

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 759 023**

51 Int. Cl.:

G16B 20/00 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **19.07.2013 PCT/US2013/051331**

87 Fecha y número de publicación internacional: **23.01.2014 WO14015273**

96 Fecha de presentación y número de la solicitud europea: **19.07.2013 E 13742577 (3)**

97 Fecha y número de publicación de la concesión europea: **04.09.2019 EP 2888688**

54 Título: **Genotipificación completa de FMR1**

30 Prioridad:

20.07.2012 US 201261674167 P

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

07.05.2020

73 Titular/es:

**ASURAGEN, INC. (100.0%)
2150 Woodward Street
Austin, TX 78744, US**

72 Inventor/es:

**CHOUDHARY, ASHISH;
LATHAM, GARY y
SAH, SACHIN**

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 759 023 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

Genotipificación completa de FMR1

5 **[0001]** Esta solicitud reivindica el beneficio de prioridad bajo 35 USC § 119 para la Solicitud Provisional de Estados Unidos N° 61 / 674.167, presentada el 20 de julio de 2012.

LISTADO DE SECUENCIAS

10 **[0002]** La presente solicitud contiene una lista de secuencias.

[0003] La presente descripción se refiere en general a un aparato y un software para el análisis de ácidos nucleicos automatizada, en particular, a métodos para analizar y regiones de genotipificación en composiciones de ácido nucleico, tales como productos de ácido nucleico aisladas, fragmentos, y plantillas, que contienen repeticiones de nucleótidos.

[0004] Loci genéticos que comprenden regiones de repeticiones de nucleótidos (por ejemplo, repeticiones de dinucleótidos, repeticiones de trinucleótidos, etc.) son comunes en el genoma humano o animal. Loci genéticos que tienen contenido GC enriquecido también son comunes. En algunas circunstancias, la expansión de las regiones ricas en GC, o la expansión de repeticiones de nucleótidos, puede estar asociada con diversos estados de enfermedad. Por ejemplo, la expansión de repeticiones CGG en la región no traducida 5 (UTR) de la X Retardo Mental-1 gen Frágil (*FMR1*), localizado en el cromosoma X, está asociado con el síndrome del cromosoma X frágil (FXS) y trastornos relacionados y fenotipos. En la mayoría de las personas, el trinucleótido CGG se repite aproximadamente 5-44 veces en la región no traducida 5 (UTR) del gen *FMR1* ("región de repetición" CGG). Ampliaciones en esta región a mayor que alrededor de 45 repeticiones CGG, y en particular a más de aproximadamente 200 repeticiones CGG, se han asociado con SXF. fenotipos FXS pueden incluir retraso mental, autismo, ansiedad y otros trastornos cognitivos o de comportamiento. (. J. Mol Diag 10 (6): 496-501 (2008)). Del mismo modo, la expansión de la repetición región CCG trinucleótido ("región de repetición CCG") en UTR del gen 5' *FMR2* se asocia con discapacidad intelectual ligados al cromosoma X, y en particular con síndrome X Frágil E (FRAXE). FRAXE es una forma común de retraso mental ligado al cromosoma X.

[0005] Métodos para la genotipificación de secuencias ricas en GC y secuencias que comprenden repeticiones de nucleótidos, tales como las repeticiones de CGG y CCG en *FMR1* y *FMR2*, incluyen estrategias de enzima de restricción de digestión y reacción en cadena de polimerasa (PCR). El análisis de digestión de restricción puede proporcionar una medida bruta del tamaño de una región de repetición de triplete. Sin embargo, el análisis de digestión de restricción se limita en la resolución, no detecta fácilmente interrupciones cortas (tales como interrupciones AGG solo codón dentro de una región de repetición CGG) y no se puede determinar el estado de metilación.

[0006] Estrategias de PCR pueden proporcionar una mayor precisión en la reconstrucción de diversos genotipos. Sin embargo, existen limitaciones en la amplificación y secuenciación de los loci genéticos que comprenden secuencias repetidas largas o contienen secuencias ricas en GC que dificultan la capacidad de genotipos de reconstrucción para estos loci. Los esfuerzos para optimizar los procedimientos de PCR para el análisis de las repeticiones CGG en *FMR1*, por ejemplo, se han intentado, e incluyen modificaciones en las condiciones convencionales del ensayo de PCR. (Ver Res genoma. 6 (7): 633-8, (1996); J. Mol. Diag. 8: 544-550, (2006); y Am. J. Med. Gineta. 51 (4): 527-34, (1994)). Más recientemente, las técnicas de PCR se han desarrollado que permitir la amplificación más fiable de loci genómicos que tienen más de 200 CGG CCG o repeticiones. Ver Solicitud de Estados Unidos N° 2010/0209970, 2010/0243451, y 2012/0107824, que describen métodos de PCR para la secuenciación de regiones de repetición de GC-enriquecido. Estos métodos a menudo proporcionan información que caracteriza una región de repetición, tal como la longitud de la región, el número de repeticiones dentro de la región, y la presencia de cualquier interrupción dentro de la región de repetición. Sin embargo, la PCR sola no permite un rápido o de alta reconstrucción rendimiento del genotipo completo de la información de parámetro que caracteriza la región rica en GC, en especial cuando la región contiene interrupciones cortas, tales como las secuencias individuales codón AGG observados en el *FMR1* 5' UTR.

[0007] Estudios recientes han indicado que las interrupciones AGG dentro de la región de repetición CGG en el 5' UTR de *FMR1* pueden conferir la estabilidad del ADN y pueden reducir el riesgo de expansión en la región de repetición de triplete en la descendencia que se asocia con FXS y trastornos relacionados, incluyendo síndrome de temblor/ataxia asociado al X frágil (FXTAS), insuficiencia ovárica primaria relacionada con X frágil-(FXPOI) y parkinsonismo sensible a dopamina. (Ver Eichler et al., Nat. Gineta. 8: 88-94 (1994); Nolin et al., Am. J. Hum. Gineta. 72: 454-64 (2003); Yrigollen et al., Genet. Med., en prensa). Aunque es posible reconstruir genotipos, tales como la región de repetición CGG en el 5' UTR de *FMR1*, mediante interpretación manual, este proceso puede ser complejo y requiere mucho tiempo. Por lo tanto, un método para un genotipado rápido, preciso y de alto rendimiento de regiones que tienen repeticiones de nucleótidos o alto contenido de GC, tales como la región de repetición CGG en el 5' UTR de *FMR1* y la región de repetición CCG en UTR de 5' *FMR2*, podría ofrecer beneficios diagnósticos. Por ejemplo, la automatización de la interpretación de los datos podría mejorar el flujo de trabajo en el laboratorio de diagnóstico al permitir la validación rápida de los datos de ensayo y asegurando sólidos y precisos resultados de la secuenciación. Liangjing Chen et al, en un artículo titulado " An Information-Rich CGG Repeat Primed PCR that detects the full range of fragile

X expanded alleles and minimizes the need for southern blot analysis" en el Journal of Molecular Diagnosis, vol. 12, nº 5, 1 de septiembre de 2010 describe un método para la amplificación por PCR de regiones de repetición CGG en el gen *FMR1* utilizando un conjunto de tres cebadores de PCR y un protocolo de electroforesis capilar. Esto genera los datos relativos a la longitud de repetición CGG total y presencia de interrupciones AGG. No se hace uso de estos datos.

[0008] En consecuencia, en el presente documento se divulgan métodos para la reconstrucción automatizada de un genotipo que comprende una o más regiones ricas en GC o una o más regiones de repetición de nucleótidos.

[0009] De acuerdo con la invención, se proporciona un método de reconstrucción automatizada de un genotipo, que comprende (a) proporcionar un ácido nucleico aislado de una muestra del paciente, donde el ácido nucleico tiene al menos una región de repetición de nucleótidos rica en GC que comprende una o más interrupciones; (B) determinar información de parámetros para el ácido nucleico en el que determinar información de parámetros comprende la medición de la longitud total de la al menos una región de repetición de nucleótidos rica en GC, la distancia desde el comienzo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región, y la distancia desde el extremo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región; y (c) utilizar un aparato para aplicar enumeración exhaustiva de la información de parámetros para generar un genotipo reconstruida, en el que la enumeración automatizada exhaustiva comprende

i) utilizar la longitud total de al menos una región de repetición de nucleótidos rica en GC y, o bien la distancia desde el inicio de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región o la distancia desde el extremo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región para generar un conjunto de genotipos posibles que comprenden todas las disposiciones posibles de la una o más interrupciones en la región; y ii) evaluar el conjunto de genotipos potenciales para determinar una solución de genotipo que satisface toda la información del parámetro.

En ciertas realizaciones, el aparato utilizado para llevar a cabo la enumeración automatizada exhaustiva comprende un procesador y un dispositivo de memoria acoplado comunicativamente al procesador, en el que el dispositivo de memoria ha almacenado instrucciones en la misma máquina ejecutable que, cuando son ejecutadas por el procesador, hacen que el procesador reciba información de parámetros y lleve a cabo un análisis de enumeración exhaustiva. En algunas realizaciones, el genotipo reconstruido se puede confirmar comparando manualmente el genotipo para la información de parámetros, mediante la realización de una digestión de restricción, o mediante la secuenciación del ácido nucleico que tiene al menos una región de repetición de nucleótidos rica en GC.

[0010] En diversas realizaciones, los métodos descritos anteriormente pueden ser utilizados para reconstruir un genotipo cuando un genotipo de los padres del paciente no se conoce.

[0011] En diversas realizaciones, los métodos descritos anteriormente pueden ser utilizados para reconstruir un genotipo para los genes *FMR1* o *FMR2* o fragmentos de los mismos, o las regiones de repetición CGG y CCG de aquellos genes y fragmentos. En estas realizaciones, la determinación de la información de parámetros comprende la determinación de la longitud total de al menos una región de repetición CGG o CCG, la distancia desde el inicio de la región de repetición al menos una CGG o CCG a una o más interrupciones AGG en la región, y la distancia desde el extremo de al menos una región de repetición CGG o CCG a una o más interrupciones AGG en la región. En algunas realizaciones, la información de parámetros se determina por reacción en cadena de la polimerasa y electroforesis capilar.

[0012] En ciertas realizaciones, los métodos descritos anteriormente se pueden utilizar para detectar un genotipo o fenotipo asociado con un trastorno *FMR1* o *FMR2*. En algunas realizaciones, el trastorno *FMR1* o *FMR2* es el síndrome de X frágil (FXS), síndrome X frágil E (FRAXE), Fragile temblor X-asociado/ataxia síndrome (FXTAS), frágil insuficiencia ovárica primaria relacionada X-(FXPOI), o dopamina sensible parkinsonismo.

[0013] En diversas realizaciones, se proporciona un aparato para la reconstrucción automatizada de un genotipo, que comprende (a) un procesador; y (b) un dispositivo de memoria acoplado comunicativamente al procesador, el dispositivo de memoria que tiene almacenado en el mismo instrucciones ejecutables por máquina que, cuando se ejecuta por el procesador, hacen que el procesador (i) reciba información de parámetros para un ácido nucleico, donde el ácido nucleico tiene al menos una región de repetición de nucleótidos rica en GC que comprende una o más interrupciones; y en donde la información de parámetros comprende la longitud total de la al menos una región de repetición de nucleótidos rica en GC, la distancia desde el comienzo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región, y la distancia desde el extremo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región; (ii) generar un conjunto de genotipos posibles que comprenden todas las disposiciones posibles de la una o más interrupciones en la región; (iii) evaluar el conjunto de genotipos potenciales para producir una solución de genotipo que satisface todos los parámetros de información; y (iv) almacenar la solución de genotipo en el dispositivo de memoria. En algunas realizaciones, el aparato comprende además un monitor comunicativamente acoplado al dispositivo de procesador y memoria, en el que las instrucciones ejecutables por máquina almacenadas en el dispositivo de memoria instruyen al

procesador para mostrar la solución de genotipo en el monitor. En algunas realizaciones, el aparato comprende además una impresora comunicativamente acoplada al dispositivo de procesador y memoria, en la que las instrucciones ejecutables por máquina almacenadas en el dispositivo de memoria instruyen al procesador para imprimir la solución de genotipo en la impresora.

5 [0014] En diversas realizaciones, se proporciona un medio legible por máquina, que comprende instrucciones ejecutables por máquina que, cuando es ejecutado por un procesador, hace que el procesador (a) reciba información de parámetros para un ácido nucleico, donde el ácido nucleico tiene al menos una región de repetición de nucleótidos rica en GC que comprende una o más interrupciones; y en donde la información de parámetros comprende la longitud total de la al menos una región de repetición de nucleótidos rica en GC, la distancia desde el comienzo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región, y la distancia desde el extremo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región; (b) genere un conjunto de genotipos posibles que comprenden todas las disposiciones posibles de la una o más interrupciones en la región; (c) evalúe el conjunto de genotipos potenciales para producir una solución de genotipo que satisfice todos los parámetros de información; y (d) almacenar la solución de genotipo en un dispositivo de memoria. En algunas realizaciones, las instrucciones ejecutables por máquina instruyen al procesador para que muestre la solución de genotipo en un monitor. En algunas realizaciones, las instrucciones ejecutables por máquina instruyen al procesador para que se imprima la solución de genotipo en una impresora.

20 **BREVE DESCRIPCIÓN DE LOS DIBUJOS**

[0015]

25 Fig. 1 muestra un ejemplo de datos de una longitud de ensayo CGG, un ensayo Anch_A, y un ensayo Anch_T generado a partir de una muestra, donde el genotipo se identificó correctamente utilizando enumeración exhaustiva. La longitud de la secuencia PCR del producto se muestra en el eje horizontal, y el número de productos de PCR que tienen una longitud particular se muestra en el eje vertical.

30 Fig. 2 muestra un ejemplo de datos de una longitud de ensayo CGG, un ensayo Anch_A, y un ensayo Anch_T generado a partir de una muestra en la enumeración exhaustiva no produjo una solución de genotipo. La longitud de la secuencia PCR del producto se muestra en el eje horizontal, y el número de productos de PCR que tienen una longitud particular se muestra en el eje vertical.

35 Fig. 3 muestra un ejemplo de datos de una longitud de ensayo CGG, un ensayo Anch_A, y un ensayo Anch_T generado a partir de una muestra en la enumeración exhaustiva identificó dos posibles soluciones de genotipo. La longitud de la secuencia PCR del producto se muestra en el eje horizontal, y el número de productos de PCR que tienen una longitud particular se muestra en el eje vertical.

40 Fig. 4A muestra un ejemplo de datos de altura de pico de un ensayo de longitud CGG representado frente a la ubicación de pico (cuadrados sólidos) y un mejor ajuste de la curva de los datos de altura de pico predichos sobre la base de una solución de genotipo generada por enumeración exhaustiva que ha sido superpuesta en los datos originales (círculos abiertos). Fig. 4B muestra un segundo mejor ajuste de la curva basado en una segunda solución de genotipo generada por enumeración exhaustiva superpuesta en los mismos datos que en la Fig. 4A.

DESCRIPCIÓN DE CIERTAS REALIZACIONES EJEMPLARES

45 [0016] Ahora se hará referencia en detalle a ciertas realizaciones ejemplares de acuerdo con la presente descripción, ciertos ejemplos de los cuales se ilustran en los dibujos adjuntos.

50 [0017] Los encabezados de sección usados en este documento son sólo para fines de organización y no deben interpretarse como limitantes del objeto descrito.

[0018] Para ayudar en la comprensión de la presente invención, ciertos términos se definen primero. Definiciones adicionales se proporcionan en toda la aplicación.

55 [0019] En esta aplicación, el uso del singular incluye el plural a menos que se especifique lo contrario. También en esta aplicación, el uso de "o" significa "y/o" a menos que se indique lo contrario. Además, el uso del término "que incluye", así como otras formas, tales como "incluye" y "incluido", no son limitativos. Cualquier intervalo descrito en el presente documento se entiende que incluye los puntos finales y todos los valores entre los puntos finales.

60 [0020] Como se usa en este documento, un "ácido nucleico" es cualquier residuo de nucleobases contiguas o análogos que han sido retirados de su entorno natural y aislados de un sujeto y/o para el que se solicita una reconstrucción de genotipo. Un ácido nucleico puede comprender un gen, fragmento de gen o región genómica aislada de un sujeto. Como se usa en este documento, un "genotipo" es la secuencia de nucleobase de un ácido nucleico.

65 [0021] Como se usa en este documento, "riqueza en GC" es la fracción o porcentaje de los residuos totales de nucleobases en un ácido nucleico o un fragmento de ese ácido nucleico que son los residuos de guanina, residuos de citosina, o análogos de los mismos. Por ejemplo, una secuencia de 100 nucleótidos que contiene exactamente 30

5 citosinas, exactamente 30 guaninas, exactamente un análogo de la citosina, y exactamente un análogo de guanina tiene una riqueza en GC de 62%. En algunas realizaciones, un ácido nucleico "rico en GC" o región de un ácido nucleico es una que contiene más de aproximadamente el 50% de residuos de guanina, residuos de citosina, o análogos de los mismos (por ejemplo, más de aproximadamente 50, 51, 55, 60, 65, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99, o 99,5% de residuos de guanina, residuos de citosina, o análogos de los mismos, o cualquier porcentaje intermedio).

10 [0022] Tal como se utiliza aquí, una "repetición" o "repetición de nucleótidos" se refiere a un ácido nucleico o una región de un ácido nucleico que comprende cualquier secuencia corta de 1-20 residuos de nucleobases de longitud (por ejemplo, una secuencia de dinucleótido, trinucleótido, tetranucleótido, pentanucleótido, hexanucleótido, etc.) en donde la secuencia corta se repite 2 o más veces (por ejemplo, 2, 3, 4, 5, 10, 15, 20, 50, 100, 200, 500, o más repeticiones). Por ejemplo, una repetición de nucleótidos abarcaría una región de un ácido nucleico en donde la secuencia corta CGG se repite dos o más veces. Un ácido nucleico o una región de un ácido nucleico puede ser tanto una repetición como una región rica en GC. Por ejemplo, el ácido nucleico o una región de un ácido nucleico pueden comprender repeticiones de di-, tri-, tetra-, penta- o hexa-nucleótidos de los residuos de guanina, residuos de citosina, o análogos de los mismos.

15 [0023] Un ácido nucleico puede comprender uno o más nucleótidos repeticiones o regiones ricas en GC que contienen una o más interrupciones. Como se usa en el presente documento, una "interrupción" en un ácido nucleico se refiere a la presencia de uno o más residuos de nucleobases o análogos en el ácido nucleico que son incompatibles con el patrón de repetición o, en una región rica en GC, comprenden una nucleobase distinta de G o C (o análogos de los mismos). Por ejemplo, una, región de repetición de nucleótidos rica en GC podría abarcar una secuencia que comprende 40 repeticiones de trinucleótidos CGG con dos secuencias AGG intercaladas dentro de las 40 repeticiones CGG.

20 [0024] *FMR1* es un gen en el cromosoma X que codifica la proteína de retraso mental del X frágil (FMRP). Esta proteína, más comúnmente encontrada en el cerebro, es esencial para el desarrollo cognitivo normal y la función reproductora femenina. La región no traducida (5' UTR 5') del gen *FMR1* puede comprender una o más regiones (por ejemplo, 1, 2, 3, 4, 5, o más regiones) que tienen repeticiones CGG. La presencia de una región que comprende 40 o más repeticiones CGG en el 5' UTR se cree que está asociada con trastornos relacionados con *FMR1*. El 5' UTR del gen *FMR1* también puede comprender una o más secuencias AGG que interrumpen una región de repetición CGG.

