(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 620 012**

(51) Int. CI.:

C12Q 1/68 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **16.09.2009 PCT/US2009/057136**

(87) Fecha y número de publicación internacional: **25.03.2010 WO2010033578**

(96) Fecha de presentación y número de la solicitud europea: **16.09.2009 E 09815105 (3)**

(97) Fecha y número de publicación de la concesión europea: **21.12.2016 EP 2334812**

(54) Título: **Diagnóstico no invasivo de la aneuploidia fetal por secuenciación**

(30) Prioridad:

20.09.2008 US 98758 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
27.06.2017

(73) Titular/es:

THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY (100.0%)
Office of the General Counsel Building 170, 3rd Floor, Main Quad P O Box 20386
Stanford, CA 94305-2038, US

(72) Inventor/es:

FAN, HEI-MUN, CHRISTINA y
QUAKE, STEPHEN, R.

(74) Agente/Representante:

IZQUIERDO BLANCO, María Alicia

Diagnóstico no invasivo de la aneuploidia fetal por secuenciación

Descripción

Campo de la invención

[0001] La presente invención se refiere al campo del diagnóstico molecular, y más particularmente al campo del diagnóstico genético prenatal.

Técnica relacionada

[0002] A continuación se presenta la información de fondo sobre determinados aspectos de la presente invención, ya que pueden estar relacionados con las características técnicas mencionadas en la descripción detallada, pero no necesariamente se describen en detalle. Es decir, ciertos componentes de la presente invención se pueden describir con mayor detalle en los materiales discutidos a continuación. La discusión siguiente no debe interpretarse como una admisión en cuanto a la relevancia de la información para la invención reivindicada o el efecto de la técnica anterior del material descrito.

[0003] Aneuploidía fetal y otras aberraciones cromosómicas afectan 9 de cada 1000 nacidos vivos (1). El patrón de oro para el diagnóstico de las anomalías cromosómicas es el cariotipo de las células fetales obtenido a través de procedimientos invasivos como la vellosidad coriónica de muestreo y amniocentesis. Estos procedimientos imponen riesgos pequeños pero potencialmente significativos tanto al feto como a la madre (2). La detección no invasiva de la aneuploidía fetal utilizando marcadores séricos maternos y ultrasonidos están disponibles, pero tienen una fiabilidad limitada (3-5). Por lo tanto, existe el deseo de desarrollar pruebas genéticas no invasivas para anomalías cromosómicas fetales.

[0004] Desde el descubrimiento de las células fetales intactas en la sangre materna, ha habido un interés intenso en el intento de utilizarlas como una ventana de diagnóstico en la genética fetal (6-9). Si bien esto todavía no se ha aplicado en la práctica (10), el descubrimiento posterior de que cantidades significativas de ácidos nucleicos fetales libres de células también existen en la circulación materna ha llevado al desarrollo de nuevas pruebas genéticas prenatales no invasivas para una variedad de rasgos (11, 12). Sin embargo, la medición de la aneuploidía sigue siendo un reto debido al fondo alto de ADN materno; El ADN fetal a menudo constituye <10% del ADN total en el plasma libre de células maternas (13).

[0005] Los métodos desarrollados recientemente para la aneuploidía se basan en el enfoque de detección de la variación alélica entre la madre y el feto. Lo et al. demostraron que las proporciones alélicas de ARNm específico de la placenta en el plasma materno podrían utilizarse para detectar la trisomía 21 en ciertas poblaciones (14).

[0006] Del mismo modo, también mostraron el uso de relaciones alélicas de los genes impresos en el ADN del plasma materno para diagnosticar la trisomía 18 (15). Dhallan et al. utilizaron alelos fetales específicos en el ADN del plasma materno para detectar la trisomía 21 (16). Sin embargo, estos métodos se limitan a poblaciones específicas porque dependen de la presencia de polimorfismos genéticos en loci específicos. Nosotros y otros argumentamos que debería ser posible en principio utilizar la PCR digital para crear una prueba universal, independiente del polimorfismo para la aneuploidía fetal usando ADN plasmático materno (17-19).

[0007] Un método alternativo para lograr la cuantificación digital del ADN es la secuenciación de escopeta directa seguida de mapeo para el cromosoma de origen y el recuento de fragmentos por cromosoma. Los recientes avances en la tecnología de secuenciación de ADN permiten una secuenciación masiva paralela (20), produciendo decenas de millones de secuencias cortas en un solo recorrido y permitiendo un muestreo más profundo de lo que puede lograrse mediante PCR digital. Como se conoce en la técnica, el término "etiqueta de secuencia" se refiere a una secuencia de ácido nucleico relativamente corta (por ejemplo, 15-100) que puede usarse para identificar una cierta secuencia mayor, por ejemplo, mapearse a un cromosoma o región genómica o gene. Estos pueden ser ESTs o etiquetas de secuencias expresadas obtenidas a partir de ARNm.

Patentes y Publicaciones Específicas

[0008] Science 309: 1476 (2 de septiembre de 2005) News Focus "An Earlier Look at Baby's Genes" describe los intentos de desarrollar pruebas para el síndrome de Down utilizando la sangre materna. Los primeros intentos de detectar el síndrome de Down utilizando células fetales de sangre materna fueron llamados "sólo modestamente prometedores". El informe también describe el trabajo de Dennis Lo para detectar el gen Rh en un feto donde está ausente en la madre. Otras mutaciones transmitidas por el padre también se han detectado, como fibrosis quística, beta-talasemia, una clase de enanismo y la enfermedad de Huntington. Sin embargo, estos resultados no siempre han sido reproducibles.

[0009] Venter et al, "The sequence of the human genome," Science, 2001 Feb 16;291(5507):1304-51, describe la secuencia del genoma humano, cuya información está disponible públicamente de NCBI. Otra secuencia genómica

de referencia es una construcción NCBI actual obtenida a partir de la pasarela de genoma UCSC.

[0010] Wheeler et al, "The complete genome of an individual by massively parallel DNA sequencing," Nature, 2008 Apr 17;452(7189):872-6 describe la secuencia de ADN de un genoma diploide de un solo individuo, James D. Watson, secuenciado a redundancia de 7,4 veces en dos meses usando secuenciación masivamente paralela en recipientes de reacción de tamaño picolitro. La comparación de la secuencia con el genoma de referencia condujo a la identificación de 3,3 millones de polimorfismos de un solo nucleótido, de los cuales 10,654 causan la sustitución de aminoácidos dentro de la secuencia codificante.

[0011] Quake et al., US 2007/0202525 titulada "Non-invasive fetal genetic screening by digital analysis", publicado 30 de agosto 2007 (y la aplicación relacionada WO 2007/092473), describe un proceso en el que la sangre materna contiene ADN fetal se diluye hasta un valor nominal de aproximadamente 0,5 equivalentes de genoma de ADN por muestra de reacción.

[0012] Chiu et al., "Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic DNA sequencing of DNA in maternal plasma," Proc. Natl. Acad. Sci. 105 (51): 20458 - 20463 (23 de diciembre de 2008) describe un método para determinar la aneuploidía fetal usando secuenciación masivamente paralela. La determinación del estado de la enfermedad (aneuploidia) se hizo calculando una "puntuación z". Puntuaciones Z se compararon con valores de referencia, de una población restringida a fetos masculinos euploides. Los autores señalaron de pasada que el contenido G/C afectó el coeficiente de variación.

[0013] Lo et al., "Diagnosing Fetal Chromosomal Aneuploidy Using Massively Parallel Genomic Sequencing" de Estados Unidos 2009/0029377, publicada el 29 de enero de, 2009 (y la aplicación relacionada WO 2009/013496), describe un método en el cual respectivas cantidades de una cromosoma clínicamente relevante y los cromosomas de fondo se determinan a partir de resultados de secuenciación masivamente paralelos. Se encontró que el porcentaje de representación de las secuencias asignadas al cromosoma 21 es más alto en una mujer embarazada portadora de un feto de trisomía 21 cuando se compara con una mujer embarazada que lleva un feto normal. Para las cuatro mujeres embarazadas que portaban un feto euploide, una media del 1,345% de sus secuencias de ADN plasmático se alinearon con el cromosoma 21.

[0014] Lo et al., "Determining a Nucleic Acid Sequence Imbalance", US 2009/0087847, publicada el 2 de abril de 2009 describe un método para determinar si existe un desequilibrio de secuencia de ácido nucleico, tal como una aneuploidía, comprendiendo el método el derivo de un primer valor de corte a partir de una concentración media de una secuencia de ácido nucleico de referencia en cada una de una pluralidad de reacciones, en donde la secuencia de ácido nucleico de referencia es la secuencia de ácido nucleico clínicamente relevante o la secuencia de ácido nucleico de fondo; comparar el parámetro con el primer valor de corte; y basándose en la comparación, determinando una clasificación de si existe un desequilibrio de secuencia de ácido nucleico.

[0015] Lo et al, "Noninvasive prenatal diagnosis of fetal chromosomal aneuploidies by maternal plasma nucleic acid analysis," Clinical Chemistry. 54: 3, 461-466 (Enero 2008) enseña que la aneuploidía fetal ha sido explorada a través del uso del análisis de la relación alélica de marcadores epigenéticos y ARN en plasma fetal, y que se ha demostrado que la PCR digital ofrece alta precisión para la relación alélica y los análisis de dosis cromosómicas relativas.

[0016] El documento WO 2007/100911 (Mitchell) describe polimorfismos de tándem de nucleótido único y métodos para su uso, por ejemplo, en el diagnóstico de Síndrome de Down.

[0017] US 2008/050739 (Stoughton) se relaciona con el diagnóstico de anomalías fetales utilizando polimorfismos incluyendo repeticiones cortas en tándem.

[0018] WO 2005/039389 (Shimkets) enseña métodos Sequence-Based Karyotyping' para la detección de anomalías genómicas.

BREVE RESUMEN DE LA INVENCIÓN

[0019] El siguiente breve resumen no pretende incluir todas las características y aspectos de la presente invención.

[0020] La presente invención proporciona un método de ensayo para una distribución anormal de un cromosoma especificado en una muestra mixta de porciones cromosómicas normalmente y anormalmente distribuidas obtenidas de un sujeto, en el que la muestra es una mezcla de ADN materno y fetal en una muestra de plasma materno, que comprende:

(a) obtener secuencias, mediante secuenciación masivamente paralela, a partir de múltiples porciones cromosómicas de la muestra mezclada para obtener un número de etiquetas de secuencia de longitud suficiente de secuencia determinada para asignarse a una ubicación de cromosoma dentro de un genoma y de número suficiente para reflejar una distribución anormal de la parte cromosómica especificada;

(b) asignar las etiquetas de secuencia a las porciones de cromosomas correspondientes incluyendo al menos la parte de cromosoma especificada comparando las etiquetas de secuencia con una secuencia genómica de referencia;

(c) determinar valores para los números de las etiquetas de secuencia que asignan a porciones cromosómicas normales y anormalmente distribuidas por:

(i) contar etiquetas de secuencia dentro de una serie de ventanas predefinidas de igual longitud dentro de al menos una parte de cromosoma distribuida normalmente para obtener un primer valor; y

(ii) contar etiquetas de secuencia dentro de una serie de ventanas predefinidas de longitudes iguales dentro de la parte de cromosoma especificada para obtener un segundo valor; y

(d) usar los valores del paso (c) para determinar un diferencial, entre el primer valor y el segundo valor, que es determinante de si existe o no la distribución anormal, en el que la distribución anormal es una aneuploidía fetal.

[0021] También se describe en este documento un método para analizar una muestra materna, por ejemplo, a partir de sangre periférica. No es invasivo en el espacio fetal, como lo es la amniocentesis o el muestreo de vellosidades coriónicas. En el método preferido, se utiliza ADN fetal que está presente en el plasma materno. El ADN fetal está en un aspecto de la descripción enriquecido debido al sesgo en el procedimiento hacia fragmentos de ADN más cortos, que tienden a ser ADN fetal. El método es independiente de cualquier diferencia de secuencia entre el genoma materno y fetal. El ADN obtenido, preferiblemente de una extracción de sangre periférica, es una mezcla de ADN fetal y materno. El ADN obtenido está al menos parcialmente secuenciado, en un método que da un gran número de lecturas cortas. Estas lecturas cortas actúan como etiquetas de secuencia, en el sentido de que una fracción significativa de las lecturas son suficientemente únicas para ser mapeadas a cromosomas específicos o localizaciones cromosómicas que se sabe que existen en el genoma humano. Se mapean exactamente, o se pueden asignar con un desajuste, como en los ejemplos a continuación. Mediante el recuento del número de marcas de secuencia asignadas a cada cromosoma (1-22, X e Y), se puede detectar la sobre-representación o sub-representación de cualquier parte de cromosoma o cromosoma en el ADN mezclado aportado por un feto aneuploide. Este método no requiere la diferenciación de secuencias del ADN fetal o materno, ya que la contribución sumada de las secuencias materna y fetal en un cromosoma o parte cromosómica particular será diferente entre un cromosoma diploide intacto y un cromosoma aberrante, es decir, con una copia extra, parte faltante o similar. En otras palabras, el método no se basa en una información de secuencia a priori que distinguiría el ADN fetal del ADN materno. La distribución anormal de un cromosoma fetal o parte de un cromosoma (es decir, una deleción o inserción macroscópica) se puede determinar en el presente método por enumeración de etiquetas de secuencia como mapeadas a diferentes cromosomas. El recuento mediano de valores autosómicos (es decir, el número de etiquetas de secuencia por autosoma) se utiliza como una constante de normalización para tener en cuenta las diferencias en el número total de etiquetas de secuencia para la comparación entre muestras y entre cromosomas. El término "parte cromosómica" se usa en la presente memoria para designar un cromosoma entero o un fragmento significativo de un cromosoma. Por ejemplo, el síndrome de Down moderado se ha asociado con la trisomía parcial 21q22.2 qter. Mediante el análisis de la densidad de etiquetas de secuencias predefinidas en las subsecciones de cromosomas (por ejemplo, 10 a 100 kb ventanas), una constante de normalización se puede calcular, y las subsecciones cromosómicas cuantificadas (por ejemplo, 21q22.2). Con un número de etiquetas de secuencia suficientemente grande, el presente método puede aplicarse a fracciones arbitrariamente pequeñas de ADN fetal. Se ha demostrado que es exacto hasta un 6% de concentración fetal de ADN. A continuación se muestra el uso exitoso de la secuenciación de escopeta y la cartografía del ADN para detectar trisomía fetal 21 (síndrome de Down), trisomía 18 (síndrome de Edward) y trisomía 13 (síndrome de Patau), llevada a cabo de forma no invasiva utilizando ADN fetal libre de células en Plasma materno. Esto constituye la base de una prueba diagnóstica no invasiva universal, independiente del polimorfismo, para la aneuploidía fetal. Los datos de la secuencia también nos permiten caracterizar el ADN plasmático en un detalle sin precedentes, lo que sugiere que está enriquecido para los fragmentos unidos a nucleosomas. El método también puede emplearse para que los datos de secuencia obtenidos puedan analizarse adicionalmente para obtener información sobre polimorfismos y mutaciones.

[0022] Por lo tanto, la presente descripción comprende, en ciertos aspectos, un método de ensayo para una distribución anormal de una parte de cromosoma se especifica en una muestra mixta de porciones de cromosomas distribuidos normalmente y anormalmente obtenidos a partir de un único sujeto, tal como una mezcla de ADN fetal y materno en una muestra de plasma materno. Uno lleva a cabo determinaciones de secuencia sobre los fragmentos de ADN en la muestra, obteniendo secuencias de múltiples porciones cromosómicas de la muestra mezclada para obtener una serie de etiquetas de secuencia de longitud suficiente de secuencia determinada para asignarse a una localización cromosómica dentro de un genoma y de número suficiente para reflejar la distribución anormal. Usando una secuencia de referencia, se asignan las etiquetas de secuencia a sus correspondientes cromosomas incluyendo al menos el cromosoma especificado comparando la secuencia con la secuencia genómica de referencia. A menudo habrá en el orden de millones de etiquetas de secuencia corta que se asignan a ciertos cromosomas y, lo que es más importante, ciertas posiciones a lo largo de los cromosomas. Se puede entonces determinar un primer número de etiquetas de secuencia asignadas a al menos una parte de cromosoma distribuida normalmente y un segundo

número de etiquetas de secuencia asignadas a la parte de cromosoma especificada, estando ambos cromosomas en una muestra mixta. El presente procedimiento también implica la corrección de las etiquetas de secuencia de distribución no uniforme en diferentes porciones cromosómicas. Esto se explica en detalle a continuación, en el que se crean varias ventanas de longitud definida a lo largo de un cromosoma, siendo las ventanas del orden de kilobasas de longitud, por lo que una serie de etiquetas de secuencia caerá en muchas de las ventanas y ventanas que cubren cada cromosoma entero en cuestión, con excepciones para regiones no informativas, por ejemplo, regiones de centrómero y regiones repetitivas. Varios números medios, es decir, valores medianos, se calculan para diferentes ventanas y se comparan. Mediante el recuento de etiquetas de secuencia dentro de una serie de ventanas predefinidas de igual longitud a lo largo de cromosomas diferentes, se pueden obtener resultados más robustos y estadísticamente significativos. El presente método también implica calcular una diferencia entre el primer número y el segundo número que es determinante de si existe o no la distribución anormal.

[0023] En ciertos aspectos, la presente descripción puede comprender un ordenador programado para analizar los datos de secuencia obtenidos a partir de una mezcla de ADN cromosómico materno y fetal. Cada autosoma (Cr. 1-22) se segmenta computacionalmente en ventanas contiguas que no se solapan. (También se podría utilizar una ventana corredera). Cada ventana es de longitud suficiente para contener un número significativo de lecturas (etiquetas de secuencia, que tienen aproximadamente 20-100 pb de secuencia) y no tienen todavía un número de ventanas por cromosoma. Típicamente, una ventana estará entre 10kb y 100kb, más típicamente entre 40 y 60 kb. Por lo tanto, habría, por ejemplo, aproximadamente entre 3.000 y 100.000 ventanas por cromosoma. Las ventanas pueden variar ampliamente en el número de etiquetas de secuencia que contienen, en función de la ubicación (por ejemplo, cerca de un centrómero o región repetida) o contenido G/C, como se explica a continuación. Se selecciona el recuento mediano (es decir, valor medio en el conjunto) por ventana para cada cromosoma; entonces la mediana de los valores autosómicos se utiliza para dar cuenta de las diferencias en el número total de etiquetas de secuencias obtenidas para diferentes cromosomas y distinguir la variación intercromosomal de sesgo de secuenciación de aneuploidía. Este método de mapeo también se puede aplicar para discernir deleciones o inserciones parciales en un cromosoma. El presente método también proporciona un método para corregir el sesgo resultante del contenido de G/C. Por ejemplo, se encontró que el método de secuenciación de Solexa produjo más etiquetas de secuencia de fragmentos con mayor contenido de G/C. Al asignar un peso a cada etiqueta de secuencia basada en el contenido G/C de una ventana en la que cae la lectura. La ventana para el cálculo del GC es preferiblemente menor que la ventana para el cálculo de la densidad de la etiqueta de secuencia.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0024]

La Figura 1 es un diagrama de dispersión gráfico que muestra las densidades de secuencia de etiqueta de dieciocho muestras, teniendo cinco genotipos diferentes, como se indica en la leyenda de la figura. La aneuploidía fetal es detectable por la sobre-representación del cromosoma afectado en la sangre materna. **La Figura 1A** muestra etiqueta de secuencia de densidad relativa para el correspondiente valor de control de ADN genómico; los cromosomas se ordenan aumentando el contenido de G/C. Las muestras mostradas como se indica son plasma de una mujer con un feto T21; plasma de una mujer con un feto T18; plasma de un macho adulto normal; plasma de una mujer que lleva un feto normal; plasma de una mujer con un feto T13. Las densidades de las etiquetas de secuencia varían más con el aumento del contenido cromosómico G/C. **La Figura 1B** es un detalle de la **Fig. 1A,** que muestra la densidad de etiqueta de secuencia de cromosoma 21 relativa de la densidad de etiqueta de secuencia de cromosoma mediano 21 de los casos normales. Obsérvese que los valores de 21 casos de 3 disomía se superponen a 1,0. La línea discontinua representa el límite superior del intervalo de confianza del 99% construido a partir de todas las muestras de disomía 21. Los cromosomas se enumeran en la **Figura 1A** con el fin de contenido de G/C, de bajo a alto. Esta cifra sugiere que se prefiere utilizar como un cromosoma de referencia en la muestra mixta con un nivel medio de contenido G/C, ya que se puede ver que los datos están más fuertemente agrupados. Es decir, los cromosomas 18, 8, 2, 7, 12, 21 (excepto en el síndrome de Down sospechado), 14, 9 y 11 pueden ser utilizados como el cromosoma diploide nominal si se busca una trisomía. **La Figura 1B** representa una ampliación del cromosoma 21 de datos.

La Figura 2 es un gráfico de diagrama de dispersión que muestra la fracción de ADN fetal y la edad gestacional. La fracción de ADN fetal en el plasma materno se correlaciona con la edad gestacional. La fracción de ADN fetal se estimó de tres maneras diferentes: 1. A partir de la cantidad adicional de secuencias de cromosomas 13, 18 y$_2$1 para T13, T18 y T21 respectivamente. 2. Del agotamiento en la cantidad de secuencias de cromosoma X para los casos masculinos. 3. De la cantidad de secuencias de cromosoma Y presentes para los casos masculinos. La línea horizontal discontinua representa la fracción mínima estimada de ADN fetal requerida para la detección de aneuploidía. Para cada muestra, se promediaron los valores de fracción de ADN fetal calculados a partir de los datos de diferentes cromosomas. Existe una correlación estadísticamente significativa entre la fracción fetal promedio de ADN y la edad gestacional ($p = 0,0051$). La línea discontinua representa la línea de regresión lineal simple entre la fracción fetal promedio de ADN y la edad gestacional. El valor R2 representa el cuadrado del coeficiente de correlación. **La Figura 2** sugiere que el presente método se puede emplear en una fase muy temprana del embarazo. Los datos se obtuvieron a partir de la etapa de 10 semanas y más tarde porque es la etapa más temprana en la que se realiza el muestreo de vellosidades coriónicas. (La amniocentesis

se hace más tarde). Desde el nivel del intervalo de confianza, se esperaría obtener datos significativos a partir de las 4 semanas de edad gestacional, o posiblemente antes.

La Figura 3 es un histograma que muestra la distribución de tamaño que muestra el ADN materno y fetal en el plasma materno. Muestra la distribución de tamaño de los fragmentos totales y cromosómicos Y específicos obtenidos a partir de la secuenciación 454 del ADN plasmático materno de un embarazo masculino normal. La distribución se normaliza a suma a 1. Los números de lecturas totales y lecturas asignadas al cromosoma Y son 144992 y 178 respectivamente. Inserción: Fracción de ADN fetal acumulada en función del tamaño del fragmento secuenciado. Las barras de error corresponden al error estándar de la fracción estimada suponiendo que el error de los recuentos de fragmentos secuenciados sigue las estadísticas de Poisson.

La Figura 4 es un par de gráficos de líneas que muestran la distribución de las etiquetas de secuencias alrededor de los sitios de inicio de transcripción (TSS) de genes ReSeq en todos los autosomas y el cromosoma X de la muestra de ADN de plasma de un embarazo masculino normal (parte superior, **Fig. 4A)** y control de ADN genómico aleatoriamente esquilado (abajo, **Fig. 4B**). El número de etiquetas dentro de cada ventana de 5 pb fue contado dentro de la región $\pm$ 1000pb alrededor de cada TSS, teniendo en cuenta la hebra de cada etiqueta de secuencia asignada. Los conteos de todos los sitios de inicio de la transcripción para cada ventana de 5 pb se sumaron y se normalizaron al recuento mediano entre las 400 ventanas. Se usó un promedio móvil para suavizar los datos. Un pico en la cadena de sentido representa el comienzo de un nucleosoma, mientras que un pico en la cadena de antisentido representa el final de un nucleosoma. En la muestra de ADN de plasma que se muestra aquí, cinco nucleosomas bien posicionados se observan aguas abajo de los sitios de inicio de la transcripción y se representan como óvalos grises. El número inferior dentro de cada óvalo representa la distancia en pares de bases entre picos adyacentes en las cadenas de sentido y antisentido, correspondiente al tamaño del nucleosoma inferido. No se observa patrón obvio para el control del ADN genómico.

