(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 708 376**

(51) Int. CI.:

C12Q 1/68 (2006.01)
G06F 19/18 (2006.01)
G06F 19/22 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **23.06.2016 E 16734594 (1)**

(97) Fecha y número de publicación de la concesión europea: **07.11.2018 EP 3283647**

(54) Título: **Un método para detección prenatal no-invasiva de aneuploidía cromosómica fetal de sangre materna**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
09.04.2019

(73) Titular/es:

TRISOMYTEST, S.R.O. (100.0%)
Ilkovicova 8
841 04 Bratislava - Karlova Ves, SK

(72) Inventor/es:

DURIS, FRANTISEK;
BUDIS, JAROSLAV;
SZEMES, TOMÁS y
MINÁRIK, GABRIEL

(74) Agente/Representante:

EZCURRA ZUFIA, Maria Antonia

DESCRIPCIÓN

Un método para detección prenatal no-invasiva de aneuploidía cromosómica fetal de sangre materna

Campo de la invención

La invención se refiere al campo de la imagen y diagnóstico prenatal no invasivo. La invención facilita un método para la detección de la presencia o ausencia de aneuploidías, concretamente trisomía del cromosoma 21, 18 y 13, en fetos desde la muestra de sangre tomada de la madre en una etapa temprana del embarazo.

Antecedentes

Hoy en día, las pruebas prenatales son una parte integrante de la práctica obstétrica. El principal objetivo de las pruebas prenatales es la imagen de aneuploidías fetales, tales como la trisomía del cromosoma 21 (T21, síndrome de Down), trisomía 18 (T18, síndrome Edwards), y trisomía 13 (T13, síndrome Patau). Aunque la mayor parte de los fetos con aneuploidía concluyen en terminación durante el desarrollo del feto, T21 tiene la tasa de supervivencia más alta, por lo tanto, la detección prenatal del T21 se considera la prueba genética prenatal más importante con el objetivo de la imagen y diagnóstico prenatal. Hay disponibles pruebas prenatales invasivas fiables, sin embargo, debido al riesgo de su naturaleza, se llevan a cabo en la actualidad sólo en embarazos de alto riesgo. Desarrollar un método fiable para pruebas prenatales no invasivas (PPNI) para aneuploidías fetales, en concreto trisomías, parece ser un reto reciente de la mayor importancia en el cuidado prenatal.

Desde el punto de vista médico, también otras trisomías, por ejemplo, la trisomía del cromosoma 8, 9, 16 o 22 podrían ser interesantes. Asimismo, pruebas prenatales para monosomías fetales, por ejemplo, monosomía del cromosoma X (síndrome Turner), podría ser importante.

El descubrimiento de ADN fetal libre de células circulante (cffDNA) en sangre materna (Lo et al., 1997) ha ofrecido la posibilidad de desarrollar procesos no invasivos que usen ácidos nucleicos fetales desde una muestra de sangre periférica materna para determinar anomalías cromosómicas fetales. CffDNA constituye aproximadamente menos del 10% del ADN libre de células circulante total (cfDNA) en el plasma materno, sin embargo, se ha descubierto recientemente que el genoma fetal completo, bajo la forma de cffDNA, está presente en la sangre materna y por lo tanto es material muy prometedor para PPNI.

Varios documentos (Chiu et al., 2008, Fan at al., 2008, Chiu et al., 2011, Sehnert et al., 2011, Lau et al., 2101 y Bianchi et al., 2012, EP 2183693, EP 2366031, US 8296076) divulgaron los métodos en los que la secuenciación de ADN en sangre materna es utilizada para obtener información sobre la dosis cromosómica anómala en el feto. Todos los métodos previamente mencionados utilizan análisis del cfDNA total en plasma materno sin la necesidad de aislar ADN específico fetal, cffDNA. Estos métodos se basan en la detección de la copia extra de cromosoma para distinguir casos normales de casos de trisomía. En el caso de un feto con trisomía, el número de copias de cromosoma trisómico en la sangre materna es ligeramente superior en comparación con otros cromosomas. Un enfoque similar puede utilizarse para la detección de las monosomías, en las que el número de copias de cromosoma monosómico es inferior. La aparición de secuenciación de ADN paralelo en masa (SPM) permitió la secuenciación de enormes cantidades de moléculas de ADN y por lo tanto la secuenciación de la siguiente generación de ADN (SGS) ha sido recientemente utilizada para detectar aneuploidía fetal no invasiva de la sangre materna.

Normalmente, la detección de trisomía fetal usando SGS se realiza a través del siguiente proceso. Primero, una pequeña región al final de cada molécula de ADN de plasma materno es secuenciada y las lecturas de secuencia obtenidas se trazan frente a la referencia de genoma humano para determinar el origen cromosómico de cada secuencia. Después, la cantidad de indicadores (i.e. lecturas trazadas) del cromosoma de interés (p. ej. cromosoma 21) es comparada con algún tipo de referencia cromosómica (i.e. con la cantidad de indicadores de cromosoma particular o grupo seleccionado de cromosomas) normalmente por medio de un ratio de dichos valores y el ratio resultante es comparado entre el ejemplo examinado y las muestras de control euploides.

EP 2183693 (Lo, Y.M., D. et al., atribuido a la Chinese University de Hong Kong, HK) divulgó un método en el que cantidades correspondientes de un cromosoma clínicamente relevante y de cromosoma origen se determinan desde resultados de SPM y desde la representación de porcentaje calculado de secuencias trazadas al cromosoma 21. Se descubrió que una representación de porcentaje similar es significativamente más alta en una muestra de mujer embarazada portando un feto con una trisomía 21 comparada con una mujer con un feto normal. No se determinó ningún nivel z como un valor decisivo en este método.

US 8296076 (Fan, Ch. H.M. and Quake, S.R. atribuido al Consejo de Administración de Leland Stanford Junior University, US) divulgó un método de diagnóstico de aneuploidía fetal basado en la secuenciación. Al contar el número de indicadores trazados a una ventana predefinida en cada cromosoma, la excesiva o escasa representación de cualquier cromosoma en plasma materno del ADN en el que ha contribuido un feto aneuploide puede ser detectada.

Los métodos de Chiu et al., 2008, Chiu et al., 2011, Sehnert et al., 2011, Lau et al., 2012 se basan en la completa secuenciación del genoma utilizando SPM y determinación del nivel z. Aunque el nivel z está ampliamente aceptado como el parámetro estándar usado para la detección de muestras aneuploides, hay diferencias en su determinación. Chiu et al., 2011 divulgaron un enfoque usando indicadores trazados a todos los cromosomas utilizados como referencia para determinación del nivel z, Lau et al., 2012 divulgaron un enfoque usando indicadores trazados a algún cromosoma específico, p.ej. 14 como referencia para T21, y Sehnert et al., 2011 eligió el cromosoma 9 para que fuera la referencia interna óptima para el cromosoma 21.

Normalmente, hay tres etapas principales de la determinación de aneuploidía del feto de una muestra de sangre, plasma o suero materno: 1) preparación de muestra de ADN y librería de ADN, 2) secuenciación, y 3) análisis de la información de la secuencia. La secuenciación realizó progresos destacados en los últimos años, sin embargo, en la primera y tercera etapa hay un espacio para mejorar que podría tener bajo coste y gran impacto en la calidad de la prueba.

Hay varios enfoques sobre cómo mejorar la primera etapa, tal y como aparece por ejemplo en EP 2366031 (Raba, R.P. et al., atribuido a Verinata Health, Inc., US). El documento divulgó un método para imagen y diagnóstico prenatal de aneuploidía cromosómica fetal sobre la base de SGN comprendiendo un protocolo novedoso para la preparación de librería de secuenciación desde una muestra materna. El nuevo enfoque a la hora de preparar librerías de secuenciación incluye los pasos consecutivos de fin-de-reparación, poliadenilación y adaptador ligando dichos ácidos nucleicos, y donde dichos pasos consecutivos excluyen purificar los productos finalmente reparados antes del paso poliadenilación y excluye purificar los productos de poliadenilación antes del paso ligar-adaptador. El método permite determinar una copia de variaciones de números (CVN) de cualquier secuencia de interés. Ningún nivel z fue determinado como un valor decisivo en este método.

Otra mejora fue divulgada en WO 2011/051283 (Benz, M. et al., atribuido a Lifecodexx, AG, DE). El método para diagnóstico no invasivo de aneuploidía cromosómica presentada en este documento es mejorado por el enriquecimiento y cuantificación de secuencias seleccionadas cfDNA en una muestra de sangre materna.

A pesar de la existencia de varios métodos para la detección no invasiva de aneuploidía fetal aún se necesita un método mejorado o método alternativo que sería al menos tan sensible y específico como los presentes métodos.

Descripción de la invención

La presente invención ofrece un método alternativo y fiable aplicable a la práctica de imagen prenatal no invasiva para aneuploidías, tales como trisomías o monosomías, preferiblemente para trisomía de cromosoma 13, 18 o 21. El presente método se ha mejorado comparado con los anteriores procedimientos especialmente en la parte del procesamiento de los datos de secuenciación. Debido al nuevo e ingenioso enfoque este método ofrece un método alternativo nivel-z, que permite una mejor distinción entre muestras euploides y aneuploides, por. ej. mejora la fiabilidad de la prueba. Además, el método necesita una cantidad relativamente baja de datos de secuenciación y por lo tanto tal método es relativamente barato y sería asequible incluso para pequeñas instituciones de atención a la salud.

La siguiente descripción explicará las principales características del método de la presente invención; sin embargo, no implica que la invención deba incluir todas las características y aspectos descritos en el mismo. El alcance de la protección será definido por las reclamaciones de la patente adjuntas a esta descripción. La persona especializada obtendrá una comprensión plena de la presente invención desde la siguiente descripción junto con los ejemplos, donde algunas características específicas serán explicadas en mayor detalle, y con gráficos adjuntos.

Los términos técnicos y científicos utilizados aquí tienen el mismo significado comprendido de forma común por las personas especializadas en el tema de la medicina, biología molecular, genética, bioinformática y diagnóstico prenatal, salvo que se especifique lo contrario en el presente documento.