25 [0025] *FMR2* (*AFF2*) es un gen en el cromosoma X que codifica la proteína de retraso mental 2 del X frágil. El 5' UTR del gen *FMR2* puede comprender una o más regiones (por ejemplo, 1, 2, 3, 4, 5, o más regiones) que tienen repeticiones de CCG. La expansión de una región de repetición CCG (o CGG) se ha asociado con discapacidad intelectual ligada al cromosoma X, y en particular con síndrome del X Frágil E (FRAXE). FRAXE es una forma común de retraso mental ligado al cromosoma X.

40 I. Gentificación completa

45 [0026] En este documento se describen métodos para la reconstrucción automatizada de un genotipo que comprenden una o más regiones de repetición de nucleótidos o regiones ricas en GC. Por ejemplo, los métodos descritos en este documento pueden ser utilizados para reconstruir un genotipo a partir de un ácido nucleico o fragmento del mismo que comprende repeticiones CGG e interrupciones AGG intercaladas, o para reconstruir un genotipo que comprende repeticiones de CCG. En algunas realizaciones, los métodos se pueden utilizar para reconstruir la región de repetición del gen de *FMR1* o *FMR2*, o fragmentos de los mismos, o el 5' UTR de *FMR1* o *FMR2*, o fragmentos de los mismos, aislados de un sujeto. En ciertas realizaciones, los métodos descritos en este documento se utilizan para ayudar en la reconstrucción del genotipo de *FMR1*, incluyendo el patrón de repetición CGG y la ubicación y organización de las interrupciones AGG dentro de los 5' UTR de *FMR1*. En otras realizaciones, los métodos descritos en este documento se utilizan para ayudar en la reconstrucción del genotipo de *FMR2*, incluyendo el patrón de repetición de CCG.

50 [0027] En algunas realizaciones, los procedimientos divulgados en el presente documento se utilizan para determinar el genotipo de un ácido nucleico o fragmento del mismo a partir de una muestra del paciente, donde el ácido nucleico tiene al menos una región de repetición rica en GC, y en donde el genotipo relacionado desde al menos una de los padres del paciente no se conoce. En ciertas realizaciones, los métodos descritos en este documento se utilizan para determinar un genotipo para la región de repetición CGG CCG o en *FMR1* o *FMR2* a partir de una muestra del paciente, en donde el genotipo relacionado *FMR1* o *FMR2* a partir de al menos uno de los padres del paciente no se conoce.

55 [0028] En diversas realizaciones, un método para la reconstrucción automatizado de un genotipo comprende proporcionar una muestra de un paciente, en donde la muestra contiene un ácido nucleico o fragmento del mismo que tiene una o más regiones de repetición o regiones ricas en GC. En algunas realizaciones, la información que caracteriza el ácido nucleico (es decir, "información de parámetros") se recoge, tales como la longitud total de la repetición o región rica en GC y la distancia en las direcciones de avance y retroceso a las interrupciones en la repetición o región rica en GC. En algunas realizaciones, la información recogida se analiza automáticamente usando enumeración exhaustiva, utilizando un aparato que comprende un procesador programado para llevar a cabo un análisis de enumeración exhaustiva automatizada. En algunas realizaciones, la enumeración exhaustiva reconstruye

una solución de genotipo (es decir, un genotipo que satisface todos los parámetros de información). En ciertas realizaciones, la exactitud de la solución de genotipo se puede evaluar mediante el análisis manual del genotipo para confirmar que se comporta con toda la información de parámetros, o mediante la realización de cualquier otro ensayo de confirmación (por ejemplo, enzima de digestión de restricción, secuenciación de Sanger, u otras formas de secuenciación de alto rendimiento). En algunas realizaciones, la solución de genotipo se puede mostrar o almacenar electrónicamente en un ordenador, o puede ser impreso con fines de diagnóstico y terapéuticos posteriores.

[0029] En ciertas realizaciones, el genotipo reconstruido se puede utilizar para detectar una mutación o genotipo, o para diagnosticar o ayudar en el diagnóstico de un trastorno asociado con una mutación en una región de repetición o región rica en GC, tal como una mutación, genotipo, o trastorno relacionados con *FMR1* o *FMR2*.

[0030] En diversas realizaciones, la información de parámetro que caracteriza un ácido nucleico se puede obtener utilizando cualquier procedimiento adecuado, tal como análisis de restricción de PCR o digestión. En ciertas realizaciones, la información de parámetros incluye la longitud total de la región de repetición o rica en GC, así como la distancia desde el inicio de la región de repetición o rica en GC a una primera o subsiguiente interrupción en la dirección hacia delante y en la dirección reversa. En algunas realizaciones, se proporciona un aparato, que comprende un procesador programado para analizar la información de parámetros y para reconstruir un genotipo de la información. En ciertas realizaciones, el aparato se utiliza para reconstruir el genotipo del ácido nucleico a partir de la información que caracteriza el ácido nucleico. En algunas realizaciones, el aparato utiliza enumeración exhaustiva para evaluar todas las posibles reconstrucciones de genotipo basado en la longitud de la repetición o región rica en GC y las interrupciones en la dirección de avance o de retroceso para seleccionar la reconstrucción que satisface toda la información de parámetros (por ejemplo, el genotipo que coloca las interrupciones en las posiciones correctas en ambas direcciones directa e inversa). En ciertas realizaciones, el aparato proporciona un informe del genotipo reconstruido que se pueden mostrar en una pantalla, guarda digitalmente para uso futuro, o impreso como un registro de papel.

[0031] En diversas realizaciones, la información de parámetros con respecto a un ácido nucleico se puede obtener usando cualquier método conocido en la técnica, con tal de que incluye información con respecto a la longitud total de una repetición o la región rica en GC en el ácido nucleico y la distancia desde el principio ("dirección hacia adelante") y final ("dirección inversa") de la región rica en GC a interrupciones. En algunas realizaciones, las enzimas de restricción que escinden un ácido nucleico de manera específica de sitio se pueden usar para analizar una región de repetición o rica en GC y de ese modo generar información de parámetros. Por ejemplo, la presencia de interrupciones AGG dentro de un tracto de repetición CGG de *FMR1* puede ser detectada mediante la digestión de un ácido nucleico con la enzima de restricción *EciI* (New England Biolabs Inc., Ipswich, MA, EE.UU.). En otras realizaciones, los métodos de PCR pueden ser usados para generar la información necesaria. Por ejemplo, digestión de restricción y/o métodos de PCR se pueden utilizar con un gen *FMR1* o *FMR2* o fragmentos del mismo aislados de un paciente con el fin de determinar la longitud de una o más repeticiones CGG o CCG. Los métodos también se pueden utilizar para determinar la distancia en las direcciones de avance y retroceso de los primeros y últimos tripletes CGG o CCG a cualesquiera interrupciones internas, tales como interrupciones AGG intercaladas dentro de una región de repetición CGG en el gen *FMR1*.

[0032] En algunas realizaciones, los métodos adecuados para la información de los parámetros de generación incluyen la reacción en cadena de la polimerasa (PCR), PCR en tiempo real (RT-PCR), amplificación de base de ácido nucleico de la secuencia (NASBA), reacción en cadena de la ligasa, amplificación de la sonda ligable de multiplex, tecnología de invasor (Third Wave), amplificación por círculo rodante, la transcripción in vitro, filamento de amplificación del desplazamiento, la amplificación mediada por transcripción (TMA), amplificación de ARN (por ejemplo, Eberwine), amplificación mediada por bucle isotérmico, o cualesquiera otros métodos que son conocidos para un experto en la técnica. Por ejemplo, información de parámetros *FMR1* se puede generar utilizando un enfoque de PCR de dos niveles con una imprimación de enlazador CGG y el kit PCR humano *FMR1* (Asuragen Inc., Austin, TX, EE.UU.). Ver Tassone et al., J Mol Diagn. 10 (1): 43-49 (2008); Chen et al., J Mol Diagn. 12 (5): 589-600 (2010); Yrigollen et al, PLoS One 6 (7): e21728 (2011). Por ejemplo, un ácido nucleico que comprende al menos una región rica en GC puede ser analizada por (a) proporcionar al menos dos cebadores de PCR, incluyendo un primer cebador que comprende repeticiones CGG, CCG, GCG, CGC, GCC, o GGC, y un segundo cebador que hibrida con una posición fuera de la región rica en GC; (B) realizar la PCR en el ácido nucleico con los al menos dos cebadores diferentes, en donde la PCR produce un conjunto de productos de la PCR; (C) resolver el conjunto de productos de PCR con una técnica de alta resolución (tal como electroforesis capilar) para generar una representación de tamaño del producto PCR y la abundancia; y (d) se deriva de las dimensiones del producto y de la abundancia de información PCR de la longitud de la región rica en GC y si o en donde se encuentra la región rica en GC de una interrupción.

[0033] En ciertas realizaciones, un genotipo de *FMR1* es reconstruido, y algunas o todas las informaciones de parámetros *FMR1* (por ejemplo, la longitud de repetición CGG y/o la distancia en el avance y/o direcciones inversas a las interrupciones AGG) se pueden generar utilizando cebadores directos e inversos de PCR que comprenden repeticiones de CGG y CCG, respectivamente. En ciertas realizaciones, la información de parámetros *FMR1* puede ser generada usando un cebador directo que comprende la secuencia (CGG)₅A (SEQ ID NO: 42) y/o un cebador inverso que comprende la secuencia de TCC (GCC)₄ (SEQ ID NO: 46). En algunas realizaciones, algunas o la totalidad de las informaciones de parámetros *FMR1* (por ejemplo, la longitud de repetición CGG y/o la distancia en las

direcciones de avance y/o retroceso a interrupciones AGG) se pueden generar usando cebadores alternativos. Por ejemplo, los cebadores que tienen secuencias de repetición CGG CCG o que son más cortas o más largas que las secuencias (CGG)₅A (SEQ ID NO: 42) y TCC (GCC)₄ (SEQ ID NO: 46) mencionadas anteriormente, u otras secuencias que son específicas para la región en o cerca de una interrupción AGG, pueden ser utilizadas.

[0034] En algunas realizaciones, la información de parámetros *FMR1* se puede generar utilizando un cebador de PCR directo que comprende la secuencia (CGG)₅A (SEQ ID NO: 42) y/o un cebador de PCR inverso que comprende la secuencia de TCC (GCC)₄ (SEQ ID NO: 46) para identificar la ubicación de las interrupciones AGG en el patrón de repetición CGG de un alelo *FMR1*, mientras que la longitud total de la repetición CGG se puede determinar usando un primer cebador de PCR que comprende repeticiones de CGG o CCG, y un segundo cebador de PCR que hibrida con una posición fuera de la región de repetición CGG. Por ejemplo, la información de parámetros se puede generar utilizando un kit PCR de *FMR1* humano (Asuragen Inc., Austin, TX, EE.UU.). En algunas realizaciones, los cebadores pueden usarse para amplificar por PCR los segmentos del locus *FMR1*, y la información de amplitud (es decir, el número de copias de un segmento del locus de *FMR1* de una longitud particular producida por la reacción PCR) se puede medir usando electroforesis capilar. Véase, por ejemplo, los métodos de la solicitud de EE.UU. N^{os} 2010/0209970, 2010/0243451, y 2012/0107824.

[0035] En diversas realizaciones, los ácidos nucleicos amplificados por PCR se analizan para generar información de parámetros, por ejemplo usando instrumentos de electroforesis capilar (EC) familiares para los expertos en la técnica, tal como el modelo ABI instrumentos EC 3100, 3130, 3730, o 3500 (Applied Biosystems, Carlsbad, CA). Otras implementaciones, incluyen cualquier instrumento capaz de dimensionar electroforéticamente o de otra manera y/o secuenciación de un ácido nucleico amplificado, también se pueden utilizar. Métodos PCR y análisis ejemplares para la utilización con repetición o regiones ricas en GC, tales como la repetición de loci CGG y CCG en los UTRs 5 de *FMR1* y *FMR2*, se describen en la Solicitud de Estados Unidos N^{os} 2010/0243451 y 2012/0107824 y la patente de los Estados Unidos N^o 8.409.805. Cualquier otro método de recoger información de parámetro también se puede utilizar (por ejemplo, secuenciación Sanger o de otras formas de secuenciación de alto rendimiento).

[0036] Por ejemplo, la información de parámetro que caracteriza el loci de repetición CGG y CCG en los UTRs 5 de *FMR1* y *FMR2* o fragmentos de los mismos se pueden generar usando los métodos descritos en la Solicitud de Patente de Estados Unidos N^o 2010/0243451, incluyendo los cebadores, polimerasa, reactivos y condiciones de reacción descritas en los párrafos [0040]-[0051], [0056]-[0060], [0065]-[0067], [0089], [0094] y [0104].

Por ejemplo, en algunas realizaciones, informaciones de parámetros *FMR1* y *FMR2* se pueden generar utilizando un cebador recocado fuera de una región de repetición y un cebador que hibrida con secuencias de repetición, permutaciones de secuencia, o complementos inversos de las secuencias (GCG, CCG, CGC, GCC, o GGC). Los cebadores que pueden hibridar fuera de (aguas arriba o aguas abajo) de la región de repetición puede ser hacia adelante o cebadores inversos. Los cebadores pueden hibridarse a secuencias que flanquean la región de repetición. Ejemplos de cebadores directos incluyen CGG TGG AGG GCC GCC TCT GAG C (SEQ ID NO: 1), CAG GCG CTC AGC TCC GTT TCG GTT T (SEQ ID NO: 2), CAG TCA GGC GCT CAG CTC CGT TTC G (SEQ ID NO: 3), TCC GGT GGA GGG CCG CCT CTG AGC (SEQ ID NO: 4), GGT TCG GCC TCA GTC AGG CGC TCA GCT CCG TTT CG (SEQ ID NO: 5), GGG TTC GGC CTC AGT CAG GCG CTC AGC TCC GTT TCG (SEQ ID NO: 6), GCG GGC CGG GGG TTC GGC CTC AGT CA (SEQ ID NO: 7), CAG CGG CCG GGG GGT TCG GCC TCA G (SEQ ID NO: 8), GCA GCG GGC CGG GGG TTC GGC CTC A (SEQ ID NO: 9), GGG CCG GGG GGT TCG CCT CAG TCA G (SEQ ID NO: 10), GGG GTT CCG CCT CAG TCA GGC GCT CA (SEQ ID NO: 11), GGG GTT CCG CCT CAG TCA GGC GCT CAG (SEQ ID NO: 12), GGC GCT CAG CTC CGT TTC GGT TTC ACT TCC (SEQ ID NO: 13), TCA GGC GCT CAG CTC CGT TTC GGT TTC A (SEQ ID NO: 14), CAC TTC CGG TGG AGG GCC GCC TCT GA (SEQ ID NO: 15), TTC CGG TGG AGG GCC GCC TCT GAG C (SEQ ID NO: 16), y TCA GGC GCT CAG CTC CGT TTC GGT TTC ACG GCG GCG GCG GCG GA (SEQ ID NO: 44). Ejemplos de cebadores inversos incluyen CGC ACT TCC ACC ACC AGC TCC TCC A (SEQ ID NO: 17), GGA GCC CGC CCC CGA GAG GTG (SEQ ID NO: 18), GGG AGC CCG CCC CCG AGA GGT (SEQ ID NO: 19), CGC ACT TCC ACC ACC AGC TCC TCC AT (SEQ ID NO: 20), CGG GAG CCC GCC CCC GAG AGG TG (SEQ ID NO: 21), CCG GGA GCC CGC CCC CGA GAG GT (SEQ ID NO: 22), CCG GGA GCC CGC CCC CGA GAG GTG (SEQ ID NO: 23), CGC CGG GAG CCC GCC CCC GAG AGG TG (SEQ ID NO: 24), GCG CCG GGA GCC CGC CCC CGA GAG GT (SEQ ID NO: 25), CGC CGG GAG CCC GCC CCC GAG AGG T (SEQ ID NO: 26), GCG CCA TTG GAG CCC CGC ACT TCC ACC A (SEQ ID NO: 27), GCG CCA TTG GAG CCC CGC ACT TCC A (SEQ ID NO: 28), AGC GCC ATT GGA GCC CCG CAC TTC C (SEQ ID NO: 29), CGC CAT TGG AGC CCC GCA CTT CCA C (SEQ ID NO: 30), TTG GAG CCC CGC ACT TCC ACC ACC A (SEQ ID NO: 31), AGC CCC GCA CTT CCA CCA CCA GCT CCT C (SEQ ID NO: 32), GAG CCC CGC ACT TCC ACC ACC AGC TCC T (SEQ ID NO: 33), CAT TGG AGC CCC GCA CTT CCA CCA CCA G (SEQ ID NO: 34), CCC GCA CTT CCA CCA CCA GCT CCT CCA TCT (SEQ ID NO: 35), TAG AAA GCG CCA TTG GAG CCC CGC ACT TCC (SEQ ID NO: 36), AAG CGC CAT TGG AGC CCC GCA CTT CC (SEQ ID NO: 37), AAG CGC CAT TGG AGC CCC GCA CTT CCC CGC CGC CGC CGC CG (SEQ ID NO: 43), y AAG CGC CAT TGG AGC CCC GCA CTT CCC CGC CGC CGC CGC CT (SEQ ID NO: 45).

[0037] En algunas realizaciones informaciones de parámetros *FMR1* y *FMR2* se pueden generar utilizando los cebadores TCAGGCGCTCAGTCCGTTTCGGTTTCACTTCCGGT (SEQ ID NO: 38), AGCGTCTACTGTCTCGGCACTTGCCTCCGCGCCGCGG (SEQ ID NO: 39), TCA GGC GCT CAG CTC CGT TTC GGT TTC A (SEQ ID NO: 40), y TCAGGCGCTCAGTCCGTTTCGGTTTCA CGGCGCGGCGGCGG (SEQ ID NO: 41). Los métodos además pueden implicar el uso de cebadores que comprenden la secuencia de cualquiera de SEQ ID

NOs 1-38 o 40 y que comprende repeticiones adicionales de CGG o las permutaciones y complementos inversos de los mismos (por ejemplo, GCG, CCG, CGC, GCC, o GGC) anexos al extremo 3'. En algunas realizaciones, el número de repeticiones CGG o permutaciones en el cebador es de cuatro o cinco. En algunas realizaciones, el cebador contiene una secuencia de repeticiones CGG (o permutaciones de los mismos) de estiramiento durante 12-15 nucleótidos o más. En algunas realizaciones, el cebador contiene una secuencia de repeticiones CGG (o permutaciones de los mismos) que van desde 3 a 10 repeticiones. El cebador puede contener 3, 4, 5, 6, 7, 8, 9, o 10 repeticiones, y opcionalmente una repetición parcial adicional de 1 o 2 residuos C y/o G. Cebadores adicionales se pueden proporcionar, por ejemplo, para asegurar la unión en un sitio polimórfico, o para amplificar una región de tamaño conocido con el fin de servir como un estándar interno.

