5

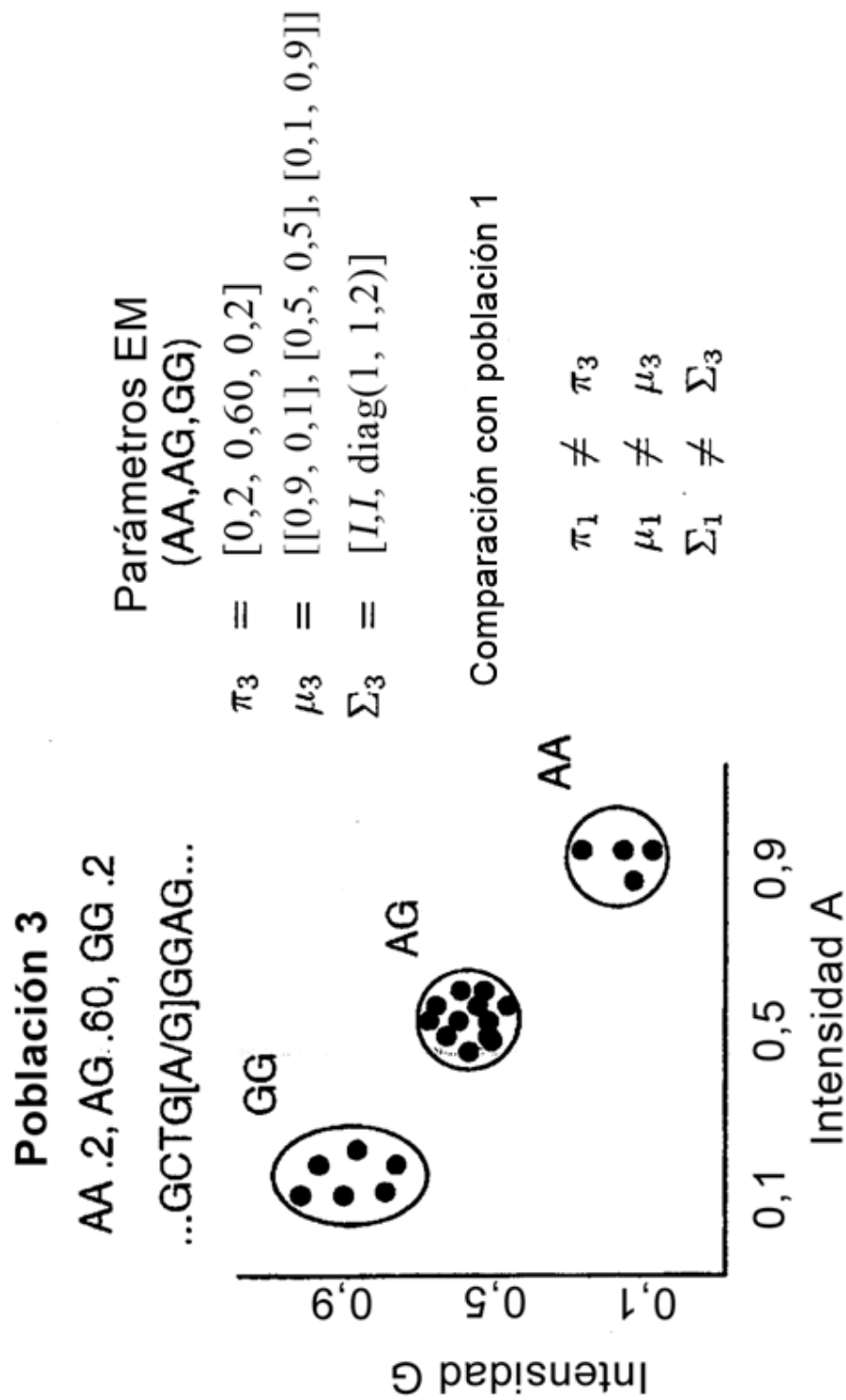


FIGURA 6

Población 4
A- 0,50, G- 0,50
...ACTG[AIG/-]G TAG...

Parámetros EM
(AA,AG,GG)

$\pi_4 = [0,50, 0,50]$
 $\mu_4 = [[0,6, 0,1], [0,1, 0,6]]$
 $\Sigma_4 = [\text{diag}(1,2, 1), \text{diag}(1, 1,2)]$