La Figura 5A es un gráfico de diagrama de dispersión que muestra la densidad de etiqueta media de secuencia para cada cromosoma de todas las muestras, incluyendo el ADN de plasma libre de células de las mujeres embarazadas y donante masculino, así como el control de ADN genómico a partir de donante varón, se traza anteriormente. Las excepciones son los cromosomas 13, 18 y $_2$1, donde se excluyen las muestras de ADN libres de células de mujeres portadoras de fetos aneuploides. Las barras de error representan la desviación estándar. Los cromosomas están ordenados por su contenido G/C. El contenido de G/C de cada cromosoma en relación con el genoma de todo el valor (41%) también se representa. **La Figura 5B** es un gráfico de diagrama de dispersión de la densidad de etiqueta media de secuencia para cada cromosoma frente al contenido de G/C del cromosoma. El coeficiente de correlación es 0,927, y la correlación es estadísticamente significativa ($p < 10^{-9}$).

La Figura 5C es un gráfico de diagrama de dispersión de la desviación estándar de la densidad de etiqueta de secuencia de cada cromosoma frente al contenido de G/C del cromosoma. El coeficiente de correlación entre la desviación estándar de la densidad de la etiqueta de secuencia y la desviación absoluta del contenido G/C cromosómico del contenido G/C genómico es 0,963, y la correlación es estadísticamente significativa (p <10-12).

La Figura 6 es un gráfico de diagrama de dispersión que muestra la diferencia porcentual de la densidad de etiqueta de secuencia de cromosoma X de todas las muestras, en comparación con la densidad de etiqueta mediana de secuencia de cromosoma X de todos los embarazos femeninos. Todos los embarazos masculinos muestran una sub-representación del cromosoma X.

La Figura 7 es un gráfico de diagrama de dispersión que muestra una comparación de la estimación de la fracción de ADN fetal para muestras de ADN libres de células a partir de 12 embarazos masculinos utilizando los datos de secuenciación de los cromosomas X e Y. La línea discontinua representa una línea de regresión lineal simple, con una pendiente de 0,85. El valor R2 representa el cuadrado del coeficiente de correlación. Existe una correlación estadísticamente significativa entre la fracción de ADN fetal estimada de los cromosomas X e Y (p = 0,0015).

La Figura 8 es un gráfico de líneas que muestra la distribución de longitud de fragmentos secuenciados de muestra de ADN de plasma libre de células madre de un embarazo normal de sexo masculino a resolución 1 pb. La secuenciación se realizó en la plataforma 454/Roche. Se retienen lecturas que tienen al menos un 90% de mapeo al genoma humano con una precisión mayor o igual al 90%, totalizando 144992 lecturas. El eje Y representa el número de lecturas obtenidas. La longitud media es de 177 pb, mientras que la longitud media es de 180 pb.

La Figura 9 es un esquema que ilustra cómo se utiliza la distribución de etiqueta de secuencia para detectar la sobre e infrarrepresentación de cualquier cromosoma, es decir, una trisomía (sobre la representación) o un cromosoma que falta (típicamente un cromosoma X o Y, al ser autosomas que faltan generalmente letales). Como se muestra en los paneles A y C de la izquierda, se traza primero el número de lecturas obtenidas frente a una ventana que se correlaciona con una coordenada cromosómica que representa la posición de la lectura a lo largo del cromosoma. Es decir, se puede ver que el cromosoma 1 (panel A) tiene aproximadamente 2,8 x 108 pb. Tendría este número dividido por ventanas de 50kb. Estos valores se repliegan (paneles B y D) para mostrar la

distribución del número de etiquetas de secuencia/ventana de 50kb. El término "bin" es equivalente a ventana. A partir de este análisis, se puede determinar un número mediano de lecturas M para cada cromosoma que, a efectos de ilustración, se puede observar a lo largo del eje x en el centro aproximado de la distribución y se puede decir que es mayor si hay más etiquetas de secuencia atribuibles a ese cromosoma. Para el cromosoma 1, ilustrado en los paneles A y B, se obtiene una M1 mediana. Tomando la M mediana de los 22 autosomas, se obtiene una constante de normalización N que puede usarse para corregir las diferencias en las secuencias obtenidas en diferentes series, como se puede ver en la Tabla 1. Así, la densidad de secuencia de secuencia normalizada para el cromosoma 1 sería M1/N; para el cromosoma 22 sería M22/N. Un examen detallado del panel A, por ejemplo, mostraría que hacia el extremo cero del cromosoma, este procedimiento obtuvo aproximadamente 175 lecturas por ventana de 50 kb. En el centro, cerca del centrómero, no había lecturas, porque esta parte del cromosoma está mal definida en la biblioteca del genoma humano.

[0025] Es decir, en los paneles de la izquierda (A y C), se traza la distribución de lecturas por coordenada de cromosoma, es decir, la posición cromosómica en términos de número de lecturas dentro de cada ventana deslizante de 50 kb que no se solapa. Después, se determina la distribución del número de etiquetas de secuencia para cada ventana de 50 kb y se obtiene un número medio de etiquetas de secuencia por cromosoma para todos los autosomas y el cromosoma X (Ejemplos de Cr 1 [arriba] y Cr 22 [abajo] se ilustran aquí). La mediana de los 22 valores de M (de todos los autosomas, cromosomas 1 a 22) se utiliza como normalización de la N constante. La densidad de la etiqueta de secuencia normalizada de cada cromosoma es M/N (por ejemplo, Cr 1: M1/N; Cr 22: M22/N). Tal normalización es necesaria para comparar diferentes muestras de pacientes ya que el número total de las etiquetas de secuencia (por lo tanto, la densidad de la etiqueta de secuencia) para cada muestra de paciente es diferente (el número total de etiquetas de secuencia fluctúa entre ~ 8 a ~ 12 millones). El análisis fluye así de la frecuencia de lecturas por coordenada (A y C) a # lecturas por ventana (B y D) a una combinación de todos los cromosomas.

[0026] La Figura 10 es un gráfico de diagrama de dispersión que muestra datos de diferentes muestras, como en la **Figura 1,** excepto que el sesgo para el muestreo G/C ha sido eliminado.

[0027] La Figura 11 es un gráfico de diagrama de dispersión que muestra el peso dado a las diferentes muestras de secuencias según el porcentaje de contenido de G/C, con menor peso dado a las muestras con un mayor contenido de G/C. El contenido de G/C oscila entre aproximadamente 30% y aproximadamente 70%; el peso puede variar en un factor de aproximadamente 3.

[0028] La Figura 12 es un gráfico de diagrama de dispersión que ilustra los resultados de pacientes seleccionados, como se indica en el eje x, y, para cada paciente, una distribución de la representación del cromosoma en el eje Y, como se desvía de un estadística t representativo, indicado como cero.

[0029] La Figura 13 es un gráfico de diagrama de dispersión que muestra el porcentaje de ADN fetal mínimo del cual sobre- o subrepresentación de un cromosoma se pudo detectar con un nivel de confianza del 99,9% para los cromosomas 21, 18, 13 y Cr. X, y un valor para todos los otros cromosomas.

[0030] La Figura 14 es un gráfico gráfico de diagrama de dispersión que muestra una relación lineal entre log10 de porcentaje de ADN fetal mínimo que se necesita frente a log10 del número de lecturas necesario.

DESCRIPCIÓN DETALLADA DE LA REALIZACIÓN PREFERENTE

Descripción general

Definiciones

[0031] A menos que se defina lo contrario, todos los términos técnicos y científicos usados en este documento tienen el mismo significado que el comúnmente entendido por aquellos de experiencia ordinaria en la técnica a la que pertenece esta invención. Aunque se pueden usar cualesquiera métodos y materiales similares o equivalentes a los descritos aquí en la práctica o pruebas de la presente invención, se describen los métodos y materiales preferidos. Generalmente, las nomenclaturas utilizadas en relación con, y las técnicas de, biología celular y molecular y química son las bien conocidas y comúnmente usadas en la técnica. Ciertas técnicas experimentales, no definidas específicamente, se realizan generalmente de acuerdo con métodos convencionales bien conocidos en la técnica y como se describen en varias referencias generales y más específicas que se citan y discuten a lo largo de la presente memoria descriptiva. A los efectos de la claridad, los siguientes términos se definen a continuación.

[0032] "Densidad de etiqueta de secuencia" significa que el valor normalizado de etiquetas de secuencias para una ventana definida de una secuencia en un cromosoma (en una realización preferida, la ventana es de aproximadamente 50 kb), donde se utiliza la densidad de etiqueta de secuencia para la comparación de diferentes muestras y para análisis subsiguiente. Una "etiqueta de secuencia" es una secuencia de ADN de longitud suficiente que puede asignarse específicamente a uno de los cromosomas 1-22, X o Y. No necesariamente tiene que ser, pero puede ser no repetitivo dentro de un solo cromosoma. Se puede permitir cierto grado pequeño de desajuste (0-1)

para explicar los polimorfismos menores que pueden existir entre el genoma de referencia y los genomas individuales (materno y fetal) que se están siendo mapeados. El valor de la densidad de la etiqueta de secuencia se normaliza dentro de una muestra. Esto puede hacerse contando el número de etiquetas que caen dentro de cada ventana en un cromosoma; obtener un valor mediano del número total de etiquetas de secuencia para cada cromosoma; obtener un valor mediano de todos los valores autosómicos; y utilizando este valor como una normalización constante para tener en cuenta las diferencias en el número total de secuencias de etiquetas obtenidas para diferentes muestras. Una densidad de etiqueta de secuencia calculada de esta manera sería idealmente de aproximadamente 1 para un cromosoma disómico. Como se describe adicionalmente a continuación, las densidades de las etiquetas de secuencia pueden variar de acuerdo con los artefactos de secuenciación, más notablemente el sesgo G/C; esto se corrige como se describe. Este método no requiere el uso de un estándar externo, sino, más bien, proporciona una referencia interna, derivada de todas las etiquetas de secuencias (secuencias genómicas), que pueden ser, por ejemplo, un único cromosoma o un valor calculado a partir de todos los autosomas.

"T21" significa la trisomía 21.

"T18" significa la trisomía 18.

"T13" significa la trisomía 13.

[0033] "Aneuploidía" se utiliza en un sentido general para referirse a la presencia o ausencia de un cromosoma entero, así como la presencia de duplicaciones parciales cromosómicas o deleciones o kilobasas o mayor tamaño, en lugar de mutaciones genéticas o polimorfismos donde las diferencias de secuencia existe.

[0034] "Secuenciación masivamente paralela" significa técnicas para la secuenciación de millones de fragmentos de ácidos nucleicos, por ejemplo, mediante la unión de ADN genómico fragmentado al azar a un plano, la superficie ópticamente transparente y en amplificación de fase sólida para crear una célula de flujo de secuenciación de alta densidad con millones de grupos, cada uno con -1,000 copias de molde por centímetro cuadrado. Estas moldes se secuenciaron utilizando la tecnología del ADN de cuatro colores de secuenciación por síntesis. Véase, productos ofrecidos por Illumina, Inc., San Diego, California. En el presente trabajo, se obtuvieron secuencias, como se describe a continuación, con un Illumina/Solexa 1G Genome Analyzer. El método Solexa/Illumina enumerado a continuación se basa en el del implemento de origen genómico fragmentado al azar de ADN a una superficie plana, ópticamente transparente. En el presente caso, el ADN de plasma no tiene que cortarse. Los fragmentos de ADN unidos se extienden y se amplifican por puente para crear una célula de flujo de secuenciación de densidad ultra-alta con ε 50 millones de grupos, cada uno contiene -1,000 copias de la misma molde. Estas moldes se secuenciaron utilizando una tecnología robusta de secuenciación por síntesis de ADN de cuatro colores que emplea terminadores reversibles con tintes fluorescentes extraíbles. Este novedoso enfoque garantiza una alta precisión y la verdadera base de secuenciación de bases por caso, la eliminación de errores específicos de secuencia de contexto y permite la secuenciación a través de homopolímeros y secuencias repetitivas.

[0035] Detección de fluorescencia de alta sensibilidad se consigue utilizando excitación láser y la óptica de reflexión interna total. Lecturas de secuencia corta se han alineado contra un genoma de referencia y las diferencias genéticas se denominan utilizando el software de análisis de tuberías de datos desarrollado especialmente.

[0036] Las copias del protocolo para la secuenciación del genoma completo utilizando la tecnología Soelxa se pueden encontrar en Bio Techniques ® Protocol Guide 2007 publicado en diciembre de 2006: p 29, www.biotechniques.com/default.asp? page=protocol&subsection=article_display&id=112378. Adaptadores de oligonucleótidos de Solexa se ligaron a los fragmentos, produciendo una biblioteca genómica totalmente representativa de moldes de ADN sin clonación. La amplificación clonal de molécula única consiste en seis pasos: Hibridación de molde, amplificación de molde, linealización, bloqueo de los extremos 3', desnaturalización e hibridación del cebador. La secuenciación por síntesis de Solexa utiliza cuatro nucleótidos de propiedad que poseen fluoróforo reversible y propiedades de terminación. Cada ciclo de secuenciación se produce en presencia de los cuatro nucleótidos.

[0037] La secuenciación utilizada actualmente se lleva a cabo preferiblemente sin una preamplificación o etapa de clonación, pero se puede combinar con los métodos basados en la amplificación en un chip microfluídico que tiene cámaras de reacción tanto para la PCR y la secuenciación basada en una molde microscópica. Sólo se necesitan aproximadamente 30 pb de la información de la secuencia aleatoria para identificar una secuencia como perteneciente a un cromosoma humano específico. Las secuencias más largas pueden identificar de forma exclusiva objetivos más particulares. En el presente caso, las lecturas se obtuvo un gran número de 25 pb, y debido al gran número de lecturas obtenidas, la especificidad 50% permitió suficiente representación de etiqueta de secuencia.

[0038] Una descripción más detallada de un método de secuenciación masiva en paralelo, que emplea el método 454 referenciado a continuación se encuentra en Rogers y Venter, "Genomics: Massively parallel sequencing," Nature, 437, 326-327 (15 de septiembre de 2005). Como se describe allí, Rothberg y otros (Margulies, M. et al,

Nature 437, 376-380 (2005)), han desarrollado un sistema altamente paralelo capaz de secuenciación de 25 millones de bases en un período de cuatro horas - alrededor de 100 veces más rápido que la corriente de secuenciación del estado de la técnica de Sanger y la plataforma basada en electroforesis capilar. El método podría permitir a un individuo preparar y secuenciar un genoma entero en unos pocos días. La complejidad del sistema se encuentra principalmente en la preparación de la muestra y en la plataforma microfabricada, masivamente paralela, que contiene 1,6 milliones reactores de tamaño de picolitro en una diapositiva 6,4-cm^2. La preparación de la muestra comienza con la fragmentación del ADN genómico, seguido por la unión de secuencias adaptadoras a los extremos de las piezas de ADN. Los adaptadores permiten que los fragmentos de ADN se unan a pequeñas cuentas (alrededor de 28µ de diámetro). Esto se realiza en condiciones que permiten que sólo una pieza de ADN se una a cada perla. Las perlas están encerradas en gotitas de aceite que contienen todos los reactivos necesarios para amplificar el ADN usando una herramienta estándar llamada la reacción en cadena de la polimerasa. Las gotas de aceite forman parte de una emulsión de manera que cada perla se mantenga aparte de su vecino, lo que garantiza que la amplificación no esté contaminada. Cada cuenta termina con aproximadamente 10 millones de copias de su fragmento de ADN inicial. Para llevar a cabo la reacción de secuenciación, el ADN de la molde de transporte de perlas se cargan en los pocillos del reactor de pico litro - teniendo cada pocillo espacio para una sola perla. La técnica utiliza un método de secuenciación por síntesis desarrollado por Uhlen y colegas, en el que se sintetiza el ADN complementario a cada cadena de molde. Las bases de nucleótidos se utilizan para la secuenciación liberan un grupo químico a medida que la base forma un enlace con la creciente cadena de ADN, y este grupo impulsa una reacción de emisión de luz en presencia de las enzimas y luciferina específicas. Lavados secuenciales de cada uno de los cuatro nucleótidos posibles se ejecutan sobre la placa, y un detector detecta cuales de los pozos emiten luz con cada lavado para determinar la secuencia de la cadena creciente. Este método ha sido adoptado comercialmente por 454 Life Sciences.

[0039] Otros ejemplos de secuenciación masiva en paralelo se dan en US 20070224613 por Strathmann, publicada el 27 de septiembre de 2007, titulada "Massively Multiplexed Sequencing." Además, para una descripción más detallada de la secuenciación masiva en paralelo, véase el documento US 2003/0022207 a Balasubramanian, et al., publicada el 30 de enero de 2003, titulada " Arrayed polynucleotides and their use in genome analysis."

Descripción general del método y materiales

Visión de conjunto

[0040] Diagnóstico prenatal no invasivo de la aneuploidía ha sido un problema difícil porque el ADN fetal constituye un pequeño porcentaje del total de ADN en la sangre materna (13) y las células fetales intactas son aún más raras (6, 7, 9, 31, 32). Mostramos en este estudio el desarrollo exitoso de una prueba verdaderamente universal, independiente de polimorfismo no invasivo para la aneuploidía fetal. Al secuenciar directamente el ADN del plasma materno, se podría detectar la trisomía 21 fetal ya en la semana 14 de gestación. El uso de ADN libre de células en lugar de células intactas permite evitar complejidades asociadas con microquimerismo y células extrañas que podrían haber colonizado la madre; estas células se producen en cantidades tan bajas que su contribución al ADN libre de células es insignificante (33, 34). Además, hay pruebas de que el ADN fetal libre de células se borra de la sangre hasta niveles indetectables dentro de unas pocas horas después del parto y por lo tanto no se arrastrará de un embarazo al siguiente (35-37).

[0041] Formas raras de aneuploidía causadas por translocaciones desequilibradas y duplicación parcial de un cromosoma son en principio detectables por el método de secuenciación aleatoria, ya que la densidad de las etiquetas de secuencias en la región triplicada del cromosoma sería más alta que el resto del cromosoma. La detección de aneuploidía incompleta causada por mosaicismo también es posible en principio, pero puede ser más difícil, ya que no sólo depende de la concentración de ADN fetal en el plasma materno sino también el grado de mosaicismo fetal. Se requieren más estudios para determinar la eficacia de la secuenciación de escopeta en la detección de estas formas raras de aneuploidía.

[0042] El presente método es aplicable a grandes deleciones cromosómicas, como el síndrome 5p-(cinco p menos), también conocido como Síndrome del maullido o Síndrome de Cri du Chat. Síndrome 5p se caracteriza en el momento del nacimiento por un llanto agudo, bajo peso al nacer, falta de tonicidad muscular, microcefalia y posibles complicaciones médicas. Del mismo modo trastornos similarmente susceptibles descritos por los presentes métodos son p-, monosomía 9P, también conocido como síndrome de Alfi o 9P-, síndrome de deleción 22q11.2, Síndrome de Emanuel, también conocido en la literatura médica como el Síndrome Supernumerario Der (22), trisomía 22, translocación desequilibrada 11/22 o trisomía parcial 11/22, microdeleción y microduplicación en 16p11.2, que está asociado con el autismo, y otras supresiones o desequilibrios, incluyendo aquellos que actualmente se desconocen.

[0043] Una ventaja de utilizar la secuenciación directa para medir la aneuploidía no invasiva es que es capaz de hacer un uso completo de la muestra, mientras que los métodos basados en la PCR analizan sólo unas pocas secuencias específicas. En este estudio, se obtuvo en promedio 5.000.000 lecturas por muestra en un solo paso, de los cuales -66.000 se mapearon en el cromosoma 21. Al representar las 5.000.000 lecturas sólo una parte de un genoma humano, en principio, menos de un equivalente genómico de ADN es suficiente para la detección de aneuploidía mediante secuenciación directa. En la práctica, se utilizó una mayor cantidad de ADN ya que no hay

pérdida de la muestra durante la preparación de la biblioteca de secuenciación, pero puede ser posible reducir más la cantidad de sangre necesaria para el análisis.

[0044] Información de secuencia de escopeta de mapeo (es decir, información de secuencia de un fragmento cuya posición genómica física no se desconoce) puede hacerse en un número de formas, que implican la alineación de la secuencia obtenida con una secuencia correspondiente en un genoma de referencia. Véase, Li et al., "Mapping short DNA sequencing reads and calling variants using mapping quality score," Genome Res., 2008 agosto 19. [Epub en avance de la edición impresa].

[0045] Hemos observado que ciertos cromosomas tienen grandes variaciones en los recuentos de fragmentos secuenciados (a partir de una muestra a otra, y que esto depende en gran medida del contenido de G/C **(Figura 1A).** No está claro en este momento si esto se deriva de artefactos PCR durante la preparación de la biblioteca de secuenciación o la generación de clúster, el proceso de secuenciación, o si se trata de un verdadero efecto biológico relativo a la estructura de cromatina. Sospechamos que es un artefacto ya que también observamos sesgo G/C en el control de ADN genómico, y tales sesgos en la plataforma de secuenciación Solexa recientemente se ha informado (38, 39). Tiene una consecuencia práctica ya que la sensibilidad a la detección de aneuploidía variará de cromosoma en el cromosoma; afortunadamente las aneuploidías humanas más comunes (por ejemplo, 13, 18, $y_2$1) tienen una baja variación y por lo tanto sensibilidad de detección alta. Tanto este problema como las limitaciones de volumen de muestra, posiblemente, pueden resolverse mediante el uso de tecnologías de secuenciación de molécula individuales, que no requieren el uso de PCR para la preparación de la biblioteca (40).

[0046] Las muestras de ADN de plasma usadas en este estudio se obtuvieron de 15 a 30 minutos después de la amniocentesis o de vellosidades coriónicas. Debido a que estos procedimientos invasivos interrumpen la interfaz entre la placenta y la circulación materna, se ha discutido si la cantidad de ADN fetal en la sangre materna podría incrementarse después de procedimientos invasivos. Ninguno de los estudios realizados hasta la fecha han observado un efecto significativo (41, 42).

[0047] Nuestros resultados apoyan esta conclusión, ya que usando el ensayo de PCR digitales se estimó que el ADN fetal constituye menos de o igual a 10% del ADN total libre de células en la mayoría de nuestras muestras de plasma materno. Esto está dentro del rango de valores previamente reportados en muestras de plasma materno obtenidas antes de procedimientos invasivos (13). Sería valioso contar con una medición directa que se refiriese a este punto en un estudio futuro.

[0048] La fracción de ADN fetal promedio estimada a partir de los datos de secuenciación es superior a los valores estimados de los datos digitales de PCR por un factor promedio de dos ($p < 0{,}005$, prueba t en par en todos los embarazos masculinos que tienen un conjunto completo de datos). Una posible explicación de esto es que la etapa de PCR durante la preparación de la biblioteca Solexa amplifica preferentemente fragmentos más cortos, que otros han encontrado estar enriquecido para el ADN fetal (22, 23). Nuestras propias mediciones de distribución de la longitud de una muestra no son compatibles con esta explicación, pero tampoco podemos rechazarla por el momento. También debe señalarse que mediante el uso de las etiquetas de secuencias encontramos alguna variación de la fracción fetal incluso en la misma muestra, dependiendo en qué cromosoma utilizamos para hacer el cálculo **(Figura 7,** Tabla 1). Esto es más probable debido a los artefactos y errores en los procesos de secuenciación y cartografía, que son sustanciales - recordar que sólo la mitad de las etiquetas de secuencia se asignan al genoma humano con un error o menos. Por último, también es posible que las mediciones de la PCR están sesgadas ya que sólo son el muestreo de una pequeña fracción del genoma fetal.