[0038] En algunas realizaciones, el cebador que hibrida con la repetición o región rica en GC tiene una actividad de unión preferencial para los sitios en la región que comprende un elemento interruptor. La unión preferencial al sitio de un elemento interruptor puede resultar en la amplificación selectiva de al menos un producto que comprende el elemento de interruptor, por ejemplo, usando el cebador en una reacción de PCR con un segundo cebador orientada opuestamente que se une fuera de la repetición o región rica en GC. La actividad de unión preferencial puede ser específica, por ejemplo, para los sitios que comprenden CGG y elementos AGG, o las permutaciones y/o complementos inversos de los mismos, tales como un sitio que comprende (1) un elemento de AGG o una parte de un elemento AGG que comprende un A, y (2) tres, cuatro, cinco, o seis elementos CGG y opcionalmente un elemento CGG parcial adicional.

[0039] En algunas realizaciones, el cebador que se hibrida a la repetición o región rica y se une preferentemente a un sitio o sitios que comprenden un elemento interruptor puede comprender un residuo A, T, o U dentro de o en el extremo de la parte de la imprimación GC que hibrida con la repetición o secuencias ricas en GC. Por ejemplo, el cebador puede tener una A, T, o U entre o al final de un tramo de repeticiones CGG, CCG, GCG, CGC, GCC, o GGC; véase, por ejemplo, SEQ ID NO 44 y 45 anteriores. El residuo A, T, o U se puede producir en el extremo 3' del cebador. Cuando el residuo A, T, o U se produce al final de las repeticiones CGG, CCG, GCG, CGC, o CCG, GGC, puede o no ser una repetición parcial CGG, CCG, GCG, CGC, GCC, o GGC entre el residuo de A, T, o U y la última repetición completa CGG, CCG, GCG, CGC, GGC, o GGC. Es posible sustituir residuos de nucleótidos no naturales que preferentemente par de bases con residuos T/U o A en relación con otros residuos de nucleótidos naturales para el residuo A, T, o U. Del mismo modo, también es posible sustituir uno o más residuos no naturales de nucleótidos que preferentemente se acoplan de base con residuos C o G en relación con otros residuos de nucleótidos naturales para uno o más residuos C y/o G y que componen la repeticiones CGG, CCG, GCG, CGC, GCC, o GGC. La presencia de uno o más de tales residuos no naturales dentro de una secuencia hizo lo contrario de repeticiones CGG, CCG, GCG, CGC, GCC, o GGC (opcionalmente con una A, T, U, o un residuo no natural como se discutió anteriormente correspondiente) no niega la identidad de dicha secuencia en el contexto de la presente descripción como una secuencia de repeticiones CGG, CCG, GCG, CGC, GCC, o GGC. Los residuos de nucleótidos no naturales son residuos de nucleótidos que comprenden una nucleobase otra que adenina, timina, guanina, citosina y uracilo (A, T, G, C y U, respectivamente). Ejemplos de residuos no naturales de nucleótidos que preferentemente se acoplan de base con residuos de A o T/U incluyen, sin limitación, productos de adición de residuos de T, U, o A que preferentemente se acoplan de base con residuos A o T/U en relación con otros residuos naturales (por ejemplo, análogos de uracilo 5-sustituido); y residuos que comprenden nucleobases tales como, por ejemplo, pseudouracilo y diaminopurina.

[0040] En algunas realizaciones, un primer cebador se utiliza al generar la información de parámetros que tiene una actividad de unión preferencial para los sitios en la repetición o región rica en GC que no comprenden elementos de interruptor. La presencia de un elemento interruptor se puede señalar en los resultados de este método por un nivel relativamente bajo de los productos cuya síntesis significa extensión del primer cebador unido a los sitios que comprenden el elemento interruptor. Estos bajos niveles pueden aparecer como un hueco o conjunto de picos bajos rodeados por picos más altos en un electroferograma. En algunas realizaciones, se proporciona un primer cebador que tiene una actividad de unión preferencial para los sitios en la repetición o región rica en GC que comprenden elementos de interruptor. La presencia de un elemento interruptor se señala en un ensayo anclado por un nivel relativamente alto de productos cuya síntesis significa extensión del primer cebador unido a los sitios que comprenden el elemento interruptor. El alto nivel puede aparecer como un pico rodeado por picos más bajos y/o señal de la línea de base en un electroferograma.

[0041] Los métodos de generación de información de los parámetros pueden referirse a las reacciones de amplificación que comprende proporcionar al menos dos o al menos tres cebadores diferentes. En algunas realizaciones, se proporcionan al menos tres cebadores diferentes y uno de los cebadores es un cebador que se une preferentemente fuera de la repetición o región rica en GC, un segundo cebador se une preferentemente dentro de la repetición o región rica en GC, y el tercer cebador es una subsecuencia de la primera o segunda imprimación. En algunas realizaciones, un cebador es un cebador quimérico que comprende repeticiones CGG y una secuencia de pestaña 5', y otro cebador tiene la secuencia de la secuencia de pestaña 5' del cebador quimérico. Cabe señalar que el cebador que tiene la secuencia de la secuencia de pestaña 5' del cebador quimérico puede, pero no necesariamente, tienen toda la secuencia de no repetición del cebador quimérico. En otras palabras, la secuencia de parte o la totalidad de un cebador puede estar compuesta por la secuencia de otro cebador; Por ejemplo, el cebador quimérico comprende una secuencia de pestaña, y otro cebador puede comprender la secuencia de parte o la totalidad de la pestaña 5'. En

algunas realizaciones, el cebador contiene 12-15 nucleótidos de una secuencia de repetición CGG. La secuencia de pestaña 5' puede corresponder a una secuencia adyacente a o cerca de la región de repetición CGG, o puede no estar relacionada con secuencias en y alrededor de la región de repetición CGG. En algunas realizaciones, la longitud del cebador quimérico puede ser de aproximadamente 35, 40, 45, 50, o 55 nucleótidos. En algunas realizaciones, uno o más de los cebadores tiene una temperatura de fusión que oscila entre 60°C a 75°C, por ejemplo, aproximadamente 60°C, 65°C, 70°C, o 75°C.

[0042] En algunas realizaciones, se proporcionan al menos tres cebadores diferentes y se proporciona un cebador a una concentración menor que la concentración de otro cebador. Por ejemplo, el cebador quimérico se proporciona opcionalmente a una concentración más baja que el cebador con la secuencia de la secuencia de pestaña 5' del cebador quimérico. La relación de concentraciones, expresada como una diferencia de pliegue, puede variar de 2 a 10.000 o más, por ejemplo, 10, 20, 50, 100, 200, 500, 1.000, 2.000, 5.000, o 10.000 (o cualquier valor intermedio). En tales realizaciones, el presente cebador a una concentración más baja se puede agotar en las primeras rondas de la reacción de amplificación, de tal manera que la extensión es generalmente todos, o casi todos, de los cebadores presentes (que inicialmente estaban presentes en concentraciones relativamente más altas).

[0043] En algunas realizaciones, los métodos de información de parámetros de generación comprenden proporcionar dNTPs en una relación GC/AT mayor que uno, y en una concentración total de dNTP propicia a la síntesis de ADN que comprende plantillas de repetición o ricas en GC. Ver Solicitud de EE.UU. nº 12/371.306. La relación GC/AT puede ser de aproximadamente 1,1, 1,2, 1,4, 1,6, 2, 2,5, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, o más alto. La relación GC/AT puede ser de entre 1,1 y 20, 1,1 y 15, 1,1 y 10, 1,1 y 8, 1 y 15, 1,1 y 7, 1,1 y 6, 1,1 y 5, 1,2 y 25, 1,4 y 25, 1,6 y 25, 2 y 25, 3 y 25, 4 y 25, 5 y 25, 2 y 15, 2,5 y 10, o 4 y 10. La concentración total de dNTP puede ser de aproximadamente 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 1,2, 1,5, 2, o 3 mM. La concentración de dNTP puede ser de entre 0,4 y 3 mM, 0,5 y 3 mM, 0,6 y 3 mM, 0,7 y 3 mM, 0,8 y 3 mM, 0,9 y 3 mM, 1 y 3 mM, 0,4 y 2 mM, 0,4 y 1,5 mM, 0,4 y 1,2 mM, 0,4 y 1 mM, 0,4 y 0,9 mM, 0,4 y 0,8 mM, 0,4 y 0,7 mM, 0,5 y 2 mM, 0,5 y 1 mM, o 0,6 y 0,9 mM. "GC/AT Ratio" significa la relación de la concentración de la suma de dCTP, dGTP, y todos los análogos de nucleótidos de los mismos, a la concentración de la suma de dATP, dTTP, dUTP, y todos los análogos de nucleótidos de los mismos, en una solución dada o mezcla. "DNTP" significa desoxinucleótido trifosfato y se refiere a dATP, dCTP, dGTP, dTTP, dUTP, y análogos de los mismos. "Análogos de nucleótidos" son moléculas o iones que comprenden un resto de base distinta de las bases de adenina naturales (A), citosina (C), guanina (G), timina (T) o uracilo (U), un resto de azúcar idéntico o similar a desoxirribosa, y al menos un resto de fosfato o fosfato múltiple (por ejemplo, difosfato o trifosfato). El análogo de nucleótido es un análogo de un nucleótido específico, en particular, dATP, dCTP, dGTP, dTTP, o dUTP, cuando comprende un trifosfato y un resto de azúcar, la estructura y la configuración de ambos de los cuales son adecuados para su incorporación en un ácido nucleico doble hélice por una polimerasa, y una base cuya base de emparejamiento de propiedades en una doble hélice de ácido nucleico y loci de la incorporación por las ADN polimerasas en una doble hélice de ácido nucleico son más similares a uno de los cinco nucleótidos enumerados anteriormente, con la excepción de que los análogos de dTTP también estarán generalmente análogos de dUTP y viceversa. El término "análogo" se utiliza en conjunción con los términos incluyendo, pero no limitado a "nucleósido", "base", "nucleobase", o "residuo" se ha de interpretar de la misma manera como si se utiliza en conjunción con "nucleótido".

[0044] En algunas realizaciones, los métodos de información de parámetros de generación pueden comprender además proporcionar tampones para las reacciones de amplificación de PCR. Los tampones pueden comprender, por ejemplo y sin limitación, tris(hidroximetil)aminometano (Tris), bis-tris propano, bicarbonato, fosfato, glicina, histidina, ácido 4- (2-hidroxietilo)-1-ácido piperazinanosulfónico (HEPES), 3-(N-morfolino)ácido propanosulfónico (MOPS), y varios bases/ácidos de conjugados y sales de los mismos.

[0045] En algunas realizaciones, los métodos de generación de información de parámetros pueden comprender proporcionar al menos una ADN polimerasa para sintetizar ADN a partir de dNTPs en una manera dependiente de la plantilla. La polimerasa de ADN puede comprender un tipo salvaje, modificado, termófilo, quimérico, de ingeniería, y/o una mezcla de más de una polimerasa. La polimerasa de ADN puede comprender Exact Polymerase (5 PRIME GmbH), AccuSure™ DNA Polymerase (Bioline), Phusion™ AccuPrime™ Pfx (Invitrogen), Platinum Taq DNA Polymerase High Fidelity (Invitrogen), Phire™ Hot Start DNA Polymerase (New England Biolabs), Phusion® Hot Start High-Fidelity DNA Polymerase (New England Biolabs), JumpStart™ REDTaq™ DNA Polymerase (Sigma-Aldrich), PfuUltra™ Hotstart DNA Polymerase (Stratagene), PfuTurbo® Cx Hotstart DNA Polymerase (Stratagene), PrimeSTAR™ HS DNA Polymerase (Takara), Extensor Hi-Fidelity PCR Enzyme (ABgene), ACCUZYME™ DNA Polymerase (Bioline), SAHARA™ DNA Polymerase (Bioline), VELOCITY DNA Polymerase (Bioline), GeneChoice® AccuPOL™ DNA Polymerase (GeneChoice, Inc.), GeneChoice® UniPOL™ DNA Polymerase (GeneChoice, Inc.), Elongase Enzyme Mix (Invitrogen), Pfx50™ DNA Polymerase (Invitrogen), Phusion DNA Polymerase (New England Biolabs), KOD HiFi DNA Polymerase (Novagen), KOD XL DNA Polymerase (Novagen), Expand 20 kb PLUS Thermostable DNA polymerase mixture (Roche Applied Science), Expand High Fidelity PLUS Thermostable DNA polymerase mixture (Roche Applied Science), Expand High Fidelity Thermostable DNA polymerase mixture (Roche Applied Science), Expand Long Template Thermostable DNA polymerase mixture (Roche Applied Science), Easy-A™ High-Fidelity PCR Cloning Enzyme (Stratagene), EXL™ DNA Polymerase (Stratagene), Herculase® Enhanced DNA Polymerase (Stratagene), Herculase® II Fusion DNA Polymerase (Stratagene), Kapa LongRange™ DNA Polymerase (Kapa Biosystems), Kapa HiFi™ DNA Polymerase (Kapa Biosystems), Kapa2G™ Robust DNA Polymerase (Kapa

Biosystems), Kapa2G™ Robust HotStart DNA Polymerase (Kapa Biosystems), Kapa2G™ Fast DNA Polymerase (Kapa Biosystems), Kapa2G™ Fast HotStart DNA Polymerase (Kapa Biosystems), LA TAQ DNA Polymerase (Takara), Optimase DNA Polymerase (Transgenomic, Inc.), Exo- Pfu DNA Polymerase (Stratagene), HotMaster Taq DNA Polymerase (5 PRIME GmbH), HotTaq DNA Polymerase (Abnova Corporation), AmpliTaq Gold® DNA Polymerase (Applied Biosystems), Bst DNA Polymerase Lg Frag (New England Biolabs), MasterAmp™ Tfl DNA Polymerase (EPICENTRE Biotechnologies), Red Hot DNA Polymerase (ABgene), Thermoprime Plus DNA Polymerase (ABgene), Taq-red DNA Polymerase (AppliChem GmbH), BIO-X-ACT™ Long DNA Polymerase (Bioline), BIO-X-ACT™ Short DNA Polymerase (Bioline), Bioline HybriPol™ DNA Polymerase (Bioline), BioTherm Taq DNA Polymerase (eEnzyme LLC), EU-Taq DNA Polymerase (eEnzyme LLC), Synergy Taq DNA Polymerase (eEnzyme LLC), GeneChoice® RedPOL™ DNA Polymerase (GeneChoice, Inc.), AccuPrime™ GC-Rich DNA Polymerase (Invitrogen), PyroPhage® 3173 DNA Polymerase, Exo Minus (Lucigen), 9 Degrees North (Modified) DNA Polymerase (New England Biolabs), Terminator DNA Polymerase (New England Biolabs), Pwo DNA Polymerase (Roche Applied Science), Paq5000™ DNA Polymerase (Stratagene), YieldAce™ DNA Polymerase (Stratagene), e2TAK™ DNA Polymerase (Takara), o polimerasas de ADN de origen natural de *P. kodakaraensis*, *P. furiosus*, *T. gorgonarius*, *T. zilligii*, *T. litoralis* "Vent™", *P. GB-D "Deep Vent"*, *T. 9N-7*, *T. aggregans*, *T. barossii*, *T. fomicolans*, *T. celer*, *Pyrococcus sp. strain ST700*, *T. pacificus*, *P. abyssii*, *T. profundus*, *T. siculi*, *T. hydrothermalis*, *Thermococcus sp. strain GE8*, *T. thioreducens*, *P. horikoshii* or *T. onnurineus NA1*, *Thermococcus sp. 9°N-7*, *Thermococcus sp. GI-J*, *Thermococcus sp. MAR-13*, *Thermococcus sp. GB-C*, *Thermococcus sp. GI-H*, *Thermus aquaticus*, *Thermus thermophilus*, *Thermus caldophilus*, *Thermus filiformis*, *Thermus flavus*, *Thermotoga maritima*, *Bacillus stearothermophilus*, or *Bacillus caldote*.

[0046] En algunas realizaciones, al menos uno de los cebadores comprende un resto detectable radiológicamente o electromagnéticamente. Restos radiológicamente detectables incluyen isótopos radiactivos que emiten partículas detectables, tales como partículas beta o gamma, por ejemplo, ¹⁴C, ³H, ³²P, ³³P, ³⁵S, y ¹²⁵I. Restos electromagnéticamente detectables incluyen entidades químicas que interactúan con la radiación electromagnética (incluyendo absorbanza, emisión, o ambas) de una manera detectable, tales como cromóforos y fluoróforos, por ejemplo, fluoresceína, FAM, colorantes de cianina, colorantes de rodamina, etc.

[0047] En algunas realizaciones el método de generación de información de parámetros puede comprender la realización de al menos dos ensayos para generar información de parámetros, en donde cada ensayo comprende (a) proporcionar cebadores, en donde los cebadores de los al menos dos ensayos no son idénticos; (B) realizar una reacción de amplificación para producir un conjunto de productos amplificados; (C) resolver el conjunto de productos, al menos dos representaciones de tamaño del producto y la abundancia que se produce a partir de los al menos dos ensayos; y (d) derivar información acerca de si una secuencia de interruptor está presente en la al menos una región rica en CGG o en donde dentro de al menos una región rica en CGG se encuentra una secuencia de interruptor. En algunas realizaciones, los al menos dos ensayos usan cebadores orientados de forma opuesta que comprenden repeticiones CGG, CCG, GCG, CGC, GCC, o GGC. Es decir, un primer ensayo puede comprender proporcionar cebadores que comprenden un cebador que comprende repeticiones que están orientadas aguas arriba, y el segundo ensayo puede comprender proporcionar cebadores que comprenden un cebador que comprende repeticiones que está orientado aguas abajo, o viceversa. En algunas realizaciones, las muestras que comprenden al menos dos alelos diferentes del locus que comprende la región rica en CGG (debido a ser al menos uno de heterocigotos, aneuploide o mosaico para ese locus) puede ser analizado para determinar al menos un detalle con respecto al genotipo.

[0048] En diversas realizaciones, los métodos para generar información de parámetro se pueden utilizar para analizar una plantilla que comprende secuencias repetidas ricas en GC, tales como, por ejemplo, trinucleótidos CGG o CCG. Este análisis puede comprender la realización de una reacción de amplificación y la resolución de las estructuras de los productos amplificados a alta resolución. La alta resolución puede ser una resolución suficiente para distinguir los productos amplificados que contienen, por ejemplo, 20 frente a 21, o 20 frente a 22, 20 frente a 23, 20 frente a 24, o 20 frente a 25 repeticiones de trinucleótidos, o en algunas realizaciones, los productos amplificados que difieren en longitud por 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, o 18 nucleótidos o pares de bases. Ejemplos de técnicas que se pueden emplear en los métodos para llevar a cabo la Etapa de Resolución incluyen, sin limitación, electroforesis capilar y electroforesis en gel de poliacrilamida (PAGE; algunas realizaciones de página que se denominan comúnmente en la técnica como "ejecutar un gel de secuenciación"), o sistemas de microfluídica/electroforesis de tipo "lab-on-a-chip". Instrumentación para la realización de la electroforesis capilar se puede obtener de, por ejemplo, proveedores tales como Applied Biosystems (ABI), Beckman, Agilent, y Qiagen, y los instrumentos adecuados incluyen, sin limitación ABI 310, 3100, 3130/3130xl, o 3730/3730xl; Beckman P/ACE MDQ; Agilent 2100 Bioanalyzer; y Qiagen QIAxcel. El proceso de resolución de los productos amplificados por electroforesis puede implicar el uso de polímeros líquidos, incluyendo, por ejemplo, POP-7, POP-6, POP-5, o POP-4, todos los cuales se venden por Applied Biosystems. En algunas realizaciones, un polímero que puede resolver los productos que contienen aproximadamente 250, 300, 350, 400 o repeticiones de trinucleótidos más amplificados con precisión de acuerdo con el tamaño, tales como, por ejemplo, POP-4, se utiliza en la solución de los productos. Resolviendo el conjunto de los productos amplificados en alta resolución se puede lograr usando una máquina, por ejemplo, elegido de máquinas que comprenden una fuente de tensión (por ejemplo, una fuente de alimentación DC), máquinas que comprenden una fuente de presión (por ejemplo, una bomba), y las máquinas que comprenden una columna o capilar adecuado para realizar separaciones químicas. La máquina puede comprender más de uno de los componentes anteriores.

