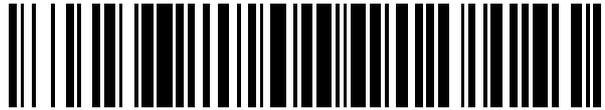


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 685 601**

51 Int. Cl.:

**G06F 19/18** (2011.01)

**G06F 19/26** (2011.01)

**G06F 19/28** (2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.08.2014 PCT/EP2014/068315**

87 Fecha y número de publicación internacional: **05.03.2015 WO15028576**

96 Fecha de presentación y número de la solicitud europea: **28.08.2014 E 14761819 (3)**

97 Fecha y número de publicación de la concesión europea: **25.07.2018 EP 3039597**

54 Título: **Haplotipado y tipado del número de copias mediante el uso de frecuencias alélicas de variantes polimórficas**

30 Prioridad:

**28.08.2013 US 201361871228 P**

**30.09.2013 US 201361960922 P**

**17.10.2013 GB 201318369**

**12.12.2013 US 201361915406 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**10.10.2018**

73 Titular/es:

**KATHOLIEKE UNIVERSITEIT LEUVEN (100.0%)  
K.U. Leuven R&D, Waaistraat 6, Box 5105  
3000 Leuven, BE**

72 Inventor/es:

**ZAMANI ESTEKI, MASOUD;  
VERMEESCH, JORIS y  
VOET, THIERRY**

74 Agente/Representante:

**ELZABURU, S.L.P**

ES 2 685 601 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Haplotipado y tipado del número de copias mediante el uso de frecuencias alélicas de variantes polimórficas

**Campo de la invención**

5 La presente invención se refiere a un método para el haplotipado y/o el tipado del número de copias de material genético. De manera más específica, la presente invención se refiere a un método para el haplotipado de todo el genoma, la determinación del perfil del número de copias de ADN específico del haplotipo, y la determinación de los orígenes meióticos/mitóticos de anomalías del ADN determinando, estableciendo la fase y segmentando fracciones alélicas de variantes polimórficas (PV) (PVAf) en células individuales, mezclas de pocas células, preparaciones de ADN multi-celulares, o incluso preparaciones de ADN exentas de células, p.ej. en el torrente sanguíneo.

**10 Antecedentes de la invención**

El desarrollo de ensayos de diagnóstico (de células individuales) dirigidos a una mutación particular para cada familia específicamente consume tiempo, es laborioso y costoso, lo que conduce además a largas listas de espera para las parejas que se someten al procedimiento. Por lo tanto, son imprescindibles nuevos métodos genéricos para el diagnóstico genético.

15 La aneuploidía cromosómica es una causa importante de aborto espontáneo y desarrollo anormal de un feto o individuo. Tales aneuploidías pueden resultar de errores meióticos, que son más prevalentes en la ovogénesis en la década que precede a la menopausia. El cribado genético preimplantacional (PGS) se ha conceptualizado para incrementar las tasas de embarazo por embrión transferido y para prevenir embarazos anormales y partos de neonatos enfermos tras IVF. Sin embargo, los métodos de PGS basados en FISH y aCGH no permiten distinguir  
20 entre errores meióticos o mitóticos. En particular, las divisiones celulares en la etapa de segmentación son proclives a errores en la segregación cromosómica mitótica, lo que no impide necesariamente un desarrollo embrionario normal.

El tipado de SNP de todo el genoma, basado en micromatrices o secuenciación de última generación, y el haplotipado posterior de todo el genoma, se ha hecho viable recientemente. Sin embargo, se ha visto impedido en  
25 gran medida el haplotipado preciso de todo el genoma de una célula individual diploide que no está en metafase, principalmente debido a errores del genotipo introducidos por artefactos de la amplificación del ADN de una célula individual como ADO aleatoria y la amplificación preferente (PA) de un alelo respecto del otro, así como por interpretaciones algorítmicas falsas de las intensidades de las sondas de SNP.

30 Wang et al. en "*PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data.*" *Genome research* 17, 1665 (2007), desarrolló un modelo oculto de Markov integrado, denominado PennCNV, para determinar las anomalías del número de copias. PennCNV usa los valores de la frecuencia del alelo B (BAF) de SNP y de logR para determinar las anomalías del número de copias. Esto funciona bien para muestras multi-celulares que no se han sometido a amplificación de todo el genoma (WGA), ya que hay patrones específicos de valores de BAF para diferentes anomalías del número de  
35 copias. Sin embargo, los valores de PVAf obtenidos de una amplificación de todo el genoma (WGA) de una célula individual puede estar significativamente distorsionados debido al sesgo en la amplificación alélica, y por lo tanto no son tan distintivos para los diferentes estados del número de copias como los valores de PVAf obtenidos de una muestra de ADN no sometida a WGA. En particular, las duplicaciones y trisomías son extremadamente difíciles de confirmar a partir de los valores de PVAf y de logR habituales en células individuales; lo que podría conducir a interpretaciones erróneas en el análisis del número de copias de células individuales, y por lo tanto incluso a diagnósticos erróneos al aplicarlo en un ámbito clínico. PennCNV, y otras aproximaciones similares conocidas en la técnica, también pueden emplear tres genotipos para determinar el origen de las anomalías en muestras de ADN obtenidas de una gran cantidad de células. Aunque estas aproximaciones funcionan bien para detectar y confirmar deleciones, mediante el uso de genotipos bi-alélicos discretos de una muestra de ADN multi-celular, no se  
40 puede determinar con exactitud ni el origen parental ni el origen mecánico de las anomalías, en particular duplicaciones. En las realizaciones de la presente invención, sin embargo, los valores de PVAf informativos se haplotipan / se determina la fase y se utilizan de manera ventajosa para la determinación de los estados del número de copias específico del haplotipo, así como su origen parental y mecánico, hasta el nivel de una célula individual.

Además de estos métodos conocidos en la técnica, también hay métodos basados en la población conocidos en la  
50 técnica para la determinación de desequilibrios alélicos y mosaicismos en muestras de ADN multi-celulares no sometidas a WGA. Vattathil et al. en "*Haplotype-based profiling of subtle allelic imbalance with SNP arrays*", *Genome research* 23, 152, (2013), usa los valores de frecuencia del alelo B (BAF) de asignaciones de SNP heterocigotos para detectar la alteración de la proporción de alelos normal 1:1. Tras la estimación estadística basada en poblaciones de los haplotipos de la línea germinal mediante el uso de fastPHASE como describió Scheet et al en "*A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase.*" en *American journal of human genetics* 78, 629 (2006), compararon estos haplotipos de la línea germinal estimados y los haplotipos "en exceso" deducidos a partir de los valores de BAF de los mismos SNPs heterocigotos que sobrepasaban un umbral (es decir, la mediana de las BAFs observadas en todos los loci  
55

heterocigotos) para determinar la concordancia de fase entre el haplotipo de la línea germinal y el haplotipo "en exceso" deducido de BAF, y así los desequilibrios alélicos específicos del haplotipo. Sin embargo, no consideran la magnitud de las BAFs por sí mismas, e ignoran las intensidades totales. En contraste, Nik-Zainal et al. en "*The life history of 21 breast cancers.*" Cell, 149, 5, 2012, aplicó elegantemente los datos del proyecto 1000 Genomas ("*A map of human genome variation from population-scale sequencing.*" Nature 467, 1061, 2010) y usó IMPUTE (Howie et al. en "*Fast and accurate genotype imputation in genome-wide association studies through pre-phasing.*" Nature Genetics 44, 8, 2012) para determinar la fase de SNPs de la línea germinal a partir de datos de secuenciación de última generación en bloques de haplotipos específicos de progenitor fragmentados y por tanto cortos. El análisis posterior de las proporciones alélicas por haplotipo mostró una sensibilidad mayor para detectar desequilibrios alélicos y anomalías que al puntuar BAFs de SNP individuales en todo el genoma. Aplicando este principio (denominado algoritmo de Battenberg) en valores de BAF de asignaciones de SNP heterocigotos, pudieron determinar la distorsión de los valores de BAF de asignaciones de SNP heterocigotos desde el valor de 0,5 esperado para determinar desequilibrios alélicos en tramos de haplotipos cortos, y por lo tanto fue posible investigar las anomalías del número de copias de ADN en mosaico en una muestra de ADN. De manera importante, sin embargo, se ha demostrado que los dos métodos anteriores conocidos en la técnica funcionan en muestras de ADN estándar extraídas de una gran cantidad de células, pero serán ineficaces con valores de BAF obtenidos de datos de células individuales, ya que los haplotipos de la línea germinal determinados en una población representan tramos cortos, y los procesos de amplificación de todo el genoma necesarios para el análisis de células individuales introducen ruido en la BAF de SNP. Además, no se pueden revelar los sitios de recombinación homóloga parentales en la muestra de células individuales. En contraste, las realizaciones de la presente invención aplican de manera ventajosa principios de determinación de la fase basados en la familia o familiares, y en particular no necesita la determinación de las asignaciones de SNP de la muestra de ADN en estudio por sí misma. Por lo tanto, el método según las realizaciones de la invención puede interpretar de manera ventajosa los valores de PVAF en haplotipos de la línea germinal amplios, y se puede usar para interpretar los valores de PVAF de una célula individual, unas pocas células o muestras de ADN que requieren amplificaciones de todo el genoma, además de muestras de ADN estándar. El método según las realizaciones de la invención aplica de manera ventajosa los genotipos de PV discretos robustos de los progenitores, y en realizaciones específicas de un familiar cercano adicional para determinar la fase de los genotipos parentales, que se aplican posteriormente al estudio de los valores de PVAF de una muestra de ADN de SNPs informativos paternos y maternos, respectivamente, de una manera específica de haplotipo. El método según las realizaciones de la invención usa bloques de haplotipos parentales largos en los que se puede tener en cuenta la influencia de los artefactos de WGA estocásticos. Además, estos métodos basados en la población no pueden trazar de manera eficaz el origen mecanístico de las anomalías genómicas. Sin embargo, la presente invención puede trazar de manera ventajosa las anomalías genómicas hasta errores de la meiosis o mitosis. Esto tiene implicaciones importantes en el diagnóstico genético preimplantacional (PGD), en particular cuando el diagnóstico se realiza en la etapa de segmentación en el desarrollo temprano. Si una célula individual de un embrión preimplantacional tiene una anomalía de origen mecanístico meiótico, ese embrión no se debería transferir, ya que la anomalía se perpetúa muy probablemente en todos los blastómeros individuales del embrión. Sin embargo, si un embrión en la etapa de segmentación tiene una anomalía de origen mecanístico mitótico, esto no significaría necesariamente que la anomalía esté presente en todos los blastómeros individuales de ese embrión, ya que la inestabilidad cromosómica es común en la embriogénesis en la etapa de segmentación. Así, la determinación del origen mecanístico meiótico o mitótico de las anomalías cromosómicas mediante la presente invención posibilita el diagnóstico genético preimplantacional para el cribado de aneuploidías (PGS) en embriones en la etapa de segmentación. La presente invención, por lo tanto, hace viable el PGD y PGS simultáneo en un ensayo.