[0049] Nuestros datos de secuenciación sugieren que la mayoría del ADN de plasma libre de células es de características de origen y comparte apoptóticos de ADN nucleosomal. Al no ser ocupación de nucleosomas en todo el genoma eucariótico necesariamente uniforme y depende de factores tales como la función, expresión, o secuencia de la región (30, 43), la representación de las secuencias de diferentes loci en el plasma materno libre de células puede no ser igual, como se espera normalmente en el ADN genómico extraído de células intactas. Por lo tanto, la cantidad de un locus particular puede no ser representativa de la cantidad de todo el cromosoma y se debe tener cuidado cuando se diseña ensayos para medir la dosis de genes en el ADN del plasma materno libre de células que se dirigen sólo a unos pocos loci.

[0050] Históricamente, debido a los riesgos asociados con el muestreo de vellosidades coriónicas y la amniocentesis, el diagnóstico invasivo de aneuploidía fetal se ofrece principalmente a las mujeres en riesgo de llevar un feto aneuploides basado en la evaluación de los factores de riesgo como la edad materna, los niveles de marcadores de suero, y los hallazgos ecográficos. Recientemente, un boletín de práctica del American College of Obstetricians and Gynecologists (ACOG) recomienda que "las pruebas de diagnóstico invasivo de aneuploidías debe estar disponible para todas las mujeres, independientemente de la edad de la madre" y que "asesoría previa debe incluir un análisis de los riesgos y beneficios de pruebas invasivas en comparación con las pruebas de detección" (2).

[0051] Un test genético no invasivo basado en los resultados descritos aquí y en futuros estudios a gran escala presumiblemente llevaría lo mejor de ambos mundos: un mínimo riesgo para el feto al tiempo que proporciona la

información genética. Los costos del ensayo ya son bastante bajos; el coste de secuenciación por muestra de este escrito es de aproximadamente $ 700 y se espera que el coste de la secuenciación se continúe disminuyendo de forma espectacular en un futuro próximo.

[0052] La secuenciación de escopeta potencialmente puede revelar muchas más características hasta ahora desconocidas de ácidos nucleicos libres de células, tales como las distribuciones de ARNm de plasma, así como características epigenéticas de ADN en plasma como la metilación del ADN y la modificación de histonas, en campos, incluyendo perinatología, oncología y trasplantes, mejorando así nuestra comprensión de la biología básica de la gestación, el desarrollo humano temprano y la enfermedad.

Métodos de secuenciación

[0053] El equipo de secuenciación disponible comercialmente se utilizó en los presentes ejemplos ilustrativos, a saber, la plataforma de secuenciación de Solexa/Illumina y la plataforma 454/Roche. Será evidente para los expertos en la técnica que un número de diferentes métodos de secuenciación y variaciones se pueden utilizar. Un método de secuenciación que se puede utilizar con ventaja en los presentes métodos implica el emparejado de secuenciación final. Cebadores de secuenciación marcados con fluorescencia se podrían utilizar para secuenciar simultáneamente ambas cadenas de un molde de ADN de doble cadena, como se describe por ejemplo, en Wiemann et al. (Anal. Biochem. 224:117 [1995]; Anal Biochem 234: 166 [1996]. Ejemplos recientes de esta técnica han demostrado co-secuenciación de multiplex usando la química de la reacción de terminación de colorantes de cuatro colores por primera vez por Prober et al. (Science 238: 336 [1987]). Solexa/Illumina ofrece un "módulo de extremos emparejados" a su Analizador de Genoma. Mediante el uso de este módulo, después de que el Analizador de Genoma haya completado la primera lectura de secuenciación, el módulo de extremos emparejados dirige la resíntesis de los moldes originales y la segunda ronda de generación de clúster. El módulo de extremo emparejado está conectado al analizador del genoma a través de una sola conexión de fluidos. Además, 454 ha desarrollado un protocolo para generar una biblioteca de lecturas de extremos emparejados. Estas lecturas de extremos emparejados son aproximadamente 84 nucleótidos de fragmentos de ADN que tienen una secuencia de adaptador de 44-mer en el medio flanqueado por una secuencia de 20-mer en cada lado. Los dos 20-meros que flanquean son segmentos de ADN que originalmente se encontraban aproximadamente 2,5 kb aparte en el genoma de interés.

[0054] Mediante el uso de lecturas de extremos emparejados en el presente método, se puede obtener más información de la secuencia de un fragmento de ADN de plasma dado, y, de manera significativa, también se puede obtener información de la secuencia de ambos extremos del fragmento. El fragmento se correlaciona con el genoma humano como se explica aquí en otro lugar. Después de la cartografía de los dos extremos, se puede deducir la longitud del fragmento de partida. Dado que el ADN fetal es conocido por ser más corto que los fragmentos de ADN maternos circulantes en el plasma, se puede utilizar esta información sobre la longitud del fragmento de ADN para aumentar eficazmente el peso dado a las secuencias obtenidas a partir de fragmentos de ADN más cortos (por ejemplo, aproximadamente 300 pb o menos). Los métodos para la ponderación se dan a continuación.

[0055] Otro método para aumentar la sensibilidad al ADN fetal es centrarse en ciertas regiones en el genoma humano. Se puede utilizar los métodos de secuenciación que seleccionan secuencias a priori, que se asignan a los cromosomas de interés (como se describe aquí en otros lugares, tales como 18, 21, 13, X e Y). También se puede optar por centrarse, utilizando este método, en deleciones cromosómicas parciales, como el síndrome de deleción 22q11. Otras microdeleciones y microduplicaciones se exponen en la Tabla 1 del documento US 2005/0181410, publicada el 18 de agosto de 2005 bajo el título "Methods and apparatuses for achieving precision genetic diagnosis."

[0056] En subsecuencias de secuencia seleccionada, se puede emplear metodologías basadas en secuencias como la secuenciación de matriz o perlas de captura con secuencias genómicas específicas utilizadas como sondas de captura. El uso de una matriz de secuenciación puede implementarse como se describe en Chetverin et al, "Oligonucleotide arrays: new concepts and possibilities". Biotechnology (NY). 1994 Nov; 12 (11): 1093-9, así como Rothberg, US 2002/0012930 A1 titulada "Method of Sequencing a Nucleic Acid," y Reeve et al., "Sequencing by Hybridization," US 6.399.364. En estos métodos, el ácido nucleico diana a secuenciarse puede ser ADN genómico, ADNc o ARN. La muestra se procesa de cadena sencilla y se captura en condiciones de hibridación con una serie de sondas de cadena simple, que se catalogan por códigos de barras o por separación física de una matriz. La emulsión PCR, tal como se utiliza en el sistema 454, el sistema de sólidos, y Polonator (Dover Systems) y otros también se pueden emplear, donde la captura se dirige a secuencias diana específicas, por ejemplo, secuencias de genoma que se trazan de forma única en el cromosoma 21 o de otro cromosoma de interés, o a una región del cromosoma como 15q11 (Prader-Willi), o repeticiones excesivas CGG en el gen FMR1 (síndrome del cromosoma X frágil).

[0057] El método de subsecuenciación es en un aspecto contrario a las metodologías convencionales de secuenciación masivamente paralelas, que tratan de obtener toda la información de la secuencia en una muestra. Este método alternativo ignora selectivamente cierta información de la secuencia utilizando un método de secuenciación que captura selectivamente moléculas de la muestra que contienen ciertas secuencias predefinidas. También se pueden usar los pasos de secuenciación tal y como se ejemplifican, pero en la cartografía de los

fragmentos de secuencia obtenidos, dar mayor peso a las secuencias que se asignan a las áreas conocidas por ser más fiables en su cobertura, como los exones. De lo contrario, el método procede, como se describe a continuación, donde se obtiene un gran número de lecturas de secuencia a partir de uno o más cromosomas de referencia, que se comparan a un gran número de lecturas obtenido a partir de un cromosoma de interés, después de considerar las variaciones derivadas de la longitud cromosómica, contenido de G/C, secuencias repetidas y similares.

[0058] También se pueden concentrar en ciertas regiones en el genoma humano, de acuerdo con los presentes procedimientos con el fin de identificar monosomías parciales y trisomías parciales. Como se describe a continuación, los presentes métodos implican el análisis de datos de secuencias en una "ventana" deslizante cromosómica definida, tales como las regiones contiguas 50kb, que no se superponen repartidas en un cromosoma. Trisomías parciales de 13q, 8p (8p23.1), 7q, 6p distal, 5p, 3q (3q25.1), 2q, 1q (1q42.1 y 1q21-qter), monosomía parcial Xpand 4q35.1 se han reportado, entre otros. Por ejemplo, las duplicaciones parciales del brazo largo del cromosoma 18 puede resultar en el síndrome de Edwards en el caso de una duplicación de 18q21.1-qter (Véase, Mewar et al., "Clinical and molecular evaluation of four patients with partial duplications of the long arm of chromosome 18," Am J Hum Genet 1993 Dec; 53 (6):. 1269-1278).

Secuenciación de escopeta de ADN de plasma libre de células

[0059] ADN de plasma libre de células de 18 mujeres embarazadas y un donante masculino, así como ADN genómico de sangre total del mismo donante masculino, se secuenciaron en la plataforma Solexa/Illumina. Se obtuvo un promedio de ~ 10 millones de etiquetas de secuencia de 25 pb por muestra. Alrededor del 50% (es decir, -5.000.000) de las lecturas mapeadas de modo único para el genoma humano con un máximo de 1 desajuste contra el genoma humano, que cubre ~ 4% del genoma entero. Un promedio de ~ 154,000, ~ 135.000, -66,000 etiquetas de secuencia asignadas a los cromosomas 13, 18, $y_2$1, respectivamente. El número de etiquetas de secuencia para cada muestra se detalla en la siguiente Tabla 1 y la Tabla 2.

Tabla 1.

Muestra	Cariotipo fetal	Edad gestacional (semanas)	Volumen de Plasma	Cantidad de ADN	Aprox cantidad de ADN de entrada *	Número total de etiquetas de secuencia
P1 plasma ADN §	47xx +21	35	1,6	761	8,0	8206694
P2 plasma ADN §	47XY 21	18	1,4	585	5,2	7751384
P6 plasma ADN §	47xx +21	14	1,6	410	4,3	6699183
P7 plasma ADN §	47XY 21	18	2,2	266	3,8	8324473
P14 plasma ADN §	47xx +21	23	3,2	57	1,2	8924944
P17 plasma ADN §	47xx +21	16	2,3	210	3,2	11599833
P19 plasma ADN §	46XY	18	3,2	333	7,0	7305417
P20 plasma ADN §	47XY 21	18	1,3	408	3,6	11454876
P23 plasma ADN §	46XY	10	1,6	258	2,7	11851612
P26 plasma ADN §	46XY	13	3,0	340	6,7	11471297
P31 plasma ADN §	46XY	20	2,2	278	4,0	8967562
P40 plasma ADN §	46XY	11	2,6	217	3,7	9205197

(continuación)

Muestra	Cariotipo fetal	Edad gestacional (semanas)	Volumen de Plasma	Cantidad de ADN	Aprox cantidad de ADN de entrada *	Número total de etiquetas de secuencia
P42 plasma ADN §	46XY	11	3,0	276	5,5	8364774
P52 plasma ADN §	47XY 21	25	1,6	645	6,8	9192596
P53 plasma ADN §	47xx +21	19	1,6	539	5,7	9771887
P57 plasma ADN §	47xx 18	23	2,0	199	2,6	15041417
P59 plasma ADN §	47XY +	21	2,0	426	5,6	11910483
P64 plasma ADN §	47XY +	17	1,8	204	2,4	12097478
Hombre de Donantes de ADN de plasma §	-	-	1,8	485	5,8	6669125
Donante masculino Sangre Total ADN genómico §	-	-	-	-	2,1	8519495
P25 plasma ADN ¶	46XY	11	5,6	132	4,9	242,599
P13 plasma ADN §	46XY	18	5,6	77	2,9	4168455

Tabla 2.

Muestra	Número de etiquetas de secuencia asignada única para el Genoma Humano (hg18) con un máximo de 1 falta de coincidencia	% de ADN fetal estimado por PCR digital con Ensayo SRY (fetos masculinos)	% de ADN fetal estimado por etiquetas de secuencia ChrY (fetos masculinos)	% de ADN fetal estimado por el agotamiento de CHRX etiquetas de secuencia (fetos masculinos)	% de ADN fetal estimado mediante la adición de etiquetas de secuencia trisómicas de cromosoma (fetos aneuploides)	Contenido de G/C general de etiquetas de secuencia (%)
P1 plasma ADN §	4632637	-	-	-	35,0	43,65
P2 plasma ADN §	4313884	6,4	15,4	21,6	15,5	48,72
P6 plasma ADN §	3878383	-	-	-	22, 9	44,78
P7 plasma ADN §	4294865	9,1	31,0	33,8	28,6	48,07
P14 plasma ADN §	3603767	-	-	-	30,5	46,38
P17 plasma ADN §	5968932	-	-	-	7,8	44,29
P19 plasma ADN §	3280521	<5,9 ‡	4,14	21,5	-	50,09
P20 plasma ADN §	6032684	10,0	15,7	11,3	11,5	44,02

(continuación)

Muestra	Número de etiquetas de secuencia asignada única para el Genoma Humano (hg18) con un máximo de 1 falta de coincidencia	% de ADN fetal estimado por PCR digital con Ensayo SRY (fetos masculinos)	% de ADN fetal estimado por etiquetas de secuencia ChrY (fetos masculinos)	% de ADN fetal estimado por el agotamiento de CHRX etiquetas de secuencia (fetos masculinos)	% de ADN fetal estimado mediante la adición de etiquetas de secuencia trisómicas de cromosoma (fetos aneuploides)	Contenido de G/C general de etiquetas de secuencia (%)
P23 plasma ADN §	6642795	5,3	12,2	9,6	-	43,80
P26 plasma ADN §	3851477	10,3	18,2	14,2	-	42,51
P31 plasma ADN §	4683777	Datos que faltan ‡	13,2	17,0	-	48,27
P40 plasma ADN §	4187561	8,6	20,0	17,1	-	42,65
P42 plasma ADN §	4315527	<4,4 ‡	9,7	7,9	-	44,14
P52 plasma ADN §	5126837	6,3	25,0	26,3	26,4	44,34
P53 plasma ADN §	5434222	-	-	-	25,8	44,18
P57 plasma ADN §	7470487	-	-	-	23,0	42,89
P59 plasma ADN §	6684871	26,4	44,0	39,8	45,1	43,64
P64 plasma ADN §	6701148	<4,4 ‡	14,0	8,9	16,7	44,21
Donante masculino de ADN de plasma §	3692931	-	-	-	-	48,30
Donante masculino Sangre Total ADN genómico §	5085412	-	-	-	-	46,53

(continuación)

Muestra	Número de etiquetas de secuencia asignada única para el Genoma Humano (hg18) con un máximo de 1 falta de coincidencia	% de ADN fetal estimado por PCR digital con Ensayo SRY (fetos masculinos)	% de ADN fetal estimado por etiquetas de secuencia ChrY (fetos masculinos)	% de ADN fetal estimado por el agotamiento de CHRX etiquetas de secuencia (fetos masculinos)	% de ADN fetal estimado mediante la adición de etiquetas de secuencia trisómicas de cromosoma (fetos aneuploides)	Contenido de G/C general de etiquetas de secuencia (%)
P25 plasma ADN ¶	144992‡	-	-	-	-	41,38
P13 plasma ADN §	2835333	9,8	5,7	n/a[i]	-	39,60
El volumen de plasma es el volumen utilizado para la Creación de Biblioteca de Secuenciación (ml). La cantidad de ADN está en el plasma (célula equivalente/ml de plasma)*. La cantidad aproximada de ADN de entrada es que el uso para la construcción de biblioteca de secuenciación (ng). * Como se cuantificó por PCR digital con ensayo TaqMan EIF2C1, convirtiendo de copias a ng asumiendo 6,6pg/equivalente celular. † Para la secuencia 454, este número representa el número de lecturas con la cobertura de al menos 90% de precisión y 90% cuando se asigna a hg18. ‡ Materiales insuficientes estaban disponibles para la cuantificación de % de ADN fetal con PCR digital para estas muestras (o bien ninguna muestra se mantenía para el análisis o bien hubo insuficiencia de muestreo). § Secuenciado en la plataforma de Solexa/Illumina; ¶ Secuenciado de 454/plataforma Roche II Muestra P13 era la primera en analizarse por secuenciación de escopeta. Era un feto normal y el valor de cromosoma era claramente disómico. Sin embargo, hubo algunas irregularidades con esta muestra y no se incluyeron en el análisis adicional. Esta muestra se secuenció en un instrumento Solexa diferente que el resto de las muestras de este estudio, y se secuenció en la presencia de un número de muestras de origen desconocido. El contenido de G/C de esta muestra era menor que el sesgo G/C del genoma humano, mientras que el resto de las muestras están por encima. Tenía el menor número de lecturas, y también el menor número de lecturas asignadas con éxito para el genoma humano. Esta muestra parecía ser atípica en la densidad de etiqueta de secuencia para la mayoría de los cromosomas y la fracción de ADN fetal calculado a partir de los cromosomas X no estaba bien definida. Por estas razones, sospechamos que las irregularidades se deben a problemas técnicos con el proceso de secuenciación.						

[0060] En la Tabla 1 y la Tabla 2, cada muestra representa un paciente diferente, por ejemplo, P1 en la primera fila. El número total de etiquetas de secuencias variadas, pero con frecuencia se estaba en el intervalo de 10 millones, utilizando la tecnología Solexa. La tecnología 454 utilizada para P25 y P13 dio un número menor de lecturas.

[0061] Se observó una distribución no uniforme de las etiquetas de secuencias a través de cada cromosoma. Este patrón de variación intracromosómica era común entre todas las muestras, incluyendo el ADN genómico cortado al azar, lo que indica que la variación observada más probablemente se debía a los artefactos de secuenciación. Aplicamos una ventana deslizante arbitraria de 50kb a través de cada cromosoma y contó el número de etiquetas que caen dentro de cada ventana. La ventana se puede variar en tamaño para dar cuenta de un mayor número de lecturas (en cuyos casos una ventana más pequeña, por ejemplo, 10 kb, da una imagen más detallada de un cromosoma) o un número más pequeño de lecturas, en cuyo caso una ventana más grande (por ejemplo, 100kb) todavía puede ser utilizada y detectará supresiones, omisiones o duplicaciones cromosómicas. Se seleccionó la concentración mediana de 50 kb por ventana para cada cromosoma. La mediana de los valores autosómicos (es decir, 22 cromosomas) se utilizó como una constante para dar cuenta de las diferencias en el número total de etiquetas de secuencias obtenidas para diferentes muestras de normalización. La variación inter-cromosómica dentro de cada muestra también era consistente entre todas las muestras (incluyendo el control de ADN genómico). La densidad de etiqueta de secuencia media de cada cromosoma se correlaciona con el contenido de G/C del cromosoma ($p < 10^{-9}$) **(Figura 5A, 5B).** La desviación estándar de la densidad de etiqueta de secuencia para cada cromosoma también se correlaciona con el grado absoluto de la desviación de contenido de G/C cromosómico del contenido de G/C en todo el genoma ($p < 10^{-12}$) **(Figura 5A, 5C).** El contenido de G/C de etiquetas secuenciadas de todas las muestras (incluyendo el control de ADN genómico) era de un promedio de 10% más alto que el valor de la secuencia de genoma humano (41%) (21) (Tabla 2), lo que sugiere que hay un fuerte sesgo G/C derivado del

proceso de secuenciación. Hemos trazado en **la Figura 1A** la densidad de etiqueta de secuencia para cada cromosoma (ordenados por el aumento de contenido G/C) con respecto al valor correspondiente del control de ADN genómico para eliminar este sesgo.

Detección de aneuploidía fetal

[0062] La distribución de la densidad de etiqueta de secuencia de cromosoma 21 para los 9 embarazos T21 está claramente separada de la de los embarazos que llevan disomía 21 fetos ($p < 10^{-5}$), prueba t de Student) **(Figura 1A** y **1B).** La cobertura del cromosoma 21 para los casos T21 es aproximadamente ~ 4-18% más alto (promedio ~ 11%) que la disomía de los 21 casos. Debido a que la densidad de etiqueta de secuencia del cromosoma 21 para los casos T21 debe ser $(1+\sum/2)$ de la de disomía de 21 embarazos, donde $\sum$ es la fracción del ADN total en plasma procedente del feto, como aumento de la cobertura en el cromosoma 21 en casos T21 corresponde a una fracción de ADN fetal de ~ 8% - 35% (promedio ~ 23%) (Tabla 1, **Figura 2).** Hemos construido un intervalo de confianza de 99% de la distribución de la densidad de etiqueta de secuencia de cromosoma 21 de disomía de embarazos 21. Los valores de los 9 casos T21 se encuentran fuera del límite superior del intervalo de confianza y los 9 casos de disomía 21 se encuentran por debajo del límite **(Figura 1B).** Si se utilizó el límite superior del intervalo de confianza como un valor umbral para la detección de T21, la fracción mínima de ADN fetal que se detecta es ~ 2%.

[0063] El ADN de plasma de mujeres embarazadas con fetos T18 (2 casos) y un feto T13 (1 caso) también se secuenciaron directamente. El exceso de representación se observó para el cromosoma 18 y el cromosoma 13 en casos T18 y T13 respectivamente **(Figura 1A).** Mientras que no había suficientes muestras positivas para medir una distribución representativa, es alentador que todos estos tres aspectos positivos son los valores extremos de la distribución de los valores de disomía. Los T18 son grandes valores atípicos y son claramente estadísticamente significativos ($p < 10^{-7}$), mientras que la significación estadística de la caja T13 solo es marginal ($p < 0,05$). La fracción de ADN fetal también se calculó a partir del cromosoma sobrerrepresentado como se describe anteriormente **(Figura 2,** Tabla 1).

La fracción de ADN fetal en plasma materno

[0064] Mediante el uso digital de Taqman PCR para un único locus en el cromosoma 1, se estimó la concentración de ADN libre de células promedio en las muestras de plasma materno secuenciadas ser -equivalente de 360 células/ml de plasma (rango: equivalente de 57 a 761 células/ml de plasma) (Tabla 1), de acuerdo a los valores en bruto se informó anteriormente (13). La cohorte incluyó 12 embarazos masculinos (6 casos normales, 4 casos T21, 1 caso T18 y 1 caso T13) y 6 embarazos femeninos (5 casos T21 y 1 caso T18). DYS14, un locus de múltiples copias en el cromosoma Y, era detectable en el plasma materno por PCR en tiempo real en todos estos embarazos pero no en cualquiera de los embarazos femininos (datos no mostrados). La fracción de ADN fetal en el ADN de plasma sin células maternas generalmente se determina mediante la comparación de la cantidad de locus específico fetal (como el locus SRY en el cromosoma Y en los embarazos masculinos) a la de un locus en cualquier autosome que es común tanto a la la madre como al feto utilizando PCR cuantitativa en tiempo real (13, 22, 23). Se aplicó un ensayo de duplex similar sobre una plataforma PCR digital (véase métodos) para comparar los recuentos del locus SRY y un locus en el cromosoma 1 en embarazos masculinos. SRY locus no era detectable en ninguna de las muestras de ADN de plasma de embarazos femeninos. Encontramos con la PCR digital que para las muestras mayoritarias, el ADN fetal constituyó § 10% del ADN total en el plasma materno (Tabla 2), estando de acuerdo con los valores previamente reportados (13).

[0065] El porcentaje de ADN fetal entre el ADN libre de células totales en el plasma materno también puede calcularse a partir de la densidad de las etiquetas de secuencias de los cromosomas sexuales masculinos para embarazos. Mediante la comparación de la densidad de etiqueta de secuencia del cromosoma Y de ADN de plasma de embarazos masculinos a la del ADN en plasma adulto de sexo masculino, que calcula el porcentaje de ADN fetal a ser en promedio ~ 19% (rango: 4-44%) para todos los embarazos masculinos (Tabla 2, arriba, **Figura 2).** Debido a que los varones humanos tienen 1 cromosoma X menos que las hembras humanas, la densidad de etiqueta de secuencia del cromosoma X en los embarazos masculinos debe ser (1-e/2) de la de los embarazos femeninos, donde E es la fracción de ADN fetal. Observamos de hecho una subrepresentación de cromosoma X en embarazos masculinos en comparación con la de los embarazos femininos **(Figura 5).** Basándose en los datos del cromosoma X, se estimó que el porcentaje de ADN fetal era en promedio ~ 19% (rango: 8-40%) para todos los embarazos masculinos (Tabla 2, arriba, **Figura 2).** El porcentaje de ADN fetal se estimó a partir de los cromosomas X e Y para cada muestra de embarazo masculino correlacionada entre sí ($p = 0,0015$) **(Figura 7).**

[0066] Representamos en la **Figura 2** la fracción de ADN fetal calculada a partir de la sobrerrepresentación del cromosoma trisómico en embarazos aneuploides, y la falta de representación del cromosoma X y la presencia del cromosoma Y para los embarazos masculinos contra la edad gestacional. La fracción de ADN fetal promedio para cada muestra se correlaciona con la edad gestacional ($p = 0,0051$), una tendencia que también se informó anteriormente (13).