[0049] En algunas realizaciones, la solución de los productos amplificados a alta resolución lleva a la producción de una representación de tamaño del producto y la abundancia. Esta representación puede ser una imagen o gráfico que un experto en la técnica puede interpretar, visualmente o con la ayuda de instrumentación tal como, por ejemplo, un ordenador con el software adecuado o un densitómetro, para entender el tamaño y cantidad de los productos de la reacción. En algunas realizaciones, la representación es un electroferograma, fotografía, gráfica, trama, o autorradiograma. La representación se puede derivar o registrar de fotones o partículas beta emitidas por los productos o moléculas de colorante unidas a los productos; éstos pueden ser detectados, por ejemplo, fotográficamente o por vía electrónica, y procesados o desarrollados para generar la representación.

[0050] En algunas realizaciones, los métodos para la información de los parámetros de generación comprenden derivar información sobre el número de repeticiones CGG (es decir, el número de repeticiones CGG, o repeticiones de trinucleótidos en general, están dentro de una plantilla) a partir de las representaciones de tamaño del producto y la abundancia. Dicha derivación puede comprender contando el número de especies de productos amplificados observables en la representación para determinar el número de repetición (a partir del número de repeticiones comprendidas por el producto más pequeño, que puede, por ejemplo, ser de 4 o 5), o la estimación del número de repetición basado en la posición del producto amplificado más grande observable en la representación. En algunas realizaciones, la región rica en CGG está compuesta por el 5' UTR de *FMR1* o la 5' UTR de *FMR2*.

[0051] En algunas realizaciones, los métodos para la generación de información de los parámetros comprenden la detección de si los elementos de interruptor están presentes dentro de una región rica en CGG. En algunas realizaciones, los elementos de interruptor son trinucleótidos AGG. Detectar si estos están presentes se puede lograr, por ejemplo, mediante la determinación de posiciones en la plantilla, donde la unión del primer cebador se redujo sustancialmente, o mediante la determinación de un conjunto de longitudes en donde se reduce sustancialmente la cantidad de producto en comparación con longitudes vecinas. Por ejemplo, si el 10^o trinucleótido era AGG en una región con al menos 14 repeticiones totales de trinucleótidos y una reacción de amplificación se realizó implica el uso de un cebador que comprende 4 repeticiones CGG, los productos con repeticiones CGG 10, 11, 12, y 13 se esperaría a estar presente en una cantidad relativa reducida sustancialmente a los productos vecinos con 9 y 14 repeticiones. El grado en que se reduce la cantidad puede variar desde 25% a 95% o más, por ejemplo, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70 %, 75%, 80%, 85%, 90%, o 95%. El grado de reducción generalmente dependerá de la cigosidad de la muestra; por ejemplo, una muestra en la que la región rica en CGG es de un individuo heterocigoto para un alelo que comprende la región rica en CGG, o para una repetición AGG específica, puede mostrar una reducción en la cantidad de los productos correspondientes que van desde 25% a 75 %. Una muestra en la que la región rica en CGG es de un hemocigoto individuo o homocigoto para un alelo que comprende la región rica en CGG, o para una repetición de AGG específico, puede mostrar una reducción en la cantidad de los productos correspondientes que van desde 50% a 95% o más.

[0052] Una vez que la información de parámetros que caracteriza a una región de repetición o rica en GC se ha recogido, es teóricamente posible montar un genotipo utilizando interpretación manual. Sin embargo, esta interpretación manual puede ser un proceso lento y complejo, especialmente cuando el gen de interés comprende más de una repetición o región rica en GC y/o más de una interrupción. Por ejemplo, un gen *FMR1* o fragmento del mismo que comprende dos regiones de repetición CGG y dos interrupciones AGG requerirían la evaluación de dieciséis posibles soluciones antes de que el genotipo correcto podría ser identificado (ver más abajo para una explicación de las dieciséis combinaciones posibles). En consecuencia, en diversas realizaciones, se proporciona un método para la reconstrucción de genotipo automatizada usando enumeración exhaustiva. En algunas realizaciones, el uso de enumeración exhaustiva aumenta la exactitud y/o la velocidad de la reconstrucción de genotipos.

[0053] En diversas realizaciones, la enumeración exhaustiva comprende: (a) proporcionar información de parámetros con respecto a la longitud de una o más de repetición o regiones ricas en GC en un ácido nucleico y el número de interrupciones dentro de esas regiones, (b) utilizando la información con respecto a la longitud de una repetición o región rica en GC y la distancia desde el inicio de la región para cada interrupción (ya sea en la dirección de avance o de retroceso) para generar un conjunto de genotipos potenciales que comprende todas las disposiciones posibles de las interrupciones, y (c) la evaluación del conjunto de genotipos potenciales para determinar la solución de genotipo que satisface toda la información de parámetros. En las realizaciones en las que las interrupciones en la dirección hacia adelante se utilizan para construir los genotipos posibles, cada genotipo potencial en el conjunto se evalúa a continuación, utilizando la información con respecto a las interrupciones en la dirección inversa. En las realizaciones en las que las interrupciones en la dirección inversa se utilizan para construir los genotipos posibles, cada genotipo potencial en el conjunto se evalúa a continuación, utilizando la información con respecto a las interrupciones en la dirección de avance. En ciertas realizaciones, el genotipo potencial que coincide con la información de interrupción hacia adelante y hacia atrás se selecciona como la solución de genotipo.

[0054] En una realización, el análisis de enumeración exhaustiva puede ser representado por la fórmula, $\pi|C_i|$, que indica el producto del número de genotipos potenciales de los posibles alelos, donde $|C_i|$ es el número de genotipos candidatos para el alelo *i*. Como resultado, esta fórmula proporciona el número de genotipos posibles que pueden ser reconstruidos a partir de la información de parámetros (es decir, el conjunto de genotipos potenciales que pueden ser reconstruidos usando la longitud de la repetición o región rica en GC y el número de interrupciones dentro de la región en o bien el avance o direcciones inversas). En una realización, cada uno de los genotipos C_i potenciales generados

usando la información de interrupción hacia adelante se evalúa de forma individual utilizando la información sobre la interrupción inversa para seleccionar una solución de genotipo. En otra realización, cada uno de los genotipos C_i potenciales generados usando la información de interrupción inversa se evalúa individualmente utilizando la información de interrupción hacia adelante para seleccionar una solución de genotipo.

[0055] El método de enumeración exhaustiva se puede llevar a cabo usando un aparato que comprende un procesador (por ejemplo, un ordenador) programado para llevar a cabo el análisis de enumeración exhaustiva. En algunas realizaciones, el procesador está programado para recibir información con respecto a un ácido nucleico y a continuación, aplicar enumeración exhaustiva para reconstruir una solución de genotipo para el ácido nucleico. En algunas realizaciones, el aparato comprende también un monitor para información de entrada de visualización y/o la solución de genotipo. En algunas realizaciones, la solución de genotipo se almacena electrónicamente en el aparato, y/o es capaz de ser impresa para más usos de diagnóstico o terapéuticos.

[0056] Como se describe en más detalle en los ejemplos a continuación, el método de enumeración exhaustiva se puede utilizar, en algunas realizaciones, para reconstruir un genotipo de la región de repetición CGG en el gen *FMR1*. El 5' UTR de *FMR1* puede comprender una o más regiones de repetición CGG, cada uno de los cuales puede contener una o más interrupciones AGG dentro de la región. Cuando más de una interrupción AGG está presente, estos generalmente no ocurren de forma contigua (es decir, es raro encontrar una región de repetición CGG que comprende $(AGG)_n$, donde n es mayor que o igual a 2). Diversas técnicas para el análisis del gen *FMR1* o fragmentos de los mismos, tales como los métodos de PCR descritos en la solicitud de Estados Unidos N^{os} 2010/0209970, 2010/0243451, y 2012/0107824, a menudo dió información de parámetros con respecto a la longitud de la región de repetición CGG y la distancia en las direcciones de avance y retroceso a cada interrupción AGG dentro de la región de repetición CGG. En diversas realizaciones, esta información de parámetros *FMR1* se puede utilizar para generar una lista de todos los genotipos posibles que tienen la longitud de repetición CGG correcta y las interrupciones AGG correctas, ya sea en la dirección de avance o de retroceso. Usando enumeración exhaustiva, cada solución potencial a continuación, se puede evaluar la consistencia en las direcciones de avance y retroceso, y una solución de genotipo se presentan en la posición correcta para la interrupción AGG dentro de la secuencia CGG repetición se puede seleccionar que satisfaga tanto la reenviar y requisitos de distancia inversa, y, además, no coloque ninguna interrupción AGG cerca de otra interrupción AGG.

[0057] En ciertas realizaciones, la enumeración exhaustiva puede producir más de una solución de genotipo al reconstruir el genotipo. Los genotipos de solución identificada pueden analizarse adicionalmente. Por ejemplo, la enumeración exhaustiva puede generar más de un genotipo *FMR1* se presentan en la posición correcta para la interrupción AGG dentro de la secuencia de repetición CGG y satisfacer tanto los requisitos de avance como retroceso a distancia, y también no colocar cualquier interrupción AGG cerca a otra interrupción AGG. En algunas realizaciones, cuando múltiples genotipos de solución se identifican por enumeración exhaustiva, los genotipos pueden ser evaluados manualmente en contra de la información de parámetros para seleccionar una sola solución de genotipo que mejor se ajuste a los datos. Por ejemplo, los genotipos de soluciones *FMR1* pueden evaluarse comparando manualmente los genotipos en contra de la altura de pico y el patrón de picos reportados por electroferograma capilar en un ensayo de longitud CGG (véase, por ejemplo, los datos del ensayo de longitud CGG en las Figs. 3 y 4). En algunas realizaciones, el genotipo que mejor se ajusta la altura del pico y el patrón de picos generados en el ensayo de longitud CGG se pueden seleccionar como la solución de genotipo única.

[0058] En algunas realizaciones, cuando la enumeración exhaustiva reconstruye el genotipo mediante la producción de más de una solución de genotipo, el método de enumeración exhaustiva puede contener además los pasos adicionales diseñados para automatizar el análisis de los genotipos para seleccionar una única solución. En algunas realizaciones, las instrucciones de medios legibles por máquina (por ejemplo, enumeración exhaustiva de software) se pueden programar para generar datos de parámetros predictivos basados en cada solución de genotipo, adaptarse a esa información para algunos o todos los datos de parámetros originales, y seleccionar el genotipo donde los datos predictivos mejor se ajustan a los datos originales (es decir, el genotipo que minimiza la suma del error entre los datos predictivos y originales). Por ejemplo, cuando la enumeración exhaustiva genera más de una solución de genotipo *FMR1*, el software puede generar datos de ensayo de longitud CGG predictivos basados en cada genotipo, comparar esos datos predictivos a los datos de ensayo de longitud CGG reales, y seleccionar el genotipo que minimiza la suma del error entre los dos conjuntos de datos. Un ejemplo que utiliza este método para ajustar los datos de predicción de dos diferentes soluciones de genotipos *FMR1* identificadas por enumeración exhaustiva se muestra en la Fig. 4A-4B. En el ejemplo mostrado en la Fig. 4A-4B, el software seleccionaría el genotipo modelado en la Fig. 4A como la solución de genotipo única porque reduce la suma del error, en comparación con el error en la Fig. 4B.

[0059] En algunas realizaciones, una vez que una solución de genotipo se identifica usando enumeración exhaustiva, la solución puede evaluarse adicionalmente por comparación manual para confirmar que satisface todas las informaciones de parámetros que caracterizan el ácido nucleico (tales como la longitud total de la repetición o región rica en GC y la distancia en las direcciones de avance y retroceso a cualquier interrupción), o mediante la realización de cualquier otro ensayo confirmatorio conocido para un experto en la técnica (por ejemplo, enzima de digestión de restricción, secuenciación Sanger, o secuenciación de alto rendimiento). Por ejemplo, la presencia de interrupciones AGG dentro de un tracto de repetición CGG de *FMR1* puede ser confirmada por digestión de la muestra de ácido nucleico con la enzima de restricción *EciI* (New England Biolabs Inc., Ipswich, MA, EE.UU.).

II. MUESTRAS

[0060] Los métodos proporcionados en este documento se refieren a la determinación del genotipo de un ácido nucleico en una muestra. En diversas realizaciones, una muestra se obtiene de un animal no humano o humano. Por ejemplo, la muestra puede ser una muestra del paciente. Una "muestra paciente" es cualquier muestra biológica de un paciente. El término muestra incluye, pero no se limita a, fluidos biológicos tales como sangre, suero, plasma, orina, líquido cefalorraquídeo, lágrimas, saliva, linfa, fluido de diálisis, fluido de lavado, semen, y/o otras muestras líquidas, así como células y tejidos de origen biológico. Las células y los tejidos pueden incluir células bucales, colecciones de enjuague bucal, o células de la piel, incluyendo folículos pilosos. El término también incluye células aisladas de un humano o células derivadas, incluyendo las células en cultivo, sobrenadantes de células, y los lisados celulares. Se incluye, además, órganos o tejidos fluidos de cultivo derivado, muestras de biopsia de tejido, muestras de biopsia de tumores, muestras de heces y fluidos extraídos de tejidos fisiológicos, así como células disociadas de tejidos sólidos, secciones de tejido, y los lisados celulares. También puede incluir muestras de tejido sólido post-mortem, tales como los del cerebro. El término muestra también incluye cualquier otro espécimen celular o no celular obtenido a partir de un animal humano o no humano que comprende un ácido nucleico de interés. En algunas realizaciones, la muestra contiene menos de aproximadamente 80, 100, 150, 200, 500, 1000, 1500, 2000, 2500, 3000, 4000, o 5000 ng del ácido nucleico de interés.

[0061] En algunos casos, la muestra incluye uno o más ácidos nucleicos de interés. El ácido nucleico de interés puede ser ADN genómico. El ADN genómico u otro ácido nucleico de interés pueden separarse de otros componentes de ADN y no de ADN de la muestra antes de ser sometida a los métodos de la invención. Muchos métodos de purificación de ADN y la separación son conocidos en la técnica y pueden ser utilizados con los métodos descritos.

[0062] En algunas realizaciones, el ácido nucleico de interés en la muestra puede comprender los genes *FMR1* y/o *FMR2* o fragmentos de los mismos, o al menos parte del 5' UTR de *FMR1* y/o *FMR2* (Por ejemplo, una porción que comprende las repeticiones CGG de 5' UTR de *FMR1* o las repeticiones de CCG en el 5' UTR de *FMR2*). En ciertas realizaciones, el tamaño del ácido nucleico puede ser de aproximadamente 50, 100, 200, 300, 500, o 700 pb, o 1, 1,5, 2, 2,5, 3, 4, 5, 7, o 10 kb, o cualquier valor en el medio. En algunas realizaciones, el tamaño del ácido nucleico puede ser de entre 50 pb y 10 kb, 100 pb y 10 kb, 200 pb y 10 kb, 300 pb y 10 kb, 500 pb y 10 kb, 700 pb y 10 kb, 1 kb y 10 kb, 1,5 pb y 10 kb, 2 pb y 10 kb, 3 pb y 10 kb, 50 pb y 7 kb, 50 pb y 5 kb, 50 pb y 4 kb, 50 pb y 3 kb, 50 pb y 2 kb, 50 pb y 1,5 kb, 100 pb y 7 kb, 200 pb y 5 kb, o 300 pb y 4 kb.

III. Aparato de genotipificación y medio legible por máquina

[0063] En diversas realizaciones, se describe un aparato para su uso en la reconstrucción automatizada de un genotipo. En algunas realizaciones, el aparato comprende un procesador acoplado en comunicación a un dispositivo de memoria. En algunas realizaciones, las instrucciones ejecutables por máquina se almacenan en el dispositivo de memoria que, cuando son ejecutadas por el procesador, hacen que el procesador para llevar a cabo el análisis de enumeración exhaustiva. En ciertas realizaciones, las instrucciones ejecutables por máquina causan que el procesador (a) acepte la entrada de información de parámetros con respecto a la longitud total de una repetición o la región rica en GC, la distancia en la dirección de avance a las interrupciones en la región, y la distancia en la dirección inversa para cualquier interrupción de la región; (b) genere un conjunto de genotipos posibles que comprenden todas las disposiciones posibles de las interrupciones en la región de repetición o rica en GC; (c) evalúe el conjunto de genotipos potenciales para producir una solución de genotipo que satisfice todos los parámetros de información; y (d) almacene la solución de genotipo en el dispositivo de memoria o en cualquier otro dispositivo de memoria que se acople comunicativamente al procesador. En ciertas realizaciones, el aparato comprende además un monitor comunicativamente acoplado al dispositivo de procesador y memoria, en donde las instrucciones ejecutables por máquina almacenadas en el dispositivo de memoria instruyen al procesador para mostrar la solución de genotipo en el monitor. En algunas realizaciones, el aparato comprende además una impresora comunicativamente acoplada al dispositivo de procesador y memoria, en donde las instrucciones ejecutables por máquina almacenadas en el dispositivo de memoria instruyen al procesador para imprimir la solución de genotipo en la impresora.

[0064] En diversas realizaciones, el aparato usado para reconstruir un genotipo es capaz de aceptar la entrada de información de parámetros con respecto a un ácido nucleico (por ejemplo, la longitud de cualquier repetición o regiones ricas en GC y la distancia desde el inicio de las regiones para cada interrupción, ya sea en la dirección de avance o retroceso). En algunas realizaciones, el aparato está programado para usar la información de parámetros para reconstruir el genotipo del ácido nucleico usando enumeración exhaustiva. El aparato puede ser programado para visualizar y/o archivar el genotipo reconstruido. En algunas realizaciones, el aparato comprende un medio para la visualización y/o archivar el genotipo reconstruido.