Navin et al. en "*Tumor evolution inferred by single-cell sequencing*" Nature 472, 90 (2011), desarrolló un método de análisis de profundidad de lectura de secuencia focal para calcular el gráfico del número de copias de ADN de células individuales tras la secuenciación de un producto de WGA de una célula individual. Se calculó la cantidad de lecturas de secuencias de un único extremo que correspondieron a regiones pequeñas específicas para determinar los valores de logR y los estados del número de copias subsiguientes de genomas de células individuales sometidas a WGA. De manera ventajosa, las realizaciones de la presente invención usan genotipos de PV, valores de logR y PVAF obtenidos de tecnologías de genotipado de alto rendimiento, que incluyen matrices de SNP y dispositivos de secuenciación de última generación, y valores de logR y PVAF integrados para mostrar las anomalías genómicas. Esta última realización funciona bien para la determinación de una delección, sin embargo los valores habituales de PVAF de células individuales no son robustos para la confirmación de duplicaciones. En comparación con la última realización, otras realizaciones de la invención categorizan y subcategorizan los valores de PVAF, seguido de la segmentación de estos valores. Así, las realizaciones del método de la invención pueden reconstruir de manera ventajosa los haplotipos parentales y determinan el origen mecanístico meiótico o mitótico de las anomalías.

Para el haplotipado de células individuales o cantidades bajas de células tras WGA, los métodos de la técnica anterior pueden utilizar genotipos bi-alélicos discretos de la célula. En estas aproximaciones, no es posible la determinación de las cantidades exactas de alelos y haplotipos y su origen. Por ejemplo, estos métodos conocidos en la técnica no pueden distinguir con exactitud un cromosoma diploide de una trisomía de origen mitótico. Además, cuando hay presentes mezclas de células en una muestra de ADN de pocas células o multi-celular, no se puede analizar la naturaleza de mosaico de la muestra de ADN. Además, debido a que estas aproximaciones utilizan genotipos bi-alélicos, pueden experimentar seriamente artefactos de la WGA, así como variantes auténticas del

número de copias de ADN en la muestra. Para aliviar los artefactos por pérdida de alelos en WGA hasta cierto punto, Handyside et al. en "*Karyomapping: a universal method for genome wide analysis of genetic disease based on mapping crossovers between parental haplotypes.*" J Med Genet, 47, 10, (2010), propuso el uso de genotipos heterocigotos solamente. Aunque esta aproximación puede hacer desaparecer la influencia de los artefactos ADO, no alivia los falsos genotipos heterocigotos que se generan por los artefactos por ganancia de alelos (ADI). Debido a que los SNPs heterocigotos en un producto de WGA solamente constituyen una proporción pequeña de un genotipo (de célula individual), los artefactos por ADI menores podrían tener un gran efecto, que conduciría a haplotipos falsos. Voet et al. en "*Single-cell paired-end genome sequencing reveals structural variation per cell cycle.*" Nucleic Acids Res, 41, 12, (2013), describe métodos para analizar el material genético de neuronas para identificar neuronas aneuploides así como variaciones del número de copias subcromosómicas, que comprende obtener datos de secuencia mediante amplificación de todo el genoma de células individuales y calcular la fracción de alelo B (BAF) de SNPs como el número de lecturas que incorporan el alelo B del SNP dividido por el número de lecturas que abarcan el SNP. Los valores de BAF son una realización de los valores de PVAF. Los autores fueron capaces de detectar los desequilibrios del ADN que resultaron de una herencia desequilibrada del cromosoma 16 de origen paterno y el agrupamiento coincidente de pares de lectura con una cartografía discordante. La determinación del perfil del número de copias de células individuales se llevó a cabo normalizando y segmentando los valores de logR, y ajustando una función constante por partes a los datos. Los cambios de logR también se usaron para deducir los puntos de ruptura aproximados.

El método según las realizaciones de la invención es ventajoso respecto de estos métodos de la técnica anterior, ya que se usan valores de PVAF informativos continuos. Esto tiene varias ventajas, que incluyen una sensibilidad mayor para los haplotipos reconstruidos debido a que la aplicación de los valores de PVAF y la segmentación posterior alivia el efecto de los artefactos de WGA estocásticos, que incluyen ADO incompleta, ADI, amplificación preferente (PA), y tiene en cuenta las variantes del número de copias genuinas en la muestra; la naturaleza de mosaico de las muestras de ADN de pocas células o multi-celulares se puede detectar y analizar; no solamente se pueden distinguir las trisomías y disomías mitóticas, sino que también se puede determinar la naturaleza meiótica y mitótica de las anomalías, y además se pueden detectar eventos sin cambios en el número de copias tales como UPhD y UPiD; además se pueden detectar zonas de LOH, que podrían ser el resultado de la consanguinidad, en cada uno de los progenitores o el ADN del individuo en estudio.

En particular, los métodos de la técnica anterior no permiten determinar (de manera simultánea) el número de copias y el haplotipo mediante el uso de las frecuencias alélicas de variantes polimórficas continuas. Esto es especialmente complicado al usar datos de genotipado con ruido obtenidos de muestras que comprenden cantidades bajas de material genético, tales como las muestras de células individuales.

Todavía existe la necesidad de métodos mejorados para el haplotipado y/o el tipado del número de copias.

### Sumario de la invención

Un objetivo de la presente invención es proporcionar un método implementado por ordenador para el análisis del material genético de un sujeto, y dicho método comprende:

- obtener valores de frecuencias de alelos de variantes polimórficas (PVAF) continuos del material genético del sujeto;
- obtener información del genotipo en fase de un primer progenitor e información del genotipo en fase o no en fase de un segundo progenitor;
- categorizar los valores de PVAF continuos en una primera categoría que corresponde al primer progenitor basándose en la información del genotipo del primer y segundo progenitor;
- subcategorizar los valores de PVAF continuos de la primera categoría que corresponde al primer progenitor en subcategorías;
- segmentar dichos valores de PVAF subcategorizados; y
- proporcionar los valores de PVAF segmentados para indicar una anomalía genética en el material genético del sujeto y/o la herencia del material genético del sujeto.

También se proporciona un producto de programa informático del mismo, y un medio de almacenamiento legible por ordenador no transitorio que almacena dicho producto de programa informático.

Una ventaja de la presente invención es no usar asignaciones de PV discretas de una muestra de ADN por sí misma. Esto es importante para las muestras de ADN de células individuales o de pocas células, ya que el método de la invención alivia las asignaciones de PV discretas erróneas que resultan de artefactos de WGA y/o la interpretación algorítmica de las señales determinadas con el material de WGA.

Una ventaja de la presente invención es que tiene dos características inherentes: (1) característica de paridad dentro

de cada perfil parental y (2) característica de complementariedad entre perfiles parentales.

Una ventaja de las realizaciones de la presente invención es proporcionar un método mejorado, basado en fracciones de alelos de variantes polimórficas, para el haplotipo de todo el genoma, perfil del número de copias y determinar los orígenes mecanísticos de las anomalías del ADN en muestras de ADN obtenidas de células individuales o multi-celulares.

5 Este objetivo se cumple mediante el método según las reivindicaciones independientes de la presente invención. Las reivindicaciones dependientes se refieren a las realizaciones preferidas.

En un primer aspecto, la presente descripción proporciona un método para el análisis del material genético de un sujeto, y dicho método comprende:

- 10 - obtener valores de frecuencias de alelos de variantes polimórficas (PVAf) continuos del material genético de un sujeto;
- obtener información del genotipo de un primer progenitor;
- categorizar los valores de PVAf continuos en una categoría que corresponde al primer progenitor basándose en la información del genotipo del primer progenitor;
- 15 - segmentar dichos valores de PVAf categorizados; y
- proporcionar los valores de PVAf segmentados para indicar una anomalía genética en el material genético del sujeto y/o la herencia del material genético del sujeto.

En un aspecto relacionado, la presente descripción proporciona un método para el análisis del material genético de un sujeto, y dicho método comprende:

- 20 - obtener valores de frecuencias de alelos de variantes polimórficas (PVAf) continuos del material genético de un sujeto;
- obtener información del genotipo de un primer y segundo progenitor;
- categorizar los valores de PVAf continuos en una categoría que corresponde al primer progenitor basándose en la información del genotipo del primer progenitor y segundo progenitor;
- 25 - segmentar dichos valores de PVAf categorizados; y
- proporcionar los valores de PVAf segmentados para indicar una anomalía genética en el material genético del sujeto y/o la herencia del material genético del sujeto.

En un segundo aspecto, la presente descripción proporciona un método que comprende:

- 30 - obtener valores de frecuencias de alelos de variantes polimórficas (PVAf) continuos del material genético de un sujeto;
- obtener información del genotipo en fase de un primer progenitor e información del genotipo en fase o no en fase de un segundo progenitor;
- categorizar los valores de PVAf continuos en una categoría que corresponde al primer progenitor basándose en la información del genotipo del primer y segundo progenitor;
- 35 - subcategorizar los valores de PVAf continuos de la categoría que corresponde al primer progenitor en subcategorías;
- segmentar dichos valores de PVAf subcategorizados; y
- proporcionar los valores de PVAf segmentados para indicar una anomalía genética en el material genético del sujeto y/o la herencia del material genético del sujeto.

40 En un tercer aspecto, la presente descripción proporciona un método que comprende:

- obtener valores de frecuencias de alelos de variantes polimórficas (PVAf) continuos del material genético del sujeto;
- obtener información del genotipo en fase de un primer progenitor e información del genotipo en fase de un segundo progenitor;
- 45 - categorizar los valores de PVAf continuos en una primera categoría que corresponde al primer progenitor y una

segunda categoría que corresponde al segundo progenitor basándose en la información del genotipo del primer y segundo progenitor;

- subcategorizar los valores de PVAF continuos en la primera y segunda categorías en subcategorías;
- segmentar dichos valores de PVAF subcategorizados; y

- 5 - proporcionar los valores de PVAF segmentados para indicar una anomalía genética en el material genético del sujeto y/o la herencia del material genético del sujeto.