Distribución del Tamaño de ADN de plasma libre de células

[0067] Se analizaron las bibliotecas de secuenciación con un sistema de electroforesis capilar comercial lab-on-a-chip. Hay una consistencia notable en el tamaño del fragmento de pico, así como la distribución alrededor del pico, para todas las muestras de ADN de plasma, incluidos los de las mujeres embarazadas y donante masculino. El tamaño de fragmento máximo de 261bp estaba en la media (rango: 256-264bp). Restar la longitud total de los adaptadores Solexa (92bp) de 260bp da 169bp como el tamaño de fragmento máximo real. Este tamaño corresponde a la longitud del ADN envuelto en un cromatosoma, que es un nucleosoma unido a una histona H1 (24). Debido a que la preparación de la biblioteca incluye una PCR de 18 ciclos, existe la preocupación de que la distribución puede no ser imparcial. Para verificar que la distribución de tamaño observada en el electroferograma no es un artefacto de la PCR, también secuenciado el ADN de plasma libre de células de una mujer embarazada que lleva un feto masculino utilizando la plataforma 454. La preparación de la muestra para este sistema utiliza emulsión PCR, que no requiere la amplificación competitiva de las bibliotecas de secuenciación y crea producto que es en gran medida independiente de la eficiencia de amplificación. La distribución de tamaño de las lecturas asignadas a lugares únicos del genoma humano se asemejan a los de las bibliotecas de secuenciación Solexa, con un pico predominante en 176bp, después de restar la longitud de 454 adaptadores universales **(Figura 3** y **Figura 8).** Estos hallazgos sugieren que la mayoría de ADN libre de células en el plasma se deriva de las células apoptóticas, de acuerdo con los resultados anteriores (22, 23, 25, 26).

[0068] De particular interés es la distribución de tamaños de ADN materno y fetal en el plasma libre de células madre. Dos grupos han demostrado previamente que la mayoría del ADN fetal tiene rango de tamaño del de mono-nucleosoma (<200-300bp), mientras que el ADN materno es más largo. Debido a que la secuencia 454 tiene una longitud de lectura dirigida de 250 pb, se interpretó el pequeño pico en alrededor de 250 pb **(Figura 3** y **Figura 8)** como el límite de la instrumentación de secuenciación de fragmentos de mayor peso molecular. Hemos trazado la distribución de todas las lecturas y aquellas mapeadas al cromosoma Y **(Figura 3).** Se observó una ligera disminución de lecturas de cromosoma Y en el extremo superior de la distribución. Las lecturas <220bp constituyen 94% del cromosoma Y y 87% de las lecturas totales. Nuestros resultados no están totalmente de acuerdo con los resultados anteriores en que no vemos un enriquecimiento tan dramático de ADN fetal en longitudes cortas (22, 23). Se necesitan más estudios para resolver este punto y para eliminar cualquier sesgo potencial residual en el proceso de preparación 454 de la muestra, pero merece la pena señalar que la capacidad de secuenciar muestras de plasma individuales permite la medición de la distribución de enriquecimientos de longitud a través de muchos pacientes individuales en lugar de la medición de la longitud media de enriquecimiento de muestras de pacientes agrupados.

ADN de plasma libre de células comparte características de ADN Nucleosomal

[0069] Al sugerir nuestras observaciones de la distribución de tamaños de ADN en plasma libre de células que el ADN de plasma es principalmente de origen apoptótico, se investigó si las características del ADN nucleosomal y posicionamiento se encuentran en el ADN de plasma. Una de estas características es el posicionamiento de nucleosomas en torno a los sitios de inicio de transcripción. Los datos experimentales de la levadura y humano han sugerido que los nucleosomas se agotan en los promotores aguas arriba de los sitios de inicio de transcripción y nucleosomas están bien posicionados cerca de los sitios de inicio de transcripción (27-30). Se aplicó una ventana de 5 pb que abarca +/- 1000 pb de sitios de inicio de transcripción de todos los genes RefSeq y se contó el número de etiquetas de asignación a las cadenas con sentido y antisentido dentro de cada ventana. Un pico en la cadena con sentido representa el comienzo de un nucleosoma, mientras que un pico en la cadena antisentido representa el final. Después de alisarse, vimos que para la mayoría de las muestras de ADN de plasma, al menos 3 nucleosomas muy bien ubicados pudieron detectarse aguas abajo de los sitios de inicio de transcripción, y en algunos casos, hasta 5 nucleosomas bien posicionados podrían detectarse, de conformidad aproximada a los resultados de Schones et al. (27) **(Figura 4).** Se aplicó el mismo análisis en las etiquetas de secuencias de ADN genómico cortado al azar y no observamos ningún patrón obvio en la localización de la etiqueta, aunque la densidad de etiquetas era mayor en el sitio de inicio de transcripción **(Figura 4).**

Corrección de sesgo de secuenciación

[0070] En las **Figuras 10** y **12** se muestran los resultados que se pueden obtener cuando los números de la etiqueta de secuencia se tratan estadísticamente con base en datos de la referencia del genoma humano. Es decir, por ejemplo, las etiquetas de secuencia de fragmentos con mayor contenido de GC pueden ser sobre-representadas, y sugieren una aneuploidía donde no existe. La información de etiqueta de secuencia puede no ser informativa, ya que sólo una pequeña parte del fragmento ordinariamente se secuenciará, mientras que es el contenido total de G/C del fragmento que causa el sesgo. Por lo tanto, se proporciona un método, descrito en detalle en los Ejemplos 8 y 10, para corregir este sesgo, y este método puede facilitar el análisis de muestras que de otra manera no producirían resultados estadísticamente significativos. Este método, para la corrección de sesgo G/C de lecturas de secuencia de secuenciación masiva en paralelo de un genoma, comprende la etapa de dividir el genoma en una serie de ventanas dentro de cada cromosoma y calcular el contenido de G/C de cada ventana. Estas ventanas no tienen que ser las mismas que las ventanas que se utilizan para el cálculo de la densidad de etiqueta de secuencia; que pueden ser del orden de 10 kb-30kb de longitud, por ejemplo. Entonces se calcula la relación entre la secuencia de la cobertura y el contenido de G/C de cada ventana mediante la determinación de un número de lecturas por una ventana dada y un contenido de G/C de la ventana. El contenido de G/C de cada ventana se conoce de la secuencia

de referencia del genoma humano. Ciertas ventanas serán ignoradas, es decir, sin lecturas o sin contenido de G/C. Entonces se asigna un peso a la serie de lecturas por una ventana dada (es decir, el número de etiquetas de secuencia asignadas a esa ventana) en base al contenido de G/C, donde el peso tiene una relación con el contenido de G/C de tal manera que los números crecientes de lecturas con contenido G/C creciente se traduce en la disminución de peso por el aumento del contenido de G/C.

EJEMPLOS

[0071] Los ejemplos siguientes describen la secuenciación directa de ADN libre de células a partir de plasma de mujeres embarazadas con la tecnología de secuenciación de escopeta de alto rendimiento, obteniendo un promedio de 5 millones de etiquetas de secuencia por muestra del paciente. Las secuencias obtenidas se asignan a ubicaciones cromosómicas específicas. Esto nos permitió medir la sobrerepresentación y de cromosomas de un feto aneuploide. El enfoque de secuenciación es independiente de polimorfismo y por lo tanto de aplicación universal para la detección no invasiva de aneuploidía fetal. Usando este método se identificaron con éxito los 9 casos de trisomía 21 (síndrome de Down), 2 casos de trisomía 18 y 1 caso de trisomía 13 en una cohorte de 18 embarazos normales y aneuploides; trisomía se detectó en edades gestacionales tan pronto como la semana 14. La secuenciación directa también nos permitió estudiar las características de ADN de plasma libre de células, y hemos encontrado pruebas de que este ADN se ha enriquecido para las secuencias de los nucleosomas.

EJEMPLO 1: Inscripción de sujetos

[0072] El estudio fue aprobado por El Comité de Revisión Institucional de la Universidad de Stanford. Las mujeres embarazadas con riesgo de aneuploidía fetal fueron reclutadas en el Centro de Diagnóstico Perinatal del Hospital Infantil de Lucile Packard de la Universidad de Stanford durante el período de abril de 2007 hasta mayo de 2008. Se obtuvo el consentimiento informado de cada participante antes de la extracción de sangre. Se recogió sangre de 15 a 30 minutos después de la amniocentesis o vellosidades coriónicas a excepción de 1 muestra que se recogió durante el tercer trimestre. El análisis del cariotipo se realizó a través de la amniocentesis o muestreo de vellosidades coriónicas para confirmar el cariotipo fetal. 9 trisomía 21 (T21), 2 trisomía 18 (T18), 1 trisomía 13 (T13) y 6 embarazos únicos normales fueron incluidos en este estudio. La edad gestacional de los sujetos en el momento de la extracción de sangre osciló entre 10 a 35 semanas (Tabla 1). La muestra de sangre de un donante masculino se obtuvo del Centro de Sangre de Stanford.

EJEMPLO 2: Procesamiento de muestra y cuantificación de ADN

[0073] 7 a 15 ml de sangre periférica extraída de cada sujeto y del donante se recogió en tubos con EDTA. La sangre se centrifugó a 1600g durante 10 minutos. El plasma se transfirió a tubos de microcentrífuga y se centrifugó a 16000g durante 10 minutos para eliminar las células residuales. Las dos etapas de centrifugación se llevaron a cabo dentro de las 24 horas después de la recogida de sangre. El plasma libre de células se almacenó a -80ºC hasta su procesamiento posterior y se congeló y se descongeló sólo una vez antes de la extracción de ADN. Se extrajo el ADN a partir de plasma libre de células usando Micro Kit de ADN de QAaRsmp (Qiagen) o el kit de plasma NucleoSpin (Macherey-Nagel) de acuerdo con las instrucciones del fabricante. ADN genómico fue extraído de 200 µl de sangre entera de los donantes utilizando Mini-Kit de ADN de Sangre QAaRsmp (Qiagen). PCR digital de microfluidos (Fluidigm) se utilizó para cuantificar la cantidad de ADN total y fetal mediante ensayos TaqMan de orientación en el locus EIF2C1 en el cromosoma 1 (delantero: 5 'GTTCGGCTTTCACCAGTCT 3' (SEQ ID NO: 1); Inverso: 5' CTCCATAGCTCTCCCCACTC 3' (SEQ ID NO: 2); Sonda: 5' HEX-GCCCTGCCATGTGGAAGAT-BHQ1 3' (SEQ ID NO: 3); tamaño de amplificación: 81bp) y el locus SRY en el cromosoma Y (delantero: 5' CGCTTAACATAGCAGAAGCA 3' (SEQ ID NO: 4); inverso: 5' AGTTTCGAACTCTGGCACCT 3' (SEQ ID NO: 5); Sonda: 5' FAM-TGTCGCACTCTCCTTGTTTTTGACA-BHQ1 3' (SEQ ID NO: 6); tamaño de amplificación: 84bp) respectivamente. Un ensayo de Taqman orientado a DYS 14 (delantero: 5 'ATCGTCCATTTCCAGAATCA 3' (SEQ ID NO: 6); inverso: 5' GTTGACAGCCGTGGAATC 3' (SEQ ID NO: 7); Sonda: 5 'FAM-TGCCACAGACTGAACTGAATGATTTTC-BHQ1 3' (SEQ ID NO: 8); tamaño de amplificación: 84bp), un locus de múltiples copias en el cromosoma Y, se utilizó para la determinación inicial de sexo fetal a partir de ADN de plasma libre de células con en tiempo real tradicional PCR. Las reacciones de PCR se realizaron con 1x iQ Supermix (Bio-Rad), 0,1% de Tween-20 (PCR digital microfluídico sólo), los cebadores 300 nM, y sondas 150 nM. El protocolo de ciclos térmicos de PCR era 95C durante 10 min, seguido de 40 ciclos de 95ºC durante 15s y 60C durante 1 min. Los cebadores y sondas fueron adquiridos de IDT.

EJEMPLO 3: Secuenciación

[0074] Un total de 19 muestras de ADN de plasma libres de células, incluyendo 18 de mujeres embarazadas y 1 de un donante de sangre masculino, y la muestra de ADN genómico a partir de sangre entera del mismo donante masculino, se secuenciaron en la plataforma Solexa/Illumina. ~ 1 a 8ng de fragmentos de ADN extraídos de 1,3 a 5,6 ml de plasma libre de células se usó para la preparación de biblioteca de secuenciación (Tabla 1). La preparación de biblioteca se llevó a cabo según el protocolo del fabricante, con ligeras modificaciones. Debido a que el plasma libre de células de ADN se fragmenta en la naturaleza, hay una mayor fragmentación por nebulización o sonicación se realizó sobre muestras de ADN de plasma.

[0075] El ADN genómico de sangre entera del donante masculino se sometió a ultrasonidos (Misonix XL-2020) (24 ciclos de 30s sonicación y 90 de pausa), produciendo fragmentos con un tamaño entre 50 y 400 pb, con un pico a 150pb. ~ 2ng de ADN genómico sometido a ultrasonidos se utiliza para la preparación de la biblioteca. Brevemente, las muestras de ADN eran de punta roma y se ligan a los adaptadores universales. La cantidad de adaptadores utilizados para la ligación era de 500 veces menos que los escritos en el protocolo del fabricante. Se llevaron a cabo 18 ciclos de PCR para enriquecer fragmentos con adaptadores utilizando cebadores complementarios a los adaptadores. Las distribuciones de tamaño de las bibliotecas de secuenciación se analizaron con el kit de ADN 1000 en el 2100 Bioanalyzer (Agilent) y se cuantificaron mediante PCR digital de microfluidos (Fluidigm). Las bibliotecas fueron secuenciadas utilizando el Analizador de Genoma Solexa 1 G de acuerdo con las instrucciones del fabricante.

[0076] ADN de plasma libre de células de una mujer embarazada con un feto masculino normal también se secuenció en la plataforma 454/Roche. Los fragmentos de ADN extraído de 5,6 ml de plasma libre de células (equivalente a ~ 4.9ng de ADN) se utiliza para la preparación de la biblioteca de secuenciación. La biblioteca de la secuenciación se preparó de acuerdo con el protocolo del fabricante, excepto que no nebulización se realizó en la muestra y la cuantificación se realizó con PCR digital de microfluido en lugar de electroforesis capilar. A continuación, la biblioteca se secuenció en el 454 Genome Sequencer FLX System de acuerdo con las instrucciones del fabricante.

[0077] Las bibliotecas de secuenciación de Solexa de electroferogramas se prepararon a partir de ADN de plasma libre de células obtenido a partir de 18 mujeres embarazadas y 1 macho donante. La biblioteca Solexa preparada a partir de ADN genómico de sangre entera se sometió a ultrasonidos del donante masculino también se examinó. Para las bibliotecas preparadas a partir de ADN libre de células, todas tenían picos a 261bp media (rango: 256-264bp). El tamaño máximo real de fragmentos de ADN en el ADN de plasma es ~ 168bp (después de la eliminación de adaptador universal Solexa (92bp)). Esto se corresponde con el tamaño de un cromatosoma.

EJEMPLO 4: Análisis de Datos

Análisis de la secuencia de la escopeta

[0078] La secuenciación de Solexa produjo 36 a 50 pb lecturas. La primera de 25 pb de cada lectura fue mapeada al constructo de genoma humano 36 (hg18) usando ELAND de la tubería de análisis de datos Solexa. Las lecturas que fueron asignadas de forma única con el genoma humano que tiene como máximo 1 desajuste se conservaron para el análisis. Para comparar la cobertura de los diferentes cromosomas, se aplicó una ventana deslizante de 50 kb a través de cada cromosoma, excepto en las regiones de huecos de montaje y los microsatélites, y se contó el número de etiquetas de secuencia que cae dentro de cada ventana y el valor de la mediana fue elegido para ser el representante del cromosoma. Debido a que el número total de etiquetas de secuencias para cada muestra era diferente, para cada muestra, que normalizó la densidad de etiqueta de secuencia de cada cromosoma (excepto el cromosoma Y) a la densidad de etiqueta mediana de secuencia entre autosomas. Los valores normalizados se utilizaron para la comparación entre las muestras en el análisis posterior. Se estimó la fracción de ADN fetal del cromosoma 21 para los casos T21, el cromosoma 18 de los casos T18, el cromosoma 13 de la caja T13, y los cromosomas X e Y para los embarazos masculinos. Para el cromosoma 21,18, y 13, la fracción de ADN fetal se estimó como 2 * (x-1), donde x era la relación de la densidad de etiqueta de secuencia de cromosoma sobrerrepresentado de cada caso de trisomía a la densidad de etiqueta de secuencia de cromosoma mediana de los casos de disomía. Para el cromosoma X, el ADN fetal se estimó como 2 * (1-x), donde X es la proparte de densidad de etiqueta de secuencia de cromosoma X de cada embarazo masculino a la densidad de etiqueta de secuencia mediana de cromosoma X de todos los embarazos femeninos. Para el cromosoma Y, la fracción de ADN fetal se estimó como la relación de densidad de etiqueta de secuencia de cromosoma Y de cada embarazo masculino a la de ADN de plasma de donante masculino. Debido a que se detectó un pequeño número de secuencias del cromosoma Y en los embarazos de mujeres, sólo se consideran las etiquetas de secuencias comprendidas en las regiones transcritas en el cromosoma Y y restamos el número medio de etiquetas en embarazos femininos de todas las muestras; esto equivale a una corrección de un pequeño porcentaje. La anchura de los intervalos de confianza del 99% se calculó para todos los embarazos de disomía 21 como t * s/VN, donde N es el número de disomía 21 embarazos, t es la estadística t correspondiente a a = 0.005 con grado de libertad igual a N-1 y s es la desviación estándar. Un intervalo de confianza da un rango estimado de valores, que es probable que incluya un parámetro desconocido de la población, el rango estimado se calcula a partir de un conjunto dado de datos de la muestra. (Definición tomada de Valerie J. Easton y Glosario de Estadísticas de John H. McColl v1.1)

[0079] Para investigar la distribución de las etiquetas de secuencias alrededor de los sitios de inicio de transcripción, se aplicó una ventana deslizante de 5 pb de -1000bp a + 1000 pb de los sitios de inicio de transcripción de todos los genes RefSeq en todos los cromosomas excepto cromosoma Y. El número de etiquetas de secuencia asignada a las hebras sentido y antisentido dentro de cada ventana se contó. El promedio móvil con una ventana de 10 puntos de datos se utilizó para suavizar los datos. Todos los análisis se realizaron con Matlab.

[0080] Se seleccionaron las etiquetas de secuencias que se mapearon únicamente para el genoma humano con un máximo de 1 desajuste (en promedio ~ 5000000) para el análisis. Se examinó la distribución de las lecturas a lo largo de cada cromosoma. Debido a que la distribución de las etiquetas de secuencias a través de cada cromosoma

era no uniforme (posiblemente artefactos técnicos), dividimos la longitud de cada cromosoma en la ventana deslizante no superpuesta con una anchura fija (en este análisis particular, se utiliza una ventana 50kbp), saltándose las regiones de huecos de montaje del genoma y regiones con repeticiones de microsatélites conocidos. La anchura de la ventana debe ser lo suficientemente grande de tal manera que hay un número suficiente de etiquetas de secuencias en cada ventana, y debe ser lo suficientemente pequeño tal que hay un número suficiente de ventanas para formar una distribución. Con la profundidad de secuenciación creciente (es decir, aumento del número total de etiquetas de secuencia), la anchura de la ventana se puede reducir. Se contó el número de etiquetas de secuencia en cada ventana. Se examinó la distribución del número de etiquetas de secuencia por 50 kb para cada cromosoma. El valor mediano del número de etiquetas de secuencias por 50kb (o 'densidad de etiqueta de secuencia') para cada cromosoma se eligió con el fin de suprimir los efectos de las regiones insuficientemente o excesivamente representadas dentro del cromosoma. Debido a que el número total de etiquetas de secuencias obtenidas para cada muestra era diferente, a fin de compararse entre muestras, normalizamos cada valor de densidad de etiqueta de secuencia cromosómica (excepto el cromosoma Y) por la densidad de etiqueta de secuencia mediana entre todos los autosomas (cromosomas no sexuales).

[0081] Para los datos de 454/Roche, lecturas se alinearon con el constructo de genoma humano 36 (hg18, véase el protocolo de transferencia de hipertexto (HTTP) genome.ucsc.edu/cgi-bin/hgGateway) utilizando el mapeador de referencia 454. Las lecturas que tienen exactitud mayor que o igual a 90% y la cobertura (es decir, fracción de lectura asignada) mayor que o igual a 90% se seleccionaron para su análisis. Para estudiar la distribución de tamaño de ADN total y fetal, el número de lecturas retenidas caen dentro de cada ventana de 10 pb entre 50 pb a 330 pb se contó. El número de lecturas incluidas en diferentes rangos de tamaño pueden estudiarse, es decir, lecturas de entre 50-60 pb, 60-70 pb, 70-80 pb, etc., hasta aproximadamente 320-330 pb, que es alrededor de la longitud de lectura máxima obtenida.

EJEMPLO 5: Recuperación de Datos del Genoma

[0082] La información relativa al contenido G/C, ubicación de los sitios de inicio de transcripción de los genes RefSeq, lugar de huecos de montaje y microsatélites se obtuvieron del Explorador de Genoma UCSC.

EJEMPLO 6 Enriquecimiento de Nucleosomas

[0083] Se analizó la distribución de etiquetas de secuencia alrededor de los sitios de inicio de transcripción (TSS) de genes RefSeq (datos no mostrados). Los gráficos eran similares a la **Figura 4.** Cada gráfico representa la distribución para cada muestra de plasma ADN o ADNg. Los datos se obtuvieron a partir de tres carreras de secuenciación diferentes (PI, P6, P52, P53, P26, P40, P42 se secuenciaron en conjunto; ADN masculino genómico, ADN de plasma masculino, P2, P7, P14, P19, P31 se secuenciaron juntos, P17, P20, P23, P57, P59, P64 se secuenciaron juntos). El segundo lote de muestras sufre mayor sesgo G/C como se observa a partir de la variación inter e intra-cromosómica. Sus distribuciones alrededor de TSS tienen tendencias similares con varias variables en la SAT. Dicha tendencia no es tan prominente como en la distribución de muestras secuenciadas en otras ejecuciones. No obstante, al menos 3 nucleosomas bien posicionados fueron detectables aguas abajo de los sitios de inicio de transcripción para la mayoría de las muestras de ADN de plasma, que sugiere que comparte ADN de plasma libre de células características de ADN nucleosomal, una pieza de evidencia de que este ADN es de origen apoptótico.

EJEMPLO 7: Cálculo de la fracción de ADN fetal en plasma materno de embarazos masculinos:

i. Con ensayos PCR TaqMan Digitales

[0084] La PCR digital es la amplificación de ADN de una sola molécula. La muestra de ADN se diluye y se distribuye a través de múltiples compartimentos de tal manera que, en promedio, hay menos de 1 copia de ADN por compartimento. Un compartimento que presenta fluorescencia al final de una PCR representa la presencia de al menos una molécula de ADN.