[0065] En diversas realizaciones, un aparato descrito en este documento comprende un dispositivo de procesador y memoria, en donde el dispositivo de memoria contiene instrucciones legibles por máquina que instruyen el procesador para aceptar la entrada de información de parámetros con respecto a un ácido nucleico y realizar un análisis de enumeración exhaustiva, que puede ser representado por la fórmula $\pi|C_i|$, que indica el producto del número de genotipos potenciales de los posibles alelos, donde $|C_i|$ es el número de genotipos candidatos para el alelo i . Como resultado, esta fórmula proporciona el número de genotipos posibles que pueden ser reconstruidos a partir de la

información de parámetros (es decir, el conjunto de genotipos potenciales que pueden ser reconstruidos usando la longitud de la repetición o región rica en GC y el número de interrupciones dentro de la región ya sea en el avance o direcciones inversas). En una realización, las instrucciones legibles por máquina instruyen al procesador para generar una lista de genotipos C_i posibles utilizando la información de interrupción de avance o de retroceso y luego evaluar individualmente cada genotipo potencial frente a todo el conjunto de información de parámetros para identificar una solución de genotipo. En algunas realizaciones, el aparato comprende además un medio para mostrar la solución de genotipo (por ejemplo, un monitor para visualizar el genotipo visualmente, un medio de almacenamiento de datos para guardar el genotipo en un formato digital, y/o una conexión para transmitir la solución de genotipo a una impresora u otro dispositivo de almacenamiento o pantalla electrónica).

[0066] En algunas realizaciones, el aparato es un ordenador, en donde el ordenador comprende un procesador y un dispositivo de memoria que tiene código de ordenador almacenado en él, en donde el código de ordenador instruye al procesador para aceptar la entrada de información de parámetros con respecto a un ácido nucleico y a continuación, aplicar enumeración exhaustiva para reconstruir un genotipo para el ácido nucleico. En algunas realizaciones, el ordenador comprende también un monitor para información de entrada de visualización y/o el genotipo reconstruido. En algunas realizaciones, el genotipo reconstruido se almacena electrónicamente en el equipo y/o es capaz de ser impreso para más usos de diagnóstico o terapéuticos. En diversas realizaciones, el equipo comprende un dispositivo para permitir la interacción del usuario. Por ejemplo, el equipo puede comprender un teclado y/o un dispositivo señalador (por ejemplo, un ratón o una bola de seguimiento) que permite a un usuario (tal como un paciente, médico u otro profesional de la salud) introducir información de parámetros y/o para el acceso y manipular el genotipo reconstruido.

[0067] En diversas realizaciones, las instrucciones para llevar a cabo la enumeración exhaustiva se pueden almacenar en un aparato en un medio legible por máquina (por ejemplo, instrucciones ejecutables por máquina, software, código de ordenador, programas de ordenador, etc.). Por ejemplo, el medio legible por máquina puede comprender un código de ordenador almacenado en C++, JAVA, PERL, o cualquier otro formato adecuado para el código de ordenador. El medio legible por máquina puede proporcionar instrucciones al aparato para realizar la enumeración exhaustiva usando información de parámetros con respecto a un ácido nucleico. En diversas realizaciones, las instrucciones en el medio legible por máquina pueden instruir a un aparato para (a) recibir información de parámetros con respecto a la longitud total de una repetición o la región rica en GC, la distancia en la dirección de avance a las interrupciones en la región, y la distancia en la dirección inversa a las interrupciones en la región; (b) generar un conjunto de genotipos posibles que comprenden todas las disposiciones posibles de las interrupciones en la repetición o región rica en GC; (c) evaluar el conjunto de genotipos potenciales para producir una solución de genotipo que satisface todos los parámetros de información; y (d) almacenar la solución de genotipo en un dispositivo de memoria. En algunas realizaciones, las instrucciones en el medio legible por máquina instruyen el aparato para visualizar la solución de genotipo en un monitor. En algunas realizaciones, las instrucciones en el medio legible por máquina instruyen el aparato para imprimir la solución de genotipo en una impresora.

[0068] Las instrucciones almacenadas en un medio legible por máquina pueden ser cualesquiera códigos, símbolos, u otras señales que proporcionan instrucciones, información y/o datos que pueden ser utilizados por un aparato (por ejemplo, por un procesador en un ordenador). En algunas realizaciones, las instrucciones almacenadas en el medio de legible por máquina codifican un programa que instruye al aparato para recibir información de parámetros con respecto a un ácido nucleico, realizar un análisis de enumeración exhaustiva, y almacenar o transmitir un genotipo reconstruido para el ácido nucleico.

[0069] En algunas realizaciones, las instrucciones almacenadas en el medio legible por máquina instruyen el aparato para ejecutar un programa de análisis de enumeración exhaustiva, que puede ser representado por la fórmula $\pi|C_i|$, que indica el producto del número de genotipos potenciales de los posibles alelos, donde $|C_i|$ es el número de genotipos candidatos para el alelo i . Como resultado, esta fórmula proporciona el número de genotipos posibles que pueden ser reconstruidos a partir de la información de parámetros (es decir, el conjunto de genotipos potenciales que pueden ser reconstruidos usando la longitud de la región de repetición o rica en GC y el número de interrupciones dentro de la región en o bien el avance o direcciones inversas). En una realización, el programa genera una lista de genotipos posibles C_i usando la información de interrupción de avance o de retroceso y luego evalúa individualmente cada genotipo potencial contra el conjunto completo de información de parámetros para identificar una solución de genotipo. En algunas realizaciones, el programa incluye instrucciones para mostrar y/o archivar la solución de genotipo reconstruida (por ejemplo, para mostrar el genotipo en un monitor, para guardar el genotipo a un medio de almacenamiento de datos, y/o para transmitir la solución de genotipo a una impresora u otro almacenamiento electrónico o dispositivo de visualización).

[0070] En algunas realizaciones, las instrucciones almacenadas en el medio legible por máquina codifican una interfaz de usuario que proporciona una visualización gráfica en un monitor. En algunas realizaciones, la interfaz permite a un usuario introducir información de parámetros con respecto a un ácido nucleico (por ejemplo, permitiendo al usuario cargar un archivo de datos o permitiendo al usuario que introduzca información en los campos de visualización mostrados en la interfaz de usuario). En algunas realizaciones, la interfaz de usuario proporciona al usuario opciones para analizar la información de parámetros, tales como diversos métodos para mostrar y/o guardar los datos de entrada y/o genotipos reconstruidos (por ejemplo, mediante la visualización de los datos en el monitor del usuario, el envío de

datos a un dispositivo electrónico especificado o la dirección electrónica, impresión, y/o guardar los datos en una ubicación particular).

5 [0071] En diversas realizaciones, un genotipo reconstruido puede ser almacenado como datos en un medio de almacenamiento conectado físicamente al aparato (por ejemplo, en un dispositivo de memoria interna, tales como un disco duro en un ordenador) y/o almacenado en un dispositivo de almacenamiento remoto que es comunicativamente conectado al aparato (por ejemplo, por una conexión de intranet o internet cableada o inalámbrica y similares). En algunas realizaciones, la interfaz de usuario proporciona al usuario opciones para almacenar automáticamente el genotipo reconstruido en un lugar determinado, la impresión del genotipo, y/o enviar el genotipo a un dispositivo electrónico especificado o la dirección electrónica (por ejemplo, a la dirección de correo electrónico del profesional médico que solicitó la reconstrucción de genotipo).

IV. Métodos de uso

15 [0072] En diversas realizaciones, los métodos para detectar una mutación o genotipo que comprende una región rica en GC o de repetición, o para diagnosticar o tratar un trastorno genético asociado con una región rica en GC o de repetición se proporcionan, que comprende (1) obtener una muestra de un paciente; (2) aislar un ácido nucleico de la muestra que tiene una o más repeticiones o regiones ricas de GC, tales como una región que comprende repeticiones de CGG o CCG; (3) la recogida de información de parámetros a partir del ácido nucleico aislado; (4) que entra en la información de secuencia en un aparato programado para utilizar enumeración exhaustiva para reconstruir un genotipo a partir de la información de parámetros; y (5) utilizando el genotipo reconstruido para detectar una mutación o genotipo o para diagnosticar un trastorno genético asociado con una mutación en una repetición o la región rica en GC y/o predecir el riesgo de un trastorno genético en el paciente o un descendiente del paciente y/o hacer una decisión sobre el tratamiento adecuado basado en el genotipo reconstruido. Por ejemplo, el procedimiento puede comprender aislar una *FMR1* o *FMR2* ácido nucleico o fragmentos de los mismos a partir de una muestra del paciente, la determinación de información de parámetros para la *FMR1* o *FMR2* ácido nucleico, aplicando enumeración exhaustiva de la información de parámetros para generar un genotipo reconstruido para la región de repetición en *FMR1* o *FMR2*, y con el genotipo reconstruido para detectar una mutación o genotipo asociado con un trastorno genético *FMR1* o *FMR2*, o para diagnosticar y/o predecir el riesgo de decisiones y/o tratamiento de maquillaje en relación con un trastorno asociado.

[0073] Numerosos genes y regiones genómicas comprenden regiones de repetición o ricas en GC y están asociados con trastornos genéticos, lo que los hace posibles objetivos diagnósticos y terapéuticos. En consecuencia, en diversas formas de realización los métodos de enumeración exhaustiva descritos en el presente documento se pueden usar para reconstruir genotipos para estos loci genéticos y se pueden utilizar para diagnosticar, pronosticar, y/o guiar las decisiones de tratamiento de los trastornos genéticos asociados. En algunas realizaciones, los métodos de enumeración exhaustiva descritos en este documento pueden ser utilizados para reconstruir un genotipo para la región de repetición de los genes *FMR1* o *FMR2*. En alguna realización, estos genotipos reconstruidos pueden ayudar en el diagnóstico de FXS, FRAXE, FXTAS, FXPOI, y parkinsonismo susceptible a dopamina, que están asociados con la longitud de regiones de repetición CGG en el 5' UTR de *FMR1* y regiones de repetición CCG en el 5' UTR de *FMR2*. Por ejemplo, un genotipo reconstruido de *FMR1* tiene más de aproximadamente 45 repeticiones CGG, y particularmente un genotipo tiene más de aproximadamente 200 repeticiones CGG, en el 5' UTR se puede utilizar para diagnosticar FXS y trastornos asociados.

45 [0074] En realizaciones adicionales, los métodos de enumeración exhaustiva pueden usarse para detectar genotipos asociados con otros trastornos de regiones de repetición o ricas en GC, tales como ataxia espinocerebelosa tipo 1, ataxia espinocerebelosa tipo 2, ataxia espinocerebelosa tipo 3, ataxia espinocerebelosa tipo 6, ataxia espinocerebelosa tipo 7, ataxia espinocerebelosa tipo 8, ataxia de Friedrich, epilepsia mioclónica progresiva, la esclerosis lateral amiotrófica, distrofia miotónica, enfermedad de Huntington, atrofia muscular espinobulbar, atrofia dentatorubropalidolouisiana, y/o ataxia espinocerebelosa. Loci genéticos asociados con estas condiciones se conocen en la técnica e incluyen, sin limitación, *SCA1*, *SCA2*, *SCA3*, *CACNA1A*, *SCA7*, *SCA8*, *X25*, *CSTB*, *C90RF72*, *DMPK*, *ZNF9*, *HTT*, *AR*, *ATN1*, *ATXN1-3*, *SCA7*, *ATXN10*, *CACNA1A*, *SCA8*, *PPP2R2B*, y *TBP*. Véase, por ejemplo, Nat Genet. 1996 Mayo; 13 (1): 105-8; Nat Genet. 1996 Mayo; 13 (1): 109-13. Hiperexpansión y/o hipermetilación de las regiones de repetición ricas en GC y/o en estos loci están asociados con las enfermedades. Tabla 6 muestra ejemplos de loci genéticos que se pueden utilizar con los métodos de enumeración exhaustiva descritos en el presente documento, y la relación entre regiones ricas en GC y/o de repetición en esos genotipos loci y enfermedades o fenotipos.

60

65

Tabla 6

Enfermedad	Gen	Número de repetición		Posición de repetición	Variante de repetición
		Normal	Mutante		
Síndrome de X frágil	<i>FMR1</i>	(CGG)<45	(CGG)>200	5'-UTR	AGG
Retraso mental de X frágil (FRAXE)	<i>FMR2</i>	(CCG)<35	(CCG)>200	5'-UTR	CTG
Distrofia miotónica	<i>DMPK</i>	(CTG)<35	(CTG)>50	3'-UTR	CCG, CTC
Ataxia espinocerebelar tipo 8	<i>SCA8</i>	(CTG)<40	(CTG)>110	ARN antisentido	CCG CTA, CTC, CCA o CTT
Ataxia de Friedrich	<i>X25</i>	(GAA)<35	(GAA)>100	Intron 1	GGA, GAG
Atrofia muscular espinabulbar	<i>AR</i>	(CAG)<30	(CAG)>40	Codificación	
Enfermedad de Huntington	<i>IT15</i>	(CAG)<40	(CAG)>40	Codificación	
Atrofia palidoluisiana dentatorubral	<i>DRPLA</i>	(CAG)<35	(CAG)>50	Codificación	
Ataxia espinocerebelar tipo 1	<i>SCA1</i>	(CAG)<40	(CAG)>40	Codificación	CAT
Ataxia espinocerebelar tipo 2	<i>SCA2</i>	ataxia (CAG) <30	(CAG)>35	Codificación	CAA
Ataxia espinocerebelar tipo 3	<i>SCA3</i>	(CAG)<40	(CAG)>40	Codificación	
Ataxia espinocerebelar tipo 6	<i>CACNA1A</i>	(CAG)<20	(CAG)>20	Codificación	
Ataxia espinocerebelar tipo 7	<i>SCA7</i>	(CAG)<40	(CAG)>40	Codificación	Alelo normal no tiene interrupción
Epilepsia de mioclono progresiva tipo	<i>CSTB</i>	(C4GC4GCG (SEQ ID NO: 66))<3	(C4GC4GCG (SEQ ID NO: 66))>50	Promotor	

[0075] Por ejemplo, la enumeración exhaustiva puede ser utilizada para detectar genotipos asociados con los trastornos de *SCA1* o *SCA2*, tales como ataxia espinocerebelar tipos 1 y 2, que están asociados con la expansión de sus regiones de repetición CAG. Por ejemplo, la información de parámetros se puede proporcionar con respecto a la longitud total de una o más repeticiones CAG en los genes *SCA1* o *SCA2*, así como la distancia en las direcciones de avance y retroceso a las interrupciones CAT o CAA en las repeticiones CAG. La enumeración exhaustiva, utilizando la longitud total de la una o más repeticiones CAG y, o bien la distancia en la dirección hacia adelante o inversa para cualquier interrupción, se puede aplicar para generar un conjunto de genotipos potenciales para el gen *SCA1* o *SCA2*. Los genotipos potenciales pueden ser evaluados para determinar una solución de genotipo que satisface todas las informaciones del parámetro. La solución de genotipo identificada puede usarse para detectar una mutación o un genotipo, o para diagnosticar o ayudar en el diagnóstico, una mutación, genotipo, o trastorno relacionados con *SCA1* o *SCA2*.

Ejemplos

[0076] Los siguientes ejemplos sirven para ilustrar, y de ninguna manera limitar, la presente descripción.

Ejemplo 1: Enumeración exhaustiva de un locus CGG en *FMR1*

[0077] El análisis de electroforesis estándar de PCR y capilar (CE) del gen *FMR1* puede proporcionar información de parámetro que caracteriza un locus de repetición CGG, tales como: (1) la longitud total del locus CGG de repetición, (2) anclado A ("anch A") de mapeo, y (3) anclado T ("anch T") de mapeo. Los datos de longitud total revelan la longitud total de la región de repetición CGG (incluyendo cualquier interrupción AGG). Los datos de anch T indican las ubicaciones de AGG dentro del locus CGG en la dirección hacia adelante, mientras que los datos de anch A indican las ubicaciones de los AGG en la dirección inversa.

[0078] Tabla 1 mostrada a continuación es un ejemplo de un locus *FMR1* CGG que tiene una longitud de repetición total de 23 trinucleótidos, con dos interrupciones AGG. La Tabla 1 muestra los recuentos de posición en las direcciones de avance y retroceso desde el inicio y el final del locus de repetición CGG para cada una de las interrupciones AGG.

5

Tabla 1

10

15

1	2	3	4	5	6	7	8	9	10	11	12
CGG	CGG	CGG	CGG	AGG	CGG	CGG	CGG	CGG	CGG	CGG	CGG
23	22	21	20	19	18	17	16	15	14	13	12
13	14	15	16	17	18	19	20	21	22	23	
CGG	CGG	CGG	CGG	CGG	AGG	CGG	CGG	CGG	CGG	CGG	
11	10	9	8	7	6	5	4	3	2	1	
La Tabla 1 describe SEQ ID NO: 47.											

20

25

[0079] El informe EC para este ejemplo indicaría una longitud de repetición CGG global de {23}, correspondiente a la longitud del locus de repetición CGG. El ensayo Anch T informaría {5,18}, que corresponde a las posiciones de las interrupciones AGG dentro de la región de repetición CGG, contando en la dirección hacia adelante. El ensayo Anch A informaría {6,19}, que corresponde a las posiciones de las interrupciones AGG al contar en la dirección inversa. Teniendo en cuenta este informe EC, la configuración del locus de repetición CGG se puede reconstruir de forma relativamente fácil utilizando la longitud CGG total y cualquiera de los datos Anch A o Anch T. El genotipo puede ser escrito utilizando la notación compacta como (CGG)₄(AGG)(CGG)₁₂(AGG)(CGG)₅ (SEQ ID NO: 47), donde los subíndices representan el número de repeticiones CGG continuas antes de llegar a una interrupción AGG.

30

35

[0080] Mientras que el genotipo de un gen *FMR1* que tiene un único locus CGG se puede reconstruir manualmente con bastante facilidad, como se muestra arriba, la tarea de reconstrucción se vuelve más compleja cuando se trata de más de un locus de repetición CGG. Por ejemplo, como se discute en el Ejemplo 3, los investigadores necesitan más de 150 horas para reconstruir manualmente aproximadamente 1.000 muestras *FMR1*, lo que indica una tasa de una muestra por cada 10 minutos para reconstruir manualmente un solo genotipo de *FMR1* en ese experimento. Se requiere un tiempo adicional para realizar la garantía de calidad y comprobación de errores de los genotipos reconstruidos de forma manual. Incluso con el control de calidad, un riesgo de error del usuario es posible durante la reconstrucción manual, especialmente si el proceso se lleva a cabo por los técnicos de laboratorio tan bien entrenados que pueden ser necesarios para la pantalla a través de numerosas muestras de pacientes. En contraste, el software de enumeración exhaustiva pudo reconstruir genotipos durante casi 1000 muestras en ese experimento en menos de 10 minutos, se traduce en ahorros de tiempo de aproximadamente mil veces.

40

Ejemplo 2: Enumeración exhaustiva de dos loci CGG en *FMR1*

45

[0081] Es posible tener más de un locus de repetición CGG en un gen *FMR1*, y esto puede aumentar en gran medida el número de genotipos posibles que deben ser evaluados. Por ejemplo, un gen *FMR1* podría tener los siguientes dos loci CGG:

Locus 1: (CGG)₁₉(AGG)(CGG)₉(SEQ ID NO: 48)

Locus 2: (CGG)₁₀(AGG)(CGG)₄₁(SEQ ID NO: 49).

50

[0082] El informe completo EC para esta muestra se muestra en la Tabla 2.