En un aspecto adicional, la presente descripción proporciona métodos para el haplotipado y/o tipado del número de copias del material genético de una muestra, y dicho método comprende:

- 10 - determinar los valores de frecuencias de alelos de variantes polimórficas (PVAF) continuos del material genético de la muestra;
- proporcionar genotipos de variantes polimórficas (PV) parentales en fase mediante el uso de genotipos de un familiar cercano o el genotipo de la propia muestra;
- 15 - categorizar los valores de PVAF continuos determinados del material genético de la muestra mediante el uso de los genotipos de PV parentales proporcionados (en fase o no en fase), lo que da como resultado valores de PVAF categorizados;
- subcategorizar dichos valores de PVAF categorizados de la muestra en subcategorías mediante el uso de los genotipos de PV en fase parentales proporcionados;
- segmentar dichos valores de PVAF subcategorizados, lo que da como resultado patrones de PVAF haplotipados / en fase y segmentados.

- 20 En las realizaciones preferidas de la invención, el material genético de la muestra se puede obtener de una célula individual, unas pocas células, un gran número de células o ADN exento de células. Se han determinado valores de PVAF continuos mediante el uso de una muestra que comprende material genético del sujeto. En una realización particular, dicha muestra comprende una cantidad baja del material genético de dicho sujeto, tal como una muestra que comprende solamente una célula o unas pocas células de dicho sujeto, o una muestra de plasma obtenido de una madre embarazada de dicho sujeto. En otra realización particular, se han determinado valores de PVAF continuos mediante el uso de una tecnología de matrices (en particular de todo el genoma) o de secuenciación, especialmente las tecnologías de matrices y de secuenciación descritas en la presente memoria. En otra realización particular, se han determinado valores de PVAF continuos mediante el uso de una muestra amplificada de todo el genoma (WGA).
- 25

- 30 En las realizaciones preferidas de la invención, una variante polimórfica puede ser cualquier variante genética que tenga al menos una forma alternativa en comparación con una secuencia de referencia. En una realización particular, dicha variante polimórfica es un polimorfismo de nucleótido simple (SNP). En una realización adicional, dichos valores de PVAF son las frecuencias del alelo B (BAFs).

- 35 En las realizaciones preferidas de la invención, el método puede comprender además normalizar los valores de la cantidad de ADN (tal como valores de recuentos de lecturas o de logR) del material genético según los valores de PVAF haplotipados / en fase y segmentados, también denominados en la presente memoria patrones de PVAF (segmentados). En particular, el método comprende obtener valores de cantidad de ADN y normalizar dichos valores de cantidad de ADN basándose en dichos valores de PVAF segmentados. Preferiblemente, los valores de la cantidad de ADN son valores de logR o valores de recuentos de lecturas.

- 40 En las realizaciones preferidas de la invención, proporcionar dichos genotipos de PV parentales en fase comprende:
- (a) determinar la fase de los genotipos de PV parentales basándose en los genotipos de un familiar (cercano), es decir, en un progenitor heterocigoto en dos loci sinténicos, la designación de qué alelo del primer locus está en el mismo cromosoma que el alelo del segundo locus.

- 45 En una realización particular, el método de la invención comprende además la reflexión de los valores de PVAF obtenidos respecto del eje medio antes de la segmentación. En una realización preferida adicional, el método de la invención comprende la reflexión de los valores de PVAF obtenidos, en una posición que corresponde a un genotipo parental en fase específico, respecto del eje medio antes de la segmentación. La reflexión de los valores de PVAF proporciona un beneficio, ya que mejora la segmentación (más valores de PVAF están presentes alrededor de un valor de PVAF específico para la segmentación). Además, la reflexión de los valores de PVAF ayuda a extraer información del haplotipo a partir de las representaciones de PVAF continuas.
- 50

Preferiblemente, la reflexión de los valores de PVAF en la categoría que corresponde al primer progenitor comprende:

- determinar los loci en los que dicho primer progenitor tiene un genotipo en fase particular (p.ej. AB o BA); y
- reflejar los valores de PVAF en loci determinados alrededor del eje medio.

5 En particular, la reflexión de los valores de PVAF se lleva a cabo en ambas subcategorías. En otra realización preferida, el método también comprende la reflexión de los valores de PVAF en la categoría que corresponde al segundo progenitor. Se debe indicar que la determinación de los loci en los que se deberían reflejar los valores de PVAF en la primera y segunda categoría parental no se realiza necesariamente para el mismo genotipo en fase particular. P.ej., en la categoría paterna, los valores de PVAF se podrían reflejar para los loci en los que el haplotipo paterno es AB; mientras en la categoría materna, los valores de PVAF se pueden reflejar para los loci en los que el haplotipo materno es BA. Sin embargo, dentro de una categoría parental específica, la reflexión de los valores de PVAF se lleva a cabo preferiblemente en loci en los que dicho progenitor tiene un genotipo en fase particular (p.ej. AB para un marcador bialélico).

En las realizaciones preferidas de la invención, el método comprende:

(a) categorizar los valores de PVAF continuos determinados del material genético de la muestra en categorías de PVAF parentales

15 (b) subcategorizar dichas categorías de PVAF parentales de (a) y reflejar los valores de PVAF determinados según combinaciones específicas de los genotipos de PV en fase parentales.

En las realizaciones preferidas de la invención, los patrones de PVAF haplotipados / en fase y segmentados proporcionan una asignación de bloques de haplotipos independientes.

20 En las realizaciones preferidas de la invención, las puntuaciones parentales y dichos patrones de PVAF haplotipados / en fase y segmentados se usan para normalizar los valores de la cantidad de ADN (también denominados en la presente memoria valores del número de copias (relativos)) (p.ej. logR) y determinar los cromosomas diploides de los materiales genéticos de la muestra.

25 En las realizaciones preferidas de la invención, el método puede comprender además integrar dichos valores del número de copias (relativos) normalizados (p.ej. logR) o un perfil del número de copias con los patrones de PVAF haplotipados / en fase y segmentados para revelar diferentes distintivos para diferentes anomalías en el material genético de la muestra.

En una realización particular, la categorización de los valores de PVAF continuos en una categoría que corresponde al primer progenitor comprende:

- 30
- determinar los loci que son informativos para el primer progenitor mediante el uso de la información del genotipo del primer y opcionalmente segundo progenitor; y
  - categorizar los valores de PVAF continuos de material genético del sujeto en dichos loci que son informativos para el primer progenitor en una categoría que corresponde al primer progenitor.

En otra realización particular, la subcategorización de los valores de PVAF continuos a partir de la categoría que corresponde al primer progenitor comprende:

- 35
- determinar los loci con una combinación de genotipo específico del primer y segundo progenitor;
  - subcategorizar los valores de PVAF continuos de material genético del sujeto en dichos loci con una combinación de genotipos específica del primer y segundo progenitor.

40 En las realizaciones preferidas de la invención, la segmentación se lleva a cabo mediante el uso de un método de segmentación. La persona experta conoce los métodos de segmentación que son adecuados para segmentar los valores de PVAF (sub)categorizados, tales como los métodos de segmentación por agrupamiento que incluyen un algoritmo de K-medias. En otra realización particular, el método de segmentación es un algoritmo de ajuste constante por partes. En otra realización particular, el método de segmentación es un método de segmentación binaria tal como segmentación binaria circular (CBS).

45 En las realizaciones preferidas de la invención, el método traduce valores de PVAF del material genético de la muestra a bloques de haplotipos.

En otra realización particular, indicar una anomalía genética se refiere a indicar, en particular determinar, la presencia o ausencia de una anomalía genética.

En las realizaciones preferidas de la invención, los patrones de PVAF haplotipados / en fase y segmentados confirman el número de copias de ADN y las anomalías sin cambios en el número de copias de ADN.

50 En las realizaciones preferidas de la invención, los patrones de PVAF haplotipados / en fase y segmentados definen

el origen parental y el origen mecanístico de las anomalías cromosómicas numéricas, estructurales o sin cambios en el número de copias.

En las realizaciones preferidas de la invención, el material genético se origina de un embrión en etapa de segmentación o en etapa de blastocisto.

- 5 En las realizaciones preferidas adicionales de la invención, el material genético se origina de un feto o muestras fetales exentas de células.

En las realizaciones preferidas adicionales de la invención, el material genético se origina de un tejido normal y/o un cáncer.

- 10 En las realizaciones preferidas de la invención, el haplotipado de células individuales usa preferiblemente valores de PVAF continuos de células individuales (p.ej. frecuencias del alelo B de SNP o BAF). En esta aproximación, en vez de las asignaciones del genotipo de PV discretas de células individuales, se exportan preferiblemente los valores de PVAF continuos de células individuales, p.ej., de GenomeStudio de Illumina, que preferiblemente posteriormente se haplotipan / se determina la fase basándose en genotipos de PV multi-celulares parentales robustos, en los que a su vez se ha determinado la fase mediante el uso de los genotipos de un familiar, tal como, p.ej., un hermano o abuelo,
- 15 según las realizaciones de la invención. Posteriormente, los valores de PVAF de células individuales consecutivos se dividen primero en cuatro subcategorías diferentes (P1 y P2 en la categoría paterna; M1 y M2 en la categoría materna) y se giran alrededor del eje medio (0,5) definido por las asignaciones de PV informativas en fase en los genotipos de los progenitores. Posteriormente, estos valores de PVAF de células individuales se segmentan preferiblemente mediante el uso, por ejemplo, de ajuste constante por partes. La aplicación de este principio a PVs informativas maternas y paternas por separado da como resultado asignaciones de bloques de haplotipos de PV independientes. Estos patrones de PVAF de célula individual haplotipados / en fase y segmentados según las realizaciones de la invención, no solamente confirman y mejoran independientemente los haplotipos de células individuales determinados con genotipos de SNP discretos, sino que también indican las anomalías del número de copias de ADN y las anomalías sin cambios en el número de copias de ADN. Además, en las realizaciones
- 20 preferidas adicionales, mediante la integración con los perfiles del número de copias de ADN, tienen la capacidad de distinguir las anomalías genómicas que son de origen meiótico de las que son de origen mitótico.