Resumen del método de acuerdo con la presente invención

La presente invención se refiere al método de determinación de aneuploidía del feto desde una muestra de sangre materna incluyendo una mezcla de moléculas de ácido nucleico maternales, donde la mezcla de moléculas de ácido nucleico maternales y fetales sean moléculas de ADN libres de células, tal y como se define en el punto 1. Dicho método comprende cuatro etapas principales: 1) tratamiento de muestras de sangre materna, 2) preparación de muestra de ADN y librería de ADN, 3) secuenciación, y 4) procesamiento de la información de secuencia para obtener un valor con el objetivo de tener un alcance predictivo. La parte esencial del método es la preparación y procesamiento de la información de formación, i.e. procesamiento del establecimiento de muestras euploides que son procesadas de la misma forma que las muestras de las pruebas. El resultado del procesamiento de las muestras de formación sirve como un dato de referencia en el paso 4) mencionado previamente. El término muestra de sangre se usa con amplio significado y comprende sangre, así como también muestra de plasma o suero.

El método de acuerdo con la presente invención puede ser aplicado a cualquier aneuploidía, tales como monosomía o trisomía, por ejemplo, aneuploidía de cromosomas 8, 9, 13, 16, 18, 21, 22 y X, en concreto trisomía de cromosomas 13, 18 y 21.

La principal mejora y ventaja de este método radica en el paso del procesamiento y análisis de los datos de secuencia. La determinación del valor (nivel-z) que es decisivo para el método (comparando este valor concreto con un valor límite preestablecido permite determinar la presencia de aneuploidía, como trisomía o monosomía) se realiza basándose en un modelo de distribución multinomial de indicadores a cromosomas o depósitos de genoma humano. Tal y como se demuestra en los ejemplos, valores nivel-z de muestras trisómicas determinados por el método de la presente invención son significativamente más altos que los valores determinados por cualquiera de los anteriores procedimientos (para muestras T21 obtuvimos una diferencia media de valores nivel-Z -1.895, -1.595, -2.485 al comparar Chiu et al., 2008, Sehnert et al., 2011, Lau et al., 2012, respectivamente, con nuestro método con un valor-p de diferencia < 0.05 para toda muestra anterior de acuerdo con la prueba de comparaciones múltiples de Dunn), mientras que los valores de muestras normales (no trisómicas) son sustancialmente idénticas (respecto al cromosoma 21 obtuvimos una diferencia media de nuestro método -0.15, -0.05, -0.21 para Chiu et al., 2008, Sehnert et al., 2011, Lau et al., 2012, respectivamente), lo cual significa que la distancia de niveles z de muestras trisómicas del valor límite aumenta para la presente invención al compararse con los procedimientos anteriores mientras que la distancia de las muestras euploides desde el valor límite se conserva básicamente. Como la distancia determina la seguridad de nuestro diagnóstico (para muestras trisómicas, cuanto más grandes mejor), la presente invención mejora la calidad de la prueba.

El punto de partida del presente método es procesar la muestra materna, que es sangre periférica, por lo tanto, el método no es invasivo. La sangre periférica incluye ADN libre de células (cfDNA) que es una mezcla de ADN materno y fetal (cffDNA). CffDNA es lo que importa; por lo tanto, ADN fetal puede enriquecerse (mediante la selección del fragmento más corto, ya sea in silico o selección física). Luego, la muestra de ADN (aún incluyendo tanto el ADN materno como el fetal, i.e. es una muestra mixta) está sujeta a la secuenciación paralelamente masiva, tal y como se realiza por el enfoque SGS, para obtener un enorme número de cortas lecturas de secuencia. Estas lecturas sirven como indicadores, i.e. están trazados a una región o cromosoma genómico determinado. Con el propósito de este método, el genoma humano al completo (24 cromosomas juntos representando ADN con más de 3 x 10^9 bp) se divide en depósitos 1 Mbp que no se superponen (i.e. aproximadamente 3 x 10^3 depósitos), y el presente método considera tanto a estos indicadores trazados a cromosomas como a estos depósitos. Posteriormente, hablando en claro, mediante el conteo del número de indicadores trazados a los depósitos específicos de interés (p. ej. depósitos asociados con cromosomas potencialmente triploides como el cromosoma 21), y mediante la comparación de este número con el número de indicadores trazados a algún tipo de cromosoma de referencia (o un grupo de cromosomas de referencia o sus partes, i.e. un grupo en particular de depósitos de referencia), la representación irregular del cromosoma de interés o sus partes puede ser detectada.

La parte esencial del método es la preparación y procesamiento de las muestras de formación, i.e. procesamiento del grupo de muestras euploides que son procesadas de la misma forma como muestras de prueba. El grupo de datos de capacitación sirve como base para la selección de los cromosomas de referencia o sus partes. La selección de una referencia adecuada, así como un grupo de capacitación suficientemente aleatorio es de la mayor importancia tal y como se deriva de nuestros propios experimentos, así como de anteriores documentos de procedimientos mencionados anteriormente. Por los tanto, nuestro enfoque incluye un método novedoso y mejorado de selección de referencia interna, específicamente cromosomas de referencia interna y depósitos de referencia interna.

Aún más, el presente método incluye la determinación del valor nivel z basado en un enfoque novedoso que no ha sido utilizado hasta ahora. Brevemente, el nivel z se calcula desde el medio y la desviación estándar determinado sobre la base de la distribución multinomial de indicadores a los depósitos. Finalmente, la comparación del nivel z con el valor límite prefijado (p. ej. 3) es indicativo de si la aneuploidía existe o no.

El presente enfoque multinomial es, en principio, un método alternativo nivel z; sin embargo, está basado en los datos significativamente más generales que los anteriores métodos de procedimiento. Más en concreto, la primera parte de la mejora del presente método radica en una selección más natural de grupo de cromosomas de referencia. No sólo tenemos en cuenta para la referencia interna cualquier combinación de cromosomas (excluyendo combinaciones de cromosomas 13, 18 y 21 debido a la trisomía potencial), también permitimos que sólo haya partes para los cromosomas, i.e., depósitos (con 1 Mbp resolución). De esta manera podemos caracterizar mejor la población euploide y por lo tanto encontrar más fácilmente desviaciones, i.e. muestras aneuploides. Por lo tanto, el primer beneficio de nuestro enfoque radica en que la referencia interna puede ser cualquier grupo de cromosomas o partes de los cromosomas con la resolución de depósitos 1 Mbp (en lugar de un cromosoma específico o todos los autosomas como en anteriores métodos de procedimientos). El otro beneficio del presente método es que nuestro enfoque multinomial está basado en modelado matemáticamente lógico y complejo de la distribución de indicadores a los depósitos y el ratio consecutivo del aneuploide, p. ej. trisómico, cromosoma a los depósitos de referencia interna. El enfoque multinomial lleva a nivel z significativamente más alto para muestras trisómicas al compararlo con métodos de procedimientos anteriores, mientras no se mantiene ningún cambio sustancial en los niveles z de muestras euploides. Por lo tanto, la fiabilidad de la prueba se ha mejorado. Además, el presente método permite la estimación de fracción fetal junto con el error de esta estimación. Mientras la estimación de la fracción fetal se presentó ante *Rava et al., 2014, Hudecova et al., 2014, y Yu et al., 2014*, ningún error de esta estimación ha sido presentado hasta ahora. Aún más, el presente método permite la estimación del número mínimo de indicadores trazados con los que un aneuploide, p. ej. muestra trisómica con determinada fracción fetal se espera mostrar un

nivel z al menos algún valor predeterminado (p. ej. valor límite). Esta información puede ser útil para decidir casos con baja fracción fetal y valor cercano al límite del nivel z.

Hay un requisito práctico para automatizar los métodos de las pruebas prenatales o diagnósticos. El método de la presente invención puede automatizarse en gran medida. Al menos la parte bioinformática del método (i.e. procesamiento de los datos de secuenciación y todas las determinaciones subsiguientes y cálculos) pueden llevarse a cabo utilizando cualquier sistema de ordenador adecuado, como por ejemplo PC equipado con un procesador, instrumentos de entrada/salida periféricos (p. ej. puertos, interfaces), memorias (p. ej. memoria de sistema, disco duro), teclado, monitor, ratón, etc. Preferiblemente el sistema de ordenador puede estar en comunicación con los datos con el sistema de secuenciación facilitando los datos de secuencia, preferiblemente bajo la forma de pluralidad de lecturas de secuencia (por cable o red inalámbrica, bluetooth, internet, nube, etc.). Significa que el sistema de ordenador está configurado para recibir datos de secuencia del sistema de secuenciación. Los sistemas de ordenador adecuados, así como los medios para la conexión con sistema de secuenciación son bien conocidos por las personas especializadas en la materia.

Al menos parte del método, específicamente la parte bioinformática del método, puede implementarse como un código software, i.e. una pluralidad de instrucciones (programa de ordenador) para ser ejecutada por un procesador de un sistema de ordenador. El código puede incluirse en el medio legible para el ordenador para almacenaje o transmisión tales como por ejemplo RAM, ROM, disco duro, SDS, CD, DVD, memoria flash, etc. Aún más, el código puede transmitirse a través de cualquier red por cable, óptica, inalámbrica, por ejemplo, a través de internet. Por ejemplo, el programa de ordenador completo puede ser descargado por el usuario (cliente) a través de internet. Por lo tanto, la presente invención se refiere también a un producto de programa de ordenador que comprende un medio legible para el ordenador incluyendo una pluralidad de instrucciones para controlar un sistema de ordenador para realizar todos los pasos del método de la invención siguiendo el paso (b) de realizar una secuenciación aleatoria.

El programa de ordenador mencionado previamente puede ser preferiblemente introducido en ordenador del sistema de secuenciación.

El objeto de la presente invención se define en los puntos 1-12 adjuntos.

Las características y ventajas de la presente invención las comprenderá la persona especializada en el tema desde la siguiente descripción detallada, ejemplos y gráficos.

Descripción detallada del método

EL MÉTODO

El método es un método nivel z basado en el modelo de distribución multinomial de indicadores trazados a los depósitos. Observe que en esta sección escribimos, por ejemplo, más sobre trisomía, especialmente del cromosoma 21, sin embargo, el método puede aplicarse a cualquier aneuploidía, monosomía u otra trisomía, en particular trisomía del cromosoma 13 y 18, de manera similar.