Comparación con población 1

$\pi_1 \neq \pi_4$
 $\mu_1 \neq \mu_4$
 $\Sigma_1 \neq \Sigma_4$

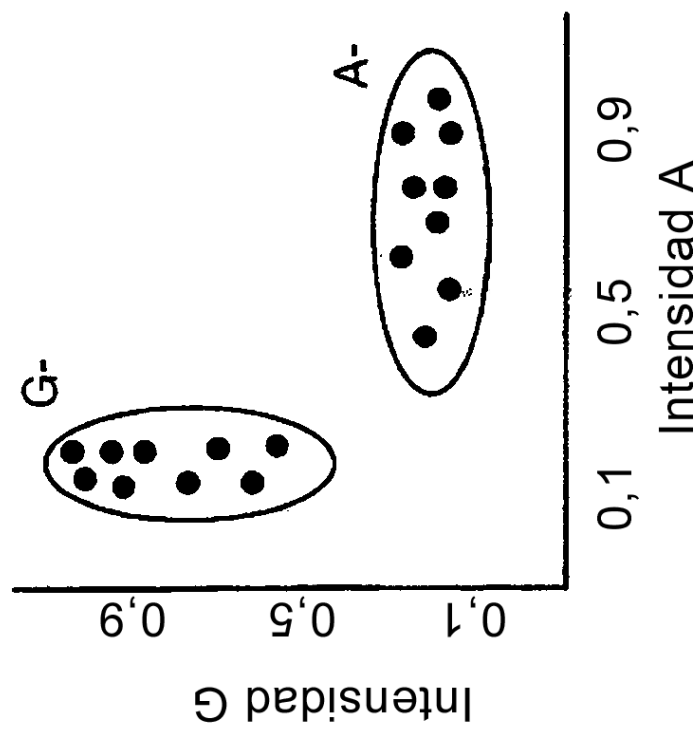
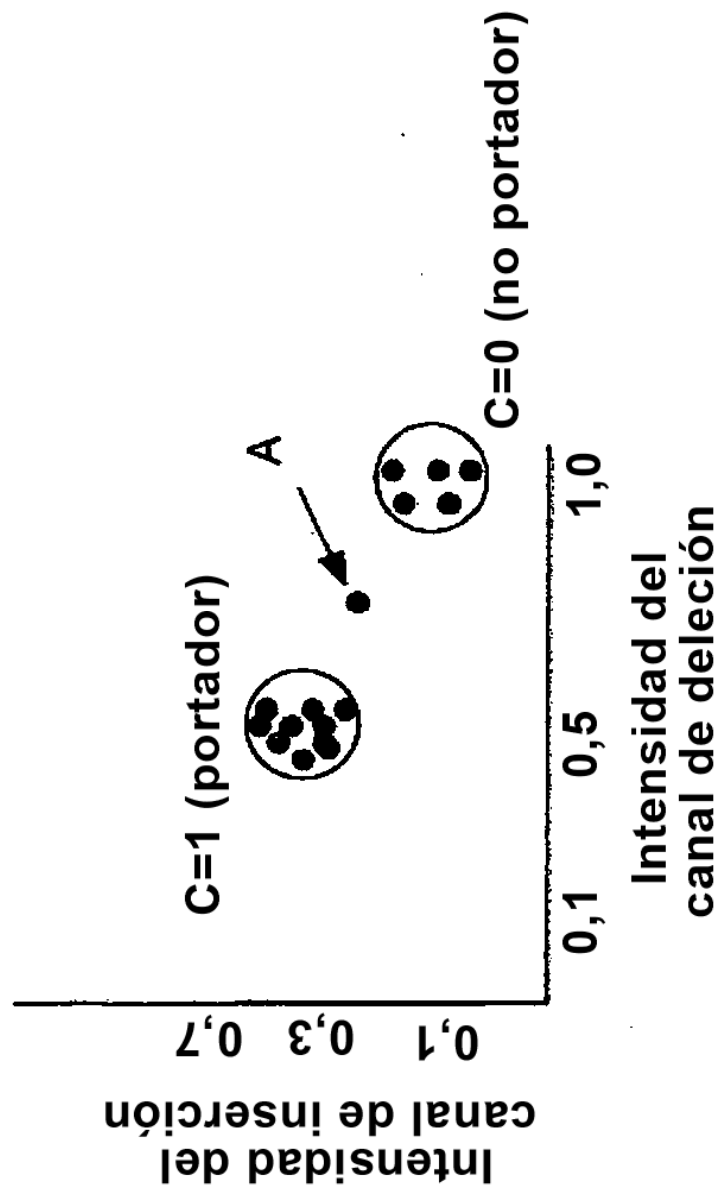


FIGURA 7

FIGURA 8

Diagrama de dispersión de intensidad
para el alelo de delección de beta-talasemia



Resolución de designaciones de frontera
con el uso de la información de ascendencia

Variables aleatorias

- $C \in 0, 1$ 0 si no portador, 1 si portador
- $X \in \mathbb{R}^2$ Vector de intensidad bidimensional
- $A \in 0, 1$ Italiano (0), europeo del norte (1)

Italiano	Europeo del norte
$\pi_{C=1 A=0} = P(C = 1 A = 0)$	$\pi_{C=1 A=1} = P(C = 1 A = 1)$
$\pi_{C=0 A=0} = P(C = 0 A = 0)$	$\pi_{C=0 A=1} = P(C = 0 A = 1)$
$\mu_{C=1 A=0} = [0,5, 0,5]$	$\mu_{C=1 A=1} = [0,5, 0,5]$
$\mu_{C=0 A=0} = [0, 0]$	$\mu_{C=0 A=1} = [0, 0]$
$\Sigma_{C=1 A=0} = I$	$\Sigma_{C=1 A=1} = I$
$\Sigma_{C=0 A=0} = I$	$\Sigma_{C=0 A=1} = I$