- Las realizaciones de la presente invención proporcionan un método ortogonal innovador que puede traducir las frecuencias de alelos de variantes polimórficas (PVAFs) en bloques de haplotipos. Debido a que las PVAFs, que son valores continuos en vez de asignaciones de AA, AB o BB discretas ejecutados por un algoritmo de genotipado,
- 30 están en fase y se interpretan según las realizaciones de la invención, se tienen en cuenta de manera ventajosa los supuestos errores del genotipado de células individuales que resultan, p.ej., de la ADO o PA incompleta que hace fracasar los métodos de determinación de la fase. Por lo tanto, las PVAFs segmentadas pero haplotipadas / en fase según las realizaciones de la invención no solamente confirman independientemente los haplotipos ortodoxos calculados a partir de las asignaciones de genotipo de PV de células individuales discretas, sino que también confirman elegantemente las anomalías del número de copias de ADN.
- 35

- Las realizaciones preferidas de la invención proporcionan un método para el haplotipado del ADN exento de células, de células individuales, de unas pocas células o multi-celular, la determinación del perfil del número de copias, o el genotipado discreto, por lo cual dicho método usa valores de la fracción de alelos de variantes polimórficas (PVAF) y/o valores del número de copias de ADN (relativos) (p.ej. logR) determinados a partir de muestras de ADN
- 40 analizadas con plataformas de tipado de polimorfismos (p.ej. matrices de SNP, secuenciación de última generación que incluye secuenciación del genoma, secuenciación de exomas y otras aproximaciones de secuenciación selectiva). Las muestras de ADN pueden comprender ADN amplificado de todo el genoma, productos de amplificación parcial del genoma y ADN no amplificado.

En las realizaciones preferidas adicionales, el método puede comprender las etapas siguientes:

- 45 - El tipado paralelo masivo (de todo el genoma) de las frecuencias de alelos del polimorfismo genético en cualquiera de dichas muestras de ADN (ADN exento de células, de células individuales, de unas pocas células o multi-celular; sometido a WGA o no);
- El tipado del número de copias mediante el uso de todas las frecuencias alélicas de polimorfismo genético en todo el genoma o una selección del mismo;
- 50 - La reconstrucción del haplotipo de la muestra de ADN mediante el uso de todas las frecuencias alélicas de variantes polimórficas genéticas en todo el genoma o una selección del mismo, por lo que dicha reconstrucción puede comprender:
- i. Determinar la fase de los genotipos de PV parentales basándose en un familiar cercano o el genotipo discreto de la propia muestra;
- 55 ii. Identificar loci de PV informativos;

- iii. Categorizar los datos de PVAF de las muestras de ADN estudiadas en dos perfiles parentales, es decir, los perfiles maternos y paternos;
- iv. Los datos de PVAF parentales se reflejan alrededor del eje medio (0,5), en donde los progenitores tienen asignaciones de PV BA (o mediante el uso de aproximaciones similares en asignaciones de PV AB);
- 5 v. Subcategorizar los datos de PVAF parentales en cuatro sub-perfiles según los genotipos parentales en fase;
- vi. Segmentar cada uno de los cuatro sub-perfiles, p.ej. mediante el uso de funciones constantes por partes (PCF) o segmentación binaria circular (CBS).
- vii. Los segmentos de PVAF en el ADN exento de células, de células individuales, de unas pocas células o multi-celular (sometido a WGA o no) revelan los haplotipos parentales hereditarios.
- 10 viii. Los patrones de perfiles de PVAF revelan diferentes distintivos para diferentes anomalías genómicas, y se pueden integrar con logR o los perfiles del número de copias de ADN (relativos).
- En las realizaciones preferidas, los valores de la frecuencia de alelos de variantes polimórficas (PVAF) se haplotipan / se determina la fase basándose en los genotipos de variantes polimórficas (PV) parentales en los que se ha determinado la fase mediante el uso de los genotipos de la descendencia o de los abuelos.
- 15 En las realizaciones preferidas adicionales, los valores de PVAF en fase obtenidos de ADN exento de células, de células individuales, de unas pocas células o multi-celular se segmentan mediante el uso de un algoritmo de segmentación, p.ej. ajuste constante por partes (PCF) o segmentación binaria circular (CBS).
- En las realizaciones preferidas, la determinación de la fase y la segmentación culminan preferiblemente con las asignaciones de bloques de PVAF-haplotipo.
- 20 En las realizaciones preferidas adicionales, los patrones de PVAF en fase y segmentados confirman las anomalías del número de copias de ADN (relativo) y las anomalías sin cambios en el número de copias.
- En las realizaciones preferidas, el método comprende además integrar dicho número de copias de ADN (relativo) (p.ej. valores de logR) y el patrón de los valores de PVAF haplotipados para distinguir el origen de cada anomalía cromosómica como resultado de un evento de falta de disyunción meiótica o mitótica. En otra realización, la presente invención proporciona los métodos descritos en la presente memoria, en los que los valores de PVAF segmentados indican el origen meiótico o mitótico de una anomalía genética en el material genético del sujeto. En una realización adicional, la presente invención proporciona métodos en los que los valores de PVAF segmentados indican el haplotipo del material genético del sujeto. En otra realización adicional, la presente invención proporciona métodos en los que los valores de PVAF indican el número de copias de un cromosoma o región cromosómica en el material genético del sujeto. En otra realización particular, los valores de PVAF segmentados indican un sitio de recombinación homóloga o un punto de ruptura cromosómico.
- 25
- 30
- En las realizaciones preferidas adicionales, el método comprende además integrar dicho número de copias de ADN (relativo) (p.ej. valores de logR) para distinguir el origen de cada anomalía cromosómica como resultado de un suceso de falta de disyunción meiótica I o meiótica II. En las realizaciones preferidas, las trisomías que son de origen meiótico I se pueden distinguir de las que son de origen mitótico o meiótico II.
- 35
- En las realizaciones preferidas adicionales, las trisomías que son de origen meiótico II se pueden distinguir de las que son de origen mitótico cuando se ha dado al menos una recombinación homóloga.
- En las realizaciones preferidas, los cromosomas diploides normales se pueden distinguir de la isodisomía uniparental (UPiD) y heterodisomía uniparental (UPhD), y por tanto la UPiD se puede distinguir de UPhD.
- 40
- En las realizaciones preferidas, las UPiDs que son de origen meiótico II se pueden distinguir de las que son de origen mitótico cuando se ha dado al menos una recombinación homóloga.
- Las realizaciones de la presente invención proporcionan de manera ventajosa un método ortogonal que traduce las frecuencias de alelos de PV de ADN exento de células, de células individuales, de unas pocas células o multi-celular en bloques de haplotipos.
- 45
- Mediante el uso del método según las realizaciones de la invención, se puede hallar la pérdida de la heterocigosidad (LOH) sin cambios en el número de copias en cada uno de los progenitores o el hermano.
- De manera ventajosa, se pueden usar las realizaciones del método para detectar mosaicismos genéticos en mezclas de muestras de ADN de unas pocas células (sometidas a WGA o no), multi-celulares (sometidas a WGA o no), o exentas de células (sometidas a WGA o no), así como su origen parental.
- 50
- De manera ventajosa, las realizaciones del método se pueden usar para detectar la arquitectura alélica de ADN fetal exento de células en el torrente sanguíneo materno.

De manera ventajosa, las realizaciones del método se pueden usar en el diagnóstico prenatal no invasivo (NiPD) para detectar la arquitectura alélica (des)equilibrada y los haplotipos del genoma fetal.

Las realizaciones preferidas de la invención proporcionan métodos para el genotipado de célula individual, de unas pocas células o multi-celular, haplotipado, determinación del perfil del número de copias y la asignación de variantes patológicas asociadas determinando la cantidad, origen parental, así como el origen meiótico o mitótico de los alelos del genoma. De manera más específica, las realizaciones preferidas de la presente invención se refieren al análisis de la frecuencia de alelos de variantes polimórficas (PVAF) de todo el genoma, como un método genérico innovador para el diagnóstico genético, en el que el ADN genómico se puede obtener de muchas células, unas pocas células, una célula individual, o simplemente de ADN exento de células. El ADN se procesa preferiblemente mediante métodos de genotipado de alto rendimiento, p.ej. matrices de SNP, y secuenciación de última generación. Además, las realizaciones de la presente invención son de manera ventajosa de importancia fundamental para comprender la etiología de las enfermedades genéticas, en particular trastornos genéticos con una base mendeliana, así como mosaicismos genéticos que surgen tras la fertilización y anomalías cromosómicas de novo determinando la arquitectura alélica del genoma. La presente invención, por lo tanto, puede conducir de manera ventajosa a avances importantes en el ámbito clínico, y puede conducir a una nueva percepción de los mecanismos de los trastornos genéticos.

En comparación con los métodos conocidos en la técnica para determinar las anomalías del número de copias, el tipado o haplotipado del progenitor de origen, un método según las realizaciones preferidas de la invención utiliza genotipos familiares para recuperar valores de PVAF informativos y determinar los perfiles de PVAF específicos del haplotipo paterno y materno para las muestras de ADN. De manera ventajosa, estas muestras de ADN pueden comprender no solamente muestras de ADN convencionales extraídas de múltiples células, sino también muestras de ADN obtenidas de una célula individual o de unas pocas células que requieren una amplificación de todo el genoma (WGA), un proceso que se sabe que introduce errores de PVAF debidos a artefactos de la amplificación alélica, o ADN obtenido de ácidos nucleicos exentos de células. La integración de los perfiles de PVAF específicos de haplotipo parental y los valores del número de copias de ADN (relativo) genera distintivos únicos. Estos distintivos no solamente revelan haplotipos auténticos, lo que incluye la localización de los sitios de recombinación homóloga meiótica parentales, de las muestras de ADN, sino también descifran diversos tipos de anomalías cromosómicas y su origen parental y mecánico. Así, un método según las realizaciones de la invención también posibilita de manera ventajosa el haplotipado simultáneo, el tipado del número de copias de ADN tradicional, el tipado del número de copias de los haplotipos y la determinación del perfil del progenitor de origen de las muestras de ADN.

Además, el método de la invención facilita la detección de mosaicismos genéticos y su origen parental en muestras de ADN de unas pocas células o multi-celulares, tras WGA o no.

En un segundo aspecto, la presente invención se refiere a un producto de programa informático que comprende un medio de código de programa informático adaptado para llevar a cabo todas las etapas del método como se describió anteriormente, cuando el producto de programa informático se ejecuta en un ordenador. En particular, la presente invención proporciona un producto de programa informático que es capaz, cuando se ejecuta en un motor de procesamiento, de llevar a cabo los métodos descritos en la presente memoria.