Un prerrequisito del procesamiento de muestras de la prueba es la selección y procesamiento del grupo de capacitación adecuado (i.e. grupo de muestras de capacitación) para elegir referencia interna adecuada para la evaluación de las muestras de la prueba. Normalmente, el grupo de capacitación debería incluir muestras femeninas euploides, así como muestras masculinas euploides y las muestras deben ser tratadas necesariamente de la misma forma que en las muestras de la prueba.

MUESTRAS DE CAPACITACIÓN

Se seleccionan muestras masculinas y femeninas para formar el grupo de capacitación denominado T_h. Para cada muestra del grupo de capacitación se realizan los siguientes pasos:

1. Recoger y tratar muestras sanguíneas.

2. Aislamiento del ADN.

3. Secuenciación.

4. Para cada muestra del grupo de capacitación aplicar los siguientes pasos de procesamiento de datos de secuencia:

(a) Lecturas de secuencia se trazan al genoma humano de referencia desenmascarado (hg19, Estructura Humana Consorcio de Referencia del Genoma 37 (GRCh37), Feb. 2009, GenBank acceso al conjunto: GCA_000001405.1, RefSeq acceso al conjunto: GCF_000001405.13) utilizando un algoritmo Bowtie2 (*Langmead and Salzberg*, 2012).

(b) Sólo se retienen lecturas de secuencia que podrían trazarse a una única ubicación genómica con al menos una disparidad.

(c) Se realiza la corrección GC de acuerdo con *Liao et al*., 2014 (en nuestro caso sin la normalización interna).

(d) Opcional: Filtrar lecturas de acuerdo con su tamaño (p. ej. <150bp) (Minarik et al., 2015).

(e) El genoma de referencia hg19 se divide en 1 Mbp posterior y depósitos que no se solapan con los indicadores del ADN trazados distribuidos a depósitos de acuerdo con su ubicación de trazado.

La filtración de tamaño en el paso 4d puede utilizarse para aumentar artificialmente la fracción fetal y por este nivel z de muestras trisómicas también. Sin embargo, esto es normalmente con el coste de disminuir el indicador total de la muestra ya que los indicadores más largos del umbral que se tiene en cuenta son desechados (para mayor detalle véase Minarik et al., 2015).

Aún más, los depósitos de referencia interna, denominados IRB, deben ser elegidos antes de la determinación del valor nivel z de la muestra de la prueba, y esto se realiza por medio de una elección de autosomas de referencia interna, denominados IRA, utilizando un grupo de capacitación T_h y un algoritmo genético. Por lo tanto, el método incluye los pasos:

5. Elegir los autosomas de referencia interna, IRA, utilizando el grupo de capacitación T_h.

(a) Cualquier combinación de autosomas (excluyendo combinaciones de autosomas 13, 18 y 21 debido a la potencial aneuploidía) se tienen en cuenta como una referencia interna potencial (junto con 524 287 combinaciones candidatas).

(b) Para cada combinación de referencia interna potencial el coeficiente de variación (CV) se calcula (véase *Cálculo de coeficiente de variación de combinación de autosomas de referencia interna candidata* a continuación).

(c) Las combinaciones de referencia interna candidatas de autosomas se piden de acuerdo con su valor CV en orden creciente.

(d) Combinaciones de referencia interna cuyo valor CV sea menor o igual a 1.1 múltiple del total del valor más bajo CV se eligen como IRA. Preferiblemente al menos cien (i.e. la primera centena) de estas combinaciones son elegidas (éstas son las mejores combinaciones de autosomas internas en términos de valores VC).

6. Elegir depósitos de referencia interna, IRB, por un algoritmo genético utilizando el grupo de capacitación T_h (véase *Elegir depósitos de referencia interna* a continuación).

Cálculo de coeficiente de variación de combinación de autosomas de referencia interna candidata

Dejar que $M = (m_{ij})_{kx22}$ sea la matriz serie asociada con el grupo de capacitación T_h en el que *k* es el tamaño de T_h y m_{ij} soporta el número de indicadores trazados (y GC corregidos) al autosoma j^{th} para la muestra i^{th} de Th, y dejar que $ir = \{0, 1\}^{22}$ sea una combinación de referencia interna candidata *(ir[i]* = 1 cuando el autosoma i^{th}, $i \in \{1, 2, ..., 22\}$, se eligió como referencia interna, de otro modo *ir[i]* = 0).

También, dejar que Mi designe la cola i^{th} de la matriz *M*. El valor normalizado de la muestra ^{ith} (i.e., cola i^{th} de la matriz M) está definida como $m_{i,x}/M_i \cdot ir$, donde x es el cromosoma aneuploide que queremos examinar. Designamos el valor normalizado de la muestra *i* como v_i. Dejar que $V_{ir} = (v1, v2, ..., v_k)$. Entonces, el valor CV de la combinación de referencia interna candidata *ir* queda definida como:

$$CV_{ir} = 100 \cdot \frac{stdev(V_{ir})}{mean(V_{ir})} \qquad \text{(Eq. 1)}$$

Elegir los depósitos de referencia interna (algoritmo genético)

Dejar que $M = (m_{ij})_{kxt}$ sea una matriz asociada al conjunto de capacitación T_h, en el que *k* es el tamaño de T_h *t* es el número de depósitos 1Mbp que cubre el genoma g19, y m_{ij} soporta el número de indicadores trazados al depósito j_{th} para la muestra i_{th} de T_h. También, dejar que M_i designe la cola i^{th} de la matriz *M*. Una combinación candidata de depósitos de referencia interna (IRB) es cualquier combinación de las columnas de la matriz *M* excluyendo las combinaciones que incluyen depósitos asociados con cromosomas 13, 18 y 21 (estos depósitos no deberían estar en el IRB seleccionado debido a la trisomía potencial; otras trisomías son muy extrañas). Dejar que $ir = \{0, 1\}^t$ designe tal candidato IRB (*ir[j]* = 1 cuando el depósito j^{th}, $j \in \{1, 2, ..., t\}$, fue elegido para referencia interna, de otro modo *ir[j]* = 0). Aún más, dejar que el vector binario $_{v21}$ designe depósitos asociados con cromosoma 21 similarmente a vector *ir* (de forma análoga para cromosomas 13 y 18 u otras aneuploidías). Para cada candidato IRB calculamos valor de depósito fraccional FB_i para cada muestra en T_h desde la cola i^{th} de la matriz *M* como

$$FB_i = \frac{M_i \cdot v_{21}}{M_i \cdot ir}, \qquad \text{(Eq. 2)}$$

en el que ´ representa el producto punto. Dejar que V_{ir} = (FB_1, FB_2, ..., FB_k) soporte valores de depósitos fraccionales para todas las muestras en T_h. Asignamos a cada candidato IRB coeficiente de rango de estudiante (SRC) que es otorgado por

$$SRC_{ir} = \frac{\max(V_{ir}) - \min(V_{ir})}{stdev(V_{ir})}. \qquad \text{(Eq. 3)}$$

El valor SRC describe la expansión de los valores FB del grupo de capacitación. Además del valor SRC, también asignamos a cada candidato IRB un valor CV que se calcula análogamente al proceso de *Cálculo de coeficiente de variación de la combinación de autosomas de referencia interna candidata*. Queremos encontrar tal IRB que tiene valores SRC y CV tan pequeños como sea posible. Sin embargo, encontrar el grupo de todos los candidatos IRB tiene aproximadamente 2^{3000} miembros por lo que no es posible calcular los valores SRC y CV de todos ellos. Por lo tanto, para seleccionar un buen IRB utilizamos un algoritmo genético (algoritmos genéticos son bien conocidos por la persona especializada en el tema) y utilizamos un programa DEAP tal y como ha publicado *Fortin et al*., 2012. Concretamente, la población inicial de candidatos IRB utilizada en el algoritmo genético consiste en dos grupos. El primer grupo es una selección aleatoria de depósitos de todos los autosomas, excluyendo los depósitos de autosomas 13, 18 y 21 debido a la potencial trisomía, mientras que el segundo grupo está basado en los mejores autosomas de referencia interna previamente determinados (IRA). De esta forma, ya hay buenos individuos en la población de IRBs candidatos (es decir, aquellos del grupo IRA), y el algoritmo genético sólo puede fomentar las combinaciones previas de la selección IRA, por ejemplo, al seleccionar sólo algunos depósitos de las combinaciones IRA.

El algoritmo genético posiblemente no seleccionará el mejor IRB en términos de valores SRC y CV, pero el IRB resultante será mejor (a menudo considerablemente) que el IRA. Esto se debe a que el IRB permite que allí haya sólo partes de cromosomas en la referencia al contrario que el total de cromosomas en el caso IRA.

Nos gustaría recalcar que el IRB seleccionado depende del cromosoma de interés seleccionado (p. ej. cromosoma 21), muestras en el grupo de capacitación, así como alguna probabilidad debida a la naturaleza de los algoritmos genéticos. Por lo tanto, múltiples turnos de esta parte del proceso resultarán probablemente en diferentes IRB. Por esta razón es aconsejable repetir este proceso múltiples veces (el número de veces sólo está limitado por el tiempo que queremos pasar en esta parte del proceso) y elegir el mejor IRB encontrado hasta entonces.

MUESTRA DE PRUEBA

La muestra de prueba se procesa exactamente de la misma forma que las muestras de capacitación en los pasos 1 a 4 tal y como aparece descrito en detalle anteriormente. Aún más, utilizando los depósitos de referencia interna, IRB, tal y como se determina anteriormente, se llevan a cabo los siguientes pasos:

5. Los indicadores trazados de la muestra de prueba se transforman en valor de depósito fraccional (FB) utilizando el grupo IRB elegido (véase *Elegir depósitos de referencia interna*)

$$FB = S_{21}/S_{IRB} \qquad \text{(Eq. 4)}$$

donde S_{21} designa el número de indicadores de la muestra de prueba trazados a los depósitos de cromosoma 21 (u otro cromosoma de interés) y S_{IRB} designa el número de indicadores de la muestra de prueba trazados a depósitos IRB.