FIGURA 9

Probabilidad condicional de ser un portador (C=1) dada la intensidad medida (X=x) y la ascendencia conocida (A=0, italiano)

$$P(C = 1|X = x, A = 0) = \frac{P(X = x|C = 1)P(C = 1|A = 0)}{P(X = x|C = 1)P(C = 1|A = 0) + P(X = x|C = 0)P(C = 0|A = 0)}$$

Valores ejemplares si $X=[0,25, 0,25]$, equidistante entre los dos centro de grupo

$$\begin{aligned} P(C = 1|X = x, A = 0) &= 0,047 \\ P(C = 1|X = x, A = 1) &= 0,00077 \\ \frac{P(C = 1|X = x, A = 0)}{P(C = 1|X = x, A = 1)} &\approx 61 \end{aligned}$$

FIGURA 10

```

Código fuente N° R para el ejemplo
Biblioteca (mvtnorm)
pi1 = 0,06
pi2 = °-pi1
mu1 = c(0,1)
mu2 = c(0,5, 0,5)
sigma1 = matriz(c(1,0,0,1),2,2)
sigma2 = sigma 1
x = c(0,25, 0,25)

t1 = dmvnorm(x, mu1, sigma1)*pi1
t2 = dmvnorm(x,mu2,sigma2)*pi2
p.italiano = t1/(t1+t2)
n°[1]0,04735656

pi1 = 0,001
pi2 = 1-pi1
t1 = dmvnorm(x,mu1,sigma1)*pi1
t2 = dmvnorm(x,mu2,sigma2)*pi2
p.cauc = t1/ (t1+t2)
N°[1] 0,0007789731

```

FIGURA 11

La información de población probabilística (por ejemplo proporcionada por AIM) puede incorporarse por una simple modificación a esta ecuación

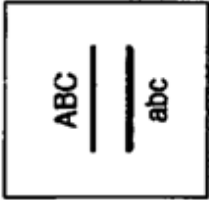
- $C \in 0, 1$ 0 si no portador, 1 si portador
- $X \in \mathbb{R}^2$ Vector de intensidad de bidimensional
- $Y \in \mathbb{R}^n$ Vector de intensidad de n-dimensional
 con todos los datos de AIM
- $A \in 0, 1$ Italiano (0), europeo del norte (1)

$$P(C = 1|X = x, Y = y) = \frac{\sum_{a=0,1} P(X = x|C = 1)P(C = 1|A = a)P(A = a|Y = y)}{\sum_{c=0,1} \sum_{a=0,1} P(X = x|C = c)P(C = c|A = a)P(A = a|Y = y)}$$

$P(A = a|Y = y)$ Este término es la probabilidad condicional con que esta región del genoma se hereda desde un individuo de ascendencia A=a dada la intensidad de señal de AIM Y=y

FIGURA 12

1) Genotipo real
organizado de la madre



Genotipado en chip
→
1) Elimina la fase
2) introduce errores
de falso positivo,
falso negativo

2) Genotipos
medidos de la madre

Genotipo medido			Rep. Binaria		Probabilidad
Aa	Bb	Cc	1	1	$(1 - \beta)^3$
Aa	Bb	CC	1	1	$\beta(1 - \beta)^2$
Aa	BB	Cc	1	0	$\beta(1 - \beta)^2$
Aa	BB	CC	1	0	$\beta^2(1 - \beta)$
AA	Bb	Cc	0	1	$\beta(1 - \beta)^2$
AA	Bb	CC	0	1	$\beta^2(1 - \beta)$
AA	BB	Cc	0	0	$\beta^2(1 - \beta)$
AA	BB	CC	0	0	β^3

Tasa de falso positivo: β (probabilidad de alelo de riesgo de designación como sano)
Tasa de falso negativo: α (probabilidad de alelo sano de designación como de riesgo)