En un tercer aspecto, la presente invención proporciona un soporte de datos que almacena un producto de programa informático según el cuarto aspecto de la presente invención. La expresión "soporte de datos" es igual a las expresiones "medio de soporte" o "medio legible por ordenador", y se refiere a cualquier medio que participa proporcionando instrucciones a un procesador para la ejecución. Tal medio puede tomar muchas formas, que incluyen, pero sin limitación, medios no volátiles, medios volátiles, y medios de transmisión. Los medios no volátiles incluyen, por ejemplo, discos ópticos o magnéticos, tales como un dispositivo de almacenamiento que es parte de un almacenamiento masivo. Los medios volátiles incluyen la memoria dinámica, tal como la RAM. Las formas habituales de medios legibles por ordenador incluyen, por ejemplo, un disco flexible, un disco duro, una cinta magnética, o cualquier otro medio magnético, un CD-ROM, cualquier otro medio óptico, tarjetas perforadas, cintas de papel, cualquier otro medio físico con patrones de orificios, una RAM, una PROM, una EPROM, una FLASH-EPROM, cualquier otro chip o cartucho de memoria, una onda portadora como se describe a continuación, o cualquier otro medio legible por ordenador. Se pueden emplear diversas formas de medios legibles por ordenador para transportar una o más secuencias de una o más instrucciones a un procesador para la ejecución. Por ejemplo, las instrucciones se pueden transportar inicialmente en un disco magnético de un ordenador remoto. El ordenador remoto puede cargar las instrucciones en su memoria dinámica y enviar las instrucciones a través de una línea telefónica mediante el uso de un módem. Un módem local del sistema informático puede recibir los datos de la línea telefónica y usar un transmisor infrarrojo para convertir los datos en una señal infrarroja. Un detector infrarrojo acoplado a un bus puede recibir los datos portados en la señal infrarroja, y colocar los datos en el bus. El bus transporta los datos a la memoria principal, de la cual un procesador recupera y ejecuta las instrucciones. Las instrucciones recibidas por la memoria principal se pueden almacenar opcionalmente en un dispositivo de almacenamiento antes o después de la ejecución en un procesador. Las instrucciones se pueden transmitir también a través de una onda portadora en una red, tal como una LAN, una WAN o Internet. Los medios de transmisión pueden tomar la forma de ondas acústicas u ópticas, tales como las generadas durante las comunicaciones de datos mediante ondas de radio e infrarrojas. Los medios de transmisión incluyen cables coaxiales, cable de cobre y fibra óptica, lo que incluye los cables que forman un bus dentro de un ordenador.

En una realización particular, la presente invención proporciona un medio de almacenamiento legible por ordenador no transitorio que almacena un producto de programa informático como se describe en la presente memoria. En otra realización particular, la presente invención proporciona un medio de almacenamiento legible por ordenador no transitorio que almacena los valores de PVAF segmentados obtenidos mediante el método de la invención, que indican una anomalía genética en el material genético del sujeto y/o la herencia del material genético de un sujeto.

5

Además, la presente descripción proporciona una interfaz gráfica de usuario adaptada para el uso de los métodos de la invención.

En otra realización particular, la presente descripción proporciona una estructura de datos o base de datos de estructuras de datos para almacenar:

- 10
- información del genotipo de al menos un primer progenitor de un sujeto;
  - valores de PVAF continuos del material genético de dicho sujeto, en el que dichos valores de PVAF continuos se han categorizado en una categoría que corresponde a dicho primer progenitor; y
  - información de segmentación para dichos valores de PVAF categorizados.

15 En particular, una estructura de datos o base de datos de estructuras de datos está adaptada además para el uso de los métodos de la invención.

En un cuarto aspecto, la presente descripción proporciona la transmisión de un producto de programa informático según el tercer aspecto de la presente invención por una red.

20 En las realizaciones preferidas, la presente descripción proporciona el uso de un método según las realizaciones de la invención, para descubrir la naturaleza meiótica o mitótica de las anomalías del ADN en el material genético de la muestra.

25 Los valores o patrones de PVAF segmentados, según las realizaciones de la invención, se pueden aplicar al haplotipado y/o la detección de variaciones del número de copias de ADN y/o la detección de variaciones del número de copias en haplotipos y/o la detección del origen parental del material genético de una muestra. Preferiblemente dicho haplotipado, tipado del número de copias o determinación del perfil del progenitor de origen comprende el cribado de todo el genoma del material genético de una muestra.

Las realizaciones de la presente invención se pueden usar de manera ventajosa como una aproximación genérica para el cribado de aneuploidías.

30 En una realización particular, la presente descripción proporciona un método para determinar una trisomía en el material genético de un sujeto. En otra realización particular, la presente invención proporciona un método para determinar el número de copias de los cromosomas 13, 18 y/o 21. En una realización adicional, el número de copias del cromosoma 13. En otra realización adicional, el número de copias del cromosoma 18. En otra realización adicional, el número de copias del cromosoma 21. En otra realización particular, la presente invención proporciona un método para detectar un trastorno genético seleccionado del grupo que comprende síndrome de Down, síndrome de Edwards, síndrome de Patau, síndrome de Klinefelter, 47XXX, 47XYY, síndrome de Turner, triploidia, síndrome de DiGeorge, síndrome Cri du Chat, síndrome de Angelman, síndrome de Prader-Willi, síndrome de Wolf-Hirschhorn, síndrome de Smith-Magenis, síndrome de Williams-Beuren, síndrome de Phelan-McDermid, síndrome de Sotos, fibrosis quística, distrofia muscular, atrofia muscular espinal, X frágil, enfermedad de Tay-Sachs, enfermedad de Gaucher, distonía de torsión, enfermedad de Niemann-Pick, mucopolisidosis, anemia de Fanconi, enfermedad de Canavan, anemia de células falciformes, síndrome de Bloom.

40 Estos y otros aspectos de la invención serán evidentes y se dilucidarán con referencia a la(s) realización(es) descrita(s) más adelante en la presente memoria.

### Breve descripción de los dibujos

Las características adicionales de la presente invención serán evidentes a partir de los ejemplos y figuras, en las que:

45 La Fig. 1 ilustra los distintivos de PVAF de disomías normales (euploides) tras un análisis de PVAF. La figura representa un par de cromosomas homólogos con un evento hipotético de recombinación homóloga, y se muestran las cuatro composiciones posibles de los gametos resultantes del primer progenitor (I, II, III y IV). Por simplicidad, solamente se consideran 10 loci informativos categorizados en la primera categoría parental para abarcar los escenarios verosímiles. Tras la combinación con un gameto hipotético de un segundo progenitor (que en la figura tiene el haplotipo ABBAABBAAB), se muestra la célula resultante. Los valores de BAF teóricos se representan debajo, y las flechas indican el giro de valores de BAF particulares. El patrón de BAF que resulta de la subcategorización y segmentación se muestra directamente al lado de la representación de BAF. Los segmentos de color gris oscuro indican los valores de BAF segmentados que se han categorizado en una primera subcategoría, y las barras de color gris claro indican los valores de BAF segmentados que se han categorizado en una segunda

50

subcategoría.

La Fig. 2 ilustra los distintivos de PVAF de monosomías tras el análisis de PVAF. Además, la figura representa un par de cromosomas homólogos con un evento hipotético de recombinación homóloga. Por simplicidad, solamente se consideran 10 loci informativos para abarcar los escenarios verosímiles.

5 La Fig. 3 ilustra el distintivo de PVAF de isodisomías uniparentales (UPiD) tras el análisis de PVAF. Además, la figura representa un par de cromosomas homólogos con un evento hipotético de recombinación homóloga. Por simplicidad, solamente se consideran 10 loci informativos para abarcar los escenarios verosímiles.

10 La Fig. 4 ilustra los distintivos de PVAF de heterodisomías uniparentales (UPhD) tras el análisis de PVAF. Además, la figura representa un par de cromosomas homólogos con un evento hipotético de recombinación homóloga. Por simplicidad, solamente se consideran 10 loci informativos para abarcar los escenarios verosímiles.

La Fig. 5 ilustra los distintivos de PVAF de trisomías meióticas I tras el análisis de PVAF. Además, la figura representa un par de cromosomas homólogos con un evento hipotético de recombinación homóloga. Por simplicidad, solamente se consideran 10 loci informativos para abarcar los escenarios verosímiles.

15 La Fig. 6 ilustra los distintivos de PVAF de trisomías meióticas II tras el análisis de PVAF. Además, la figura representa un par de cromosomas homólogos con un evento hipotético de recombinación homóloga. Por simplicidad, solamente se consideran 10 loci informativos para abarcar los escenarios verosímiles.

La Fig. 7 ilustra los distintivos de PVAF de trisomías mitóticas tras el análisis de PVAF. Además, la figura representa un par de cromosomas homólogos con un evento hipotético de recombinación homóloga. Por simplicidad, solamente se consideran 10 loci informativos para abarcar los escenarios verosímiles.

20 La Fig. 8 ilustra los distintivos de PVAF de UPDs en mosaico tras el análisis de PVAF. Además, la figura representa un par de cromosomas homólogos con un evento hipotético de recombinación homóloga. Por simplicidad, solamente se consideran 10 loci informativos para abarcar los escenarios verosímiles. Si la UPD en mosaico es de origen materno, la Fig. 8a muestra los distintivos de PVAF en el perfil materno y la Fig. 8b muestra los distintivos de PVAF en el perfil paterno. Además, si la UPD en mosaico es de origen paterno, la Fig. 8a muestra los distintivos de PVAF en el perfil paterno y la Fig. 8b muestra los distintivos de PVAF en el perfil materno.

25 La Fig. 9 ilustra diferentes ejemplos de datos reales para anomalías de cromosomas completos detectadas mediante la presente invención en blastómeros individuales. La Fig. 9a representa una nulisomía, la Fig. 9b una monosomía materna (es decir, se conserva el alelo materno), la Fig. 9c una disomía normal y la Fig. 9d una trisomía paterna mitótica (es decir, dos copias del alelo paterno y una copia del alelo materno).

30 La Fig. 10 ilustra diferentes ejemplos de datos reales y demuestra la diferencia de las trisomías meióticas y mitóticas detectadas mediante la presente invención en blastómeros individuales. La Fig. 10a representa una trisomía mitótica materna mientras la Fig. 10b representa una trisomía meiótica I materna.

35 Las Figs. 11a-e ilustran perfiles de PVAF haplotipados parentales usados en las realizaciones de la invención, que ilustran la capacidad de las realizaciones de la presente invención de detectar diferentes anomalías cromosómicas. La Fig. 11a es una monosomía paterna, la Fig. 11b es una disomía normal, la Fig. 11c es una UPiD materna, la Fig. 11d es una trisomía mitótica materna, y la Fig. 11e es una trisomía meiótica materna.