Ensayo para el ADN total: EIF2C1 (cromosoma 1)
Ensayo de ADN fetal: SRY (cromosoma Y)

[0085] El número de compartimentos positivos desde el chip PCR digital de microfluidos de cada ensayo se convierte en el recuento más probable de acuerdo con el método descrito en la información de apoyo de la referencia siguiente: Warren L, Bryder D, Weissman IL, Quake SR (2006) Factor de transcripción de perfiles en progenitores hematopoyéticos individuales por RT-PCR digital. Proc Nat Acad Sci, 103: 17807-12.

$$\text{Fracción de ADN fetal } \varepsilon = (\text{conteo SRY}) / (\text{conteo EIF2C1 conteo} / 2)$$

ii. Con marcadores de secuencias

[0086] De CHRX:

Siendo la fracción de ADN fetal ε

	Contribución materna	Contribución feto masculino	Contribución feto femenino
CHRX	2 (1- ∑)	∑	2 ∑

Embarazos masculinos densidad de etiqueta de secuencia ChrX (fetal y maternal) = 2(1-ε) + ε = 2 - ε

Embarazos femininos densidad de etiqueta de secuencia ChrX (fetal y maternal) = 2(1-ε) + 2 ε = 2

[0087] Siendo x la relación entre la densidad de etiqueta de secuencia de CHRX de embarazos macho a hembra. En este estudio, el denominador de esta relación se toma como la densidad de etiqueta media de secuencia de todos los embarazos femeninos.

[0088] Por lo tanto, la fracción de ADN fetal ∑ = 2 (1-x)

De ChrY:

[0089] Fracción de ADN fetal ∑ = (densidad de etiqueta de secuencia de ChrY en la densidad de etiqueta de plasma materno/secuencia de ChrY en plasma macho)

[0090] Nótese que en estas derivaciones, se supone que el número total de etiquetas de secuencia obtenida es la misma para todas las muestras. En realidad, el número total de etiquetas de secuencias obtenidas para diferente muestra es diferente, y hemos tenido en cuenta estas diferencias en nuestra estimación de la fracción de ADN fetal mediante la normalización de la densidad de etiqueta de secuencia de cada cromosoma mediana de las densidades de etiqueta de secuencia autosómica para cada muestra.

Cálculo de la fracción de ADN fetal en plasma materno de embarazos aneuploides (trisomia):

[0091] Siendo la fracción de ADN fetal ε

	Contribución materna	Contribución del feto con trisomía	Contribución de feto disómico
Cromosoma trisómico	2 (1- ∑)	3 ∑	2 ∑

Embarazos trisómicos conteos de secuencia de cromosoma trisómico (fetal y materno)

= 2(1- ε) + 3 ε = 2 + ε

Embarazos disómicos conteos de secuencia de cromosoma trisómico (fetal y materno)

= 2(1- ε) + 2 ε = 2

[0092] Siendo x la relación de conteos de secuencia de cromosoma trisómico (o densidad de etiqueta de secuencia) de embarazos trisómicos a disómicos. En este estudio, el denominador de esta relación se toma para ser la densidad de etiqueta de secuencia mediana de todos los embarazos disómicos.

[0093] Por lo tanto, la fracción de ADN fetal ∑ = 2 (x-1).

EJEMPLO 8: Corrección de sesgo de densidad de etiqueta de secuencia resultante de contenido G/C o A/T entre los diferentes cromosomas en una muestra

[0094] Este ejemplo muestra un refinamiento de resultados que indican secuencias de asignación a diferentes cromosomas y que permite la determinación del conteo de diferentes cromosomas o regiones de los mismos. Es decir, los resultados como se muestran en la **Figura 1A** pueden ser corregidos para eliminar las variaciones en la densidad de etiqueta de secuencia que se muestran para los cromosomas más altos en contenido de G/C, que se muestra hacia la derecha de la figura. Esta difusión de los valores resulta del sesgo de secuenciación en el método utilizado, donde un mayor número de lecturas tienden a ser obtenido dependiendo del contenido de G/C. Los resultados del método de este ejemplo se muestran en la **Figura 10. La Figura 10** es una superposición que muestra los resultados de un número de diferentes muestras, como se indica en la leyenda. Los valores de densidad de etiqueta de secuencia en las **Figs 1** y **10** se normalizaron a los de un control de ADN genómico masculino, ya que los valores de densidad no son siempre 1 para todos los cromosomas (incluso después de la corrección GC)

pero son consistentes entre una muestra. Por ejemplo, después de la corrección GC, los valores de todas las muestras para chr19 se agrupan alrededor de 0,8 (no mostrado). El ajuste de los datos a un valor nominal de 1 se puede hacer mediante el trazado del valor relativo al control ADNg masculina. Esto hace que los valores para todos los cromosomas se agrupen alrededor de 1.

[0095] Densidades de etiquetas de secuencia de cromosoma periféricas pueden considerarse como significativamente por encima de una densidad de etiqueta media de secuencia; cromosomas disómicos se agrupan alrededor de una línea que recorre un valor de densidad de alrededor de 1. Como se puede ver allí, los resultados del cromosoma 19 (a la derecha, más alto en contenido G/C), por ejemplo, muestran un valor similar cuando disómico como otros cromosomas disómicos. Las variaciones entre los cromosomas con contenido bajo y alto G/C se eliminan de los datos que han de examinarse. Las muestras (tales como P13 en el presente estudio) que no podrían haberse interpretado de forma inequívoca ahora pueden serlo. Dado que el contenido de G/C es lo contrario del contenido A/T, el presente método corregirá para ambos. Cualquiera de sesgo G/C o sesgo A/T pueden resultar de diferentes métodos de secuenciación. Por ejemplo, se ha informado por otros que el método Solexa da como resultado un mayor número de lecturas de secuencias en las que el contenido de G/C es alto. Véase, Dohm et al., "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing", Nuc. Acids Res. 36 (16), e105; doi: 10.1093/NAR/gkn425. El procedimiento del presente ejemplo sigue los siguientes pasos:

a. Calcular el contenido de G/C del genoma humano. Calcular el contenido de G/C de cada ventana 20kb que no se solapan de cada cromosoma del genoma humano (HG18) utilizando la secuencia de comandos hgG/CPercent del "árbol de fuentes de kent," del Explorador de Genoma de UCSC que contiene diferentes programas de utilidades, a disposición del público bajo licencia. El fichero de salida contiene las coordenadas de cada bin 20kb y el correspondiente contenido de G/C. Se encontró que un gran número de lecturas se obtuvieron rangos G/C más altos (aproximadamente 55-70%) y muy pocas lecturas se obtuvieron en porcentajes de contenido G/C más bajo, con esencialmente ninguno por debajo de aproximadamente 30% G/C (datos no mostrados). Debido a que la longitud real de un fragmento de ADN secuenciado no se conoce (sólo secuenciamos el primer 25bp de un extremo de una porción de ADN en la célula de flujo), y es el contenido de G/C de toda la porción de ADN que contribuyó al sesgo de secuenciación, se elige una ventana arbitraria de la secuencia de ADN genómico humano conocido para determinar el contenido de G/C de diferentes lecturas. Elegimos una ventana 20kb para examinar la relación entre el número de lecturas y contenido de GC. La ventana puede ser mucho más pequeña, por ejemplo, 10 kb o 5 kb, pero un tamaño de 20 kb hace que el cálculo sea más sencillo.

b. Calcular la relación entre la secuencia de la cobertura y el contenido de G/C. Asignar peso a cada lectura de acuerdo con el contenido de G/C. Para cada muestra, se cuenta el número de lectura por bin 20kb. El número de lectura se representa frente a contenido de G/C. El número medio de lectura se calcula para cada contenido de 0,1% G/C, haciendo caso omiso de los bins sin lecturas, bins con cero por ciento de G/C, y bins con lecturas sobre-abundantes. El recíproco del número promedio de lecturas para un por ciento G/C particular con respecto a la mediana del número global de la lectura se calcula como el peso. Cada lectura se asigna entonces un peso en función del por ciento G/C de la ventana de 20kb en el que caiga.

c. Investigar la distribución de lecturas a través de cada autosoma y el cromosoma X. En este paso, el número de lecturas, tanto no ponderado como ponderado, en cada ventana 50kb que no se solapa se registra. Para el conteo, se optó por una ventana de 50kb con el fin de obtener un número razonable de lecturas por la ventana y número razonable de ventanas por cromosoma para examinar las distribuciones. Tamaño de la ventana puede ser seleccionado basándose en el número de lecturas obtenido en un experimento dado, y puede variar en un amplio rango. Por ejemplo, 30K-100K puede utilizarse. Regiones de microsatélites conocidas se ignoran. Un gráfico que muestra los resultados de chr1 de P7 se muestra en la **Figura 11,** la cual ilustra la distribución de peso de esta etapa (c) a partir de la muestra P7, en la que el peso asignado a diferentes contenidos de G/C se muestra; Lecturas con mayor contenido de G/C están excesivamente representadas de la media y por lo tanto se les da menos peso.

d. Investigar la distribución de las lecturas a lo largo de chrY. Calcular el número de lecturas chrY en las regiones transcritas después de aplicar el peso a lecturas en chrY. El cromosoma Y es tratado de forma individual, ya que es corto y tiene muchas repeticiones. Incluso los datos de la secuencia del genoma femenino se asignarán en alguna parte en el cromosoma Y, debido a errores de secuenciación y de alineación. El número de lecturas chrY en las regiones transcritas después de aplicar peso a lecturas en chrY se utiliza para calcular el porcentaje de ADN fetal en la muestra.

EJEMPLO 9: Comparación de diferentes muestras de pacientes utilizando análisis estadísticos (estadística t)

[0096] Este ejemplo muestra otro refinamiento de los resultados tal como se obtiene usando los ejemplos anteriores. En este caso, múltiples muestras de pacientes se analizan en un solo proceso. **Figura 12** ilustra los resultados de un análisis de pacientes P13, P19, P31, P23, P26, P40, P42, P1, P2, P6, P7, P14, P17, P20, P52, P53, P57, P59 y P64, con sus respectivos cariotipos indicaron, como en la **Tabla 1,** anteriormente. La línea punteada muestra el intervalo de confianza del 99%, y los valores extremos se puede identificar con rapidez. Se puede ver consultando debajo de la línea que los fetos masculinos tienen menos cromosoma X (triángulos sólidos). Una excepción es P19, donde se

cree que no había suficientes lecturas totales para este análisis. Se puede ver al mirar por encima de la línea que los pacientes de trisomía 21 (círculos sólidos) son P 1, 2, 6, 7, 14, 17, 20, 52 y 53. P57 y 59 tienen trisomía 18 (rombos abiertos) y P64 tiene trisomía 13 (estrella). Este método puede ser presentado por el siguiente proceso de tres pasos:

Paso 1: Calcular en estadística t para cada cromosoma en relación con todos los demás cromosomas en una muestra. Cada estadística t comunica el valor de cada cromosoma mediano con respecto a otros cromosomas, teniendo en cuenta el número de lecturas asignado a cada cromosoma (ya que la variación de la escala mediana con el número de lecturas). Como se describió anteriormente, los presentes análisis arrojaron sobre 5.000.000 lecturas por muestra. Aunque uno puede obtener 3-10 million lecturas por muestra, estas son lecturas cortas, por lo general sólo alrededor de 20 a 100 pb, por lo que uno en realidad sólo ha secuenciado, por ejemplo aproximadamente 300 millones de 3000 millones pb en el genoma humano. Por lo tanto, se usan métodos estadísticos donde se tiene una pequeña muestra y la desviación estándar de la población (3.000 millones, o 47 millones para el cromosoma 21) es desconocida y que se desea estimar desde el número de la muestra de lecturas con el fin de determinar el significado de una variación numérica. Una forma de realizar esto es mediante el cálculo de la distribución t de Student, que se puede utilizar en lugar de una distribución normal esperadq de una muestra más grande. La estadística t es el valor obtenido cuando se calcula la distribución t. La fórmula utilizada para este cálculo es la siguiente. Utilizando los métodos que aquí se presentan, otras pruebas t se pueden utilizar.

Paso 2: Calcular la matriz promedio de estadística t promediando los valores de todas las muestras con cromosomas disómicos. Cada dato de muestra del paciente se coloca en el matriz, en donde la fila es chr1 a chr22, y la columna también es chr1 a chr22. Cada célula representa el valor t cuando se comparan los cromosomas en la fila y la columna correspondiente (es decir, la posición (2,1) de la matriz es el valor t del examen de chr2 y chr1) la diagonal de la matriz es 0 y la matriz es simétrica. El número de lecturas de mapeo a un cromosoma se compara individualmente a cada uno de chr1-22.

Paso 3: Reste la matriz promedio de estadística t de la matriz de estadística t de cada muestra. Para cada cromosoma, la mediana de la diferencia en la estadística t se selecciona como el valor representativo.

[0097] La estadística t de confianza del 99% para el gran número de muestras es de 3,09. Cualquier cromosoma con una estadística t representativa era de -3,09 a 3,09 se determina como no disómico.

EJEMPLO 10: Cálculo del número requerido de lecturas de secuencia después de la corrección de sesgo G/C

[0098] En este ejemplo, se presenta un método que se utilizó para calcular la concentración mínima de ADN fetal en una muestra que sería necesario para detectar una aneuploidía, en base a un cierto número de lecturas obtenidas para ese cromosoma (excepto el cromosoma Y). **Figura 13** y **la Figura 14** muestran los resultados obtenidos a partir de muestras de ADN de plasma de 19 pacientes, 1 donante muestra de ADN de plasma, y ejecuciones en duplicado de una muestra de ADNg donante. Se estima en **la Figura 13** que el % de ADN fetal mínimo del cual se puede detectar exceso de representación de Chr21 a la mejor tasa de muestreo (~ 70K lecturas asignadas a Chr21) es ~ 6%. (indicado por líneas continuas en la **Fig. 13).** Las líneas se dibujan entre aproximadamente 0,7 $X10^5$ lecturas y concentración de ADN fetal 6%. Se puede esperar que con los números más altos de lecturas (no se ejemplifican aquí) el porcentaje de ADN fetal necesario se reducirá, probablemente a aproximadamente 4%.

[0099] En la **Figura 14,** los datos de la **Figura 13** se presentan en una escala logarítmica. Esto muestra que la concentración de ADN fetal mínimo requerido escala de forma lineal con el número de lecturas en una relación de raíz cuadrada (pendiente de -.5). Estos cálculos se realizaron

Para valores grandes de n (n> 30), estadística t $t = \frac{\overline{y_2} - \overline{y_1}}{\sqrt{\frac{s_2^2}{n_2} + \frac{s_1^2}{n_1}}}$, donde $\overline{y_2} - \overline{y_1}$ es la diferencia en las medias (o la cantidad de sobrerepresentación o subrepresentación de un cromosoma particular) a medirse; *s* es la desviación estándar del número de lecturas por 50kb en un cromosoma particular; n es el número de muestras (es decir, el número de ventanas 50kb por cromosoma). Al ser fijo el número de ventanas de 50kb por cromosoma, $n_1 = n_2$. Si suponemos que s_1 H s_2, $\overline{y_2} - \overline{y_1} \approx t\sqrt{\frac{2s_1^2}{n_1}} = \text{sqrt}(2)^* \text{ mitad}$ anchura del intervalo de

confianza al nivel de confianza gobernado por el valor de *t.* Por lo tanto, $\frac{\overline{y_2}}{y_1} - 1 \approx \frac{t\sqrt{\frac{2s_1^2}{n_1}}}{\overline{y_1}}$. Para todos los cromosomas en todas las muestras, podemos calcular el valor $\frac{t\sqrt{\frac{2s_1^2}{n_1}}}{\overline{y_1}}$, que corresponde al mínimo exceso o falta de representación que se puede resolver con el nivel de confianza regido por el valor *de t.* Nota que $2*(\frac{\overline{y_2}}{y_1} - 1)*100\%$ corresponde al % de ADN fetal mínimo de los cuales se puede detectar cualquier representación excesiva o insuficiente de cromosomas. Esperamos que el número de lecturas asignadas a cada cromosoma juegue un papel en la determinación de la desviación estándar s_1, ya que de acuerdo a la distribución de Poisson, la desviación estándar es igual a la raíz cuadrada de la media. Mediante el trazo $2*(\frac{\overline{y_2}}{y_1} - 1)*100\%$ vs. número de lecturas asignadas a cada cromosoma en todas las muestras, podemos evaluar el % de ADN fetal mínimo del cual cualquier exceso o falta de representación de los cromosomas se puede detectar debido a la frecuencia de muestreo actual.

[0100] Después de la corrección del sesgo de G/C, el número de lecturas por ventana de 50kb para todos los cromosomas (excepto el cromosoma Y) se distribuye normalmente. Sin embargo, hemos observado valores atípicos en algunos cromosomas (por ejemplo, una sub-región en el cromosoma 9 tiene representación cerca de cero; una sub-región en el cromosoma 20, cerca del centrómero tiene representación inusualmente alta) que afectan el cálculo de la desviación estándar y la media. Por lo tanto, decidimos calcular un intervalo de confianza de la mediana en lugar de la media para evitar el efecto de los valores atípicos en el cálculo del intervalo de confianza. No esperamos que el intervalo de confianza de la mediana y la media sea muy diferente si el pequeño número de valores atípicos se ha eliminado. El intervalo de confianza del 99,9% de la mediana para cada cromosoma se estima a partir del proceso de arranque de 5000 muestras de la distribución de lecturas 50kb utilizando el método de percentil. La anchura media del intervalo de confianza se calcula como intervalo de confianza 0,5*. Se traza 2* (anchura media de intervalo de confianza de la mediana)/media* 100% frente al número de lecturas asignadas a cada cromosoma para todas las muestras.

[0101] Remuestreo de proceso de arranque y otros cálculos implementados en ordenador descritos aquí se llevaron a cabo en MATLAB®, disponible de The MathWorks, Natick, MA.

REFERENCIAS

[0102]

1. Cunningham F, et al. (2002) in Williams Obstetrics (McCrraw-Hill Professional, New York), p. 942.

2. (2007) ACOG Practice Bulletin No. 88, December 2007. Invasive prenatal testing for aneuploidy. Obstet Gynecol, 110: 1459-1467.

3. Wapner R, et al. (2003) First-trimester screening for trisomies 21 and 18. N Engl JMed, 349: 1405-1413.

4. Alfirevic Z, Neilson JP (2004) Antenatal screening for Down's syndrome. Bmj 329: 811-812.

5. Malone FD, et al. (2005) First-trimester or second-trimester screening, or both, for Down's syndrome. N Engl JMed, 353: 2001-2011.

6. Herzenberg LA, et al. (1979) Fetal cells in the blood of pregnant women: detection and enrichment by fluorescence- activated cell sorting. Proc Natl Acad Sci USA, 76: 1453-1455.

7. Bianchi DW, et al. (1990) Isolation of fetal DNA from nucleated erythrocytes in maternal blood. Proc Natl Acad Sci USA, 87: 3279-3283.

8. Cheung MC, Goldberg JD, Kan YW (1996) Prenatal diagnosis of sickle cell anaemia and thalassaemia by analysis of fetal cells in maternal blood. Nat Genet, 14: 264-268.

9. Bianchi DW, et al. (1997) PCR quantitation of fetal cells in maternal blood in normal and aneuploid pregnancies. Am JHum Genet, 61: 822-829.

10. Bianchi DW, et al. (2002) Fetal gender and aneuploidy detection using fetal cells in maternal blood: analysis of NIFTY I data. National Institute of Child Health and Development Fetal Cell Isolation Study. Prenat Diagn, 22: 609-615.

11. Lo YM, et al. (1997) Presence of fetal DNA in maternal plasma and serum. Lancet, 350: 485-487.

12. Dennis Lo YM, Chiu RW (2007) Prenatal diagnosis: progress through plasma nucleic acids. Nat Rev Genet, 8: 71-77.

13. Lo YM, et al. (1998) Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. Am JHum Genet, 62: 768-775.

14. Lo YM, et al. (2007) Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection. Nat Med, 13: 218-223.

15. Tong YK, et al. (2006) Noninvasive prenatal detection of fetal trisomy 18 by epigenetic allelic ratio analysis in maternal plasma: Theoretical and empirical considerations. Clin Chem, 52: 2194-2202.

16. Dhallan R, et al. (2007) A non-invasive test for prenatal diagnosis based on fetal DNA present in maternal blood: a preliminary study. Lancet, 369: 474-481.

17. Fan HC, Quake SR (2007) Detection of aneuploidy with digital polymerase chain reaction. Anal Chem, 79: 7576-7579.

18. Lo YM, et al. (2007) Digital PCR for the molecular detection of fetal chromosomal aneuploidy. Proc Natl Acad Sci USA, 104: 13116-13121.

19. Quake SR, Fan HC. (2006). Non-invasive fetal genetic screening by digital analysis. USA Provisional Patent Application No. 60/764,420 (published for WO 2007/092473).

20. Mardis ER (2008) Next-Generation DNA Sequencing Methods. Annu Rev Genomics Hum Genet, 9: 387-402.

21. Lander ES, et al. (2001) Initial sequencing and analysis of the human genome. Nature, 409: 860-921.

22. Chan KC, et al. (2004) Size distributions of maternal and fetal DNA in maternal plasma. Clin Chem, 50: 88-92.

23. Li Y, et al. (2004) Size separation of circulatory DNA in maternal plasma permits ready detection of fetal DNA polymorphisms. Clin Chem, 50: 1002-1011.

24. Cooper G, Hausman R (2007) in The cell: a molecular approach (Sinauer Associates, Inc, Sunderland), p. 168.

25. Jahr S, et al. (2001) DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res, 61: 1659-1665.

26. Giacona MB, et al. (1998) Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. Pancreas, 17: 89-97.

27. Schones DE, et al. (2008) Dynamic regulation of nucleosome positioning in the human genome. Cell, 132: 887-898.

28. Ozsolak F, Song JS, Liu XS, Fisher DE (2007) High-throughput mapping of the chromatin structure of human promoters. Nat Biotechnol, 25: 244-248.

29. Yuan GC, et al. (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. Science, 309: 626-630.

30. Lee W, et al. (2007) A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet, 39: 1235-1244.

31. Sohda S, et al. (1997) The proportion of fetal nucleated red blood cells in maternal blood: estimation by FACS analysis. Prenat Diagn, 17: 743-752.

32. Hamada H, et al. (1993) Fetal nucleated cells in maternal peripheral blood: frequency and relationship to gestational age. Hum Genet, 91: 427-432.

33. Nelson JL (2008) Your cells are my cells. Sci Am, 298: 64-71.

34. Khosrotehrani K, Bianchi DW (2003) Fetal cell microchimerism: helpful or harmful to the parous woman? Curr Opin Obstet Gynecol, 15: 195-199.

35. Lo YM, et al. (1999) Rapid clearance of fetal DNA from maternal plasma. Am J Hum Genet, 64: 218-224.

36. Smid M, et al. (2003) No evidence of fetal DNA persistence in maternal plasma after pregnancy. Hum Genet, 112: 617-618.

37. Rijnders RJ, Christiaens GC, Soussan AA, van der Schoot CE (2004) Cell-free fetal DNA is not present in plasma of nonpregnant mothers. Clin Chem, 50: 679-681; author reply 681.

38. Hillier LW, et al. (2008) Whole-genome sequencing and variant discovery in C. elegans. Nat Methods, 5: 183-188.

39. Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res.

40. Harris TD, et al. (2008) Single-molecule DNA sequencing of a viral genome. Science, 320: 106-109.

41. Samura O, et al. (2003) Cell-free fetal DNA in maternal circulation after amniocentesis. Clin Chem, 49: 1193-1195.

42. Lo YM, et al. (1999) Increased fetal DNA concentrations in the plasma of pregnant women carrying fetuses with trisomy 21. Clin Chem, 45: 1747-1751.