FIGURA 13

3) Haplotipos inferidos de la madre

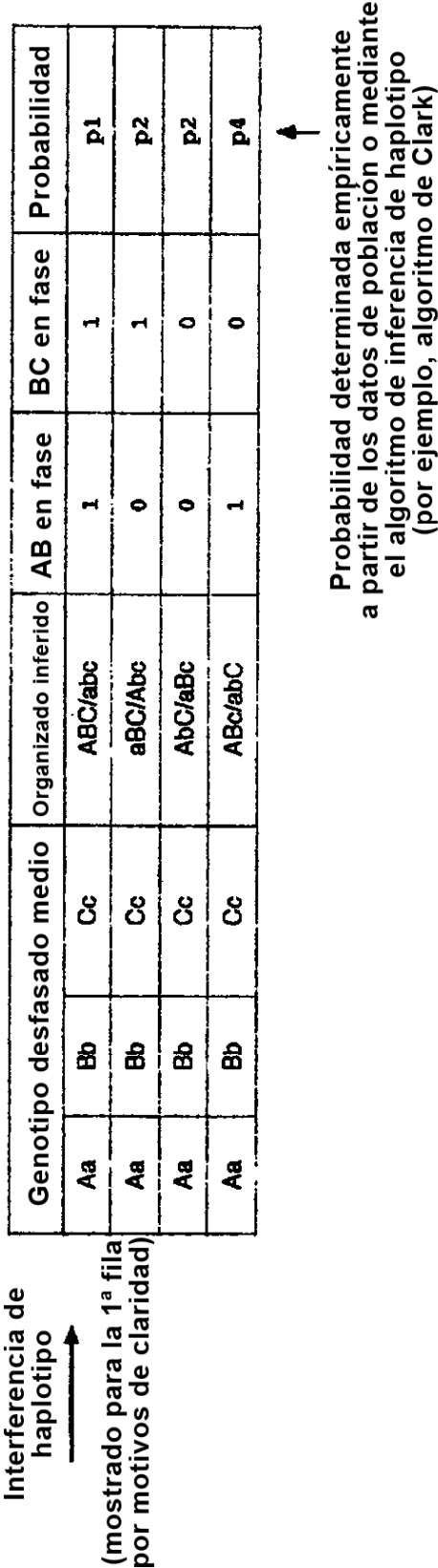


FIGURA 14

4) Gametos posibles de la madre

4a) Tabla de probabilidad de recombinación

Fase inferida	Recomb. AB	Recomb. BC	Gametos	Probabilidad
ABC/abc	0	0	ABC, abc	$(1 - r_{AB})(1 - r_{BC})$
ABC/abc	0	1	ABc, aBC	$(1 - r_{AB})(r_{BC})$
ABC/abc	1	0	AbC, aBC	$(r_{AB})(1 - r_{BC})$
ABC/abc	1	1	Abc, aBc	$(r_{AB})(r_{BC})$

Generación de gametos
oara haplotipo inferido

→

1) mostrada para la 1ª fila
2) Uso de distancias de
mapa genético de
Hapman/NCBI

4b) Tabla de probabilidad de gametos

Gametos	Probabilidad
ABC	$.5(1 - r_{AB})(1 - r_{BC})$
ABC	$.5(1 - r_{AB})(r_{BC})$
AbC	$.5(r_{AB})(1 - r_{BC})$
AbC	$.5(r_{AB})(r_{BC})$
abc	$.5(1 - r_{AB})(1 - r_{BC})$
abc	$.5(1 - r_{AB})(r_{BC})$
abc	$.5(r_{AB})(1 - r_{BC})$
abc	$.5(r_{AB})(r_{BC})$

FIGURA 15

5) Uniones de gametos de los niños =
posibles genotipos organizados

4) Gametos posibles de la madre

Muestra de combinaciones
de gametos para generar
posibles niños; aquí se
supone que la madre y el
padre tienen ambos las
mismas combinaciones de
gametos

4) Gametos posibles del padre

Gametos para padre/madre	ABC	ABC	Abc	Abc	Abc	Abc	abc	abc	abc
ABC									
ABc									
Abc									
AbC									
abc									
abC									
aBC									
aBc									

$$P_{ABC, \text{madre}} \times P_{ABc, \text{padre}} = [.5(1 - r_{AB})(1 - r_{BC})] \times [.5(1 - r_{AB})r_{BC}]$$

Ejemplo de
probabilidad de genotipo organizado
del niño = multiplicar la probabilidad
de muestreo de estos dos gametos

FIGURA 16

Interpretación de fenotipo	Valor de fenotipo (condicional en geno)	Genotipo del niño
Tipo silvestre	0,0	AABCC
Portador	1,0	AABcC
Portador	1,0	AABCa
Afectado	2,0	aaBBCC

6b) Otro mapa geno/feno: mirar la tabla sobre haplotipos de madre/padre

Interpretación de fenotipo	valor de fenotipo (condicional en haplos)	Haplotipo de la madre 1	Haplotipo del padre 2
Tipo silvestre	0,0	ABC	ABC
Portador	1,0	ABC	ABc
Portador	1,0	ABc	ABC
Afectado	2,0	aBc	aBc

6c) Mapa geno/feno más complejo: distribución condicional dados los haplotipos de madre/padre

Nota: en este ejemplo, cada alelo en letra minúscula desplaza a la derecha la distribución hacia más altura
Nota adicional: por simplicidad, se representa la distribución condicional aquí como un vector de 101 valores puntuales flotantes

Valor de fenotipo (por ejemplo, altura) (percentil 0 = min)	Valor de fenotipo percentil 1	Valor de fenotipo percentil 99	Valor de fenotipo percentil 100 (=max)	Haplotipo de la madre 1	haplotipo del padre 2
50in	53in	77in	79in	ABC	ABC
52in	54in	78in	80in	ABC	ABc
53in	55in	79in	80in	ABc	ABC
54in	56in	80in	91in	aBc	aBc