Las Figs. 12a y 12b ilustran un método según las realizaciones de la invención para una muestra de ADN multi-celular que tiene un mosaicismo de UPiD paterno del cromosoma 17q. La Fig. 12a representa el resumen de todo el genoma (cromosomas 1 a 22 y X), mientras la Fig. 12b proporciona una vista más detallada del cromosoma 17.

40 Las Figs. 13a-c ilustran un diagrama de flujo que comprende un módulo según un método según las realizaciones de la invención.

La Fig. 14 ilustra la capacidad de la presente invención tras la secuenciación de una representación de una biblioteca reducida del genoma humano, p.ej. secuenciación de exomas. La Fig. 14a representa perfiles de PVAF haplotipados obtenidos de datos de secuenciación de exomas de una muestra de ADN multi-celular.

45 La Fig. 14b representa perfiles de PVAF haplotipados obtenidos de datos de matrices de SNP de la misma muestra de ADN. La comparación de los perfiles de PVAF haplotipados tras la secuenciación de exomas (Fig. 14a) y matrices de SNP (Fig. 14b) muestra la aplicación de la presente invención sobre datos de secuenciación.

50 La Fig. 15 ilustra la exactitud de la detección de sitios de recombinación homóloga (sitios HR) tras la transformación de perfiles parentales de PVAF haplotipados / en fase y segmentados en haplotipos parentales discretos. La Fig. 15a representa la exactitud (%) de los haplotipos de células individuales transformados en PVAF deducidos a partir de valores de PVAF de 5 células individuales transformadas con EBV que se hacen coincidir con el haplotipo correcto en comparación con su referencia de haplotipo multi-celular. Las frecuencias de concordancia de haplotipos de células individuales se representan para 500 SNPs antes y después de los sitios HR presentes en dos individuos. La Fig. 15b representa la exactitud (%) de los haplotipos de células individuales deducidos a partir de las asignaciones

de SNP discretas de las mismas 5 células individuales transformadas con EBV que se hacen coincidir con el haplotipo correcto en comparación con su referencia de haplotipo multi-celular. Las frecuencias de concordancia de haplotipos de células individuales se representan para 500 SNPs antes y después de los sitios HR presentes en dos individuos. Los haplotipos en bruto se muestran en el fondo, y los haplotipos interpretados se muestran en gris.

5 La Fig. 16 ilustra una muestra de ADN multi-celular con un cariotipo XXXY provocado por errores meióticos I y mitóticos. La Fig. 16a muestra los perfiles de PVAF haplotipados del cromosoma X deducidos a partir de datos de matrices de SNP de esta muestra de ADN. La Fig. 16b muestra la representación esquemática de cómo se rastreó el origen de la anomalía hasta errores que ocurrieron en la meiosis I y mitosis de este individuo.

10 La Fig. 17 proporciona una visión esquemática del método de la invención según una realización preferida. La información del genotipo paterno se indica con cuadrados, mientras la información materna se indica con círculos.

Los dibujos son solamente esquemáticos y no limitantes. En los dibujos, se puede exagerar el tamaño de algunos de los elementos, y no están dibujados a escala con fines ilustrativos. Cualquier signo de referencia en las reivindicaciones no se debe considerar como limitante del alcance. En los diferentes dibujos, los mismos signos de referencia se refieren a los mismos elementos o a elementos análogos.

### 15 Descripción detallada de las realizaciones preferidas

La presente invención se describirá con respecto a realizaciones particulares y con referencia a ciertos dibujos, pero la invención no se limita a ello, sino solamente por las reivindicaciones. Los dibujos descritos son solamente esquemáticos y no limitantes. En los dibujos, se puede exagerar el tamaño de algunos de los elementos, y no están dibujados a escala con fines ilustrativos. Cuando se usa la expresión "que comprende" en la presente descripción y las reivindicaciones, no excluye otros elementos o etapas. Cuando se usa un artículo indefinido o definido al hacer referencia a un nombre singular, p.ej. "un" o "uno/una", "el/la", incluye el plural de ese nombre, a menos que se indique específicamente otra cosa. La expresión "que comprende", usada en las reivindicaciones, no se debería interpretar como limitada a los medios enumerados a continuación; no excluye otros elementos o etapas. Así, el alcance de la expresión "un dispositivo que comprende medios A y B" no se debería limitar a los dispositivos que consisten solamente en los componentes A y B. Significa que, con respecto a la presente invención, los únicos componentes relevantes del dispositivo son A y B. Además, los términos primero, segundo, tercero y similares, en la descripción y en las reivindicaciones, se usan para distinguir entre elementos similares, y no necesariamente para describir un orden secuencial o cronológico. Se debe entender que los términos así usados son intercambiables en circunstancias adecuadas, y que las realizaciones de la invención descritas en la presente memoria son capaces de funcionar en otras secuencias distintas de las descritas o ilustradas en la presente memoria. Además, los términos superior, inferior, sobre, bajo y similares, en la descripción y las reivindicaciones, se usan con fines descriptivos y no necesariamente para describir posiciones relativas. Se debe entender que los términos así usados son intercambiables en circunstancias adecuadas, y que las realizaciones de la invención descritas en la presente memoria son capaces de funcionar en otras orientaciones distintas de las descritas o ilustradas en la presente memoria.

En los dibujos, los números de referencia iguales indican características iguales; y un número de referencia que aparece en más de una figura se refiere al mismo elemento. Los dibujos y las descripciones detalladas siguientes muestran realizaciones específicas de un método o dispositivo para el haplotipado y/o tipado del número de copias de un material genético, más específicamente métodos o dispositivos para el haplotipado, tipado del número de copias, y genotipado mediante análisis de la frecuencia de alelos de variantes polimórficas.

45 Cuando en las realizaciones de la presente invención se hace referencia a una "variante polimórfica" (PV), se hace referencia a las variantes polimórficas que pueden ser variantes genéticas bi-alélicas o multi-alélicas que segregan en una familia o población, sin ningún umbral específico sobre la frecuencia del alelo menor en la población. Los "genotipos de PV o asignaciones de PV discretas" se refieren a los diferentes valores que representan las PVs. Estos valores podrían ser descriptivos (p.ej., en el caso de PV bi-alélica: 'AA', 'AB' y 'BB'; o los genotipos de bases 'AT', 'CG', 'TG', 'AA', etc.) o valores numéricos (p.ej., en el caso de genotipos de PV bi-alélica se codifican como: '11', '10' y '00'). Los "valores de PV continuos" se refieren a valores numéricos que no se limitan a valores diferentes, y que podrían estar en un intervalo continuo entre dos valores numéricos discretos consecutivos. Estos valores pueden tener decimales.

50 Cuando en las realizaciones de la presente invención se hace referencia a la "frecuencia de alelos de variante polimórfica" (PVAF) o la "fracción de alelos de variante polimórfica", se hace referencia a la fracción de un alelo respecto de la cantidad total de alelos en la muestra de ADN tras el genotipado. De manera convencional, para las PVs bialélicas, la PVAF se refiere a la frecuencia de alelos B (BAF), es decir, la fracción de alelos B en los datos de tipado de PV, que se puede obtener de una muestra de ADN mediante métodos de genotipado de alto rendimiento, p.ej. matrices de SNP o tecnologías de secuenciación de última generación. En una realización preferida, una frecuencia de alelos de variante polimórfica es una frecuencia de alelos B. Evidentemente, cuando una reivindicación o realización de la invención se refiere a una frecuencia de alelos B, también se podrían usar las frecuencias de alelos A. Las frecuencias de alelos B comprenden información de frecuencias de alelos A, y viceversa.

En general, se expresa un valor de PVAF mediante el uso de un valor de 0 a 1, ya que se refieren a la frecuencia o fracción. En principio, los valores de PVAF se pueden expresar mediante el uso de un múltiplo de dicho valor, p.ej. mediante el uso de un valor de 0 a 100. Por ejemplo, un valor de PVAF de 0,5, que indica que la mitad de la cantidad total de los alelos tiene el alelo de variante polimórfica, se puede expresar, p.ej., como 50. En ese caso, un valor de PVAF de 1 (es decir, todos los alelos tienen el genotipo particular) se expresará como 100. Como se menciona en la presente memoria, PVAF<sub>max</sub> indica el valor de PVAF máxima (es decir, todos los alelos tienen el genotipo particular) y PVAF<sub>min</sub> indica el valor de PVAF mínima (es decir, ninguno de los alelos tienen el genotipo particular). A lo largo de la presente solicitud, los valores de PVAF (en particular BAF) se indican mediante el uso de un valor de 0 a 1, así PVAF<sub>min</sub> es 0 y PVAF<sub>max</sub> es 1. Sin embargo, las realizaciones de la invención no se limitan a los valores de PVAF expresados mediante el uso de este intervalo particular. Además, el eje medio de los valores de PVAF se refiere al eje que atraviesa los valores que indica que la mitad de la cantidad total de los alelos tienen la variante polimórfica. Por lo tanto, el eje medio también se denomina en la presente memoria como eje 0,5. Cuando se usa, p.ej., un intervalo de valores de PVAF de 0 a 100, el eje medio se refiere al eje que atraviesa 50, etcétera. En particular, a lo largo de la presente solicitud, si se menciona un valor de PVAF de 1, se refiere a PVAF<sub>max</sub>. Si se menciona un valor de PVAF de 0,5, se refiere al valor de PVAF medio (mediano) (es decir, la media de PVAF<sub>min</sub> y PVAF<sub>max</sub>).

La reflexión de un valor respecto del eje medio (también denominado "reflejo" o "giro") se refiere a un proceso en el que a los valores se les reasigna el valor de la reflexión del valor respecto del eje medio. Cuando los valores de PVAF se expresan mediante el uso de un valor de 0 a 1, y el eje 0,5 es el eje medio, los valores reflejados ejemplares son los siguientes:

Valor de PVAF original	Valor de PVAF reflejado
0	1
0,3	0,7
0,5	0,5
0,8	0,2
1	0

20 Cuando en las realizaciones de la presente invención se hace referencia a las tecnologías de genotipado de alto rendimiento, se hace referencia a cualquier tecnología de secuenciación en paralelo masiva, matriz de SNP o secuenciación de última generación. Una tecnología de genotipado de alto rendimiento proporciona, después del procesamiento inicial, datos de PV en bruto, que incluye asignaciones de genotipo, valores de cantidad de ADN y valores de PVAF.