43. Segal E, et al. (2006) A genomic code for nucleosome positioning. Nature, 442: 772-778.

LISTA DE SECUENCIAS

[0103]

<110> El Consejo de Administración de la Universidad de Leland Stanford Junior

<120> DIAGNÓSTICO NO INVASIVO DE ANEUPLOIDÍAS FETALES POR SECUENCIACIÓN

<130> 3815-63-1PCT S08-309

<140> No asignado todavía
<141> 2009-09-16

<150> US 61/098.758
<151> 2008-09-20

<150> US Nat'1 Stage App
<151> 2009-09-16

<160> 9

<170> versión de patentina 3.5

<210> 1
<211> 19
<212> ADN
<213> artificial

<220>
<223> oligonucleótido sintético

<400> 1
gttcggcttt caccagtct 19

<210> 2
<211> 20
<212> ADN
<213> artificial

<220>
<223> oligonucleótido sintético

<400> 2
ctccccactc ctccatagct 20

<210> 3
<211> 19
<212> ADN
<213> artificial

<220>
<223> oligonucleótido sintético

<400> 3
gccctgccat gtggaagat 19

<210> 4
<211> 20
<212> ADN
<213> artificial

<220>
<223> oligonucleótido sintético

<400> 4
agcagaagca cgcttaacat 20

<210> 5
<211> 20
<212> ADN
<213> oligonucleótido sintético

<400> 5
agtttcgaac tctggcacct 20

<210> 6
<211> 25
<212> ADN
<213> artificial

<220>
<223> oligonucleótido sintético

<400> 6
tgtcgcactc tccttgtttt tgaca 25

<210> 7
<211> 20
<212> ADN
<213> oligonucleótido sintético

<400> 7
tccagaatca atcgtccatt 20

<210> 8
<211> 18
<212> ADN
<213> artificial

<220>
<223> oligonucleótido sintético

<400> 8
gttgacagcc gtggaatc 18

<210> 9
<211> 27
<212> ADN
<213> artificial

<220>
<223> oligonucleótido sintético

<400> 9
tgccacagac tgaactgaat gattttc 27

Reivindicaciones

1. Un método de ensayo para una distribución anormal de una parte de cromosoma se especifica en una muestra mixta de porciones de cromosomas distribuidos normalmente y anormalmente obtenidos de un sujeto, en el que la distribución anormal es una aneuploidía fetal, y en el que la muestra es una mezcla de ADN materno y fetal en una muestra de plasma materno, que comprende:

(a) la obtención de secuencias, por secuenciación masiva en paralelo, de múltiples porciones de cromosomas de la muestra mezclada para obtener un número de etiquetas de secuencias de suficiente longitud de la secuencia determinada a asignarse a una localización cromosómica dentro de un genoma y en número suficiente para reflejar la distribución anormal de la parte de cromosoma especificado;
(b) la asignación de las etiquetas de secuencia de porciones de cromosomas que incluyen al menos la parte de cromosoma especificado mediante la comparación de las etiquetas de secuencia a una referencia de secuencia genómica correspondiente;
(c) la determinación de los valores para los números de secuencia de las etiquetas de asignación a porciones de cromosoma distribuidas normal y anormalmente por:

(i) las etiquetas de secuencia de conteo dentro de una serie de ventanas predefinidas de longitudes iguales dentro de al menos una parte de cromosoma distribuido normalmente para obtener un primer valor; y
(ii) las etiquetas de secuencia de conteo dentro de una serie de ventanas predefinidas de longitudes iguales dentro de la parte de cromosoma especificado para obtener un segundo valor; y

(d) el uso de los valores de la etapa (c) para determinar una diferencia, entre el primer valor y el segundo valor, que es determinante de la sobrerrepresentación o infrarrepresentación de la porción de cromosoma especificada en la mezcla de ADN materno y fetal.

2. El método de la reivindicación 1, en el que las ventanas están predefinidas 10 kb a 100 kb de longitud.

3. El método de la reivindicación 1, en el que para determinar un diferencial incluye la etapa de comparar una densidad de etiqueta de secuencia normalizada de la parte de cromosoma especificada a una densidad de etiqueta de secuencia normalizada de otra parte del cromosoma en dicha muestra mixta, en la que todos los autosomas se utilizan para calcular la densidad de etiqueta de secuencia normalizada.

4. El método de la reivindicación 1, en el que:

la distribución anormal es una aneuploidía de al menos uno de los cromosomas 13, 18 y 21; o
la parte de cromosoma especificado es cualquiera de los cromosomas X, Y, 18, 21, 17 o 13.

5. El método de la reivindicación 1 en el que la etapa de asignar las etiquetas de secuencias a las porciones de cromosomas correspondientes permite una falta de coincidencia.

6. El método de la reivindicación 1 en el que las etiquetas de secuencia son de 25 a 100 pb de longitud.

7. El método de la reivindicación 1 o la reivindicación 6, en el que se obtienen al menos 1 millón de etiquetas de secuencia.

8. El método de la reivindicación 6 que comprende además la etapa de comparar una densidad de etiqueta de secuencia normalizada de la parte especificada de cromosoma a una densidad de etiqueta de secuencia normalizada de otra parte del cromosoma en dicha muestra mixta.

9. El método de la reivindicación 8 en el que la etapa de determinar un diferencial incluye la etapa de comparar una densidad de etiqueta de secuencia normalizada de la parte de cromosoma especificado a una densidad de etiqueta de secuencia normalizada de otra porción de cromosoma en dicha muestra mixta, en la que todos los autosomas se utilizan para calcular la densidad de etiqueta de secuencia normalizada.

10. El método de la reivindicación 1 o la reivindicación 9 que comprende además la etapa de medición de sobrerepresentación y subrepresentación de un cromosoma mediante la determinación de una densidad de etiqueta de secuencia para cada cromosoma en la muestra, es decir, los cromosomas 1-22, X y también el cromosoma Y si está presente.

11. El método de la reivindicación 1, en el que:

dicha determinación de un diferencial comprende la obtención de una densidad de etiqueta de secuencia del cromosoma anormalmente distribuido y comparándolo con un valor de un cromosoma disómico;

y/o
dichas ventanas se componen de ventanas correderas no superpuestas de 10 a 100 kb que se extienden sustancialmente a lo largo de un cromosoma entero; y/o
que comprende además la etapa de medición de una serie de etiquetas de secuencias dentro de sitios de inicio transcripcional.

12. El método de la reivindicación 1, el cual comprende:

la determinación de los números de las etiquetas de secuencias asignadas a cada ventana en al menos cada autosoma;
la determinación de una media de dichos números para cada autosoma y un segundo medio de al menos todos los autosomas;
el cálculo de un valor normalizado de todos los autosomas, utilizando dicha segunda media; y
la comparación entre los valores normalizados entre autosomas para determinar cualquiera porción de cromosomas autosómica anormalmente distribuida de interés.

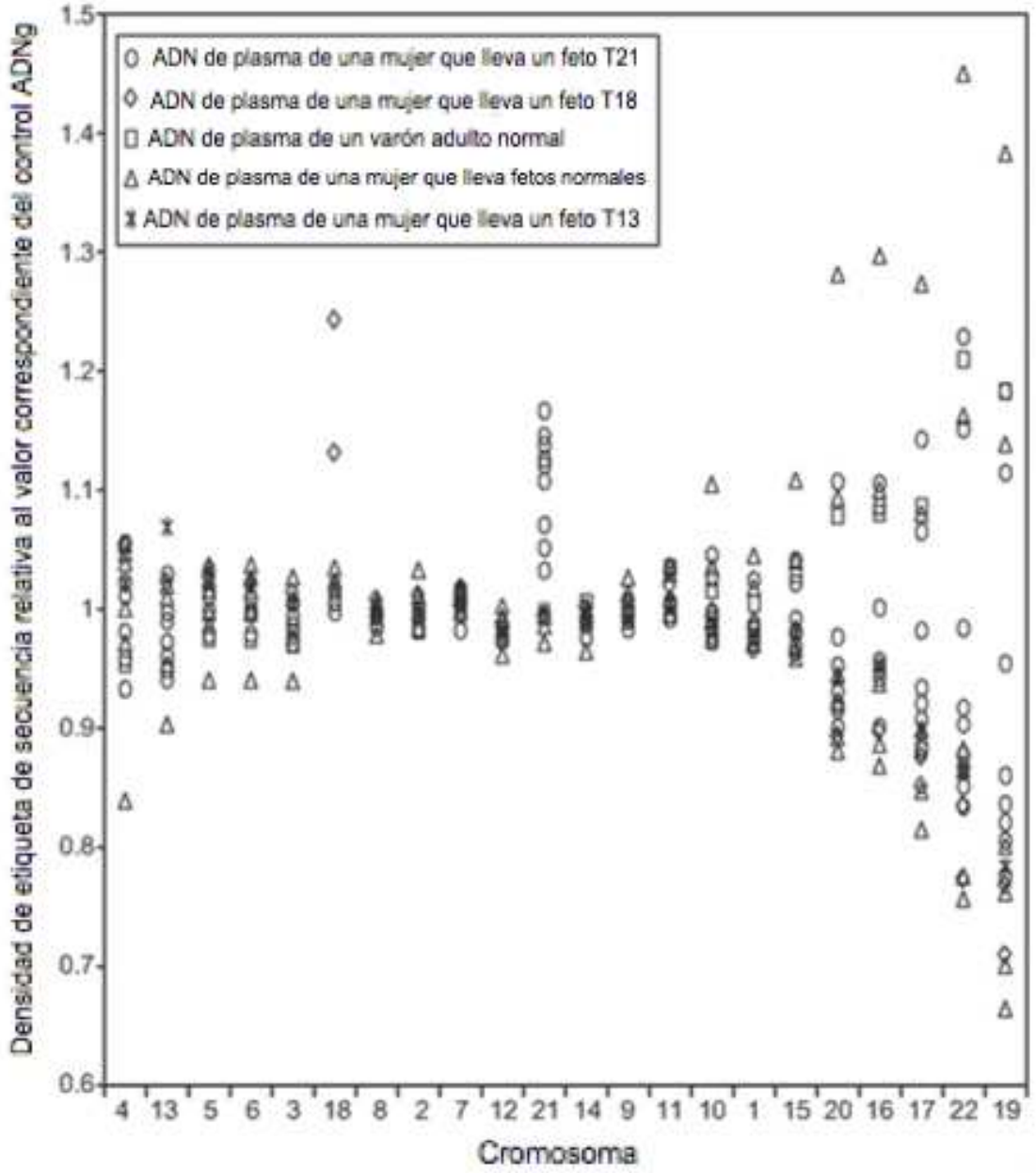
ADN de plasma de una mujer que lleva un feto T21
ADN de plasma de una mujer que lleva un feto T18
ADN de plasma de un varón adulto normal
ADN de plasma de una mujer que lleva fetos normales
ADN de plasma de una mujer que lleva un feto T13
Densidad de etiqueta de secuencia relativa al valor correspondiente del control ADNg
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
4 13 5 6 3 18 8 2 7 12 21 14 9 11 10 1 15 20 16 17 22 19
Cromosoma

FIG. 1A

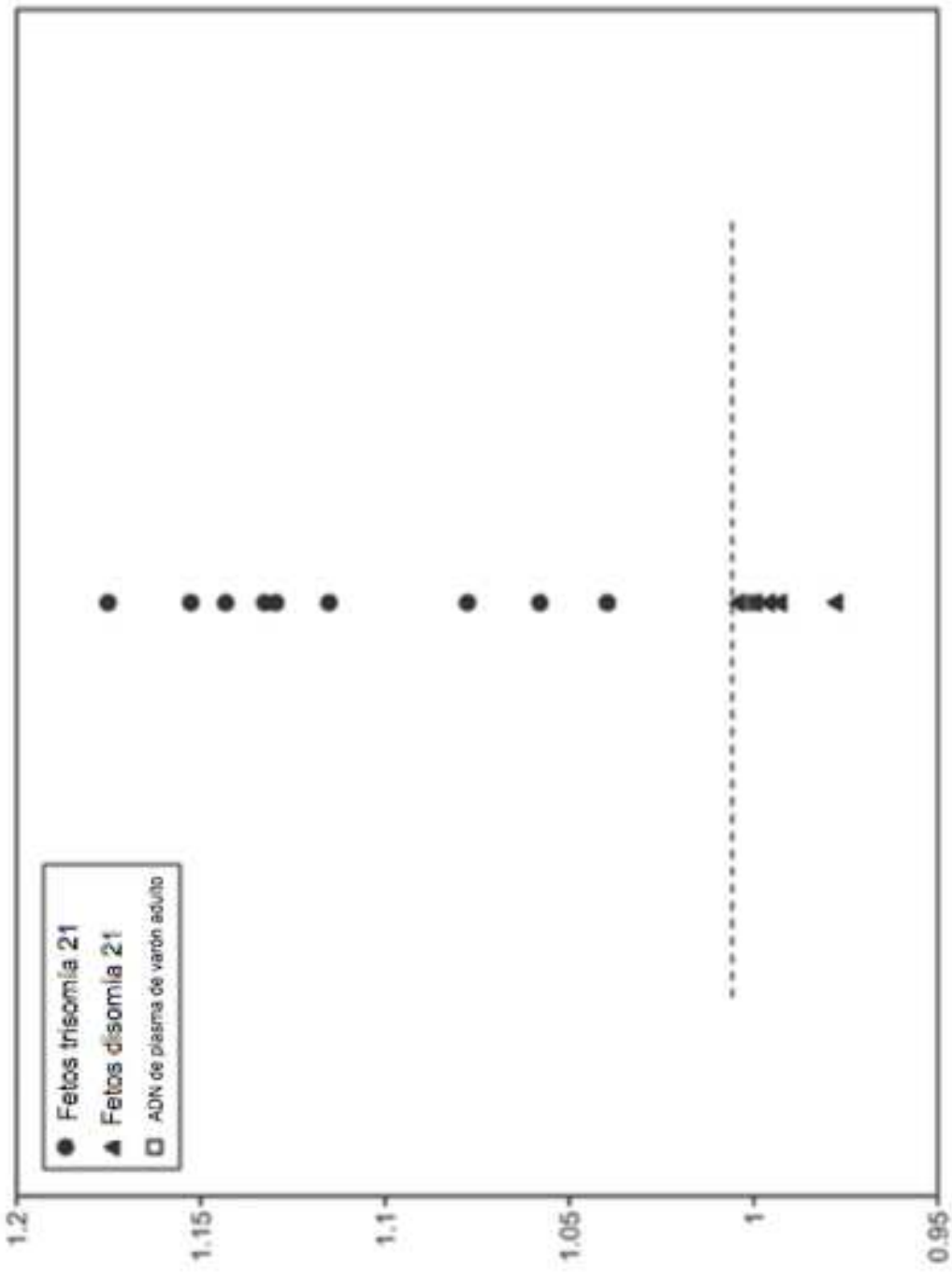
Fetos trisomía 21
Fetos disomía 21
ADN de plasma de varón adulto
1.2
1.15
1.1
1.05
1
0.95
Densidad de etiqueta de secuencia de cromosoma 21 relativa al valor mediano de casos de disomía 21
Cromosoma 21

FIG. 1B

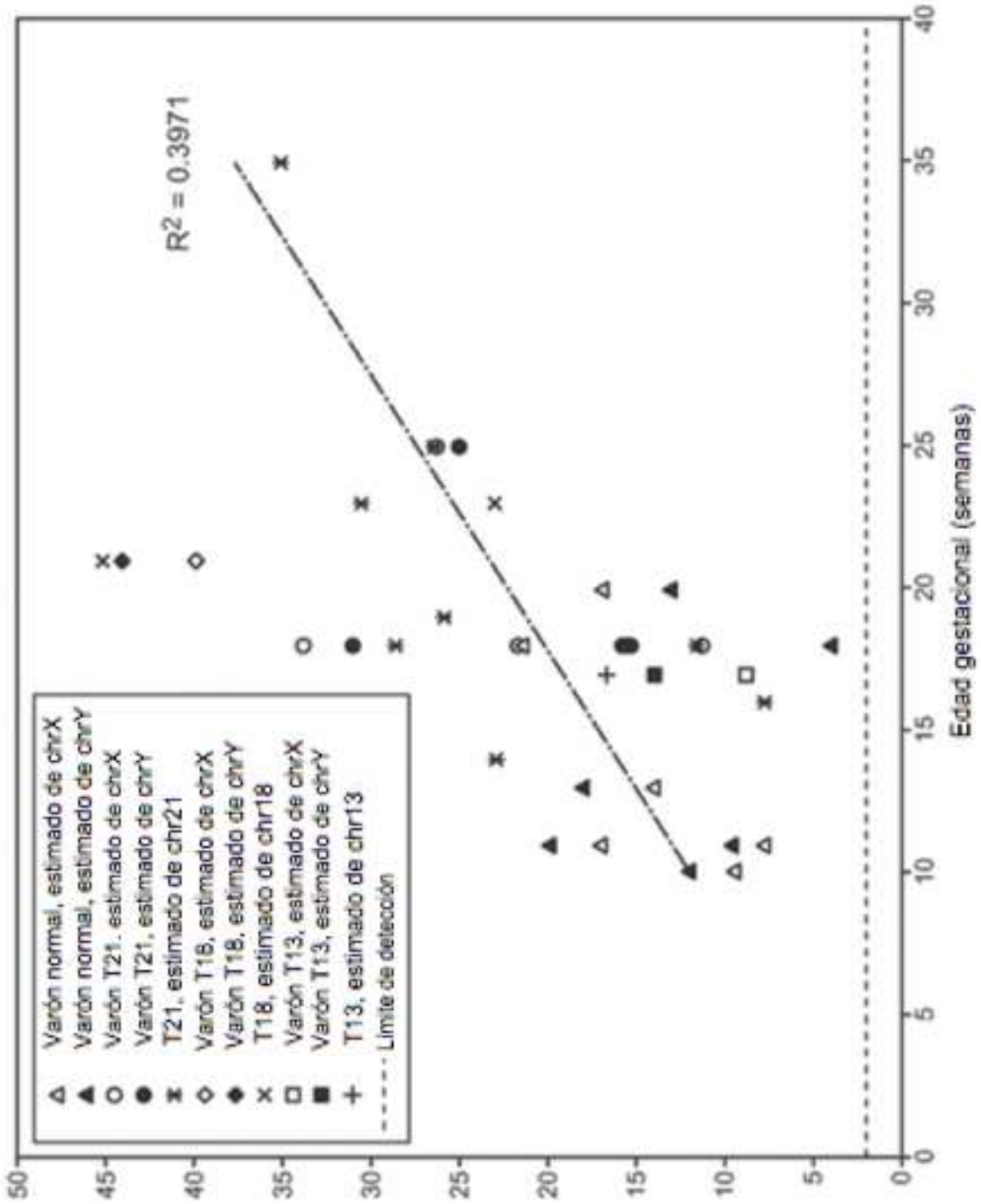
Varón normal, estimado de chrX
Varón normal, estimado de chrY
Varón T21, estimado de chrX
Varón T21, estimado de chrY
T21, estimado de chr21
Varón T18, estimado de chrX
Varón T18, estimado de chrY
T18, estimado de chr18
Varón T13, estimado de chrX
Varón T13, estimado de chrY
T13, estimado de chr13
Límite de detección
R2 = 0.3971
Porcentaje de ADN materno libre de células que se origina en el feto (%)
Edad gestacional (semanas)
0
5
10
15
20
25
30
35
40
45
50

FIG. 2

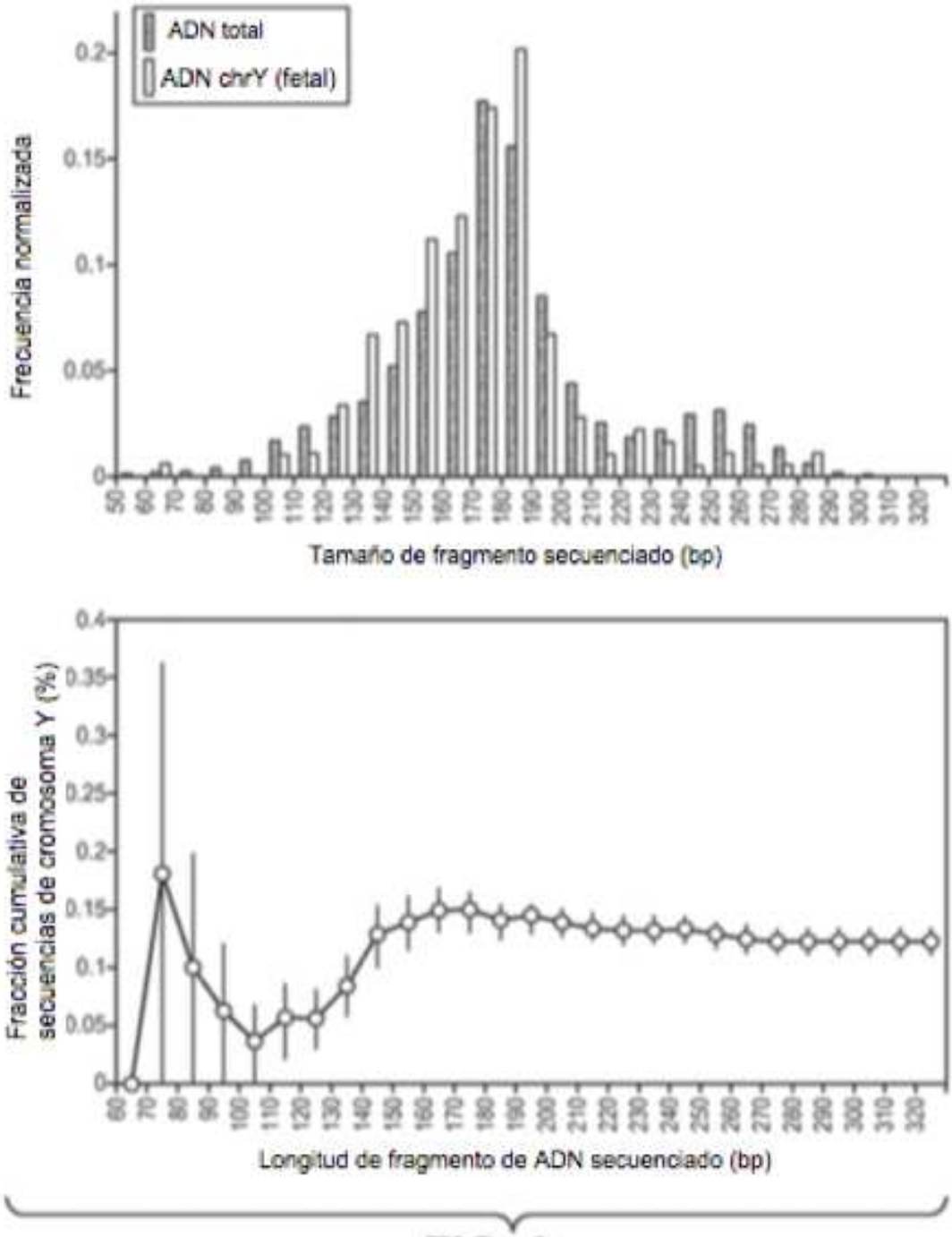
ADN total
ADN chrY (fetal)
Frecuencia normalizada
0.2
0.15
0.1
0.05
0
50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320
Tamaño de fragmento secuenciado (bp)
Fracción cumulativa de secuencias de cromosoma Y (%)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320
Longitud de fragmento de ADN secuenciado (bp)

FIG. 3

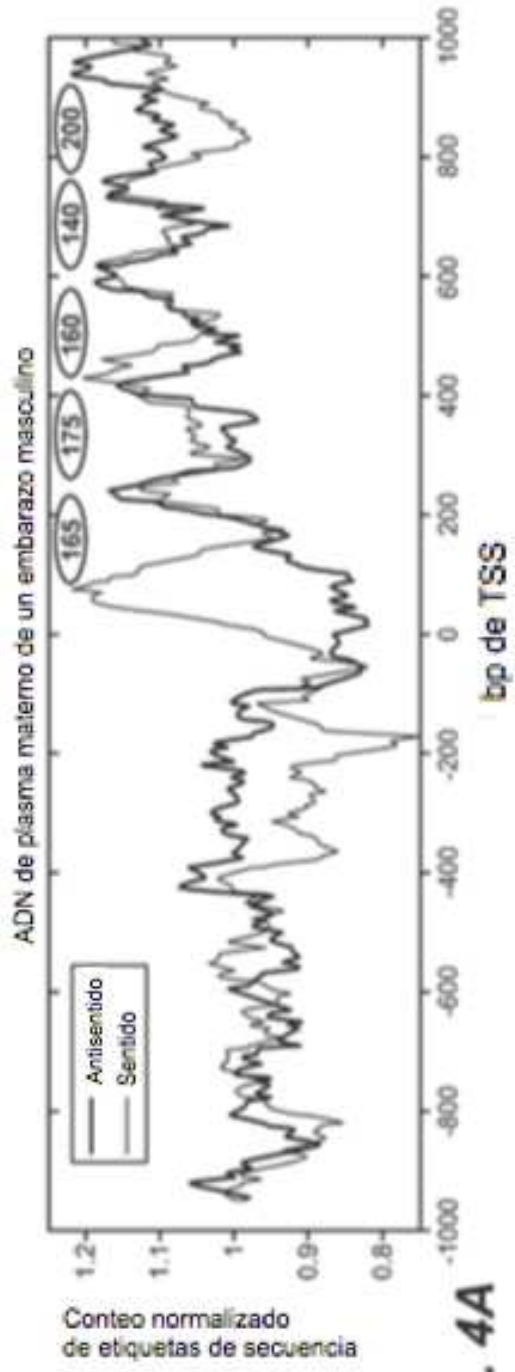
ADN de plasma materno de un embarazo masculino
Antisentido
Sentido
165
175
160
140
200
1.2
1.1
1
0.9
0.8
-1000
-800
-600
-400
-200
0
200
400
600
800
1000
Conteo normalizado
de etiquetas de secuencia
bp de TSS