6d) Finalmente, el mapa geno/feno conceptualmente más general es una función que da la distribución condicional del fenotipo dado el genotipo del niño "o" los haplotipos de la madre y el padre para el niño

FIGURA 17

Cálculo de riesgo totalmente probabilístico
Justifica todas las fuentes de incertidumbre

Variables aleatorias

Y_c = Fenotipo del niño
 H_c = Haplotipo del niño
 G_m = Gameto de la madre
 R_m = Recombinaciones de la madre
 H_m = Haplotipos de la madre
 $\hat{X}_m$ = Genotipo inferido de la madre
 X_m = Genotipo real de la madre
 (Sustituir un subíndice f para el padre)

La expresión completa para la distribución sobre el fenotipo estimado del niño
 (por ejemplo, enfermedad ri dados los genotipos de los padres es:

$$P(Y_c | X_m, X_f) = \sum_{h_c} \sum_{g_m, g_f} P(Y_c | H_c) P(H_c | G_m, G_f) P(G_m | X_m) P(G_f | X_f)$$

$P(Y_c | X_m, X_f)$ = CD del fenotipo del Niño dados los genotipos reales de los padres
 $P(Y_c | H_c)$ = CD del fenotipo del niño dados los haplotipos del niño
 $P(H_c | G_m, G_f)$ = CD de los haplotipos del niño dados los gametos de la madre/padre
 $P(G_m | X_m)$ = CD de los gametos de la madre dado el genotipo real de la madre
 $P(G_f | X_f)$ = CD de los gametos del padre dado el genotipo real del padre

La subexpresión para la distribución sobre los posibles gametos de la madre
 dado su genotipo real se expande como:

$$P(G_m | X_m) = \sum_{r_m} \sum_{h_m} P(G_m | R_m) P(R_m | H_m) P(H_m | \hat{X}_m) P(\hat{X}_m | X_m)$$

$P(G_m | R_m)$ = CD de gametos dados recombinantes
 $P(R_m | H_m)$ = CD de recombinantes dado haplotipo
 $P(H_m | \hat{X}_m)$ = CD de haplotipos dado genotipo estimado
 $P(\hat{X}_m | X_m)$ = CD de genotipo estimado dado genotipo real

La expresión correspondiente para el padre puede obtenerse sustituyendo f por m. **FIGURA 18**