6. Se calcula el medio multinomial *μ(n)* y desviación estándar σ(n) para la muestra de prueba (véase a continuación *Cálculo de medio multinomial y desviación estándar*)

7. La muestra de prueba es evaluada (véase a continuación *Evaluación de la muestra de prueba*)

Cálculo de medio multinomial y desviación estándar

Primero observe que por el medio multinomial y desviación estándar no queremos decir el medio y la desviación estándar de la distribución multinomial. En su lugar, es el medio y desviación estándar de una variable aleatoria nueva derivada de la distribución estándar que definimos en la sección *El modelo matemático* a continuación en esta descripción.

Dejar que IRB sea un grupo de depósitos de referencia interna elegidos. Dejar que T_h sea un grupo de muestras euploides en cadena de tamaño *k*. Dejar que $M = (m_{ij})_{kxt}$ sea una matriz asociada al grupo en cadena T_h, donde *k* es el tamaño de T_h *t* es el número de depósitos 1 Mbp que cubren el genoma humano hg19, y m_{ij} soporta el número de indicadores trazados al depósito j^{th} para la muestra i^{th} de T_h. Dejar que m_i, $i \in \{1, 2, ..., k\}$, designe la cantidad de la fila i^{th} de la matriz *M*. Dejar que $P = (p_{ij})_{kxt}$ sea matriz tal que $p_{ij} = m_{ij}/m_i$, $i \in \{1, 2, ..., k\}$, $j \in \{1, 2, ..., t\}$, donde m_{ij} es un elemento de la matriz *M*. Aún más, dejar que p_j, $j \in \{1, 2, ..., m\}$, designe el valor medio de la columna j_{th} de la matriz *P*, y dejar que p = (p1, ..., pt}. Interpretamos el valor p_j, $j \in \{1, 2, ..., m\}$, como una probabilidad de que un fragmento de ADN trazará en el depósito j_{th}. Observe que este vector es facilitado por el grupo de capacitación T_h y también son solo los únicos valores que nuestro método necesita para capacitar. Observamos que estos valores cambian con diferente procesamiento de laboratorio de las muestras, por lo que las muestras de capacitación de T_h y las muestras de prueba necesitan ser preparadas de la misma forma. Dejar que P_x, $X \in \{13, 18, 21\}$ dependiendo de la trisomía que queremos estudiar, y P_{IRB} será la suma de probabilidades para depósitos asociados con el cromosoma X^{th} e IRB, respectivamente. Definimos el medio multinomial $\mu(n)$ y desviación estándar multinomial σ(n) como

$$\mu(n) = \frac{P_X}{P_{IRB}}\left(1 - (1 + P_{IRB})^n\right) \qquad X \in \{13, 18, 21\}, \tag{Eq. 5}$$

$$\sigma(n) = \sqrt{\frac{W_1}{n} + \frac{W_2}{n^2}}, \tag{Eq. 6}$$

donde

$$W_1 = \left(\frac{P_X}{P_{IRB}}\right)^2 \left(\frac{1}{P_X} + \frac{1}{P_{IRB}}\right), \tag{Eq. 7}$$

$$W_2 = \frac{P_X}{1 - P_{IRB}} \left(\frac{1}{P_{IRB}}\right)^3 \frac{1 - P_X - P_{IRB}}{1 - P_{IRB}}, \tag{Eq. 8}$$

y *n* es el número total de los indicadores trazados (ahora incluyendo los indicadores de depósitos asociados con cromosomas 13, 18 y 21) de la muestra que queremos estudiar. Observe que la muestra de prueba con diferente *n* tiene medio multinomial y desviación estándar diferente. De esta forma la prueba está personalizada para cada muestra.

Evaluación de los datos de prueba

Dejar que *FB* sea el valor de depósito fraccional de una muestra que queremos estudiar (véase *Elegir los depósitos de referencia interna*). Dejar que *n* sea el número total de los indicadores de ADN trazados (incluyendo indicadores de depósitos asociados con cromosomas 13, 18 y 21) para esta muestra. Dado un caso una trisomía que desearíamos estudiar, calculamos el medio multinomial adecuado $\mu(n)$ y desviación estándar multinomial σ(n) (véase anteriormente *Cálculo de medio multinomial y desviación estándar*). Posteriormente, para nuestra muestra de prueba calculamos su nivel z

$$z = \frac{FB - \mu(n)}{\sigma(n)}, \tag{Eq. 9}$$

y si z es más alto entonces que el valor prefijado, la muestra queda marcada por el presente método como trisómica (normalmente para z > 3).

Fracción fetal

Además, la fracción fetal aproximada *f* de la muestra de prueba es facilitada por

$$f \approx 2\frac{FB \cdot P_{IRB} - P_{21}}{P_{21}}, \qquad \text{(Eq. 10)}$$

donde P_{21} y P_{IRB} son del *Cálculo de medio multinomial y desviación estándar* (de forma similar para otras trisomías) y *FB* es valor de depósito fraccional de *Elegir los depósitos de referencia interna*.

El error de la estimación de fracción fetal

Dejar que *p* = (p_1, ..., p_t) sea desde el Cálculo de medio multinomial y desviación estándar. Dejar Q(f) = (q_1 (f), q_2(f), q_3(f), ..., q_t(f)) donde

$$q_i(f) = \begin{cases} \frac{p_i + \frac{f}{2}p_i}{1+\frac{f}{2}S_{21}} & \text{para depósitos asociados con el cromosoma 21} \\ \frac{p_i}{1+\frac{f}{2}S_{21}} & \text{para otros depósitos} \end{cases} \qquad \text{(Eq. 11)}$$

donde S_21 es la suma de p_j asociada con el cromosoma 21. Los rangos i desde 1 a t, y f es una variable que puede alternar de 0 a 1. Para otras trisomías la ecuación es similar. Para monosomía la ecuación cambia a

$$q_i(f) = \begin{cases} \frac{p_i + \frac{f}{2}p_i}{1+\frac{f}{2}S_{21}} & \text{para depósitos asociados con el cromosoma 21} \\ \frac{p_i}{1+\frac{f}{2}S_{21}} & \text{para otros depósitos} \end{cases} \qquad \text{(Eq. 12)}$$

Transformamos el vector Q(f) al vector Q´(f) = (Q_X, Q_IRB, 1 – Q_X – Q_IRB) donde Q_X es la suma de los valores q_i(f) asociados con el cromosoma 21 u otro cromosoma aneuploide, y Q_IRB es la suma de los valores q_i(f) asociados con los depósitos IRB. Observe que hasta que definamos el valor de f, los vectores Q(f) y Q´(f) son simbólicos. Utilizando los valores Q_X y Q_IRB en lugar de P_X y P_IRB, definimos los valores $\mu(n, f)$ en lugar de $\mu(n)$ porque aún no fijamos el valor de f.

Dejar que FB sea el valor de depósito fraccional de una muestra con la que estamos trabajando actualmente (Véase *Elegir los depósitos de referencia interna*). Al resolver la ecuación

$$FB - \mu(n, f) = 2 * \sigma(n,f) \qquad \text{(Eq. 13)}$$

para f (recuerde que n es la cuenta de indicador total de la muestra de la prueba), obtenemos el límite inferior de la fracción fetal de la muestra de prueba. Al resolver la ecuación

$$FB - \mu(n, f) = -2 * \sigma(n,f) \qquad \text{(Eq. 14)}$$

para f, obtenemos el límite superior sobre la fracción fetal de la muestra de prueba.

La naturaleza de estos límites es la siguiente. Ya que la secuenciación es un proceso aleatorio en el sentido de que los fragmentos secuenciados son elegidos de forma aleatoria, puede ocurrir que dos vueltas de secuenciación de la misma muestra terminen con un número diferente de fragmentos secuenciados de, digamos, el cromosoma 21. Por esta razón el nivel z de una muestra puede variar con diferentes turnos de secuenciación (dada la misma elección de IRB), y lo mismo se mantiene para la fracción fetal derivada. Estos límites se fijan de manera que el 95% de los turnos de secuenciación de la misma muestra se espere que resulten en fracción fetal dentro de estos límites. Por lo

tanto, para la previamente calculada fracción fetal f, estos límites nos dan el margen probable de error para este valor.

Cuenta de indicador mínima para predicción fiable

Para muestras euploides el valor del depósito fraccional esperado (FB) es aproximadamente p'_1/p'_2 (por *Teorema 1*). Para una muestra T21, como ejemplo, con fracción fetal *f* este valor *FB* es aproximadamente igual a (1+f/2) p'_1/p'_2 porque en muestras trisómicas el parámetro p'_1 aumenta por una cantidad de $f/2\ p'_1$ viniendo de la tercera copia de cromosoma 21 fetal mientras que otros cromosomas no varían (asumiendo que la muestra no tiene otras trisomías presentes). También hay un factor normalizador $1/(1+ f/2\ p'_1)$ porque al incrementar el parámetro de probabilidad p'_1 nosotros violamos la ecuación $p'_1 + p'_2 + p'_3 = 1$.

Sin embargo, este factor queda eliminado en la fracción p'_1/p'_2. Por lo tanto, por *Teoremas 1* y *2* (véase a continuación) el esperado nivel z de una muestra trisómica con fracción fetal *f* es (vea que nosotros desechamos el término $W2/n_2$ del *Teorema 2* por el bien de la simplicidad)

$$z = \frac{\frac{p'_1}{p'_2}\left(1+\frac{f}{2}\right) - \frac{p'_1}{p'_2}}{\sqrt{\frac{1}{n}\left(\frac{p'_1}{p'_2}\right)^2\left(\frac{1}{p'_1}+\frac{1}{p'_2}\right)}} = \frac{\frac{f}{2}}{\sqrt{\frac{1}{n}\left(\frac{1}{p'_1}+\frac{1}{p'_2}\right)}}. \qquad \text{(Eq. 15)}$$

Queremos que *z* sea al menos algo *k*. Además, los niveles z de muestras euploides están distribuidos alrededor de 0 en ambos lados (siendo 0 el nivel z esperado), y lo mismo se sostiene para muestras trisómicas (por supuesto, tales niveles z estarían distribuidos alrededor de diferente valor medio). De ahí, es posible para una muestra trisómica que su valor de depósito fraccional FB no sea igual a $(1+f/2)\ p'_1/p'_2$; puede ser tanto inferior como superior. Por lo tanto, nosotros queremos tener

$$z = \frac{\frac{f}{2}}{\sqrt{\frac{1}{n}\left(\frac{1}{p'_1}+\frac{1}{p'_2}\right)}} = k + l, \qquad \text{(Eq. 16)}$$

donde *k* es el nivel z que desearíamos alcanzar y *l* corrige la posibilidad de la muestra *s* llegando más abajo de lo esperado (como en un caso de muestra euploide que tiene nivel z menor de 0). Como los niveles z de muestras euploides siguen la distribución estándar normal (esta suposición fue positivamente testada con pruebas estadísticas), 95% de niveles z de estas muestras caerían dentro del intervalo [-2;2], por lo que fijar $l = 2$ parece razonable. Al resolver la última ecuación para n nosotros obtenemos la cuenta total requerida de fragmentos de ADN trazados para el nivel z esperado al menos *k* como

$$n = \frac{4(k+l)^2}{f^2}\left(\frac{1}{p'_1}+\frac{1}{p'_2}\right). \qquad \text{(Eq. 17)}$$

Con $k = 4$ y $l = 2$, obtenemos la cuenta de indicador mínima alrededor de 1.24 millones para fracción fetal $f = 0.1$. Un argumento más general aparece en el Gráfico 1.