FIG. 4A

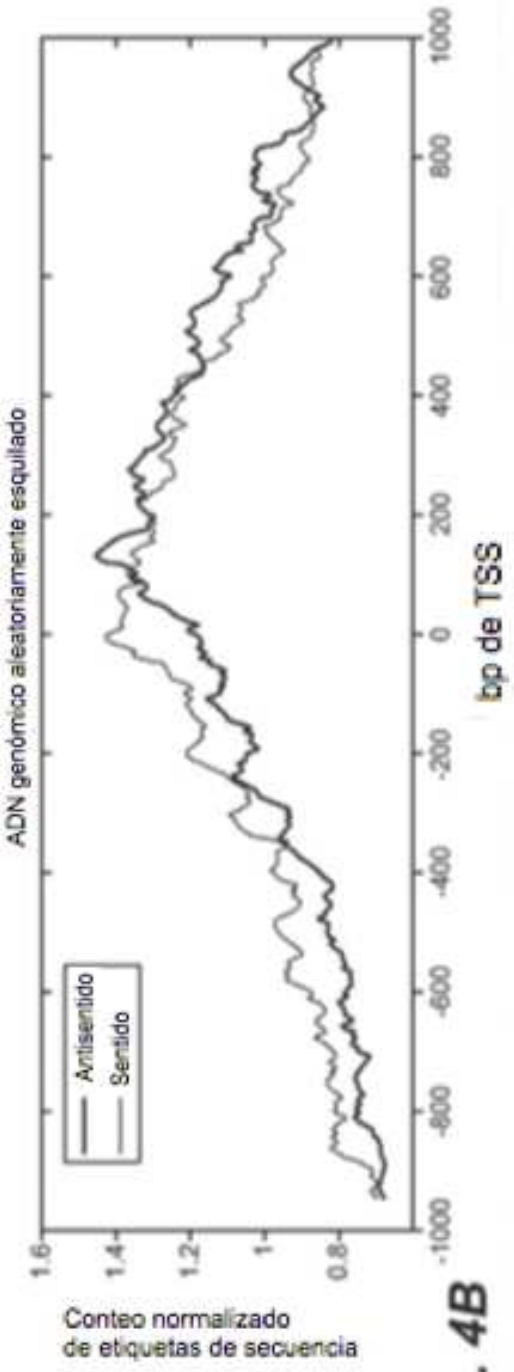
ADN genómico aleatoriamente esquilado
Antisentido
Sentido
1.6
1.4
1.2
1
0.8
-1000
-800
-600
-400
-200
0
200
400
600
800
1000
Conteo normalizado
de etiquetas de secuencia
bp de TSS

FIG. 4B

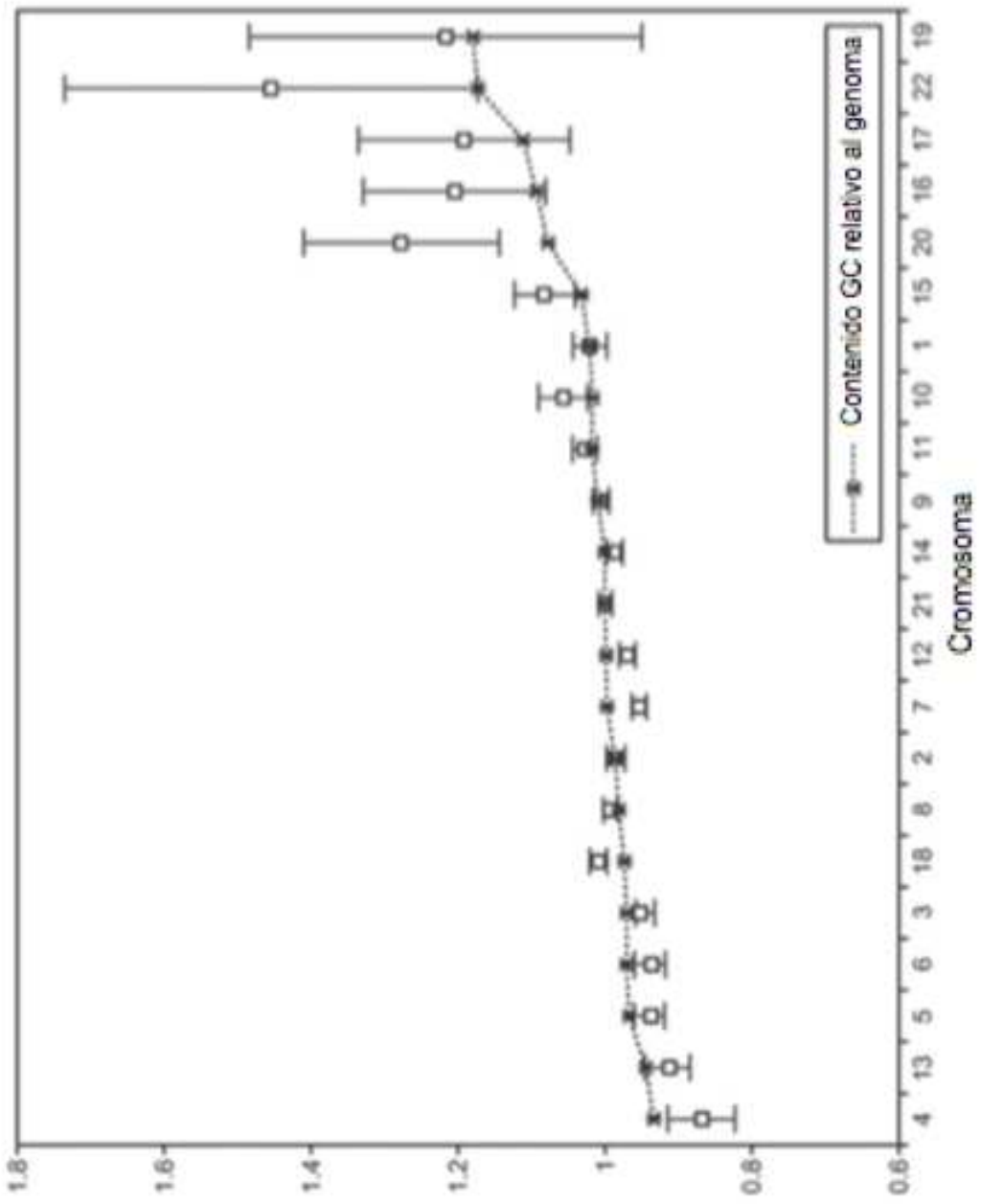
Contenido GC relativo al genoma
Densidad de etiqueta de secuencia media
Cromosoma
1.8
1.6
1.4
1.2
1
0.8
0.6
4 13 5 6 3 18 8 2 7 12 21 14 9 11 10 1 15 20 16 17 22 19

FIG. 5A

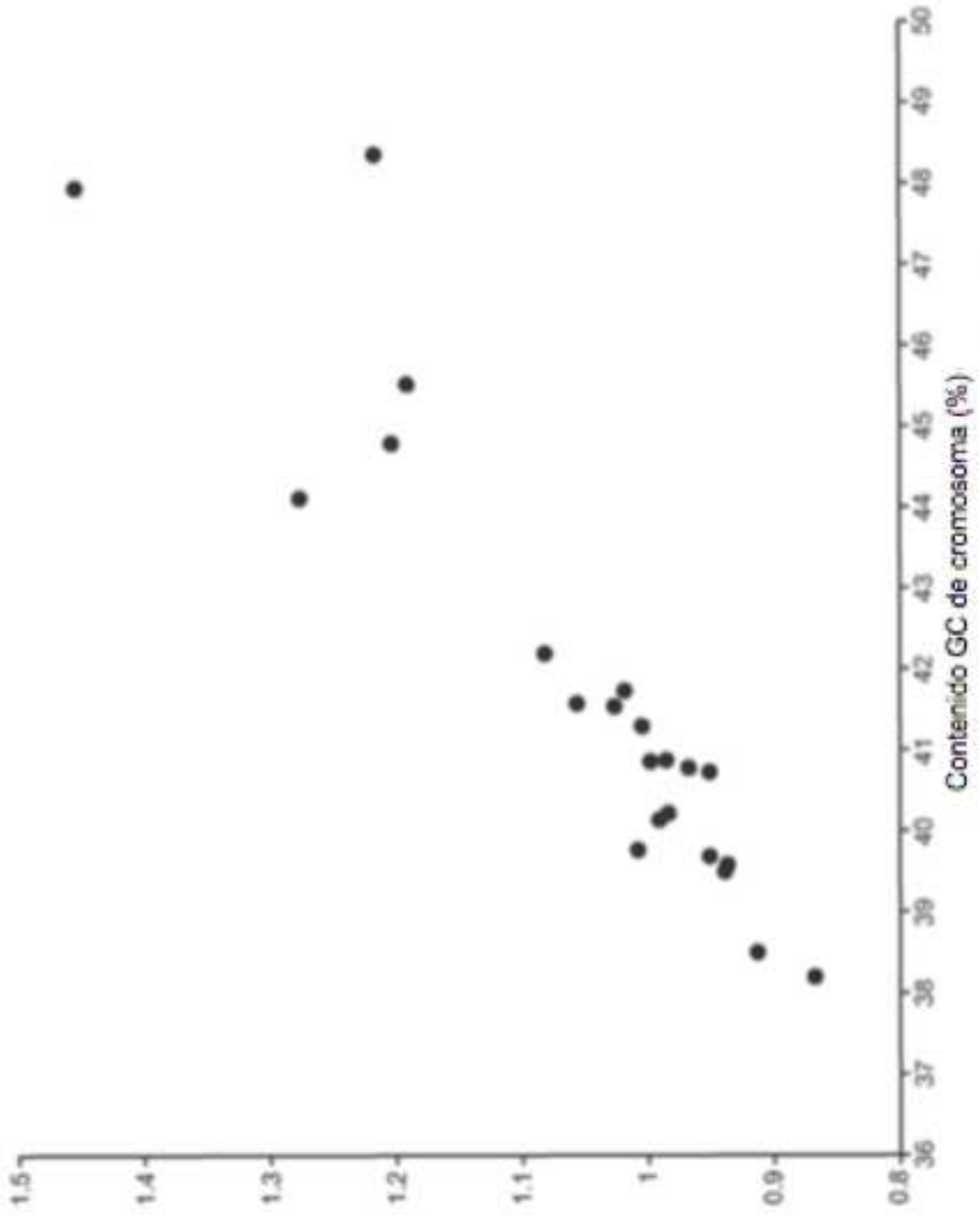
Contenido GC de cromosoma (%)
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
Densidad de etiqueta de cromosoma medio
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5

FIG. 5B

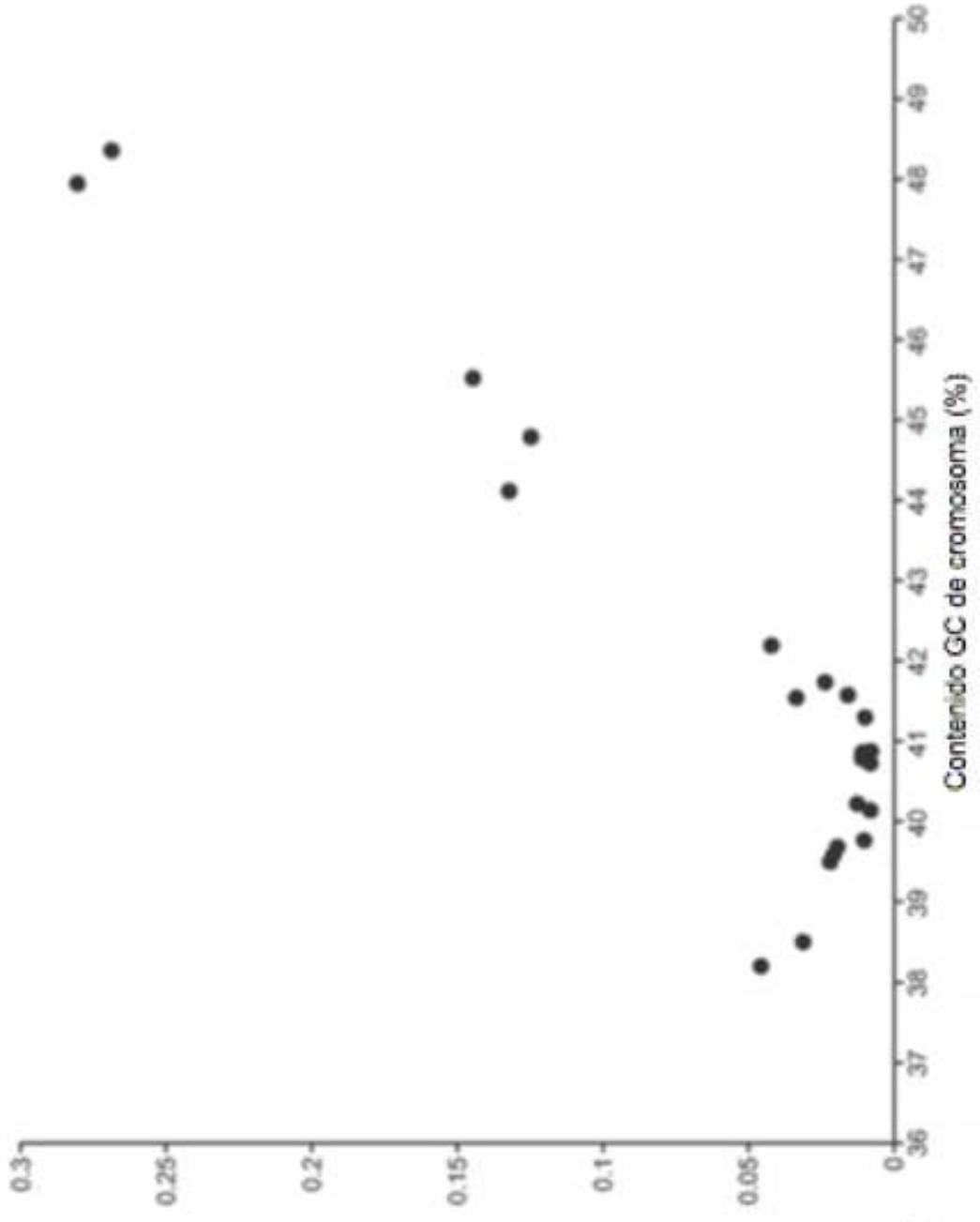
Contenido GC de cromosoma (%)
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
Desviación estándar de densidad de etiqueta de secuencia de cromosomas
0
0.05
0.1
0.15
0.2
0.25
0.3

FIG. 5C


Fetos masculinos
Fetos femeninos
ADN de plasma de varón adulto
0
-10
-20
-30
-40
-50
-60
% diferencia de densidad de etiqueta de secuencia chrX relativo a la de embarazos femeninos
Cromosoma X

FIG. 6

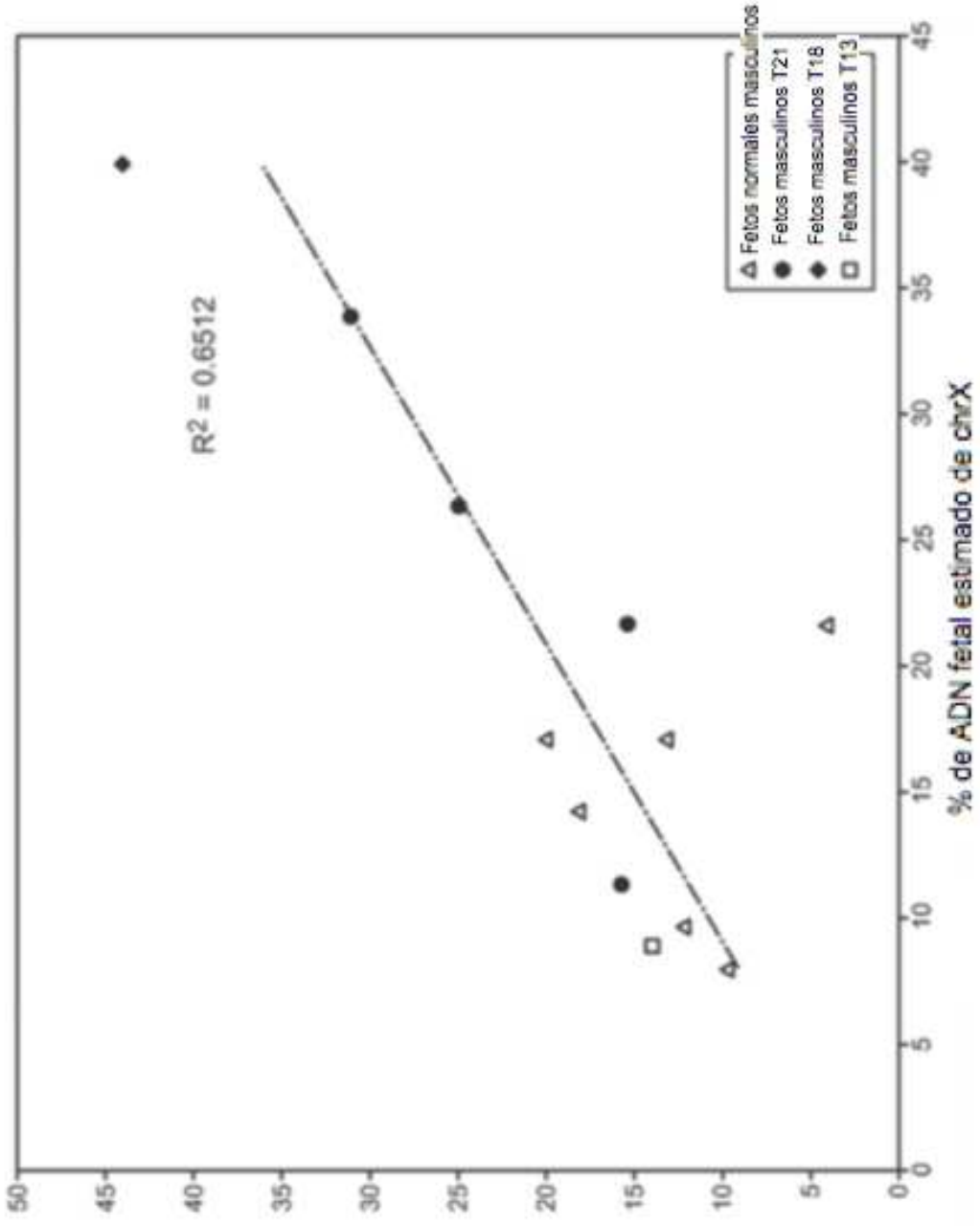
Fetos normales masculinos
Fetos masculinos T21
Fetos masculinos T18
Fetos masculinos T13
R² = 0.6512
% de ADN fetal estimado de chrY
% de ADN fetal estimado de chrX
0
5
10
15
20
25
30
35
40
45
50

FIG. 7

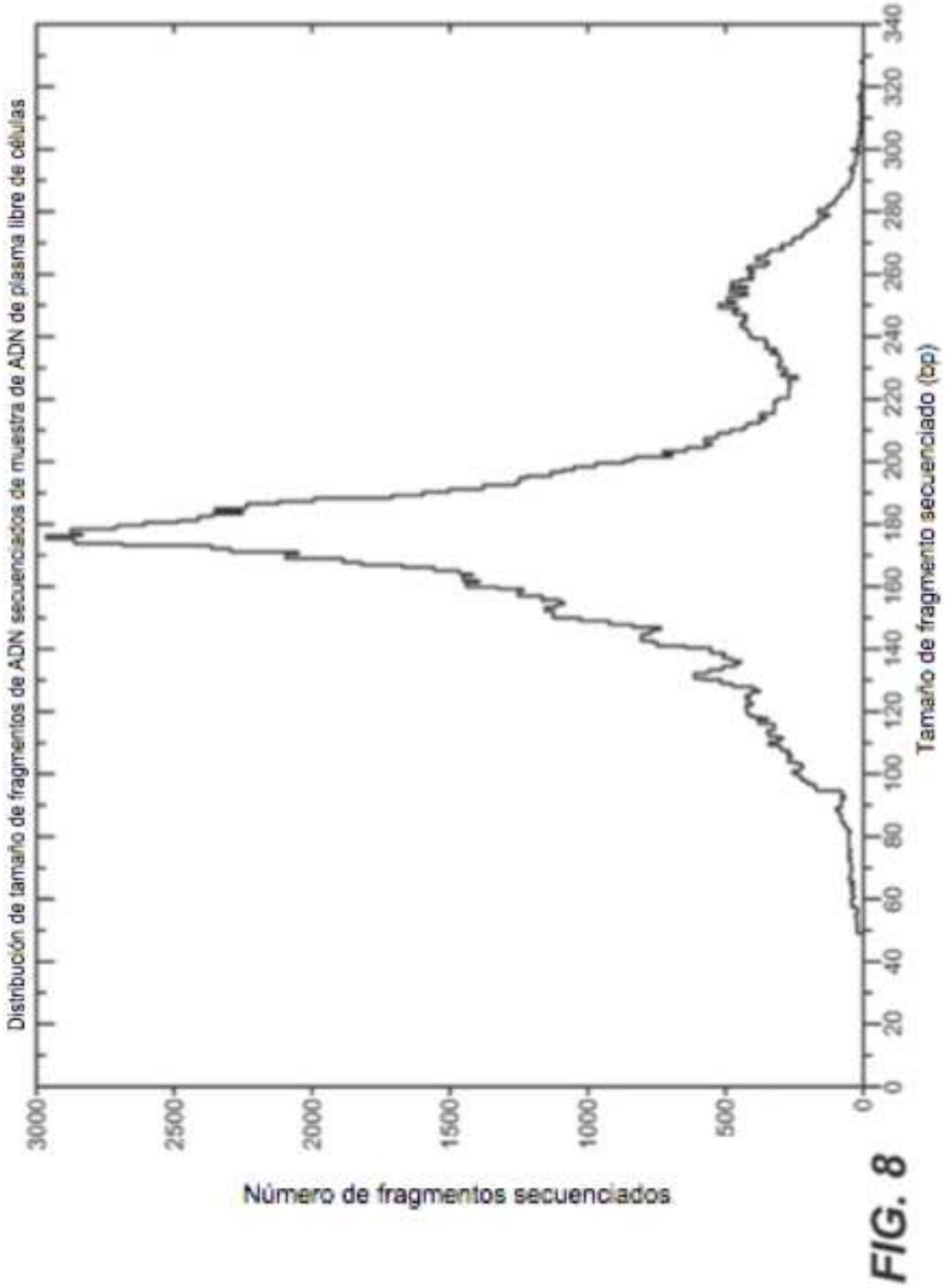
Distribución de tamaño de fragmentos de ADN secuenciados de muestra de ADN de plasma libre de células
Número de fragmentos secuenciados
Tamaño de fragmento secuenciado (bp)
0
500
1000
1500
2000
2500
3000
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340