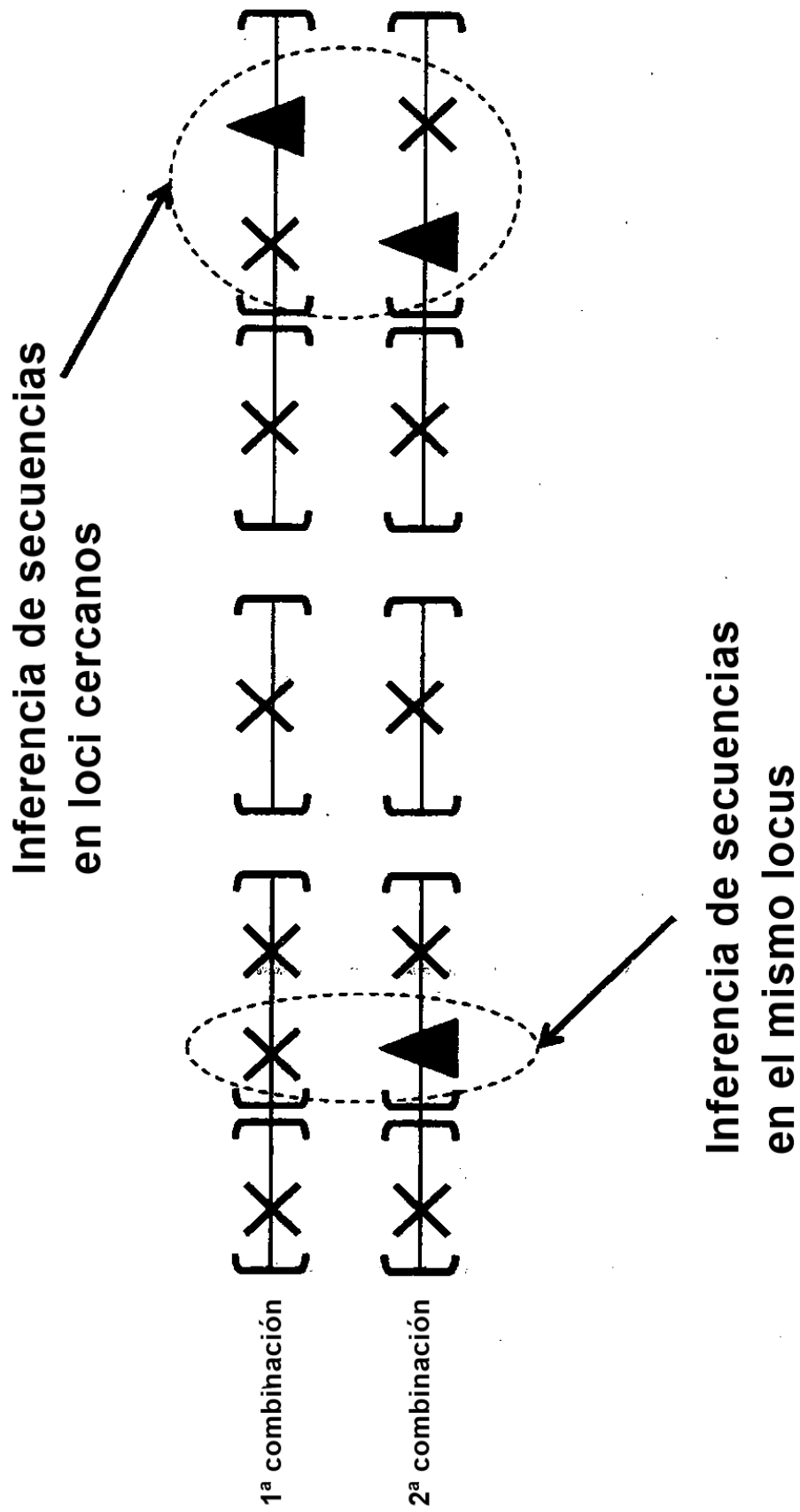


FIGURA 19

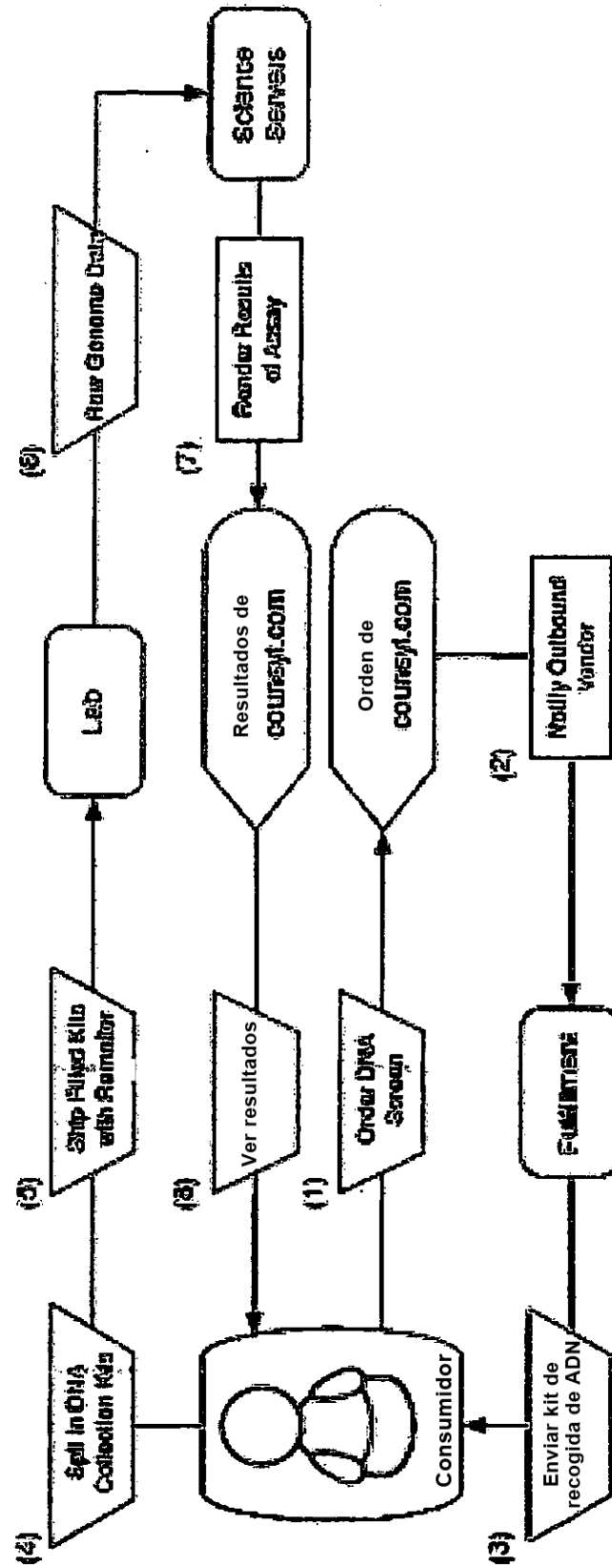


FIGURA 20

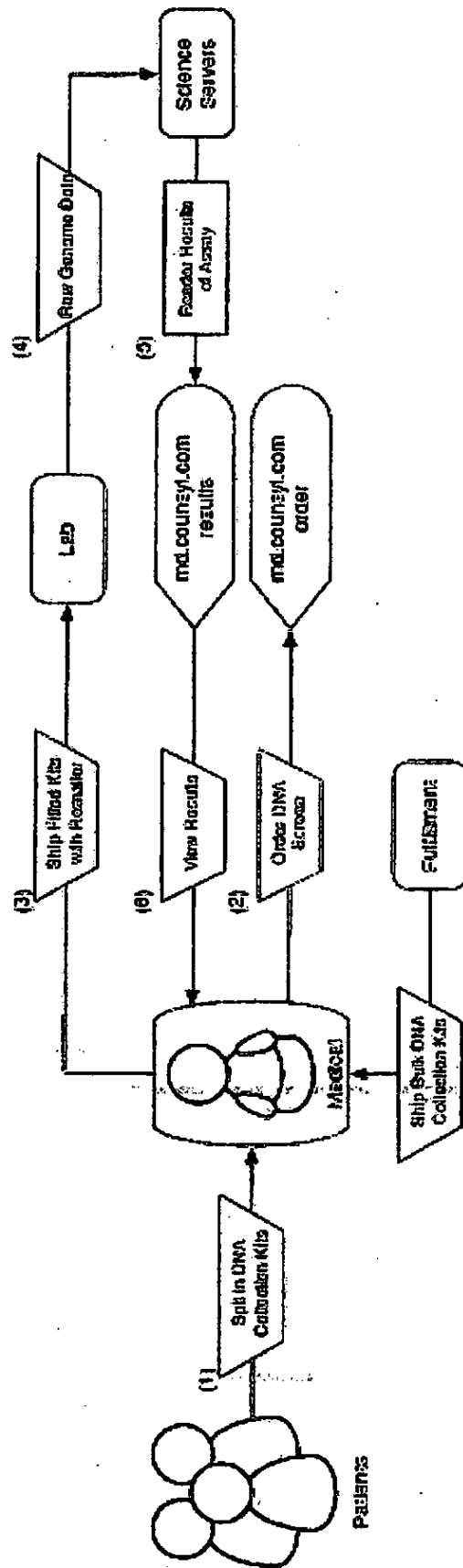