<u>*DESARROLLO DEL MODELO MULTINOMIAL*</u>

MOTIVACIÓN BIOLÓGICA

Se observó que la representación genómica fraccional (RGF) de un cromosoma concreto depende (entre otras cosas) de su tamaño (*Chiu et al., 2008; Ehrich et al., 2011*). De este podemos deducir, en el sentido matemático, que si una muestra de plasma materno es secuenciada y estas secuencias son posteriormente trazadas y contadas para cada cromosoma, entonces el resultado de este proceso (aleatorio) es un vector de 24 números, uno para cada cromosoma, y podemos ver este vector como un vector aleatorio que se dibuja desde la distribución multinomial con parámetros $(p_1, p_2, ..., p_{24}; n)$, donde p_i, $i \in \{1, 2, ..., 24\}$, es probabilidad de que un indicador concreto de ADN secuenciado trace hacia el cromosoma i^{th}, y *n* es el número de todos los indicadores trazados. Como resultado, formulamos la siguiente observación.

Observación: Los datos de secuenciación (i.e., indicadores cfADN trazados) están distribuidos de forma multinomial entre los cromosomas con parámetros de probabilidad fuertemente correlacionados con las longitudes de cromosoma.

Ir desde cromosomas a depósitos 1 Mbp es un asunto sencillo, sólo que en este caso el número de parámetros p_i no es 24 sino el número de depósitos que cubre el genoma humano hg19 (designamos este número como *t*).

EL MODELO MATEMÁTICO

Dejar que *X* sea un vector aleatorio desde la distribución multinomial. Definimos una nueva variable aleatoria *Y* como un ratio de 2 o más elementos del vector aleatorio multinomial. Por ejemplo, para $X=(x_1,x_2,x_3,...,x_{20})$ Y puede definirse como $Y = (x_1+x_2)/(x_2+x_3+x_{10})$. Para modelar la situación descrita en *Motivación biológica* dejamos el vector aleatorio $X = (x_1, x_2, ..., x_t)$ representar el número de indicadores trazados a depósitos 1 Mbp que cubren el genoma humano hg19. Para la detección de la trisomía el numerador de Y es el número de indicadores trazados a cromosoma 21 (o 13 o 18), y el denominador de *Y* es la suma de indicadores trazados a los depósitos *IRB*. Sin embargo, esta definición de *Y* no es necesariamente general y complicada. Esto es porque para obtener el valor de *Y* para una muestra concreta, no importa en realidad, por ejemplo, a qué depósito *IRB* específico quedó trazado algún indicador de ADN; sólo necesitamos distinguir tres clases de indicadores: 1) aquellos trazados a un depósito asociado con el cromosoma aneuploide, 2) aquellos trazados a uno de los depósitos de referencia interna, y 3) el resto. Por lo tanto, podemos reemplazar el vector aleatorio *X* por un nuevo vector aleatorio multinomial $X' = (x'_1, x'_2, x'_3)$, correspondiendo a las tres clases mencionadas, con los siguientes nuevos parámetros de probabilidad facilitados por

$$p'_1 = P_X, \quad X \in \{13, 18, 21\}, \tag{Eq. 18}$$

$$p'_2 = P_{IRB}, \tag{Eq. 19}$$

$$p'_3 = 1 - p'_1 - p'_2, \tag{Eq. 20}$$

donde P_X y P_{IRB} han sido definidos anteriormente en *Cálculo de medio multinomial desviación estándar*. En este sentido, nosotros podemos simplificar ahora la definición de *Y* como $Y = x'_1 / x'_2$. Finalmente, para evitar problemas con cero en el numerador de *Y*, cambiamos la definición de *Y* de la siguiente forma

$$Y = \frac{x'_1}{x'_2 + 1}. \tag{Eq. 21}$$

Como los valores aceptados por el vector aleatorio $X = (x'_1, x'_2, x'_3)$ son del tipo 10^4 y mayores, el +1 en el denominador de Y sólo tiene un efecto insignificante. En la aplicación de este modelo necesitamos saber los valores de *E[Y]* y *Var[Y].*

Teorema 1. Dejar que $X' = (x'_1, x'_2, x'_3)$ sea un vector aleatorio desde la distribución multinomial con parámetros $(p'_1, p'_2, p'_3; n)$. El valor esperado de la variable aleatoria *Y* dada por la ecuación 21 es

$$E[Y] = \frac{p'_1}{p'_2}\left(1 - (1 - p'_2)^n\right). \tag{Eq. 22}$$

Teorema 2. Dejar que $X = (x'_1, x'_2, x'_3)$ sea un vector aleatorio desde la distribución multinomial con parámetros $(p'_1, p'_2, p'_3; n)$. Para $n \geq 10^5$ y $1 > p'_1, p'_2 \geq 0.01$ la varianza de la variable aleatoria *Y* dada por la ecuación 21 puede aproximarse mediante

$$Var[Y] \approx \frac{W_1}{n} + \frac{W_2}{n^2} \tag{Eq. 23}$$

donde

$$W_1 = \left(\frac{p'_1}{p'_2}\right)^2 \left(\frac{1}{p'_1} + \frac{1}{p'_2}\right), \qquad \text{(Eq. 24)}$$

$$W_2 = \frac{p'_1}{1-p'_2} \cdot \left(\frac{1}{p'_2}\right)^3 \frac{1-p'_1-p'_2}{1-p'_2}. \qquad \text{(Eq. 25)}$$

El error de esta aproximación es menor de 0.01% del valor real.

Una muestra de prueba dada, caracterizada por su valor de depósito fraccional *FB*, se normaliza utilizando los valores *E[Y]* y *(Var[Y])*$^{-1/2}$. En la actualidad métodos de nivel z utilizados (mencionados previamente y descritos a continuación), si el valor normalizado de muestra fuera mayor de 3, se consideraba trisómico. Este valor límite venía de la aceptación de que los valores normalizados estaban distribuidos de acuerdo con la distribución normal estándar. También probamos esta aceptación de forma positiva, es decir, la variable aleatoria *z* dada por la ecuación 9 es estándar normalmente distribuida. Por esta razón, establecer el valor límite en 3, como es en la actualidad un estándar, es adecuado para nuestro método también.

Breve descripción de los gráficos

Gráfico 1: Cálculo de la profundidad de secuenciación requerida del ADN libre de células circulante en el plasma materno para la detección de aneuploidía del autosoma 21 como una función de fracción ADN fetal. El mínimo teórico se basa en la ecuación 17 y valores k = 3 (nivel z mínimo para denominar una muestra trisómica), 1 = 2 (dado por la dispersión de niveles z euploides alrededor de cero), y p'_1 = 0.0128, p'_2 = 0.128 (dado por datos en cadena MiSeq). Para IonTorrent la curva del mínimo teórico es casi idéntica y no está trazada aquí. Los puntos trazados muestran las muestras T21 observadas en MiSeq (cuadros) e IonTorrent (círculos).

Gráfico 2: El plan del diagrama de flujo de los principales pasos del método para detección prenatal no invasiva de aneuploidía del cromosoma fetal desde las muestras de sangre materna de acuerdo con la presente invención.

Ejemplos

Ejemplo 1

Técnicas de secuenciación utilizadas en la realización de la invención

Análisis de secuencia

La secuenciación masiva en paralelo es necesaria para la aplicación del método de acuerdo con la invención. El método fue específicamente desarrollado y validado para sistemas de secuenciación de la siguiente generación de pequeños bancos de laboratorio para permitir costes iniciales bajos para el montaje de servicio de laboratorio NIPT. El método fue validado en el sistema MiSeq (Illumina, Inc., San Diego, CA, EEUU) e Ion Torrent PGM (life Technologies Corp., San Francisco, CA, EEUU).

Instrumentos de secuenciación comercialmente disponibles junto con los protocolos correspondientes y reactivos recomendados por el distribuidor, fueron utilizados en el ejemplo ilustrativo, sin embargo, la persona especializada en el tema es consciente del número de varios métodos de secuenciación y sus variaciones, que también podrían ser utilizados en la práctica de la presente invención.

El kit de secuenciación Illumina Version 3 y el kit Reactivo MiSeq v3 (MS-102-3003) fueron utilizados.

En nuestros experimentos, una media de 5.829± millones y 3.098±0.951 millones de lecturas brutas fueron obtenidas tras el análisis de secuenciación en IonTorrent PGM y MiSeq, respectivamente. Debido a que 200 bp de extremo único y 2x 100 bp protocolos de extremo emparejado fueron utilizados en Ion Torrent PGM y MiSeq, fuimos capaces de determinar la distribución del tamaño del fragmento de ADN nativo. Las longitudes de lectura media de las lecturas 179±7 bp y 172±7 bp fueron grabadas en IonTorrent PGM y MiSeq, respectivamente. Para alcanzar condiciones comparables para la comparación de las dos plataformas SGS de banco testadas en el paso de validación las cuentas de lectura bruta ganadas fueron normalizadas y el máximo de 3 millones de lecturas fueron seleccionadas de forma aleatoria para todas las muestras para pasos posteriores. Tras filtrar cualitativamente las lecturas, trazado de lectura, selección de indicadores de trazado único, exclusión de indicadores trazados a regiones cubiertas más de 10 veces por encima de la media de cubierta y corrección GC de la cuenta media de los tan

denominados indicadores finales que decrecen desde 3 millones y por muestra a 2.160±0.065 millones y 2.099±0.221 millones de indicadores para IonTorrent PGM y MiSeq, respectivamente (para mayor detalle véase *Minarik et al., 2015*).