FIG. 8

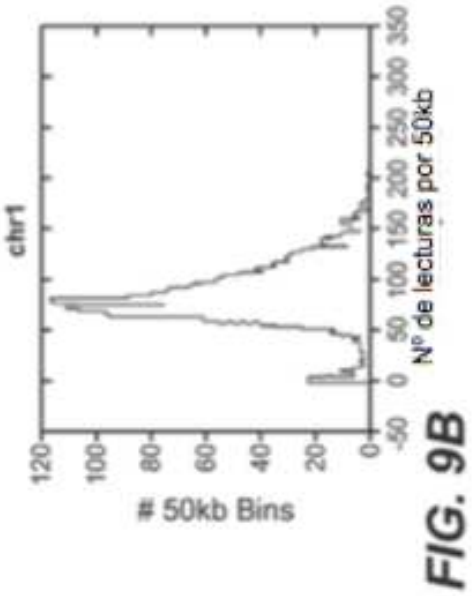

FIG. 9B

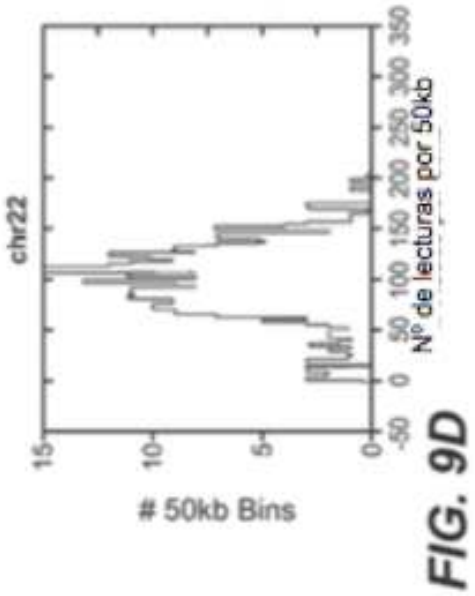

FIG. 9D

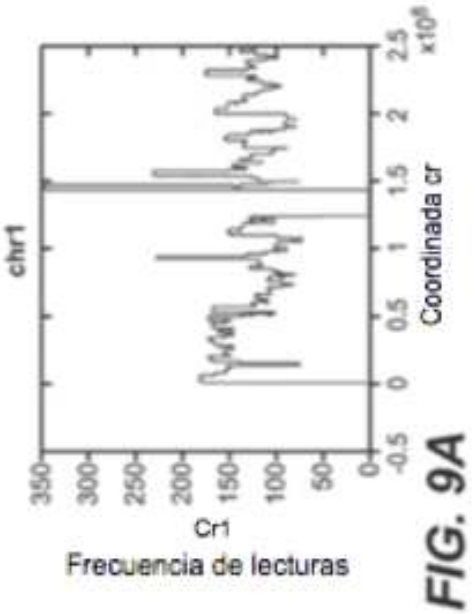

FIG. 9A

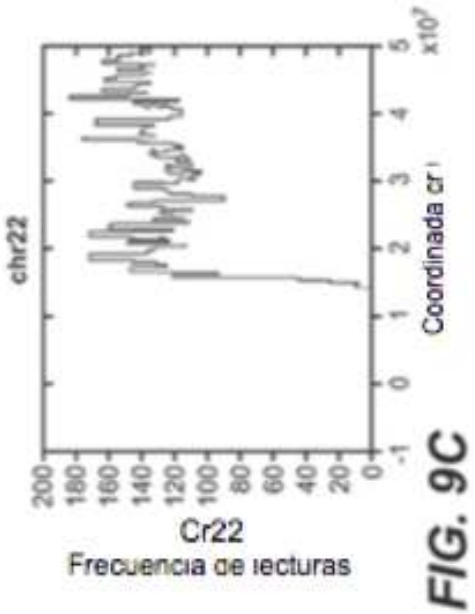

FIG. 9C

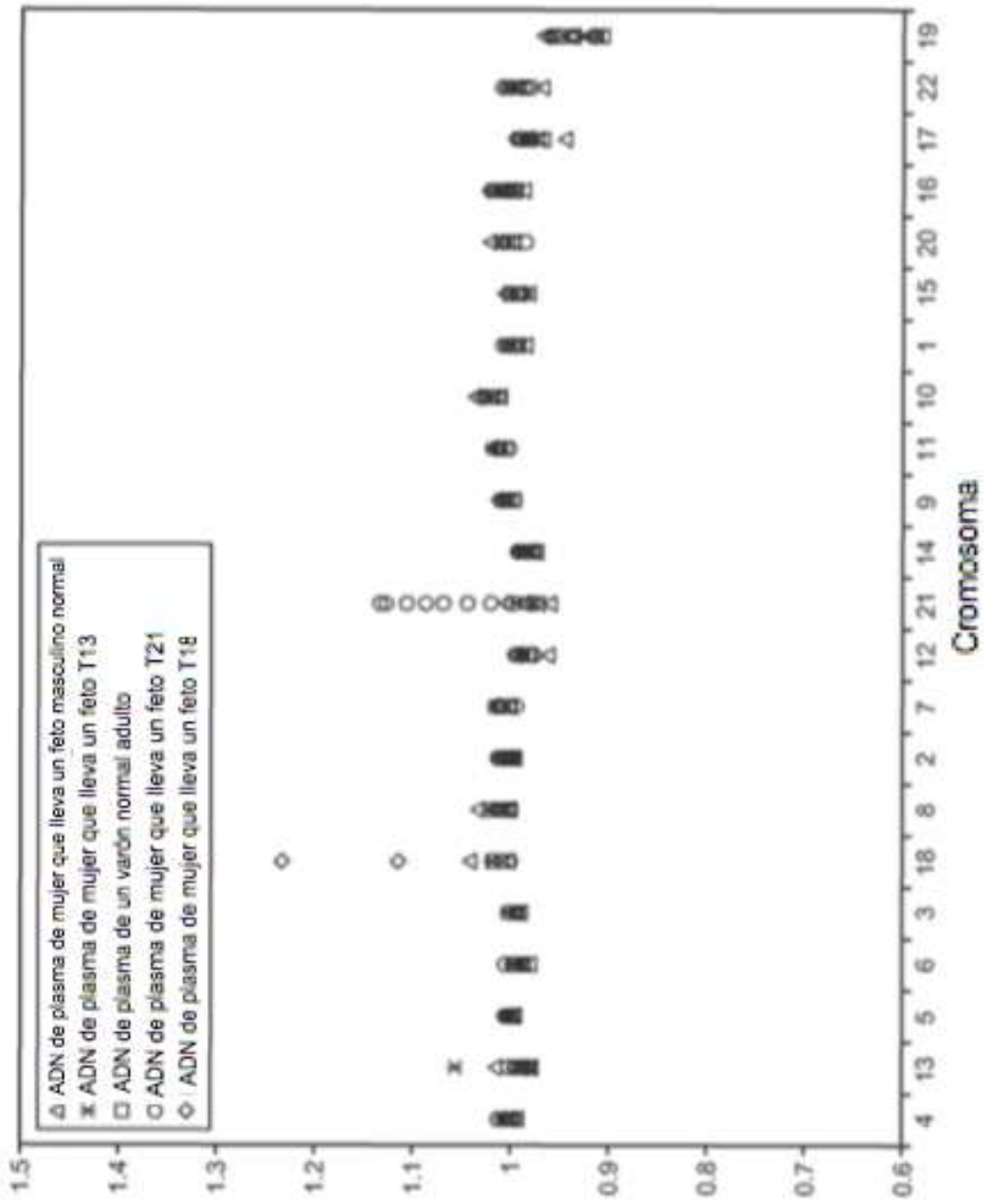
ADN de plasma de mujer que lleva un feto masculino normal
ADN de plasma de mujer que lleva un feto T13
ADN de plasma de un varón normal adulto
ADN de plasma de mujer que lleva un feto T21
ADN de plasma de mujer que lleva un feto T18
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
4 13 5 6 3 18 8 2 7 12 21 14 9 11 10 1 15 20 16 17 22 19
Cromosoma
Densidad de etiqueta de secuencia relativa al valor correspondiente del control de ADNg

FIG. 10

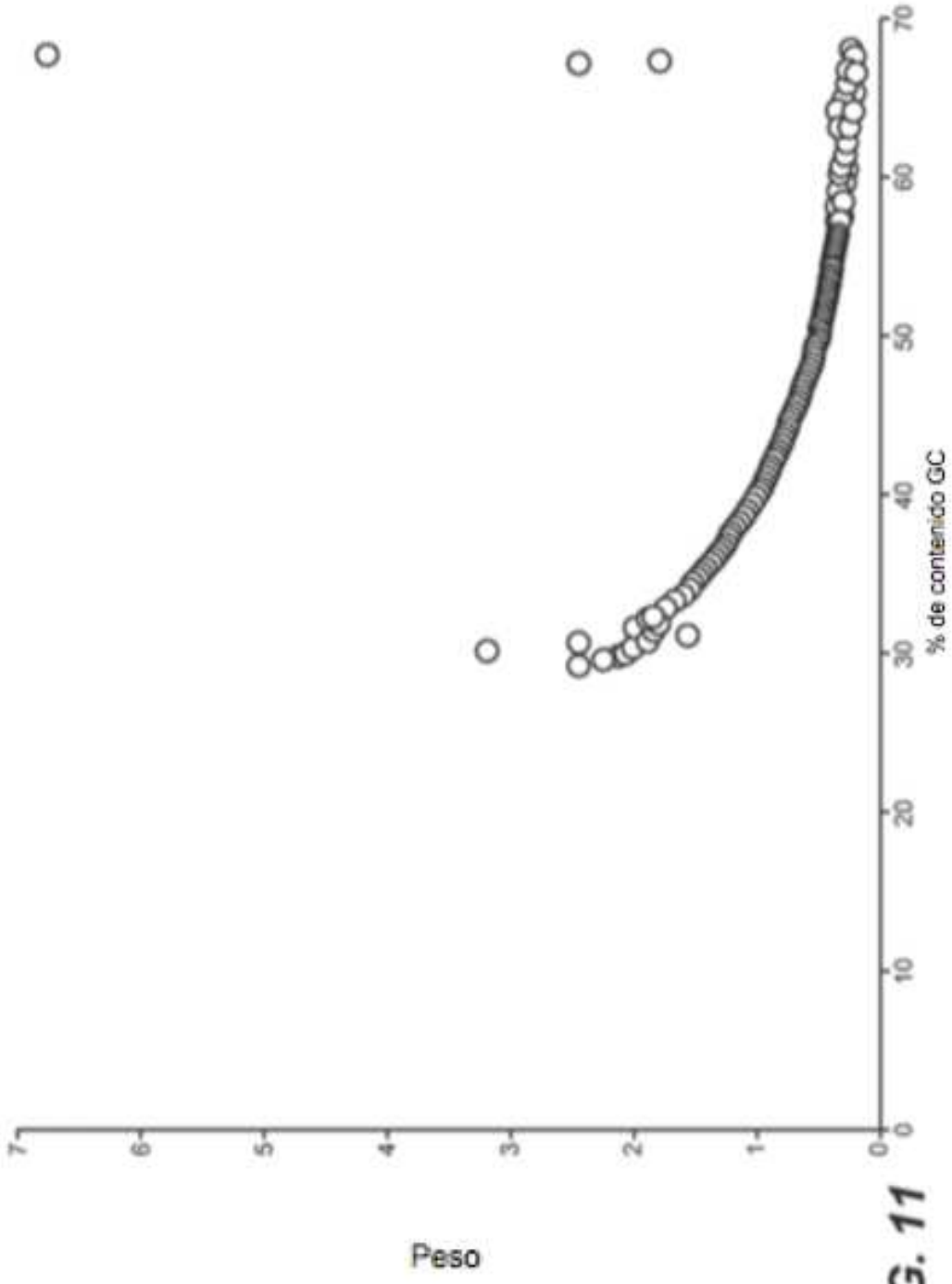

FIG. 11

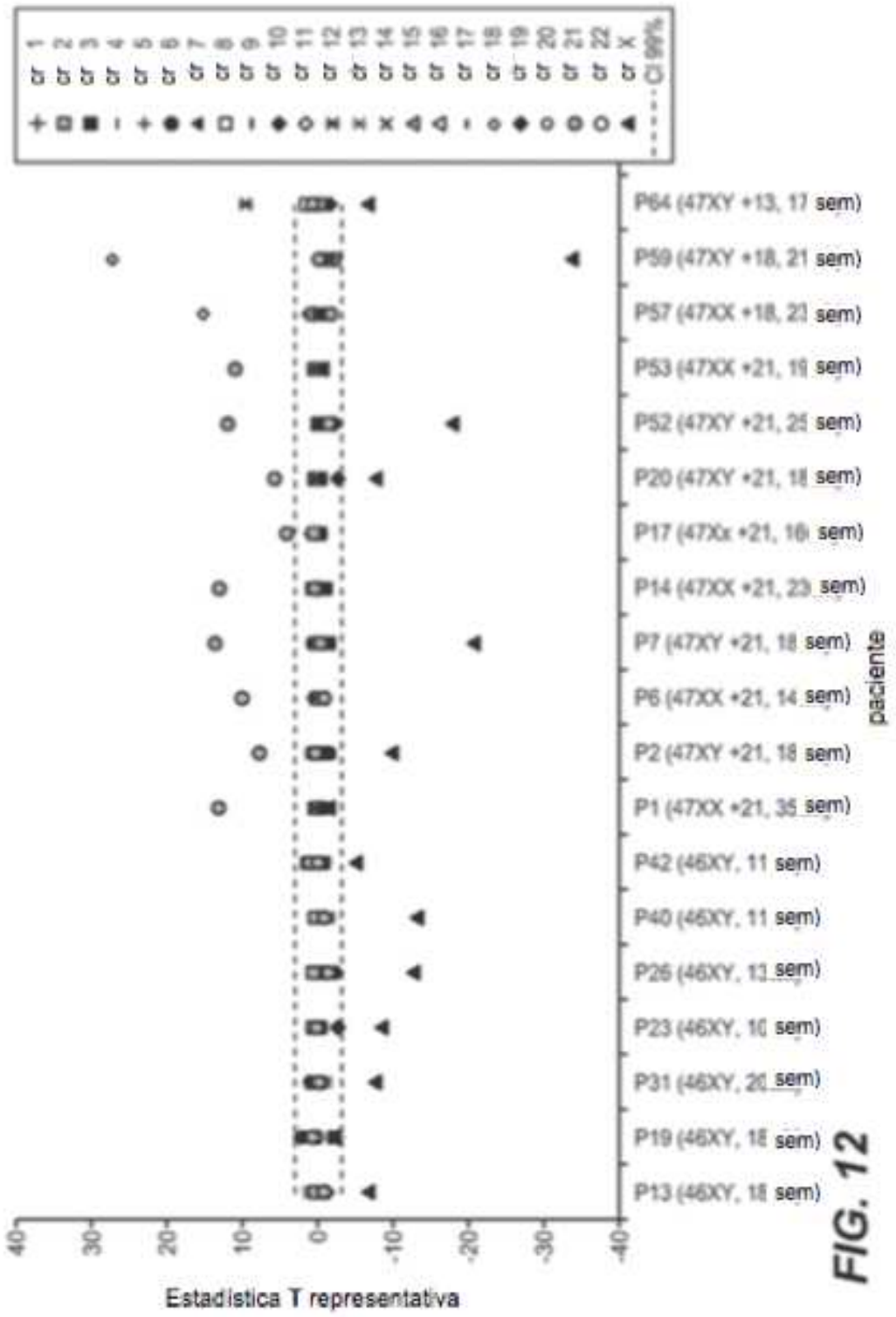
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
CI 99%
P64 (47XY +13, 17 sem)
P59 (47XY +18, 21 sem)
P57 (47XX +18, 2
P53 (47XX +21, 1
P52 (47XY +21, 2
P20 (47XY +21, 1
P17 (47Xx +21, 1
P14 (47XX +21, 2
P7 (47XY +21, 18 sem)
P6 (47XX +21, 14 sem)
P2 (47XY +21, 18 sem)
P1 (47XX +21, 35 sem)
P42 (46XY, 11 sem)
P40 (46XY, 11 sem)
P26 (46XY, 1
P23 (46XY, 1
P31 (46XY, 2
P19 (46XY, 1
P13 (46XY, 1
paciente
40
30
20
10
0
-10
-20
-30
-40
Estadística T representativa

FIG. 12

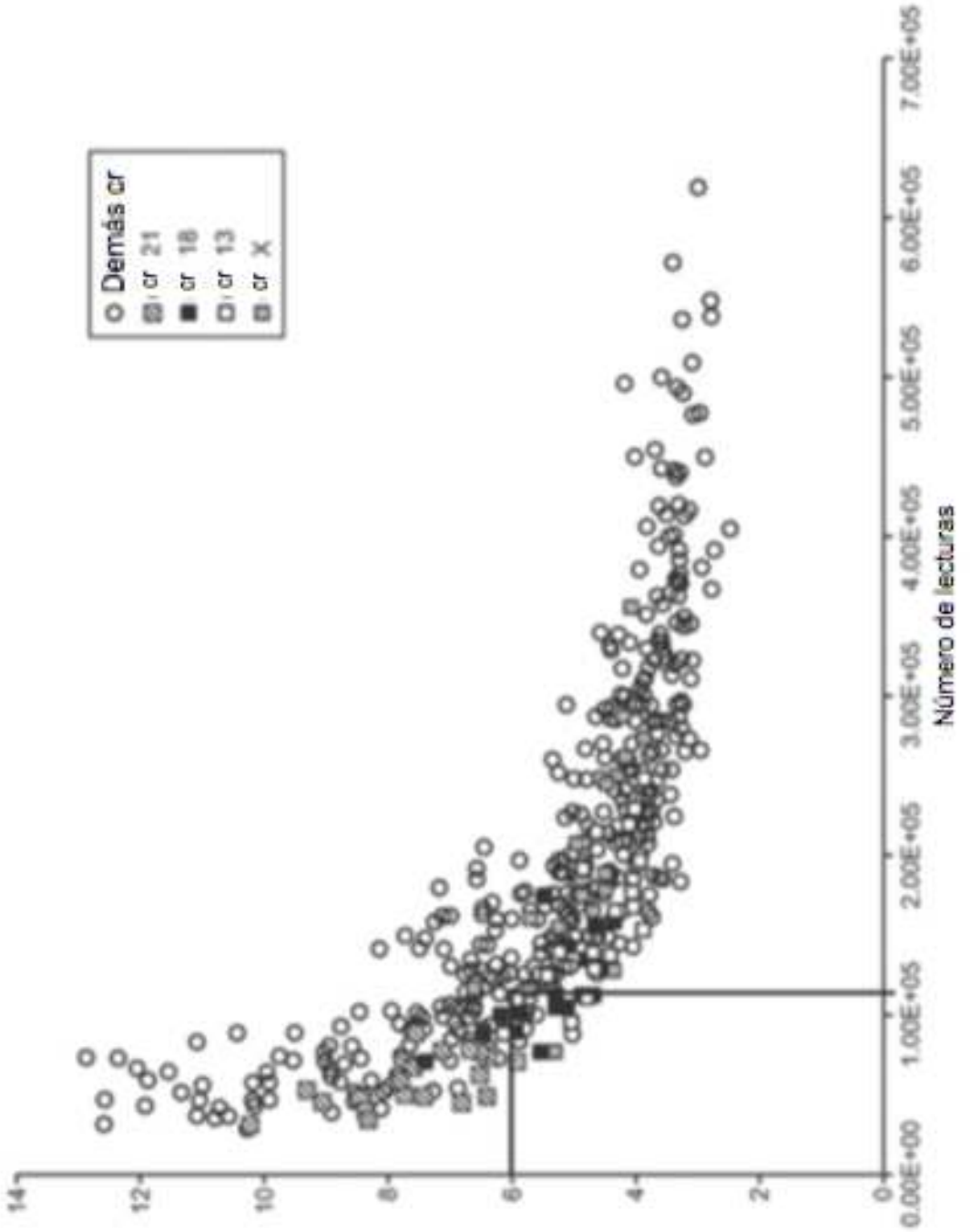
Demás cr
cr 21
cr 18
cr 13
cr X
0.00E+00
1.00E+05
2.00E+05
3.00E+05
4.00E+05
5.00E+05
6.00E+05
7.00E+05
Número de lecturas
0
2
4
6
8
10
12
14
% de ADN fetal mínimo cuya sobre o subrepresentación de cromosoma podría detectarse con 99,9% CI

FIG. 13

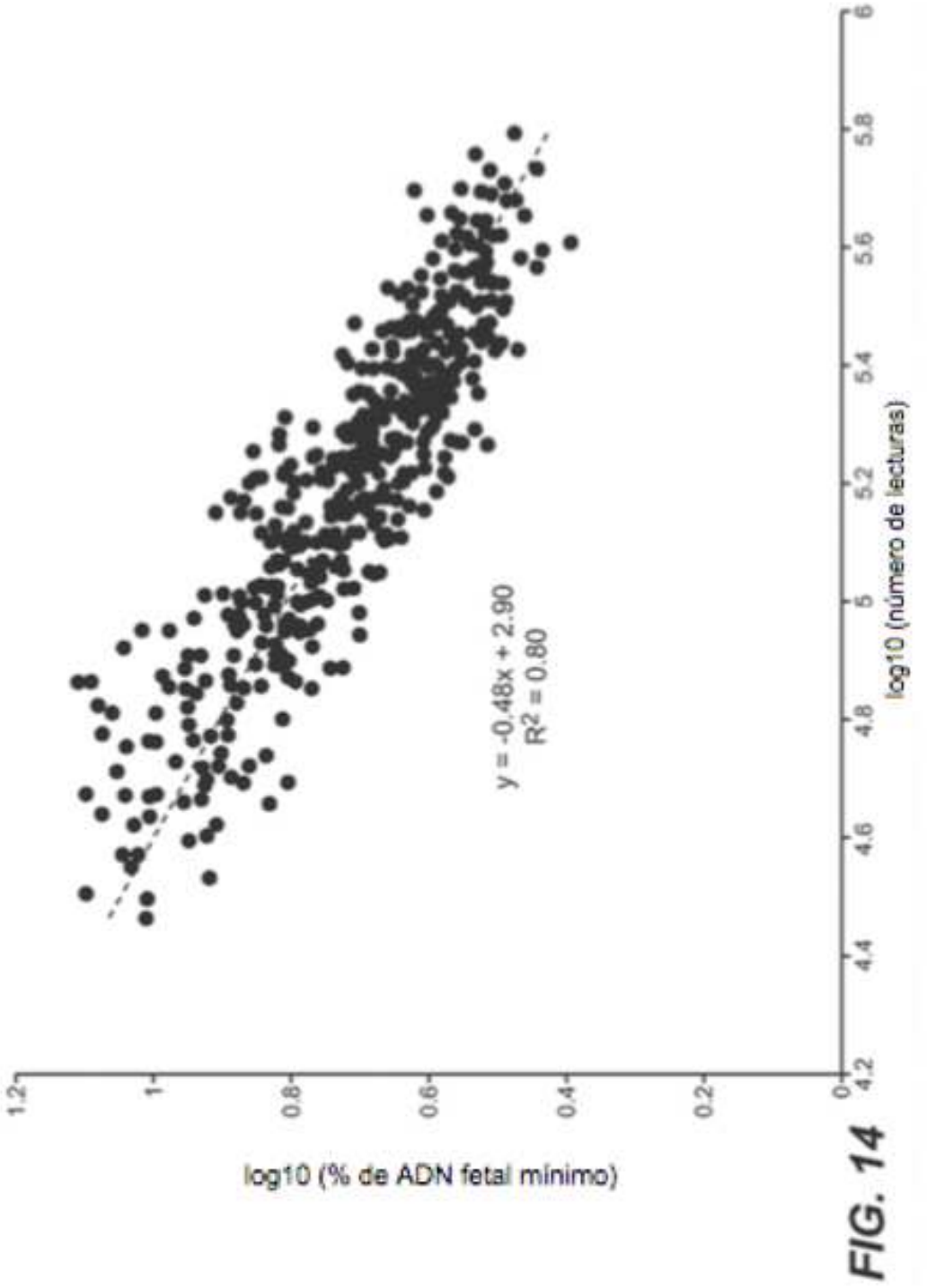
y = -0.48x + 2.90
R² = 0.80
log10 (% de ADN fetal mínimo)
log10 (número de lecturas)
1.2
1
0.8
0.6
0.4
0.2
0
4.2
4.4
4.6
4.8
5
5.2
5.4
5.6
5.8
6

FIG. 14