25 Los "valores de cantidad de ADN" se refieren a las medidas del genotipado de alto rendimiento que indican la cantidad de material genético presente en la muestra. Los valores típicos de cantidad de ADN obtenidos mediante el uso del genotipado con matriz de SNP son valores de logR. Los valores típicos de cantidad de ADN obtenidos mediante el uso de las tecnologías de secuenciación son valores de recuentos de lecturas y/o logR. Los valores de cantidad de ADN se pueden haber determinado localmente o en todo el genoma. Preferiblemente, se obtienen cantidades de ADN de todo el genoma y se usan o se normalizan en los métodos de la presente invención. La normalización de los valores de cantidad de ADN se refieren a un proceso que normaliza los valores de cantidad de ADN en bruto (p.ej. valores de logR). En general, este procedimiento puede consistir en: (1) la corrección del sesgo de %GC en los valores en bruto, (2) la detección de los cromosomas disómicos probablemente normales, y (3) la corrección de la media (o mediana) truncada basándose en los cromosomas disómicos normales detectados (véase la descripción detallada de la presente invención).

Cuando en las realizaciones de la presente invención se hace referencia a tecnologías de secuenciación (paralelas masivas), se hace referencia a cualquier metodología de secuenciación de última generación, p.ej. secuenciación de todo el genoma, secuenciación de exomas, secuenciación selectiva.

40 Amplificación de todo el genoma (WGA): WGA es un proceso que se aplica al ADN para incrementar la cantidad de ADN. Las muestras de ADN de células individuales y de pocas células puede requerir WGA para producir una cantidad suficiente de ADN para las tecnologías de genotipado de alto rendimiento. Se han descrito diferentes métodos de WGA, y se basan en general en la amplificación mediante desplazamiento múltiple (MDA) o una amplificación de todo el genoma basada en PCR, o una combinación de las mismas. Se sabe que WGA produce diferentes tipos de artefactos, que incluyen el sesgo de amplificación según la composición de bases del genoma (p.ej., % de contenido de GC), pérdida de alelos (ADO), ganancia de alelos (ADI), amplificación preferente (PA), moléculas quiméricas de ADN y errores de copia de nucleótidos.

50 Información del genotipo: La información del genotipo de un progenitor se refiere a la constitución genética del progenitor. La información del genotipo puede ser información de genotipo sin fase o en fase. La información del genotipo parental se puede obtener directamente mediante genotipado de una muestra obtenida del progenitor (p.ej. mediante secuenciación paralela masiva o tipado con matriz de una muestra de sangre o tejido parental). En otros casos, la información del genotipo del progenitor de un sujeto se puede obtener indirectamente, p.ej. mediante

genotipado de familiares cercanos (tales como hermanos y/o abuelos). Posteriormente, se puede obtener la información del genotipo del progenitor a partir de los genotipos disponibles de los familiares cercanos. El experto sabe cómo calcular un genotipo basándose en los genotipos de familiares cercanos. Preferiblemente, la información del genotipo de el/los progenitor(es) está disponible en forma de valores polimórficos discretos.

- 5 Genotipo en fase: La determinación de la fase de genotipos (PV) de el/los progenitor(es) se puede conseguir mediante el uso de un familiar cercano, p.ej. abuelos y/o un hermano y/o uno o unos pocos embriones. En el proceso de determinación de la fase, se usa al familiar cercano para establecer el orden de los genotipos de PV heterocigotos en los alelos parentales.

- 10 "Loci informativos" o "loci de PV Informativos": Los loci informativos se refieren a loci en los que los genotipos parentales permiten identificar positivamente de qué progenitor proceden los alelos del sujeto. En general, los loci informativos, tal como se usa en la presente memoria, son aquellos en los que un progenitor tiene genotipos de PV homocigotos, y para los mismos loci el otro progenitor tiene genotipos de PV heterocigotos. Sin embargo, en los métodos de la invención también se pueden usar loci semi-informativos en los que un progenitor es homocigoto y otro progenitor es homocigoto para un alelo diferente. Por ejemplo, si el primer progenitor tiene un genotipo AA y el segundo progenitor tiene un genotipo BB, un genotipo AB en el sujeto identifica el origen paterno del primer alelo (A, que procede del primer progenitor) y del segundo alelo (B, que procede del segundo progenitor). Los loci semi-informativos son especialmente útiles para el tipado adicional del progenitor de origen o para la confirmación del progenitor de origen determinado mediante el uso de loci informativos solamente (combinación homocigoto-heterocigoto).

- 20 Progenitor: En el contexto de la presente invención, un progenitor se refiere a un progenitor biológico del sujeto. En el caso de un denominado embrión/feto/niño de tres progenitores, en el que dos progenitores contribuyen al ADN cromosómico y una tercera parte contribuye al ADN mitocondrial, un progenitor se refiere a un progenitor que contribuye al ADN cromosómico.

- 25 Cuando en las realizaciones de la presente invención se hace referencia a una muestra que comprende material genético (en particular una muestra de ADN), se hace referencia en general a todas las muestras que comprenden material genético obtenido de: una célula individual, unas pocas células, un gran número de células o ADN exento de células; sometido a amplificación de todo el genoma (WGA) o no sometido a WGA. Una muestra que comprende material genético también se puede referir a una muestra exenta de células obtenida de una muestra de fluido corporal.

- 30 Cuando en las realizaciones de la presente invención se hace referencia a una muestra (de ADN) multi-celular, en general se hace referencia a muestras de ADN obtenidas de muestras diferentes con una representación fija de ADN entre todas las células o una mezcla de células con arquitectura de mosaico. Esta última podría consistir en una mezcla de células normales y anormales, p.ej. una mezcla de mosaico de células normales y anormales en muestras de tumores.

- 35 Como es evidente a partir de la descripción de la invención en la presente memoria, los métodos de la presente invención se aplican preferiblemente a muestras que contienen cantidades bajas de ácidos nucleicos objetivo, también denominado material genético. En particular, dicho material genético de interés está presente en una o unas pocas células objetivo, o como material circulante libre en la muestra. Así, en una realización particular, dicha muestra contiene una o unas pocas células objetivo. En una realización adicional, dicha muestra contiene una célula objetivo. En otra realización, dicha muestra contiene unas pocas células objetivo, en particular 1 a 30, más en particular 1 a 20, células objetivo. Por ejemplo, 1-15, 1-10, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, una o dos células objetivo. En otra realización particular, los ácidos nucleicos objetivo están presentes en una cantidad de 2 ng o menos en dicha muestra, en particular 1 ng o menos, más en particular 0,5 ng o menos. En otra realización particular, los ácidos nucleicos objetivo están presentes en una cantidad de 250 pg o menos en dicha muestra; en particular 200 pg o menos; más en particular 150 pg o menos. En otra realización particular, dichos ácidos nucleicos objetivo están presentes en una cantidad de 100 pg o menos; en particular en una cantidad de 50 pg o menos; más en particular en una cantidad de 30 pg o menos. En otra realización particular, dichos ácidos nucleicos objetivo son ácidos nucleicos circulantes exentos de células. Por ejemplo, ADN fetal circulante exento de células de una muestra materna, o ADN tumoral circulante de una muestra de un paciente. Aunque el material genético (p.ej. ADN materno) puede ser abundante en tales muestras, el ADN objetivo (es decir, el material genético del sujeto a investigar, p.ej. ADN fetal) está presente solamente en cantidades muy limitadas. En una realización particular, los ácidos nucleicos objetivo están presentes como ácidos nucleicos exentos de células en una muestra de fluido. En particular, dichos ácidos nucleicos exentos de células están presentes en una muestra de fluido que comprende ácidos nucleicos adicionales (no objetivo). En una realización particular, dicha muestra comprende una mezcla de ácidos nucleicos objetivo y no objetivo. Preferiblemente, dichos ácidos nucleicos objetivo están presentes en una cantidad entre un 0,1 y 20% de dichos ácidos nucleicos no objetivo. En otra realización particular, dicha muestra comprende una mezcla de ácidos nucleicos objetivo y no objetivo, en la que dichos ácidos nucleicos objetivo están presentes en una cantidad de 700 ng o menos, en particular 500 ng o menos, más en particular 300 ng o menos. En una realización adicional, 200 ng o menos, en particular 100 ng o menos, más en particular 50 ng o menos. En otra realización, dicha muestra comprende ácidos nucleicos exentos de células, en la que dichos ácidos nucleicos exentos de células están presentes en una cantidad como se definió anteriormente en la presente memoria.

En una realización particular, proporcionar una muestra que comprende cantidades bajas de ácidos nucleicos objetivo comprende aislar una o unas pocas células objetivo. Los métodos de la invención pueden comprender además lisar una o unas pocas células objetivo.

5 Una "muestra (de ADN) de una célula individual" se refiere a una muestra (de ADN) que se obtiene de una célula solitaria de cualquier tipo de tejido o célula. Debido a que una célula individual solamente contiene unos pocos picogramos de ADN, los métodos que implican muestras de células individuales comprenden preferiblemente la amplificación (de todo el genoma) para el genotipado de las variantes polimórficas. "Muestra (de ADN) de unas pocas células" se refiere a una muestra (de ADN) que procede de unas pocas células de cualquier tipo de tejido o célula. Dependiendo del número de células usadas para la extracción del ADN, los métodos que implican tal muestra  
10 pueden comprender la amplificación (de todo el genoma) para el genotipado de las variantes polimórficas. "Muestra (de ADN) multi-celular": una muestra de ADN que procede de un gran número de células de cualquier tipo de tejido o célula.

"Muestra (de ADN) exenta de células" se refiere a una muestra que procede de una muestra de fluido que contiene material genético circulante. Esto se puede referir a ADN fetal exento de células que circula libremente en el torrente  
15 sanguíneo materno (también denominado ADN fetal libre, ffADN). Un ejemplo de tal muestra es una muestra de fluido corporal (en particular una muestra de sangre, plasma o suero; más en particular, una muestra de plasma) obtenida de una hembra embarazada y que comprende una mezcla de material genético materno y fetal. Otra realización de una muestra exenta de células se refiere al ADN tumoral exento de células que circula libremente en el torrente sanguíneo del paciente.