Ejemplo 2

Procedimiento de preparación de muestra para secuenciador MiSeq (Illumina Inc., San Diego, CA, EEUU)

Completar el método de obtención de los datos de secuencia se describe a continuación como una sencilla lista de los principales procedimientos ya que la mayoría de los procedimientos de laboratorio rutinarios fueron utilizados. También son bien conocidos los protocolos para pasos específicos para las personas especializadas en el tema de la biología molecular, examen prenatal y bioinformática y tal persona es consciente de las posibles modificaciones de los procedimientos. Varios protocolos especiales fueron también facilitados por el fabricante de los kits o plataforma de secuenciación. Una descripción más detallada de procedimientos de laboratorio puede encontrarse en Minarik et al., 2015.

1. RECOGIDA DE MUESTRA

El método fue validado utilizando muestras de 12 semanas de embarazo. 2 x 10mL muestras de sangre fueron recogidas en tubos K3-EDTA, y procesadas en el siguiente paso dentro de las siguientes 48 horas.

2. SEPARACIÓN DEL PLASMA

Dos etapas de separación de plasma fueron llevadas a cabo para permitir resultados óptimos de la prueba. Para mayor detalle del procedimiento véase Minarik et al. 2015.

3. AISLAMIENTO DE ADN

Se utilizó el kit Mini Sangre Qiagen. Muestras de plasma fueron procesadas inmediatamente o almacenadas y procesadas en el plazo de un mes tras el procedimiento de separación del plasma sin observar efecto alguno en los resultados de la prueba. Para mayor detalle del procedimiento véase Minarik et al. 2015.

4. PREPARACIÓN DE LIBRERÍA

El procedimiento comprende los siguientes pasos:

a. MEDIR LA CONCENTRACIÓN DEL ADN AISLADO

b. FINALIZAR REPARACIÓN

c. SELECCIÓN DE TAMAÑO

d. ELIMINACIÓN DE LA POLIADENILACIÓN, LIGADURA DE ADAPTADOR

e. AMPLIFICACIÓN PCR

f. QUANTIFICACIÓN DE LIBRERÍA

5. ANÁLISIS DE SECUENCIACIÓN

El método fue validado con las siguientes configuraciones de análisis de secuenciación:

Tipo de análisis: sólo FastQ

Tipo de librería: Truseq LT

Secuenciación del extremo emparejado: Si

Longitud de secuenciación: 2 x 100

Tiempo de recorrido de secuenciación aproximado: 24h

Archivos FastQ fueron descargados en el servicio web sobre un navegador web Firefox. Informes generados pueden ser posteriormente descargados utilizando el mismo servicio web.

Procedimiento con plataforma Ion Torrent

En el procesamiento de muestras para la plataforma Ion Torrent (Life Technologies Corp., San Francisco, CA, EEUU), los pasos 1 y 2 son necesariamente idénticos, los otros son similares o diferentes, siendo específica esta plataforma de secuenciación concreta. Como los protocolos los facilita el productor o se publican en algún otro lugar, la persona especializada sabe cómo obtener datos de secuencia utilizando esta plataforma.

Ejemplo 3

Validación del método innovador con dos plataformas de secuenciación y comparación con otros métodos nivel z

En la actualidad, hay varios métodos para la detección de trisomía basados en la determinación del nivel z. Nosotros elegimos comparar el método de acuerdo con la presente invención, designada MUL, con los tres métodos más comunes ya en uso. Denominamos estos métodos basados en los autores de los artículos que los describían:

1. CHIU (método descrito en *Chiu et al.,* 2008),

2. LAU (método descrito en *Lau et al*., 2011), y

3. SEH (método descrito en *Sehnert et al*., 2011)

El método CHIU, brevemente abreviado, se basa en la representación genómica fraccional (RGF) del cromosoma 21, que es un ratio r_{21} del número de indicadores trazados (tras corrección GC) a cromosoma 21 del número total de indicadores trazados (tras la corrección GC y excluyendo indicadores trazados a cromosomas sexuales). El medio y desviación estándar de r_{21} de las muestras de control fue utilizado para normalizar el valor r_{21} de una muestra con cariotipo desconocido

$$z_{21} = \frac{r_{21} - E[r_{21}|\text{normal control}]}{\sqrt{var[r_{21}|\text{normal control}]}},$$

y si el valor normalizado r_{21} es mayor que 3, la muestra se clasifica como trisómica (esto es porque *Chiu et al*., 2008 asumió que los valores normalizados siguen la distribución normal estándar).

Sehnert et al., 2011 y *Lau et al.,* 2012 propusieron un algoritmo diferente que calcula el ratio q_k del número de indicadores trazados al cromosoma 21 hasta el número de indicadores trazados a algún otro cromosoma (tan denominado cromosoma de referencia interna). Mientras *Sehnert et al.,* 2011 eligió el cromosoma 9 para ser la referencia interna óptima para el cromosoma 21*, Lau et al*., 2012 lo eligió para que fuera el cromosoma 14. Tal fue el caso con el método CHIU, el medio y desviación estándar de q_{21} de las muestras de control fue utilizado para normalizar el valor q_{21} de una muestra con cariotipo desconocido

$$\zeta_{21} = \frac{q_{21} - E[q_{21}|\text{normal control}]}{\sqrt{var[q_{21}|\text{normal control}]}},$$

y si el valor ζ_{21} es mayor que 3, la muestra se clasifica como trisómica (fue de nuevo aceptado por *Sehnert et al*., 2011 y *Lau et al.,* 2012 que los valores normalizados ζ siguen la distribución normal estándar).

Para todos los métodos nivel z, incluyendo el método MUL de acuerdo con la presente invención (esquemáticamente representada en el Gráfico 2), los datos de cuenta de indicador de entrada fueron iguales (ver anteriormente EL MÉTODO, MUESTRAS DE CAPACITACIÓN, sin paso de selección de tamaño opcional) El grupo de capacitación consistía en 30 muestras euploides iguales para todos los métodos. Del resto de las 71 muestras de prueba, había 24 casos T21 y 47 casos euploides. Además, los depósitos de referencia interna IRB para nuestro método MUL se encontraron tal y como hemos descrito anteriormente (véase EL MÉTODO, MUESTRAS DE CAPACITACIÓN). Pestaña 1 muestra niveles z de muestras T21 de nuestro grupo de prueba T21 (los valores más altos aparecen en negrita). Con el valor límite 3 no se observaron falsos negativos en ninguno de los métodos mencionados.

Las diferencias observadas en niveles z entre MiSeq e IonTorrent pueden atribuirse a diferencias entre procesamiento de muestra biológica, desviaciones de la máquina u otros factores desconocidos.

Aún más, la tabla 2 muestra resultados para 6 muestras con trisomía de cromosoma 18 y la tabla 3 muestra resultados para 5 muestras con trisomía de cromosoma 13. Los datos de secuencia para muestras T18 y T13 se obtuvieron utilizando la plataforma MiSeq.

Asimismo, una muestra con monosomía de cromosoma 18 fue analizada. Los resultados aparecen en la tabla 4. Los valores nivel z son negativos en el caso de monosomía ya que en este caso falta una copia del cromosoma. Como resultado, la cantidad de indicadores trazados del cromosoma 18 es menor que en muestras normales con dos copias del cromosoma 18, lo cual muestra el valor de nivel z.

Naturalmente, no se puede sacar una conclusión de amplio alcance de un experimento, sin embargo, el ejemplo mostró que la monosomía puede ser detectada por nuestro método al menos tan bien como con métodos anteriores.

Tal y como aparece en las tablas 1 a 4, los valores de nivel z de muestras trisómicas determinados por el método de la presente invención son significativamente más altos que los valores determinados por cualquiera de los métodos anteriores. Por ejemplo, para muestras T21 obtuvimos una diferencia media de valores de nivel z -1.895, -1.595, -2.485 al comparar CHIU, SEH y LAU, respectivamente, con nuestro método con diferencia valor $p < 0.05$ para todo tema anterior de acuerdo con la prueba de comparaciones múltiples de Dunn, mientras que valores de muestras normales (no trisómicas) son sustancialmente idénticas (respecto al cromosoma 21 obtuvimos diferencia media de nuestro método -0.15, -0.05, -0.21 para CHIU, SEH y LAU, respectivamente), lo cual significa que la distancia de niveles z de muestras trisómicas del valor límite ha incrementado para la presente invención comparado con los anteriores métodos mientras que la distancia de las muestras euploides desde este valor límite se ha mantenido básicamente.