FIGURA 21

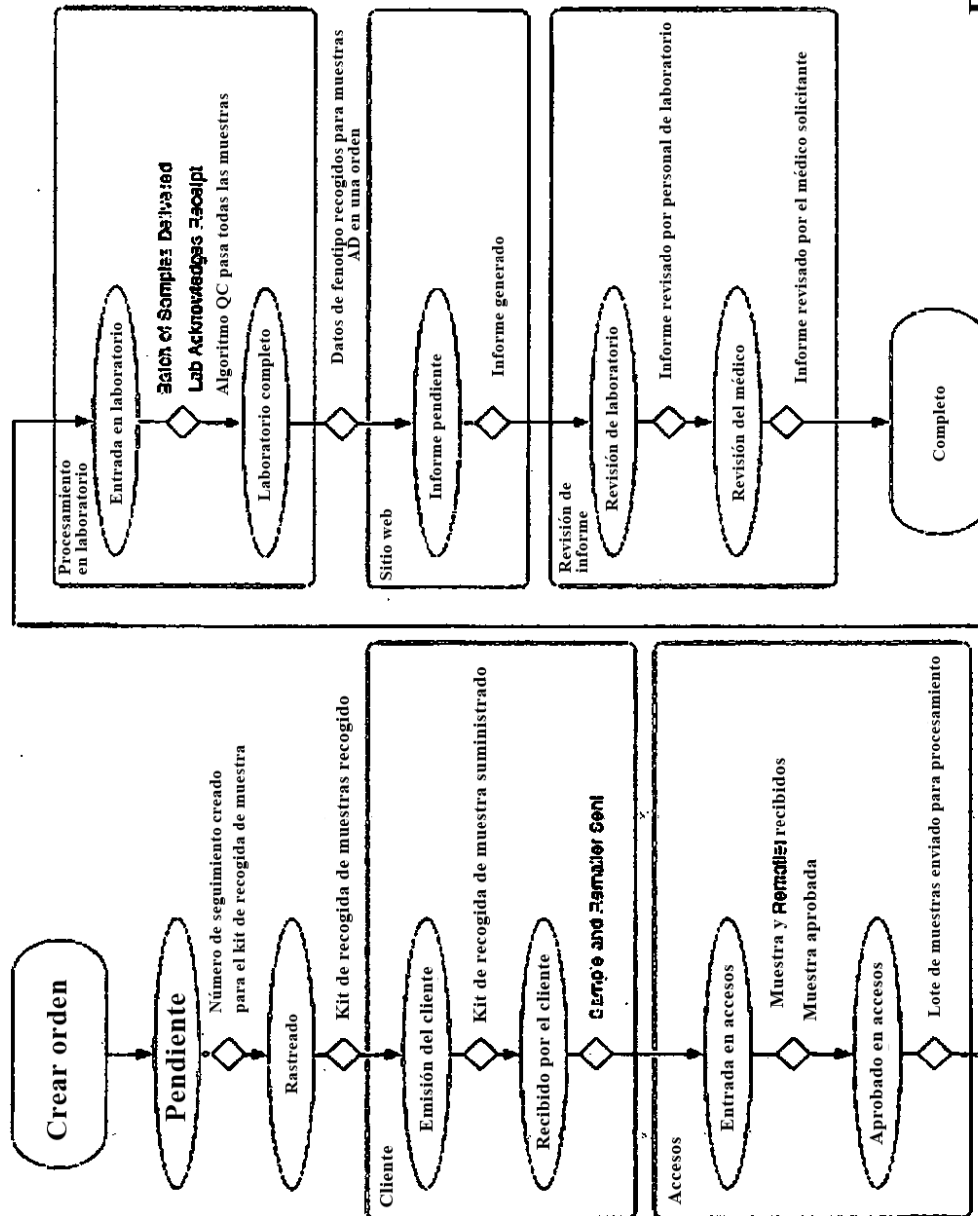


FIGURA 22

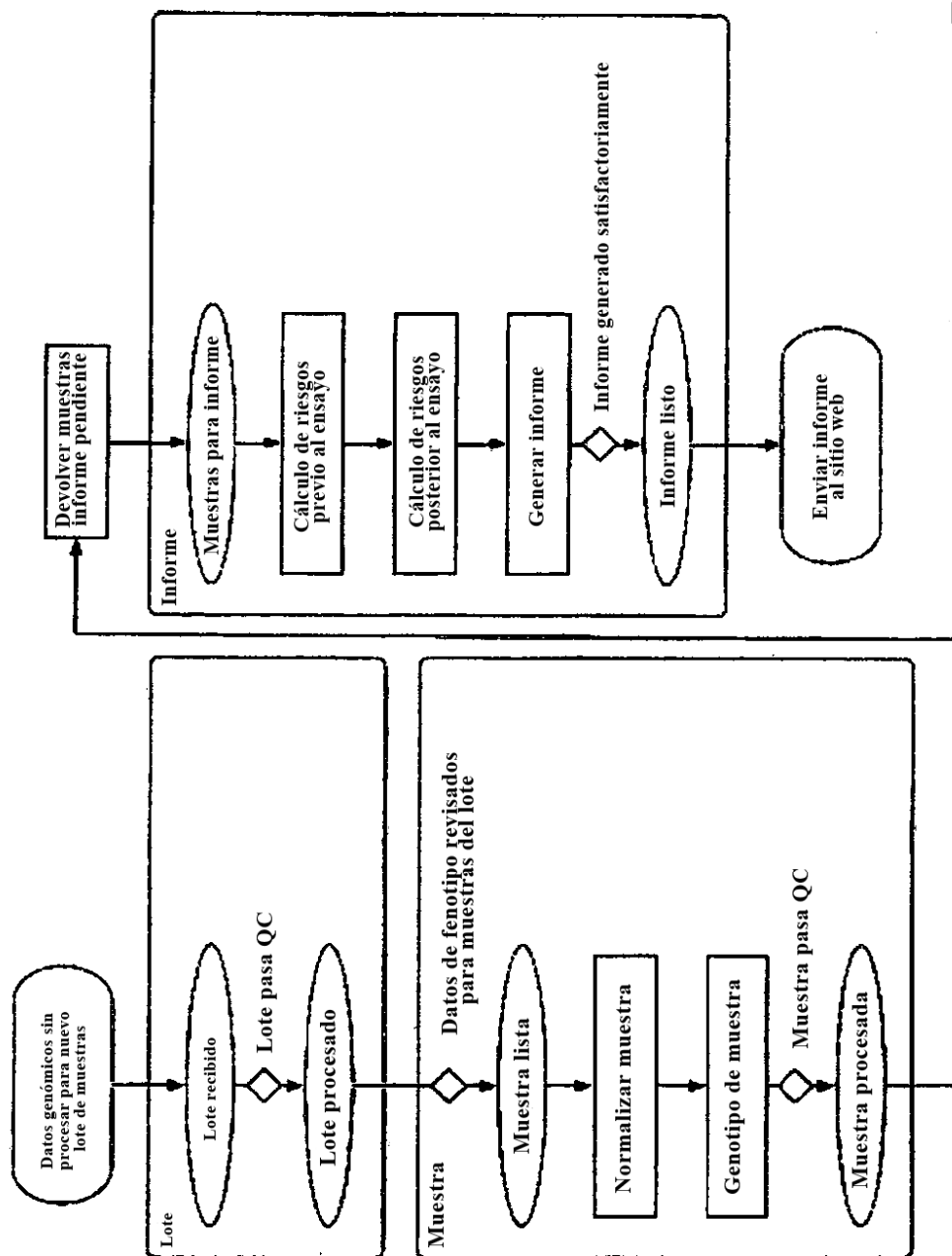


FIGURA 23

Predicción del niño usando genotipos y fenotipos de los padres, y opcionalmente geno/feno relativo

[illegible]

FIGURA 24