Tabla 2

55

CGG_Locus	Longitud	{29,52}
Anch_T		{11, 20}
Anch_A		{10, 42}

60

[0083] La mera utilización de la información que se muestra en la Tabla 2, no es una tarea sencilla para reconstruir el genotipo *FMR1*. En particular, no está claro qué componentes del informe Anch T (o el informe Anch A) corresponden a las interrupciones en cualquiera de los loci de repetición de dos CGG. En su lugar, el software de enumeración exhaustiva se utiliza para identificar el genotipo correcto.

65

[0084] En primer lugar, el software enumera todos los genotipos posibles para cada uno de los dos loci utilizando la longitud de repetición CGG global y la información Anch T. A continuación, las soluciones son evaluadas para la consistencia y compatibilidad de auto informe. Auto consistencia pregunta si la solución potencial de genotipo tiene una o más secuencias AGG consecutivas y rechazan cualquier secuencia potencial que tiene aggs consecutivos. La compatibilidad de informe pregunta si la solución de genotipo potencial es plenamente compatible con todos los datos

en el informe de la EC. En otras palabras, la compatibilidad de informe pregunta si una posible solución es consistente con la longitud de los datos de locus CGG, Anch T, y Anch A. El genotipo que satisface los tres requisitos se selecciona como la solución de genotipo.

5 **[0085]** Usando la longitud de repetición CGG global y datos Anch T mostrados en la Tabla 2, el programa de enumeración exhaustiva generaría conjuntos de secuencias potenciales para loci CGG 1 y 2, como se muestra en las Tablas 3 y 4.

Tabla 3: Secuencias potenciales para locus CGG 1

10

(CGG) ₂₉ (SEQ ID NO: 50)	correspondiente a un genotipo que carece de interrupciones
(CGG) ₁₀ (AGG)(CGG) ₁₈ (SEQ ID NO: 51)	correspondiente a un genotipo que tiene una interrupción en la posición {11}
(CGG) ₁₉ (AGG)(CGG) ₉ (SEQ ID NO: 48)	correspondiente a un genotipo que tiene una interrupción en la posición {20}
(CGG) ₁₀ (AGG)(CGG) ₈ (AGG)(CGG) ₉ (SEQ ID NO: 52)	correspondiente a un genotipo que tiene dos interrupción en las posiciones {11, 20}

15

Tabla 4: Secuencias potenciales para locus CGG 2

20

(CGG) ₅₂ (SEQ ID NO: 53)	Correspondiente a un genotipo que carece de interrupciones
(CGG) ₁₀ (AGG)(CGG) ₄₁ (SEQ ID NO: 49)	Correspondiente a un genotipo que tiene una interrupción en la posición {11}
(CGG) ₁₉ (AGG)(CGG) ₃₂ (SEQ ID NO: 54)	Correspondiente a un genotipo que tiene una interrupción en la posición {20}
(CGG) ₁₀ (AGG)(CGG) ₈ (AGG)(CGG) ₃₂ (SEQ ID NO: 55)	Correspondiente a un genotipo que tiene dos interrupciones en las posiciones {11, 20}

25

30 **[0086]** La combinación de las Tablas 3 y 4 genera un conjunto de 16 genotipos potenciales derivados de los datos de EC para el gen *FMR1* que tiene loci de repetición 2 CGG, como se muestra en la Tabla 5.

Tabla 5

35

(CGG) ₂₉ (SEQ ID NO: 50) / (CGG) ₅₂ (SEQ ID NO: 53)
(CGG) ₂₉ (SEQ ID NO: 50) / (CGG) ₁₀ (AGG)(CGG) ₄₁ (SEQ ID NO: 49)
(CGG) ₂₉ (SEQ ID NO: 50) / (CGG) ₁₉ (AGG)(CGG) ₃₂ (SEQ ID NO: 54)
(CGG) ₂₉ (SEQ ID NO: 50) / (CGG) ₁₀ (AGG)(CGG) ₈ (AGG)(CGG) ₃₂ (SEQ ID NO: 55)
(CGG) ₁₀ (AGG)(CGG) ₁₈ (SEQ ID NO: 51) / (CGG) ₅₂ (SEQ ID NO: 53)
(CGG) ₁₀ (AGG)(CGG) ₁₈ (SEQ ID NO: 51) / (CGG) ₁₀ (AGG)(CGG) ₄₁ (SEQ ID NO: 49)
(CGG) ₁₀ (AGG)(CGG) ₁₈ (SEQ ID NO: 51) / (CGG) ₁₉ (AGG)(CGG) ₃₂ (SEQ ID NO: 54)
(CGG) ₁₀ (AGG)(CGG) ₁₈ (SEQ ID NO: 51) / (CGG) ₁₀ (AGG)(CGG) ₈ (AGG)(CGG) ₃₂ (SEQ ID NO: 55)
(CGG) ₁₉ (AGG)(CGG) ₉ (SEQ ID NO: 48) / (CGG) ₅₂ (SEQ ID NO: 53)
(CGG) ₁₉ (AGG)(CGG) ₉ (SEQ ID NO: 48) / (CGG) ₁₀ (AGG)(CGG) ₄₁ (SEQ ID NO: 49)
(CGG) ₁₉ (AGG)(CGG) ₉ (SEQ ID NO: 48) / (CGG) ₁₉ (AGG)(CGG) ₃₂ (SEQ ID NO: 54)
(CGG) ₁₉ (AGG)(CGG) ₉ (SEQ ID NO: 48) / (CGG) ₁₀ (AGG)(CGG) ₈ (AGG)(CGG) ₃₂ (SEQ ID NO: 55)
(CGG) ₁₀ (AGG)(CGG) ₈ (AGG)(CGG) ₉ (SEQ ID NO: 52) / (CGG) ₅₂ (SEQ ID NO: 53)
(CGG) ₁₀ (AGG)(CGG) ₈ (AGG)(CGG) ₉ (SEQ ID NO: 52) / (CGG) ₁₀ (AGG)(CGG) ₄₁ (SEQ ID NO: 49)
(CGG) ₁₀ (AGG)(CGG) ₈ (AGG)(CGG) ₉ (SEQ ID NO: 52) / (CGG) ₁₉ (AGG)(CGG) ₃₂ (SEQ ID NO: 54)
(CGG) ₁₀ (AGG)(CGG) ₈ (AGG)(CGG) ₉ (SEQ ID NO: 52) / (CGG) ₁₀ (AGG)(CGG) ₈ (AGG)(CGG) ₃₂ (SEQ ID NO: 55)

40

45

50

55 **[0087]** Cada uno de los genotipos en la Tabla 5 se evalúa para la consistencia y compatibilidad de auto informe. Por ejemplo, el primer genotipo potencial ((CGG)₂₉ (SEQ ID NO: 50) / (CGG)₅₂ (SEQ ID NO: 53)) es auto-consistente ya que el genotipo no contiene ninguna secuencia de AGG contiguos. Sin embargo, el genotipo genera un Anch T de {null} y un Anch A de {null} porque no hay interrupciones AGG, ya sea en loci, que no es consistente con Anch T de {11, 20} y Anch A de {10, 42} se muestra en el informe de la EC. Por lo tanto, esta solución potencial es rechazada como incompatible con el informe. Del mismo modo, la siguiente solución potencial, ((CGG)₂₉ (SEQ ID NO: 50) / (CGG)₁₀(AGG) (CGG)₄₁ (SEQ ID NO: 49)) es también auto consistente, pero no compatible con el informe, ya que genera un Anch T de {11} y un Anch A de {42}. Por lo tanto, el software también rechazaría esta solución potencial como incompatibles con el informe. El software sigue examinando cada uno de los 16 genotipos posibles y rechaza todos ellos a excepción de (CGG)₁₉(AGG) (CGG)₉ (SEQ ID NO: 48)/(CGG)₁₀(AGG) (CGG)₄₁ (SEQ ID NO: 49), que es auto consistente y compatible con el informe (que genera un Anch T de {11, 20} y un Anch A de {10, 42}, que corresponden a los datos Anch A y Anch T en el informe de la EC). Esta secuencia se identifica como la solución de genotipo.

65

Ejemplo 3: Aplicación de enumeración exhaustiva en 887 muestras *FMR1*

5 [0088] Para validar la precisión del software de enumeración exhaustiva, 887 muestras (de madres, padres y descendencia en familias vinculadas a los trastornos *FMR1*) fueron analizadas independientemente para el genotipo utilizando software de enumeración exhaustiva y en comparación con los resultados de la interpretación manual. Nolin et al., Am. J. Med. Gineta. Parte A 9999: 1-8 (2013). El software produce un ahorro significativo de tiempo, requiriendo menos de 10 minutos para reconstruir genotipos para todas las 887 muestras, mientras que los investigadores reconstruyeron manualmente las mismas muestras a un ritmo mucho más lento. En promedio, sólo un solo genotipo *FMR1* fue reconstruido en un período de tiempo similar (10'). Se requiere un tiempo adicional para realizar la garantía de calidad y comprobación de errores de los genotipos reconstruidos de forma manual.

15 [0089] Se obtuvieron las muestras utilizadas en el estudio como ADN_g purificado de la sangre de los donantes y se representaron los alelos en el rango intermedio/premutación normal, (es decir, longitudes de 35-69 repeticiones CGG), y las mutaciones completas.

20 [0090] Los datos de parámetros se generaron usando tres ensayos de PCR y los rastros del electroferograma capilar para resolver el estado AGG de cada alelo. Los ensayos incluyen una PCR cebada por repetición de CGG (Tassone et al., J. Mol Diagn 10: 43-49 (2008); Chen et al., J. Mol Diagn 12: 589-600 (2010)) y dos ensayos de PCR adicionales derivados de una técnica de PCR de lectura larga previamente validada (Filipovic-Sadic et al., Clin Chem 56: 399-408 (2010)) que resuelven las ambigüedades en el ensayo de repetición cebada. Las muestras de ADN se diluyeron a 20 ng/ml y se utilizó un total de 120 ng de ADN en cada reacción PCR. Señales de electroferograma se generaron para los productos de PCR usando un analizador 3500XL genético (Life Technologies, Carlsbad, CA). En general, las muestras fueron reunidas en grupos de 94 muestras y dos controles por placa de microtitulación. Usando tres termocicladores y uno 3500XL, el tiempo de respuesta para la adquisición de datos fue de 24-36 horas por lote de la muestra. Un total de 99,0% de las muestras (878/887) se identificó correctamente utilizando el software (cuando se compara con las reconstrucciones manuales), y 96,8% (859/887) de las muestras produjo una solución única correcta.

30 [0091] Un ejemplo de los datos de parámetros generados para una muestra en la que el genotipo se identificó correctamente utilizando el software de enumeración exhaustiva se muestra en la Fig. 1. El ensayo de CGG indicó una sola región de repetición CGG que tiene una longitud total de 42 repeticiones, mientras que el ensayo Anch_A indicó una interrupción AGG en la posición 42 y el ensayo Anch_T indicó una interrupción AGG en la posición 10. El genotipo producido por el software fue (CGG)₉AGG(CGG)₃₂ (SEQ ID NO: 56) y esto coincidía con la secuencia decodificada manualmente.

35 [0092] En 1% de las muestras analizadas, el software no produjo una sola solución de genotipo. En esas muestras, una evaluación adicional reveló una interrupción AGG que ocurre después de menos de 5 repeticiones CGG, lo que refleja una limitación de detección de los métodos de PCR utilizados para analizar la muestra en este experimento. El cebador directo usado en el ensayo Anch_A contenía la secuencia (CGG)₅A (SEQ ID NO: 42). Para las muestras que contienen una interrupción AGG después de repeticiones de menos de 5 CGG, el cebador no lo hizo recocido y un producto de PCR único no fue producido para ese tramo de la región CGG (el ensayo Anch_A no informó de una interrupción, sino más bien un tramo más largo de CGG repite con una longitud dependiendo de la ubicación de la siguiente interrupción AGG). Del mismo modo, el cebador inverso usado en el ensayo Anch_T contenía la secuencia (CGG) 4CCT (SEQ ID NO: 57) y por lo tanto se requiere un tramo de al menos 4 repeticiones CGG entre interrupciones AGG con el fin de recocido.

45 [0093] Por ejemplo, la Fig. 2 muestra los datos de PCR generados para una muestra que comprende la secuencia (CGG)₆AGG(CGG)₂AGG(CGG)₄₁ (SEQ ID NO: 58). El ensayo de repetición CGG identificó correctamente la longitud total de la región de CGG a 51 repeticiones. La caja en el ensayo CGG indica una caída de pico 8 en altura de pico, más grande que la caída de pico 5 normalmente vista en el ensayo de CGG cuando una sola interrupción AGG está presente. La caída de pico 5 normal es causada por la unión del cebador utilizado en el ensayo CGG insuficiente (que contiene una secuencia de repetición (CGG)₅ (SEQ ID NO: 59)) para la secuencia que contiene AGG en cada posición posible diana a lo largo del tramo 5 de repetición. En el ensayo Anch_A, el cebador directo no se unió en la región (CGG)₂, y por lo tanto el pico en la posición 42 no fue identificado, sólo un único pico en la posición 45. De igual forma, el cebador inverso en el ensayo Anch_T no se unió en la región (CGG)₂, y por lo tanto no mostró un pico en la posición 7, sólo uno en la posición 10. El software de enumeración exhaustiva, por lo tanto no podría generar una solución de genotipo correcta. Para abordar esta limitación, cebadores alternativos pueden ser diseñados de forma que no requieren una secuencia de repetición de longitud CGG similar, pero son todavía específicos para la región cerca de una interrupción AGG.

60 [0094] En un pequeño porcentaje de casos (menos del 4%), el software de enumeración exhaustiva genera más de una solución de genotipo. Se muestra un ejemplo de los datos de PCR de una muestra donde el software produjo dos genotipos en la Fig. 3. Los datos de ensayo de CGG muestran dos alelos, uno que tiene 29 repeticiones CGG y uno que tienen 36 repeticiones CGG. Las dos solución de genotipo reportadas por el software eran

65

Primera solución de genotipo:

Alelo 1 = (CGG)₉AGG(CGG)₉AGG(CGG)₉ (SEQ ID NO: 60)

5 Alelo 2 = (CGG)₉AGG(CGG)₉AGG(CGG)₆AGG(CGG)₉ (SEQ ID NO: 61)

Segunda solución de genotipo:

10 Alelo 1 = (CGG)₉AGG(CGG)₁₉ (SEQ ID NO: 62)

Alelo 2 = (CGG)₉AGG(CGG)₉AGG(CGG)₆AGG(CGG)₉ (SEQ ID NO: 61)

15 **[0095]** La reconstrucción de genotipo manual confirmó que la primera solución de genotipo era correcta. El genotipo correcto también se puede identificar por comparación de las alturas de los picos en los datos del electroferograma del ensayo CGG contra las soluciones de genotipo identificadas por el software de enumeración exhaustiva (véase el Ejemplo 4 a continuación para una explicación de la utilización de la altura del pico y el patrón de pico en la reconstrucción de genotipo).

20 **[0096]** Mientras que el software automatizado informó más de un genotipo en un pequeño porcentaje de los casos y, además, el análisis manual identificó una única solución, la carga de trabajo manual fue todavía significativamente reducida. En lugar de evaluar manualmente cada posible genotipo sugerido por los datos de PCR, los investigadores solamente tuvieron que evaluar manualmente el conjunto más limitado de genotipos de soluciones generadas por el software donde se identificaron inicialmente múltiples genotipos de solución. Además, pasos de automatización adicionales pueden ser añadidos al software para permitir la evaluación automatizada de la altura de pico CGG y el patrón de pico para identificar la solución de genotipo única correcta, como se discute en el Ejemplo 4 a continuación. Por lo tanto, aun cuando varios genotipos son reportados por el software, la cantidad de tiempo del operador manual necesaria para identificar el genotipo correcto solo, y el riesgo de error del operador, se reduce.

30 **Ejemplo 4: Resolución de muestras con más de una solución de genotipo**

35 **[0097]** En los casos experimentales donde el software de enumeración exhaustiva produce más de una solución, los inventores resolvieron el genotipo correcto mediante la evaluación del patrón de altura de pico producido por el ensayo CGG. Los datos de electroferograma de ensayo CGG crudo proporcionan una serie de picos de señal de intensidad decreciente a medida que aumenta la longitud del producto PCR, con una caída de los niveles de señal de fondo más allá de la longitud del alelo general. El patrón de altura de pico puede ser modelado utilizando una función polinómica como un ajuste cúbico. Cuando el alelo contiene una interrupción AGG, esto se indica por un baño en la altura del pico para 5 picos en el ensayo CGG, utilizando un cebador que es complementaria a una región que contiene 5 repeticiones CGG. La caída de pico 5 es causada por la unión del cebador utilizado en el ensayo de CGG (complementaria a una secuencia de repetición insuficiente (CGG)₅ (SEQ ID NO: 59)) para la secuencia que contiene AGG-en cada posición posible diana a lo largo del tramo 5 de repetición. Mediante la evaluación de este patrón de alturas de pico en el ensayo de longitud CGG, los investigadores fueron capaces de seleccionar manualmente el genotipo correcto para esas muestras de las 887 ensayadas en el Ejemplo 3 en donde el software de enumeración exhaustiva proporciona más de una secuencia de solución.

45 **[0098]** Esta evaluación manual también puede ser automatizada y se incorpora en el software de enumeración exhaustiva. Por ejemplo, para cada uno de los posibles genotipos K generados por enumeración exhaustiva, el software puede ser programado para generar una curva polinómica que mejor se ajusta a los datos de ensayo CGG y concuerda con el genotipo particular. El error (es decir, la diferencia entre la curva polinómica y los datos reales de ensayo CGG) se puede calcular para cada curva, y el software puede entonces seleccionar la curva que minimiza el error, que identifica la solución de genotipo correcta.

50 **[0099]** El error (E) en una curva se puede calcular utilizando la fórmula $E = [e_1, \dots, e_k, \dots, e_k]$, en donde e_k representa el error entre la curva polinómica y la altura del pico CGG real en una posición particular. El e_k minimizada para un genotipo puede ser representado por la fórmula $e_k = \min_d (X_k * P - O)$ en la que P es el parámetro de curva de vector polinomio, O los datos de los picos observados, y X_k representa la matriz de estructura de la solución para cada solución de genotipo identificada por el software de enumeración exhaustiva. La matriz de estructura de la solución es la concatenación de las matrices de estructura de alelo individuales: $X_a = [X_{a1} | X_{a2} | \dots]$. La matriz de estructura de alelo X es una concatenación por columnas de los índices de posición de picos y sus potencias más altas (hasta el grado deseado) para cada uno de los alelos, es decir $X_{a1} = [l(a) | a | a^2 | a^3 | \dots]$. Para cada alelo en la solución, un vector a se define como [5,6,... longitud de alelo] para los alelos sin intercalaciones; en caso de intercalaciones, la secuencia correspondiente de 5 picos se establece en cero. El parámetro de distancia d puede ser una norma L_p simple o una función diana de regresión robusta.