Tabla 1: Muestras trisómicas T21 en MiSeq y IonTorrent. Todas las muestras fueron correctamente clasificadas por cada método al nivel límite 3. El valor más alto de nivel z (marcado en negrita) para cada muestra y máquina se consiguió por el método MUL de acuerdo con la presente invención. Cada fila se corresponde con una muestra biológica

MiSeq z-score				IonTorrent z-score			
CHIU	LAU	SEH	MUL	CHIU	LAU	SEH	MUL
31.13	28.6	28.64	**37.94**	27.25	27.3	27.12	**33.79**
19.01	17.8	17.15	**22.43**	18.26	17.47	16.35	**22.39**
17.77	16.15	16.12	**20.09**	15.51	14.8	14.35	**18.43**
14.99	14.14	13.57	**18.34**	13.28	13.27	12.31	**15.05**
14.21	13.13	12.51	**16.48**	12.95	13.03	11.6	**15.98**
12.45	12.14	11.58	**15.28**	13.08	12.34	12.59	**16.07**
11.88	10.8	10.85	**14.79**	12.64	12.85	12.25	**15.36**
12.01	10.78	9.77	**14.64**	9.72	9.47	9.29	**11.69**
10.43	9.36	9.43	**12.79**	11.56	12.17	10.52	**13.67**
10.56	9.55	9	**10.68**	8.57	7.77	8.44	**10.91**
9.87	9.05	8.47	**11.91**	6.87	6.77	6.07	**8.6**
9.44	9.55	9.11	**10.77**	10.49	10.68	9.07	**12.71**
9.43	8.75	8.8	**11.13**	10.56	11.1	9.88	**12.36**
8.87	8.25	8.58	**10.84**	8.14	7.59	7.68	**9.47**
8.4	7.71	7.95	**9.77**	5.45	6.03	5.34	**7.28**
8.45	7.88	7.87	**10.63**	8.23	7.94	7.65	**9.98**
8.8	8.73	7.55	**9.42**	6.94	7.17	7.59	**8.5**
8.13	8.08	7.07	**9.95**	7.71	6.73	7.92	**9.99**
8.41	7.66	7.43	**9.78**	6.88	7.03	5.59	**8.61**
8.15	8.29	7.76	**9.71**	7.21	6.36	6.75	**9.56**
7.28	7.15	6.31	**8.75**	5.76	5.58	5.53	**7.12**
7.02	6.31	6.91	**8.28**	5.91	6.35	5.94	**6.78**
7.15	6.25	6.13	**8.38**	5.96	5.23	5.58	**6.48**
5.02	5.48	4	**6.32**	6.12	5.93	5.58	**8.15**

Tabla 2: Nivel z para T18 (MiSeq, 6 muestras)

CHIU	LAU	SEH	MUL
17,9	15,5	17,65	**22,2**
13,36	10,54	11,82	**13,8**
9,82	8,53	9,79	**16,54**
9,42	8,57	9,29	**10,78**
5,73	4,75	5,39	**6,63**
7,23	6,35	8,78	**9,83**

Tabla 3: Nivel z para T13 (MiSeq, 5 muestras)

CHIU	LAU	SEH	MUL
22,59	21,88	26,46	**30,73**
16,91	16,8	19,76	**25,47**
22,98	22,16	26,56	**31,76**
21,41	20,4	24,65	**24,86**
3,92	4,75	4,79	**6,44**

Tabla 4: Nivel z para cromosoma monosómico 18 (MiSeq, 1 muestra)

CHIU	LAU	SEH	MUL
-17,34	-12,67	-18,67	-16,14

REFERENCIAS

Documentación distinta de las patentes

Bianchi, D.W., L.D. Platt, J.D. Goldberg, A.Z. Abuhamad, A.J. Sehnert, y R.P. Rava, *Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencin*g. Obstet Gynecol, 2012. **119**(5): p. 890-901.

Chiu, R. W., K. A. Chan, Y. Gao, V. Y. Lau, W. Zheng, T. Y. Leung, C. H. Foo, B. Xie, N. B. Tsui, F. M. Lun, et al., *Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of dna in maternal plasma*. Proceedings of the National Academy of Sciences, 105(51):20458-20433, 2008.

Chiu, R.W., R. Akolekar, Y.W. Zheng, T.Y. Leung, H. Sun, K.C. Chan, F.M. Lun, A.T. Go, E.T. Lau, W.W. To, W.C. Leung, R.Y. Tang, S.K. Au-Yeung, H. Lam, Y.Y. Kung, X. Zhang, J.M. van Vugt, R. Minekawa, M.H. Tang, J. Wang, C.B. Oudejans, T.K. Lau, K.H. Nicolaides, y Y.M. Lo, *Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study*. BMJ, 2011. **342**: p. c7401.

Ehrich, M., C. Deciu, T. Zwiefelhofer, J.A. Tynan, L. Cagasan, R. Tim, V. Lu, R. McCullough, E. McCarthy, A.O. Nygren, J. Dean, L. Tang, D. Hutchison, T. Lu, H. Wang, V. Angkachatchai, P. Oeth, C.R. Cantor, A. Bombard, and D. van den Boom, *Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting*. Am J Obstet Gynecol, 2011. **204**(3): p. 205 el-11.

Fan, H.C., Y.J. Blumenfeld, U. Chitkara, L. Hudgins, and S.R. Quake, *Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood*. Proc Natl Acad Sci USA, 2008. **105**(42): p. 16266-71.

Fortin, F.-A., et al. "DEAP: Evolutionary algorithms made easy." *The Journal of Machine Learning Research* 13.1 (2012): 2171-2175.

Hudecova, l., D. Sahota, M. Heung, Y. Jin, W.S. Lee, T.Y. Leung, Y.M.D. Lo, and R.W. Chiu*., Maternal plasma fetal dna fractions in pregnancies with low and high risks far fetal chromosomal aneuploidies.* PloS one, 9(2):e88484, 2014.

Langmead, B., Salzberg, S., *Fast gapped-read alignment with Bowtie 2*. Nature Methods. 2012, 9:357-359

Lau, T. K., F. Chen, X. Pan, R.K. Pooh, F. Jiang, Y. Li, H. Jiang, X. Li, S. Chen, and X. Zhang. *Noninvasive prenatal diagnosis of common fetal chromosomal aneuploidies by maternal plasma DNA sequencing*. Joumal of Maternal-Fetal and Neonatal Medicine, 25(8): 1370{ 1374, 2012.

Liao, C., A.H. Yin, C.F. Peng, F. Fu, J.X. Yang, R. Li, Y.Y. Chen, D.H. Luo, Y.L. Zhang, Y.M. Ou, J. Li, J. Wu, M.Q. Mai, R. Hou, F. Wu, H. Luo, D.Z. Li, H.L. Liu, X.Z. Zhang, and K. Zhang, *Noninvasive prenatal diagnosis of common aneuploidies by semiconductor sequencing.* Proc Natl Acad Sci USA, 2014. **111**(20): p. 7415-20.

Lo, Y.M., N. Corbetta, P.P. Chamberlain, V. Rai, I.L. Sargent, C.W. Redman, and J.S. Wainscoat, *Presence of fetal DNA in maternal plasma and serum*. Lancet, 1997. **350**(9076): p. 485-7.

Rava, R.P., A. Srinivasan, A.J. Sehnert, and D.W. Bianchi. *Circulating fetal cell-free DNA fractions differ in autosomal aneuploidies and monosomy X.* Clinical chemistry, 60(1):243-250, 2014.

Minarik, G., Repiska, G., Hyblova, M., Nagyova, E., Soltys, K., Budis, J. et al. *Utilization of Benchtop Next Generation Sequencing Platforms Ion Torrent PGM and MiSeq in Noninvasive Prenatal Testing far Chromosome 21 Trisomy and Testing of Impact of In Silico and Physical Size Selection on lts Analytical Performance*. PLoS One. 2015;10(12):e0144811. doi: lü.1371/joumal.pone.0144811.

Sehnert, A.J., B. Rhees, D. Comstock, E. de Feo, G. Heilek, J. Burke, and R.P. Rava, *Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood*. Clin Chem, 2011. **57**(7): p. 1042-9.

Yu, S.C. et al., *Size-based molecular diagnostics using plasma DNA far noninvasive prenatal testing*. Proceedings of the National Academy of Sciences 111.23 (2014): 8583- 8588.

Documentos de Patentes

EP 2183693

EP 2366031

US 8296076

WO 2011/051283

REIVINDICACIONES

1. Un método para determinar la aneuploidía fetal del cromosoma de interés desde una muestra de sangre materna que incluye una mezcla de moléculas de ADN libres de células y de origen materno, dicho método comprende:

a) aislamiento del ADN libre de células de la sangre materna;

b) realizar una secuenciación aleatoria en al menos una porción de una pluralidad de las moléculas de ADN libres de células contenidas en la sangre, la porción incluye ADN correspondiente al menos a un cromosoma de interés y una pluralidad de cromosomas presumiblemente euploides, por lo tanto, se obtendría una pluralidad de lecturas de secuencia;

c) trazado de lecturas de secuencia al genoma humano de referencia, así se obtendría una pluralidad de indicadores trazados;

d) dividir dicho genoma de referencia humano en 1 Mbp posterior y depósitos que no se superpusieran;

e) determinación de la distribución de los indicadores a dichos depósitos y cromosomas completos de acuerdo con su ubicación de trazado;

f) determinar el valor de depósito fraccional FB como un ratio de número de indicadores trazados a los depósitos para el cromosoma de interés en la muestra al número de indicadores trazados a depósitos de referencia interna IRB;

g) determinar medio de muestra multinomial $\mu(n)$ y desviación estándar multinomial $\sigma(n)$; de donde $\mu(n)$ y $\sigma(n)$ se calculan para una variable aleatoria definida como un ratio entre el número de indicadores trazados a depósitos de referencia interna IRB, de donde la probabilidad de distribución de trazado de los indicadores al cromosoma de interés y a los depósitos de referencia interna IRB es multinomial, y los parámetros de dicha distribución multinomial se determinan del grupo de capacitación de muestras de sangre euploides;

h) determinar el valor nivel z como una diferencia entre valor de depósito fraccional FB y medio multinomial $\mu(n)$ dividido por la desviación multinomial estándar $\sigma(n)$; y

i) comparar el valor nivel z con al menos un valor límite para determinar si una aneuploidía existe para el cromosoma de interés; así, como prerrequisito de los pasos f)-i), un grupo de capacitación de muestras de sangre euploides está sujeto a esencialmente los mismos procedimientos que consisten en pasos a)-e) mencionados y después los siguientes pasos se llevan a cabo:

f´) elegir autosomas de referencia interna IRA de autosomas de referencia interna candidata basándose en el coeficiente de variación CV del mismo, mientras que los cromosomas aneuploides quedan excluidos;

g´) elegir los depósitos de referencia interna IRB mediante un algoritmo genético de dos grupos candidatos, el primer grupo candidato se establece de depósitos seleccionados aleatoriamente, mientras que depósitos que corresponden a cromosomas potencialmente aneuploides quedan excluidos, y el segundo grupo candidato es el grupo de depósitos que corresponde con los mejores autosomas de referencia interna IRA determinados en el paso anterior f´.

2. El método conforme a la reclamación 1, en la que el cromosoma de interés es el cromosoma 8, 9, 13, 16, 18, 21, 22 y X.

3. El método conforme a la reclamación 1 o 2, donde la aneuploidía es trisomía del cromosoma 13, cromosoma 18 o cromosoma 21.

4. El método conforme a cualquiera de las reclamaciones 1-3, donde el grupo de capacitación de muestras de sangre euploides comprende muestras masculinas euploides y femeninas euploides.

5. El método conforme a cualquiera de las reclamaciones 1-4, comprende además el paso del filtrado de tamaño que consiste en la limitación del tamaño de lecturas aceptable para un procesamiento adicional.

6. El método conforme a cualquiera de las reclamaciones 1-5, donde al elegir los autosomas de referencia interna IRA, para cada combinación de referencia interna potencial el coeficiente de variación se calcula, las combinaciones de referencia interna candidata de autosomas se ordenan de acuerdo con su valor CV en orden creciente y luego combinaciones de referencia interna cuyo valor CV es menor que o igual que 1.1

múltiple del total del valor más bajo CV se eligen como IRA, de donde preferiblemente la primera centena de estas combinaciones son elegidas como IRA para procesamiento adicional.

7. El método conforme a cualquiera de las reclamaciones 1-6, donde elegir los depósitos de referencia interna IRB por medio de un algoritmo genético, se elige el primer grupo candidato del grupo candidato basándose en sus valores FB utilizando el coeficiente de rango estudiantil y al mismo tiempo basándose en sus valores CV.

8. El método conforme a cualquiera de las reclamaciones 1-7 que comprenden además la determinación de fracción fetal.

9. El método conforme a la reclamación 8 que comprende además la determinación del error de estimación de fracción fetal.

10. El método conforme a cualquiera de las reclamaciones 1-9 que comprende además la determinación de cuenta de indicador mínima para predicción fiable de aneuploidía.

11. Un producto de programa de ordenador que comprende un medio legible por el ordenador que incluye una pluralidad de instrucciones para controlar un sistema de ordenador para realizar todos los pasos del método conforme a cualquiera de las reclamaciones 1-10 siguiendo el paso b) de realizar secuenciación aleatoria.

12. Un método implementado por ordenador para determinar aneuploidía fetal que comprende todos los pasos del método de acuerdo con cualquiera de las reclamaciones 1-10 siguiendo el paso b) de realizar una secuenciación aleatoria.

Gráfico 1

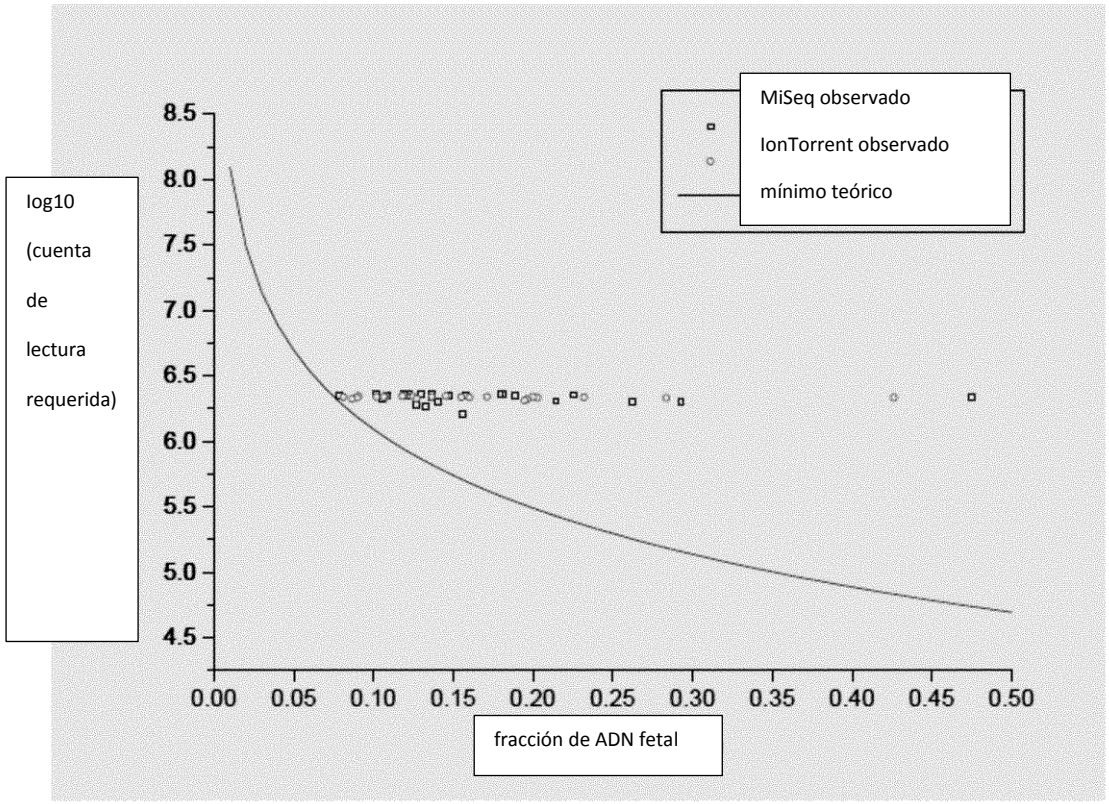
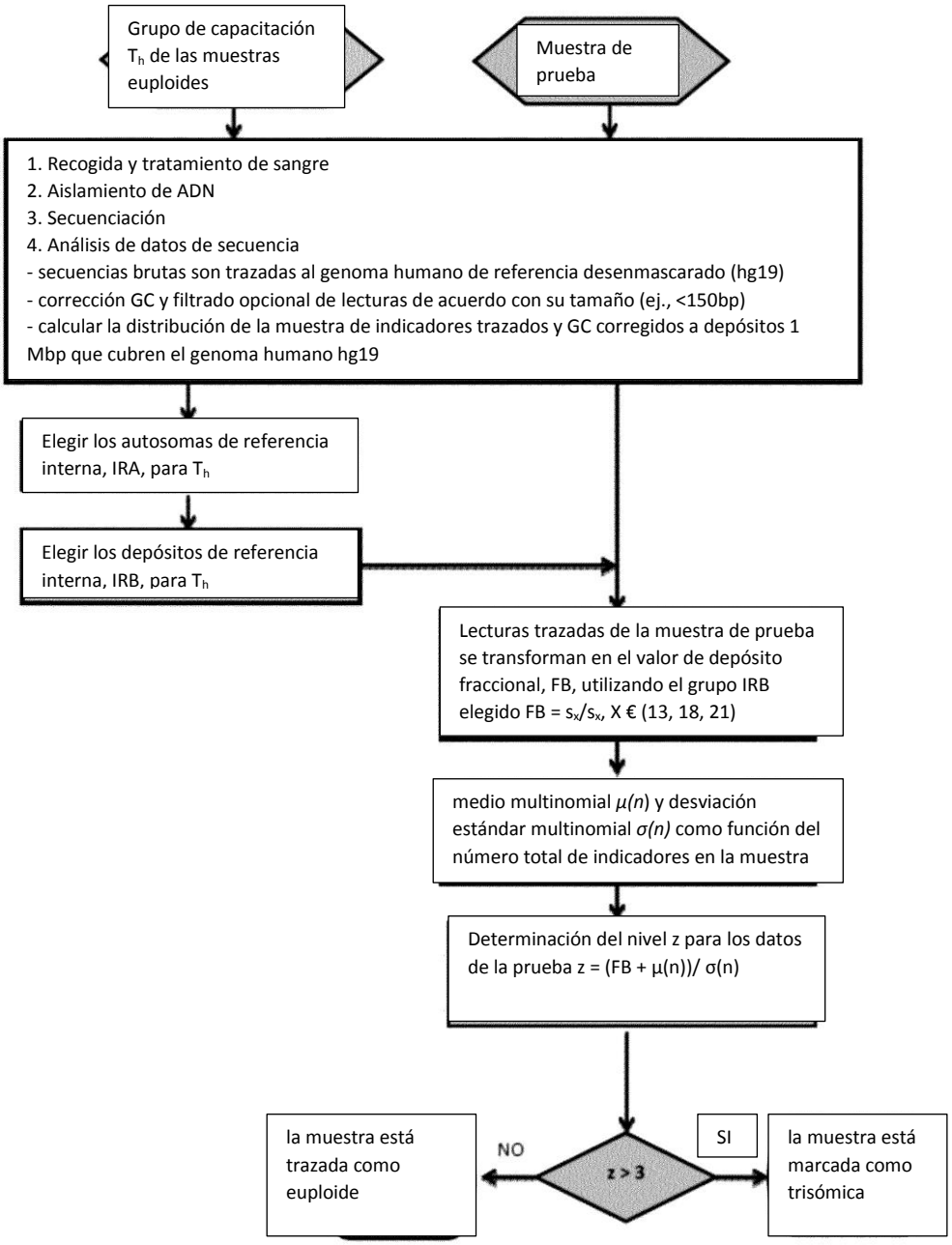

MiSeq observado
IonTorrent observado
mínimo teórico
log10 (cuenta de lectura requerida)
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
fracción de ADN fetal

Gráfico 2


Grupo de capacitación T_h de las muestras euploides
Muestra de prueba
1. Recogida y tratamiento de sangre
2. Aislamiento de ADN
3. Secuenciación
4. Análisis de datos de secuencia
- secuencias brutas son trazadas al genoma humano de referencia desenmascarado (hg19)
- corrección GC y filtrado opcional de lecturas de acuerdo con su tamaño (ej., <150bp)
- calcular la distribución de la muestra de indicadores trazados y GC corregidos a depósitos 1 Mbp que cubren el genoma humano hg19
Elegir los autosomas de referencia interna, IRA, para T_h
Elegir los depósitos de referencia interna, IRB, para T_h
Lecturas trazadas de la muestra de prueba se transforman en el valor de depósito fraccional, FB, utilizando el grupo IRB elegido FB = s_x/s_x, X € (13, 18, 21)
medio multinomial $\mu(n)$ y desviación estándar multinomial $\sigma(n)$ como función del número total de indicadores en la muestra
Determinación del nivel z para los datos de la prueba z = (FB + μ(n))/ σ(n)
NO
z > 3
SI
la muestra está trazada como euploide
la muestra está marcada como trisómica