65 **[0100]** Un ejemplo que aplica este método automatizado se describe a continuación. Por ejemplo, los siguientes datos de PCR, el software de enumeración exhaustiva producirá dos genotipos de soluciones:

ES 2 759 023 T3

ensayo de longitud CGG = {30,40}

Anch_T_ensayo = {11,12,21}

5 Anch_A_ensayo = {10,20,29}.

[0101] Las dos soluciones de genotipo son:

Genotipo 1:

10

Alelo 1: (CGG)₁₀(AGG)(CGG)₉(AGG)(CGG)₉ (SEQ ID NO: 63)

Alelo 2: (CGG)₁₁(AGG)(CGG)₂₈ (SEQ ID NO: 64)

15

Genotipo 2:

Alelo 1: (CGG)₁₀(AGG)(CGG)₉(AGG)(CGG)₉ (SEQ ID NO: 63)

Alelo 2: (CGG)₁₁(AGG)(CGG)₈(AGG)(CGG)₁₉ (SEQ ID NO: 65)

20

[0102] Para determinar la solución correcta, los datos de altura de pico en la Tabla 7, tomados a partir de los datos crudos del electroferograma de ensayo CGG, se utiliza para proporcionar un mejor ajuste de la curva para cada uno de los dos genotipos identificados por el software:

25

Tabla 7

Ubicación	Altura de pico observada
5	6477
6	5922
7	5756
8	5503
9	5295
10	3193
11	2209
12	2052
13	2124
14	2252
15	4601
16	4596
17	4327
18	4301
19	3937
20	2441
21	1795
22	1745
23	1723
24	1974
25	3477
26	3639
27	3319
28	2890
29	1971
30	1278
31	255
32	153
33	547
34	1676
35	1709
36	1716

65

(Continuación)

Ubicación	Altura de pico observada
37	1609
38	1547
39	1360
40	997

5

10 **[0103]** Para cada uno de los 2 genotipos, el software genera matrices X_a . Los valores de los parámetros (P) para cada una de las dos soluciones que minimicen el error cuadrático medio robusto e_k entre el ajuste observado y mejor de cada matriz son $P1 = [2070,52, 116,22, -7,63, 0,10, 5545,94, -432,51, 16,75, -0,22]$; y $P2 = [-1361,32, 550,63, -20,60, 0,18, 9362,48, -967,56, 37,00, -0,45]$. Los propios errores cuadráticos medios son el $e_1= 57161,39$ y $e_2= 568727,9$. La mejor curva de ajuste para cada uno de los dos genotipos se muestra en la Fig. 4. Usando este resultado, el software

15 identifica el primer genotipo que el que tiene un error cuadrático inferior medio y por lo tanto es el genotipo correcto.

[0104] Los ejemplos anteriores están destinados a ilustrar y de ningún modo limitar la presente descripción. Otras realizaciones de los dispositivos y métodos descritos serán evidentes para los expertos en la técnica a partir de la consideración de la memoria y la práctica de los dispositivos y métodos descritos en este documento.

20 LISTADO DE SECUENCIAS

[0105]

25 <110> ASURAGEN, INC.

<120> GENOTIPIFICACIÓN COMPLETA DE *FMR1*

<130> 10256,0039-00304

30 <140>

<141>

<150> 61/674.167

35 <151> 2012-07-20

<160> 66

<170> PatentIn versión 3.5

40 <210> 1

<211> 22

<212> ADN

<213> Secuencia artificial

45 <220>

<221> fuente

<223>/nota = "Descripción de la Secuencia Artificial: cebador sintético"

50 <400> 1

cggtggaggg cgcctctga gc 22

<210> 2

<211> 25

55 <212> ADN

<213> Secuencia artificial

<220>

<221> fuente

60 <223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

<400> 2

caggcgctca gctccgttc ggttt 25

65 <210> 3

<211 > 25

ES 2 759 023 T3

<212> ADN
<213> Secuencia artificial

<220>
5 <221> fuente
<223>/nota = "Descripción de la Secuencia artificial: cebador sintético"

<400> 3
10 cagtcaggcg ctcagctccg ttctg 25

<210> 4
<211> 24
<212> ADN
<213> Secuencia artificial

15 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

20 <400> 4
tccggtggag ggccgcctct gagc 24

<210> 5
<211> 35
25 <212> ADN
<213> Secuencia artificial

<220>
30 <221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

<400> 5
ggttcggcct cagtcaggcg ctcagctccg ttctg 35

35 <210> 6
<211> 36
<212> ADN
<213> Artificial Secuencia

40 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia Artificial: cebador sintético"

45 <400> 6
gggttcggcc tcagtcaggc gctcagctcc gttctg 36

<210> 7
<211> 26
<212> ADN
50 <213> Secuencia artificial

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

55 <400> 7
gcgggccggg ggttcggcct cagtca 26

<210> 8
60 <211> 25
<212> ADN
<213> Secuencia artificial

<220>
65 <221> fuente

ES 2 759 023 T3

<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

5 <400> 8
cagcgggccc ggggttcggc ctcag 25

<210> 9
<211> 25
<212> ADN
<213> Secuencia artificial

10 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

15 <400> 9
gcagcgggcc ggggttcgg cctca 25

20 <210> 10
<211 > 25
<212> ADN
<213> Secuencia artificial

<220>
<221> fuente

25 <223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

<400> 10
gggccggggg ttcggcctca gtcag 25

30 <210> 11
<211> 26
<212> ADN
<213> Artificial Secuencia

35 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia Artificial: cebador sintético"

40 <400> 11
ggggttcggc ctcagtcagg cgctca 26

<210> 12
<211> 27
<212> ADN

45 <213> Secuencia artificial

<220>
<221> fuente

50 <223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

<400> 12
ctcagtcagg ggggttcggc cgctcag 27

55 <210> 13
<211> 30
<212> ADN
<213> artificial Secuencia

<220>

60 <221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

<400> 13
ggcgctcagc tccgttcgg ttcactcc 30

65

ES 2 759 023 T3

<210> 14
<211> 28
<212> ADN
<213> Secuencia artificial

5
<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

10 <400> 14
tcaggcgctc agctccgttt cggttca 28

<210> 15
<211> 26
<212> ADN
<213> Secuencia artificial

15
<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia Artificial: cebador sintético"

20 <400> 15
cacttccggt ggagggccgc ctctga 26

25 <210> 16
<211> 25
<212> ADN
<213> Secuencia artificial

30 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

35 <400> 16
ttccggtgga gggccgcctc tgagc 25

<210> 17
<211> 25
<212> ADN
<213> Secuencia artificial

40
<220>
<221> fuente
<223> /nota = "Descripción de la secuencia artificial: cebador sintético"

45 <400> 17
cgacttcca ccaccagctc ctcca 25

50 <210> 18
<211> 21
<212> ADN
<213> artificial Secuencia

55 <220>
<221> fuente
<223> /nota = "Descripción de artificial Secuencia: cebador sintético "

60 <400> 18
ggagcccgcc cccgagaggt g 21

<210> 19
<211> 21
<212> ADN
<213> Secuencia artificial

65

ES 2 759 023 T3

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

5 <400> 19 gggagccgc cccgagagg t 21

<210> 20
<211> 26
<212> ADN
10 <213> Secuencia artificial

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia Artificial: cebador sintético"

15 <400> 20
cgcaactcca ccaccagctc ctccat 26

<210> 21
<211> 23
<212> ADN
20 <213> Secuencia artificial

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: Synthetic cebador"

25 <400> 21
cgggagcccg ccccgagag gtg 23

30 <210> 22
<211> 23
<212> ADN
35 <213> Secuencia artificial

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

40 <400> 22
ccgggagccc gccccgaga GGT 23

<210> 23
<211> 24
45 <212> ADN
<213> Secuencia artificial

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

50 <400> 23
ccgggagccc gccccgaga ggtg 24

55 <210> 24
<211> 26
<212> ADN
<213> Secuencia artificial

60 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

65 <400> 24
cgccgggagc ccgccccga gaggtg 26

ES 2 759 023 T3

<210> 25
<211> 26
<212> ADN
<213> Secuencia artificial
5
<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia Artificial: cebador sintético"
10 <400> 25
gcgccgggag cccgccccg agaggt 26

<210> 26
<211> 25
15 <212> ADN
<213> Secuencia artificial

<220>
<221> fuente
20 <223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

<400> 26
cgccgggagc ccgccccga gaggt 25
25 <210> 27
<211> 28
<212> ADN
<213> Secuencia artificial

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"
30

<400> 27
35 gcgccattgg agccccgcac ttccacca 28

<210> 28
<211> 25
<212> ADN
40 <213> Secuencia artificial

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"
45

<400> 28
gcgccattgg agccccgcac ttcca 25
50 <210> 29
<211> 25
<212> ADN
<213> Secuencia artificial

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"
55

<400> 29
60 agcgccattg gagccccgca ctcc 25

<210> 30
<211> 25
<212> ADN
<213> Secuencia artificial
65
<220>

ES 2 759 023 T3

<221> fuente
<223> /nota = "Descripción de la Secuencia Artificial: cebador sintético"

5 <400> 30
cgccattgga gccccgact tccac 25

10 <210> 31
<211> 25
<212> ADN
<213> Secuencia artificial

15 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

20 <400> 31
ttggagcccc gcacttccac cacca 25

25 <210> 32
<211> 28
<212 > ADN
<213> Secuencia artificial

30 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

35 <400> 32
agccccgcac tccaccacc agctcctc 28

40 <210> 33
<211> 28
<212> ADN
< 213> Secuencia artificial

45 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

50 <400> 33
gagccccgca ctccaccac cagctcct 28

55 <210> 34
<211> 28
<212> ADN
<213> Secuencia artificial

60 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

65 <400> 34
cattggagcc cgcacttcc accaccag 28

70 <210> 35
<211> 30
<212> ADN
<213> Secuencia artificial

75 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia Artificial: cebador sintético"

80 <400> 35 cccgcacttc caccaccagc tcctcatct 30

85 <210> 36

ES 2 759 023 T3

<211> 30
<212> ADN
<213> Secuencia artificial

5 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

<400> 36
10 tagaaagcgc cattggagcc cgcacttc 30

<210> 37
<211 > 26
<212> ADN
15 <213> Secuencia artificial

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

20 <400> 37
aagcgccatt ggagccccgc acttc 26

<210> 38
25 <211> 36
< 212> ADN
<213> Secuencia artificial

<220>
30 <221> fuente
<223> /nota = "Descripción de Secuencia artificial: cebador sintético"

<400> 38
35 tcaggcgctc agctccgttt cggtttact tccggt 36

<210> 39
<211> 36
<212> ADN
<213> Secuencia artificial

40 <220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

45 <400> 39
agcgtctact gtctcggcac ttgccgccc cgcgcc 36

<210> 40
<211> 28
50 <212> ADN
<213> Secuencia artificial

<220>
<221> fuente
55 <223> /nota = "Descripción de la Secuencia Artificial: cebador sintético"

<400> 40
agctccgttt tcaggcgctc cggtttca 28

60 <210> 41
<211> 43
<212> ADN
<213> Secuencia artificial

65 <220>
<221> fuente

ES 2 759 023 T3

<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

<400> 41
tcaggcgctc agctccgttt cggtttcacg gcggcgcgcg cgg 43

5

<210> 42
<211> 16
<212> ADN
<213> Secuencia artificial

10

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: oligonucleótido sintético"

15

<400> 42
ggcggga cggcgggcggc 16

<210> 43
<211> 41
<212> ADN
<213> Secuencia artificial

20

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

25

<400> 43
aagcgccatt ggagccccgc acttccccgc cgccgccc g 41

30

<210> 44
<211> 44
<212> ADN
<213> Secuencia artificial

35

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: cebador sintético"

40

<400> 44
tcaggcgctc agctccgttt cggtttcacg gcggcgcgcg cgga 44

<210> 45
<211> 41
<212> ADN
<213> Secuencia artificial

45

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia Artificial: cebador sintético"

50

<400> 45
aagcgccatt ggagccccgc acttccccgc cgccgccc t 41

<210> 46
<211> 15
<212> ADN
<213> Secuencia artificial

55

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: oligonucleótido sintético"

60

<400 > 46
tccgcccgg ccgcc 15

65

<210> 47

ES 2 759 023 T3

<211> 69
 <212> ADN
 <213> Secuencia artificial

5 <220>
 <221> fuente
 <223> /nota = "Descripción de la Secuencia artificial: oligonucleótido sintético"

10 <400> 47
 cggcggcggc ggagggcggc gcggcggcgg cggcggcggc ggcggcggcgg gaggcggcgg 60
 cggcggcggc 69

15 <210 > 48
 <211> 87
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <221> fuente
 <223> /nota = "Descripción de la Secuencia artificial: oligonucleótido sintético"

25 <400> 48
 cggcggcggc ggcggcggcgg gcggcggcgg cggcggcggc ggcggcggcgg gcggcggagg 60
 cggcggcggc ggcggcggcgg gcggcggcgg 87

30 <210> 49
 <211> 156
 <212> ADN
 <213> Secuencia artificial

35 <220>
 <221> fuente
 <223> /nota = "Descripción de la Secuencia Artificial: polinucleótido sintético"

40 <400> 49
 cggcggcggc ggcggcggcgg gcggcggcgg aggcggcggc ggcggcggcgg gcggcggcgg 60
 cggcggcggc ggcggcggcgg gcggcggcgg cggcggcggc ggcggcggcgg gcggcggcgg 120
 cggcggcggc ggcggcggcgg gcggcggcgg cggcggcgg 156

45 <210> 50
 <211> 87
 <212> ADN
 <213> Secuencia artificial

50 <220>
 <221> fuente
 <223> /nota = "Descripción de la Secuencia artificial: oligonucleótido sintético"

55 <400> 50
 cggcggcggc ggcggcggcgg gcggcggcgg cggcggcggc ggcggcggcgg gcggcggcgg 60
 cggcggcggc ggcggcggcgg gcggcggcgg 87

60 <210> 51
 <211> 87
 <212> ADN
 <213 > Secuencia artificial

65 <220>

ES 2 759 023 T3

	<221> fuente	
	<223> /nota = "Descripción de la Secuencia artificial: oligonucleótido sintético"	
	<400> 51	
5		
	cgggcggcggc ggcggcggcg gcggcggcggc aggcggcggc ggcggcggcg gcggcggcgg	60
	cgggcggcggc ggcggcggcg gcggcggc	87
10	<210> 52	
	<211> 87	
	<212> ADN	
	<213> Secuencia artificial	
15	<220>	
	< 221> fuente	
	<223> /nota = "Descripción de la Secuencia artificial: oligonucleótido sintético"	
	<400> 52	
20		
	cgggcggcggc ggcggcggcg gcggcggcggc aggcggcggc ggcggcggcg gcggcggcgg	60
	cgggcggcggc ggcggcggcg gcggcggc	87
25	<210> 53	
	<211> 156	
	<212> ADN	
	<213> Secuencia artificial	
30	<220>	
	<221> fuente	
	<223> /nota = "Descripción de la Secuencia Artificial: polinucleótido sintético"	
35	<400> 53	
	cggcggcggc ggcggcggcg gcggcggcggc cggcggcggc ggcggcggcg gcggcggcgg	60
	cggcggcggc ggcggcggcg gcggcggcggc cggcggcggc ggcggcggcg gcggcggcgg	120
40	cggcggcggc ggcggcggcg gcggcggcggc cggcggc	156
	<210> 54	
	<211> 156	
45	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<221> fuente	
50	<223> /nota = "Descripción de la Secuencia artificial: polinucleótido sintético"	
	<400> 54	
	cggcggcggc ggcggcggcg gcggcggcggc cggcggcggc ggcggcggcg gcggcggcgg	60
55	cggcggcggc ggcggcggcg gcggcggcggc cggcggcggc ggcggcggcg gcggcggcgg	120
	cggcggcggc ggcggcggcg gcggcggcggc cggcggc	156
60	<210> 55	
	<211> 156	
	<212> ADN	
	<213 > Secuencia artificial	
65	<220>	
	<221> fuente	

ES 2 759 023 T3

<223> /nota = "Descripción de la Secuencia artificial: polinucleótido sintético"

<400> 55

5	<p>cggcggcggc ggcggcggcg gcggcggcgg aggcggcggc ggcggcggcg gcggcggagg</p> <p>cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg</p> <p>cggcggcggc ggcggcggcg gcggcggcgg cggcgg</p>	<p>60</p> <p>120</p> <p>156</p>
---	---	---------------------------------

10

<210> 56
 <211> 126
 <212> ADN
 <213> Secuencia artificial

15

<220>
 <221> fuente
 <223> /nota = "Descripción de la Secuencia Artificial: polinucleótido sintético"

20

<400> 56

20	<p>cggcggcggc ggcggcggcg gcggcggagg cggcggcggc ggcggcggcg gcggcggcgg</p> <p>cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg</p> <p>cggcgg</p>	<p>60</p> <p>120</p> <p>126</p>
----	--	---------------------------------

25

<210> 57
 <211> 15
 <212> ADN
 <213> Secuencia artificial

30

<220>
 <221> fuente
 <223> /nota = "Descripción de la Secuencia artificial: oligonucleótido sintético"

35

<400> 57
 ccgccgccgc cgcct 15

40

<210> 58
 <211> 153
 <212> ADN
 <213> Secuencia artificial

45

<220>
 <221> fuente
 <223> /nota = "Descripción de la Secuencia artificial: polinucleótido sintético"

50

<400> 58

50	<p>cggcggcggc ggcggcggag gcggcggagg cggcggcggc ggcggcggcg gcggcggcgg</p> <p>cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg</p> <p>cggcggcggc ggcggcggcg gcggcggcgg cgg</p>	<p>60</p> <p>120</p> <p>153</p>
----	--	---------------------------------

55

<210> 59
 <211> 15
 <212> ADN
 <213> Secuencia artificial

60

<220>
 <221> fuente
 <223> /nota = "Descripción de la Secuencia artificial: oligonucleótido sintético"

65

<400> 59

ES 2 759 023 T3

	cggcggcggc ggcgg	15	
	<210> 60		
	<211> 87		
5	<212> ADN		
	<213> Secuencia artificial		
	<220>		
	<221> fuente		
10	<223> /nota = "Descripción de la Secuencia Artificial: oligonucleótido sintético"		
	<400> 60		
	cggcggcggc ggcggcggcg ggcggcggcg aggcggcggc ggcggcggcg ggcggcggcg	60	
15	cggcggcggc ggcggcggcg ggcggcgg	87	
	<210> 61		
	<211> 108		
20	<212> ADN		
	<213> Secuencia artificial		
	<220>		
	<221> fuente		
25	<223> /nota = "Descripción de la Secuencia artificial: polinucleótido sintético"		
	<400> 61		
	cggcggcggc ggcggcggcg ggcggcggcg aggcggcggc ggcggcggcg ggcggcggcg	60	
30	cggcggcggc ggcggcggcg ggcggcggcg aggcggcggc ggcggcggcg	108	
	<210> 62		
	<211> 87		
35	<212> ADN		
	<213> Secuencia artificial		
	<220>		
	<221> fuente		
40	<223> /nota = "Descripción de la Secuencia artificial: oligonucleótido sintético"		
	<400> 62		
	cggcggcggc ggcggcggcg ggcggcggcg aggcggcggc ggcggcggcg ggcggcggcg	60	
45	cggcggcggc ggcggcggcg ggcggcgg	87	
	<210> 63		
	<211> 90		
50	<212> ADN		
	<213> Secuencia artificial		
	<220>		
	<221> fuente		
55	<223> /nota = "Descripción de la Secuencia artificial: oligonucleótido sintético"		
	<400> 63		
	cggcggcggc ggcggcggcg ggcggcggcg aggcggcggc ggcggcggcg ggcggcggcg	60	
60	aggcggcggc ggcggcggcg ggcggcggcg	90	
	<210> 64		
	<211> 120		
65	<212> ADN		
	<213> Secuencia artificial		

ES 2 759 023 T3

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia Artificial: polinucleótido sintético"

5 <400> 64

cggcggcggc ggcggcggcg gcggcggcgg cggaggcggc ggcggcggcg gcggcggcgg 60

10 cggcggcggc ggcggcggcg gcggcggcgg cggaggcggc ggcggcggcg gcggcggcgg 120

<210> 65
<211> 120
<212> ADN
15 <213> Secuencia artificial

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: polinucleótido sintético"

20 <400> 65

cggcggcggc ggcggcggcg gcggcggcgg cggaggcggc ggcggcggcg gcggcggcgg 60

25 aggcggcggc ggcggcggcg gcggcggcgg cggaggcggc ggcggcggcg gcggcggcgg 120

<210> 66
<211> 12
<212> ADN
30 <213 > Secuencia artificial

<220>
<221> fuente
<223> /nota = "Descripción de la Secuencia artificial: oligonucleótido sintético"

35 <400> 66
ccccgccccg cg 12

40

45

50

55

60

65

REIVINDICACIONES

1. Un método de reconstrucción automatizada de un genotipo, que comprende

- 5 a. proporcionar un ácido nucleico aislado de una muestra del paciente, donde el ácido nucleico tiene al menos una región rica en GC de repetición de nucleótidos que comprende una o más interrupciones;
- b. determinar la información de parámetro para el ácido nucleico, en donde determinar la información de parámetro comprende la medición de la longitud total de la al menos una región de repetición de nucleótidos rica en GC, la distancia desde el inicio de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región, y la distancia desde el extremo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región; y
- 10 c. utilizando un aparato para llevar a cabo la enumeración exhaustiva automatizada en la información de parámetros para generar un genotipo reconstruido, en donde la enumeración exhaustiva comprende
 - 15 i) utilizando la longitud total de la al menos una región de repetición de nucleótidos rica en GC y, o bien la distancia desde el inicio de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región o la distancia desde el extremo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región para generar un conjunto de genotipos potenciales que comprende todas las disposiciones posibles de la una o más interrupciones en la región; y
 - 20 ii) evaluar el conjunto de genotipos potenciales para determinar una solución de genotipo que satisface toda la información del parámetro.

25 2. El método de la reivindicación 1, en donde la información de parámetros se determina por reacción en cadena de la polimerasa.

30 3. El método de la reivindicación 1, en donde el aparato comprende un procesador y un dispositivo de memoria acoplado comunicativamente al procesador, en donde el dispositivo de memoria se ha almacenado en el mismo instrucciones ejecutables por máquina que, cuando se ejecutan por el procesador, hacen que el procesador reciba información de parámetros y realice un análisis de enumeración exhaustiva.

35 4. El método de la reivindicación 1, que comprende además confirmar el genotipo reconstruido comparando manualmente el genotipo para la información de parámetros, mediante la realización de una digestión de restricción, o mediante la secuenciación del ácido nucleico que tiene al menos una región de repetición de nucleótidos rica en GC.

5. El método de la reivindicación 1, en donde el genotipo reconstruido es:
- 40 a. el gen de *FMR1* o un fragmento del mismo, en donde la muestra comprende un ácido nucleico *FMR1* o un fragmento del mismo, y en donde la información de parámetros comprende determinar la longitud total de al menos una región de repetición CGG en el ácido nucleico *FMR1* o fragmento, la distancia desde el inicio de al menos una región de repetición CGG a una o más interrupciones AGG en la región, y la distancia desde el extremo de al menos una región de repetición CGG a una o más interrupciones AGG en la región, o
 - 45 b. el gen *FMR2* o un fragmento del mismo, en donde la muestra comprende un ácido nucleico *FMR2* o un fragmento del mismo, y en donde la información de parámetros comprende determinar la longitud total de al menos una región de repetición CCG en el ácido nucleico *FMR2* o fragmento del mismo, la distancia desde el inicio de la región de repetición al menos un CCG a una o más interrupciones AGG en la región, y la distancia desde el extremo de al menos una región de repetición CCG a una o más interrupciones AGG en la región.

50 6. El método de la reivindicación 5, que comprende además usar el genotipo reconstruido para detectar un genotipo asociado con un trastorno o fenotipo *FMR1* o *FMR2* en un paciente, o para detectar un riesgo de un trastorno o fenotipo *FMR1* o *FMR2* en la descendencia del paciente.

55 7. El método de la reivindicación 6, en donde el trastorno *FMR1* o *FMR2* es el síndrome de X frágil (FXS), síndrome X frágil E (FRAXE), síndrome de temblor/ataxia asociado a X frágil (FXTAS), insuficiencia ovárica primaria relacionada con X frágil (FXPOI), o parkinsonismo sensible a dopamina.

8. Un aparato para la reconstrucción automatizada de un genotipo, que comprende

- 60 a. un procesador; y
- b. un dispositivo de memoria acoplado comunicativamente al procesador, el dispositivo de memoria que tiene almacenado en el mismo instrucciones ejecutables por máquina que, cuando se ejecuta por el procesador, hacen que el procesador
 - 65 i. reciba información de parámetros para un ácido nucleico, donde el ácido nucleico tiene al menos una región de repetición de nucleótidos rica en GC que comprende una o más interrupciones; y en

donde la información de parámetros comprende la longitud total de la al menos una región de repetición de nucleótidos rica en GC, la distancia desde el comienzo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región, y la distancia desde el extremo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región;

ii. genere un conjunto de genotipos posibles que comprenden todas las disposiciones posibles de la una o más interrupciones en la región mediante la realización de la enumeración exhaustiva automatizada sobre la información de parámetros;

iii. evalúe el conjunto de genotipos potenciales para producir una solución de genotipo que satisface todos los parámetros de información; y

iv. almacene la solución de genotipo en el dispositivo de memoria.

9. El aparato de la reivindicación 8, que comprende además un monitor comunicativamente acoplado al dispositivo de procesador y memoria, en donde las instrucciones ejecutables por máquina almacenadas en el dispositivo de memoria instruyen al procesador para mostrar la solución de genotipo en el monitor.

10. El aparato de la reivindicación 8, que comprende además una impresora comunicativamente acoplada al dispositivo de procesador y memoria, en donde las instrucciones ejecutables por máquina almacenadas en el dispositivo de memoria instruyen al procesador para imprimir la solución de genotipo en la impresora.

11. Un medio legible por máquina que comprende instrucciones ejecutables por máquina que, cuando son ejecutadas por un procesador, hace que el procesador

a. reciba información de parámetros para un ácido nucleico, donde el ácido nucleico tiene al menos una región de repetición de nucleótidos rica en GC que comprende una o más interrupciones; y en donde la información de parámetros comprende la longitud total de al menos la región de repetición de nucleótidos rica en GC, la distancia desde el comienzo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región, y la distancia desde el extremo de la al menos una región de repetición de nucleótidos rica en GC a la una o más interrupciones en la región;

b. genere un conjunto de genotipos posibles que comprenden todas las disposiciones posibles de las una o más interrupciones en la región mediante la realización de enumeración exhaustiva automatizada en la información de parámetros;

c. evalúe el conjunto de genotipos potenciales para producir una solución de genotipo que satisface todos los parámetros de información; y

d. almacene la solución de genotipo en un dispositivo de memoria.

12. El medio legible por máquina de la reivindicación 11, en donde las instrucciones ejecutables por máquina instruyen al procesador para visualizar la solución de genotipo en un monitor.

13. El medio legible por máquina de la reivindicación 11, en donde las instrucciones ejecutables por máquina instruyen al procesador para imprimir la solución de genotipo en una impresora.

14. Un kit que comprende el aparato de una cualquiera de las reivindicaciones 8-10, o el medio legible por máquina de una cualquiera de las reivindicaciones 11-13, e instrucciones para usar el aparato o medio legible por máquina para reconstruir un genotipo.

15. El kit de la reivindicación 14, en donde el kit se usa para detectar un genotipo asociado con un trastorno o fenotipo *FMR1* o *FMR2* en un paciente, o para detectar un riesgo de un trastorno o fenotipo *FMR1* o *FMR2* en la descendencia del paciente.

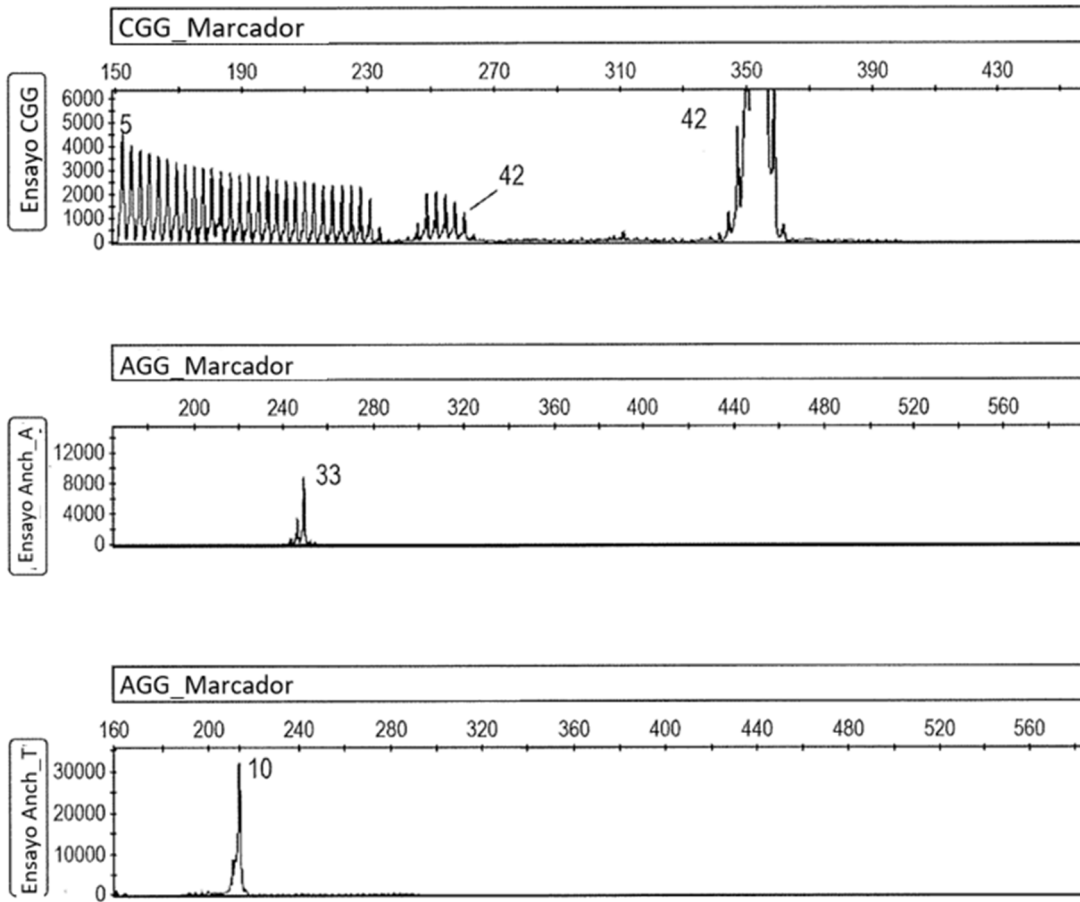


FIG. 1

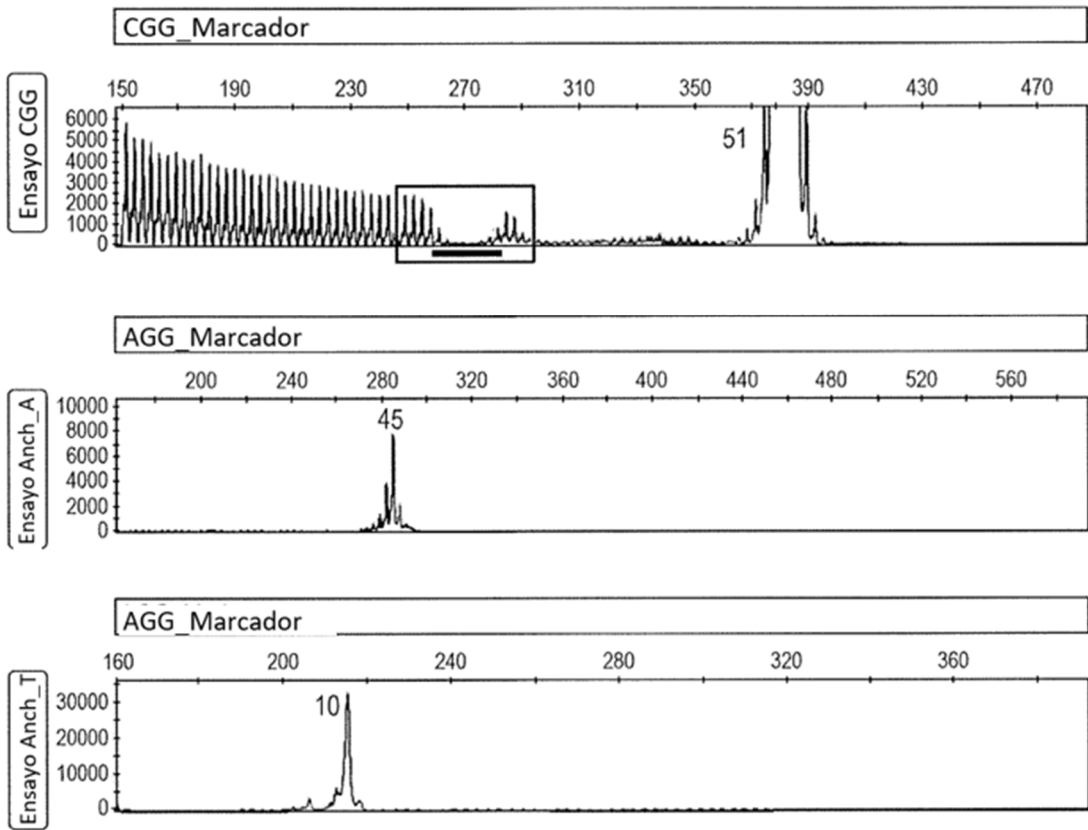


FIG. 2

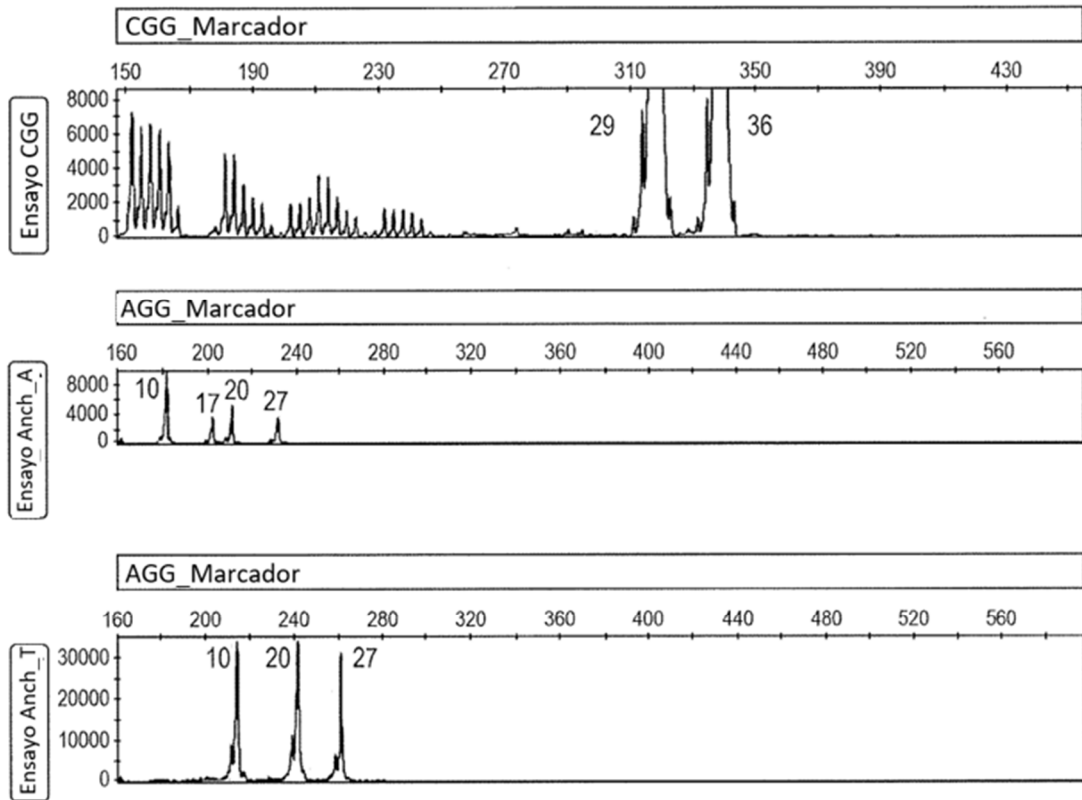


FIG. 3

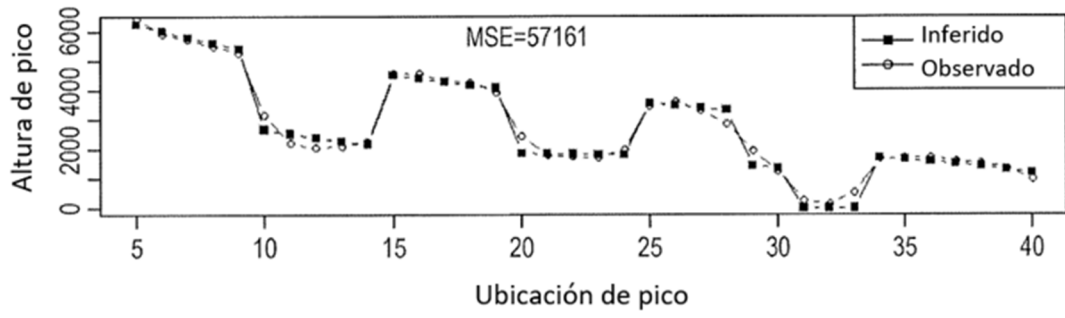


FIG. 4A

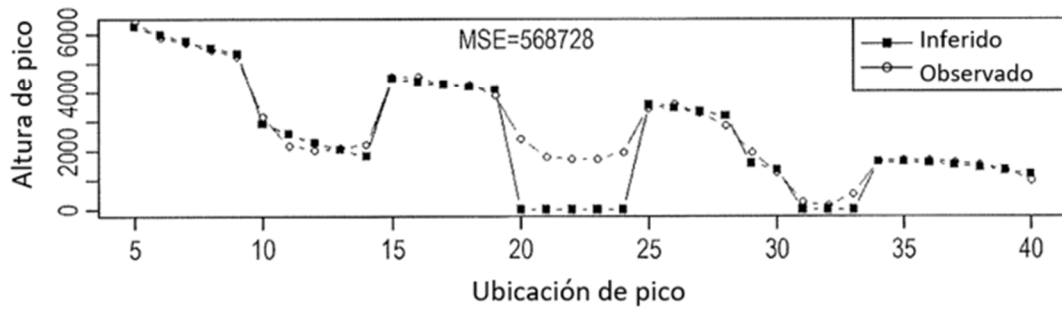


FIG. 4B