20 La muestra se obtiene preferiblemente de un organismo eucariótico, más en particular de un mamífero. En una realización preferida adicional, dicha muestra es de origen animal no humano (más adelante en la presente memoria también denominado animal) o de origen humano. En una realización particular, dicho animal es un animal domesticado o un animal usado en agricultura, tal como un caballo o una vaca. En una realización particular adicional, dicho animal es un caballo. En otra realización particular, dicha muestra es de origen humano. En otra  
25 realización particular, dicha muestra se obtiene de una mujer embarazada. En otra realización, dicha muestra se obtiene de un paciente que se sospecha que tiene un tumor o cáncer. En otra realización particular, dicha célula es una célula eucariótica, en particular una célula de mamífero. En una realización más particular, el origen de dicha célula es como se describe según las realizaciones preferidas sobre el origen de la muestra, como se describió anteriormente. En otra realización particular, dichos ácidos nucleicos objetivo son de origen eucariótico, en particular  
30 de origen mamífero. En una realización más particular, dichos ácidos nucleicos objetivo son como se describe según la realización preferida sobre el origen de la muestra. Con respecto a ello, en una realización preferida, dichos ácidos nucleicos objetivo se originan de un embrión o un feto. En otra realización preferida, dichos ácidos nucleicos objetivo se originan de un cáncer (sospechoso) o célula tumoral.

35 Los métodos de la invención son aplicables a cualquier tipo de célula. Las células preferidas son cuerpos polares, blastómeros, células trofoectodérmicas de blastocistos o muestras de vellosidad coriónica. El material genético preferido comprende ADN, más en particular ADN exento de células. Preferiblemente, el ADN fetal exento de células es de sangre, plasma o suero materno. Se pueden identificar en sangre materna tanto células fetales intactas como ácidos nucleicos fetales exentos de células (ADN, ARN). Se cree que la fuente principal de la mayoría de ácidos nucleicos fetales exentos de células en la circulación materna es la apoptosis de células placentarias. Como ya se  
40 mencionó anteriormente en la presente memoria, se aplican los métodos en un pequeño número de estos tipos de células, es decir, en unas pocas células, en particular hasta 50 células; más en particular seleccionadas de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más células hasta 30; aún más en particular en una o dos células. Cuando se aplica sobre el trofoectodermo, dichas pocas células se pueden seleccionar de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más células; en particular hasta 50 células de trofoectodermo.

45 Para la extracción de al menos una célula adecuada, la zona pelúcida en las etapas de segmentación y de blastocisto se puede atravesar mediante perforación mecánica de la zona, solución de Tirode acidificada o láser. En las realizaciones preferidas de la invención, al menos una, preferiblemente una célula individual, es un blastómero humano o animal. En las realizaciones particulares, se aplica el ensayo genético para un ensayo de diagnóstico, ensayo de portador, ensayo prenatal, ensayo preimplantacional, o ensayo predictivo y presintomático. En estas  
50 realizaciones particulares, el ensayo genético permite ayudar a los pacientes a completar con éxito la reproducción asistida. En otra realización particular, los métodos de la invención se aplican al cribado de recién nacidos. En otra realización particular, los métodos de la invención se aplican a pruebas forenses.

"En todo el genoma", tal como se usa en la presente memoria, significa que los métodos se aplican y proporcionan información sobre secuencias a lo largo del genoma. En particular, los métodos de la presente invención  
55 proporcionan información sobre todos los cromosomas de los que al menos hay presentes fragmentos en la muestra. En una realización particular, "de todo el genoma" se refiere a la información sobre al menos una variante por 100 Mb, en particular al menos una variante por 10 Mb, en particular al menos una variante por 1 Mb a lo largo del genoma. En una realización adicional, significa al menos una variante por ventana de 100 Mb, en particular al menos 1 variante por ventana de 50 Mb, más en particular al menos una variante por ventana de 10 Mb a lo largo  
60 del genoma. En otra realización particular, de todo el genoma se refiere a información sobre al menos una variante por ventana de 1 Mb. En una realización adicional, significa al menos una variante por ventana de 100 kb, en

particular al menos 1 variante por ventana de 50 kb, más en particular al menos una variante por ventana de 10 kb a lo largo del genoma. En otra realización, de todo el genoma se refiere a información sobre al menos una variante por ventana de 1 kb.

5 "Anomalía genética" se refiere a cualquier anomalía del genoma. La detección de anomalías genéticas puede implicar detectar la presencia o ausencia de cromosomas o regiones cromosómicas genéticamente normales o anormales. Las anomalías genéticas se pueden detectar directamente o indirectamente. La detección directa de anomalías comprende, p.ej., la detección directa de porciones ausentes, adicionales, o irregulares de ADN cromosómico. La detección indirecta de las anomalías comprende determinar y obtener información de la herencia para detectar si una anomalía genética ha sido heredada por un sujeto. Por ejemplo, los métodos de la presente invención permiten determinar si un defecto genético en uno o más fragmentos cromosómicos de un padre o madre han sido heredados por un embrión, un feto o un niño. Como se detalla en la presente memoria, indicar una anomalía genética también puede comprender indicar el origen meiótico o mitótico de dicha anomalía. En otra realización, indicar una anomalía genética se refiere a proporcionar un mapa que muestra el número de copias de un cromosoma o región cromosómica. En una realización adicional, dicho mapa abarca un cromosoma completo. En una realización adicional, dicho mapa abarca todos los autosomas, en particular todos los cromosomas.

La "herencia" de la información genética se refiere a cualquier información sobre la herencia de la información genética de progenitores o antepasados. En particular, la herencia se refiere a la distribución de la información genética dentro de una familia prolongada (es decir, uno o más progenitores, abuelos, hermanos, sobrinos, nietos, tías, tíos, y/o sobrinas). En una realización particular, la presente invención proporciona un método para determinar de qué progenitor y/o abuelo se ha heredado una región cromosómica. En una realización adicional, indicar la herencia de la información genética se refiere a proporcionar el haplotipo de un cromosoma o región cromosómica. En otra realización, indicar la herencia se refiere a proporcionar un mapa que muestra qué cromosoma o región cromosómica se ha heredado de qué familiar. En una realización adicional, dicho mapa abarca un cromosoma completo. En una realización adicional, dicho mapa abarca todos los autosomas, en particular todos los cromosomas. La información de la herencia no incluye necesariamente información sobre las anomalías genéticas. Por ejemplo, la presente invención permite construir haplotipos y/o determinar la herencia de genomas sin defectos genéticos en una familia.

Distancia alélica paterna ( $d_{Pat}$ ):  $d_{Pat}$  es una distancia local entre los valores P1 y P2 segmentados.

Distancia alélica materna ( $d_{Mat}$ ):  $d_{Mat}$  es una distancia local entre los valores M1 y M2 segmentados.

30 Característica de paridad: tras la segmentación, los segmentos subcategorizados resultantes en una categoría parental (p.ej. P1 y P2, o M1 y M2) tienen aproximadamente la misma longitud.

Característica de complementariedad: tras la segmentación, la distancia vertical (también denominada distancia alélica) entre los valores de PVAf segmentados subcategorizados resultantes en una primera categoría parental cambian de una manera complementaria con la distancia vertical (distancia alélica) entre los valores de PVAf segmentados subcategorizados resultantes en la otra categoría parental.

Grado de mosaicismo ( $\rho$ ):  $\rho$  indica la proporción de células anormales en una muestra de ADN en mosaico.  $(1-\rho)$  indica la proporción de células genéticamente normales (euploides) en una muestra de ADN en mosaico.

En las realizaciones preferidas de la invención, se proporciona un método que se basa en un concepto que se aplica para definir preferiblemente (a) los haplotipos de alelos en una muestra de ADN exenta de células, de células individuales, de unas pocas células o multi-celular (tras WGA o no), (b) el estado del número de copias de alelos o haplotipos en una muestra de ADN exenta de células, de células individuales, de unas pocas células o multi-celular (tras WGA o no), (b) el origen parental y mecánico supuesto de las anomalías alélicas en una muestra de ADN exenta de células, de células individuales, de unas pocas células o multi-celular (tras WGA o no).

45 Preferiblemente, la categorización los valores de PVAf continuos en una categoría que corresponde a un primer progenitor comprende:

- determinar los loci que son informativos para el primer progenitor mediante el uso de la información del genotipo del primer y segundo progenitor; y

- categorizar los valores de PVAf continuos de material genético del sujeto en dichos loci que son informativos para el primer progenitor en una categoría que corresponde al primer progenitor.

50 En particular, los loci que son informativos para el primer progenitor comprenden los loci para los que el primer progenitor es heterocigoto y para los que el segundo progenitor es homocigoto.

Se debe indicar que, a lo largo de la presente solicitud, los valores de PVAf categorizados también se denominan valores de PVAf paternos o maternos si se han categorizado en la categoría paterna o materna, respectivamente. Así, tal como se usa en la presente memoria, un valor de PVAf paterno se refiere al valor de PVAf de la muestra en la categoría paterna.

En una realización particular adicional, la subcategorización de los valores de PVAF continuos a partir de la categoría que corresponde al primer progenitor comprende:

- determinar los loci con una combinación de genotipo específico del primer y segundo progenitor;
- subcategorizar los valores de PVAF continuos de material genético del sujeto en dichos loci con una combinación de genotipos específica del primer y segundo progenitor.

En otra realización particular adicional, la subcategorización de los valores de PVAF continuos a partir de la categoría que corresponde al primer progenitor comprende:

- determinar los loci en los que el primer alelo del primer progenitor comprende un genotipo que es idéntico al genotipo de los dos alelos del segundo progenitor y subcategorizar los valores de PVAF continuos en dichos loci en una primera subcategoría; y
- determinar los loci en los que el segundo alelo del primer progenitor comprende un genotipo que es idéntico al genotipo de los dos alelos del segundo progenitor y subcategorizar los valores de PVAF continuos en dichos loci en una segunda subcategoría;

Se debe entender que, en las realizaciones de la presente invención, la subcategorización se puede invertir, en cuyo caso subcategorizar los valores de PVAF continuos de la categoría correspondiente al primer progenitor puede comprender:

- determinar los loci en los que el genotipo del primer alelo del primer progenitor comprende un genotipo que es diferente del genotipo de los dos alelos del segundo progenitor y subcategorizar los valores de PVAF continuos en dichos loci en una primera subcategoría; y
- determinar los loci en los que el segundo alelo del primer progenitor comprende un genotipo que es diferente del genotipo de los dos alelos del segundo progenitor y subcategorizar los valores de PVAF continuos en dichos loci en una segunda subcategoría;

En una realización especialmente preferida, la presente invención proporciona un método que comprende:

- obtener valores de frecuencias de alelos de variantes polimórficas (PVAF) continuos del material genético del sujeto;
- obtener información del genotipo en fase de un primer progenitor e información del genotipo en fase o no en fase de un segundo progenitor;
- determinar "loci (informativos)", es decir, loci que son informativos para el primer progenitor mediante el uso de la información de genotipo del primer y segundo progenitores;
- categorizar los valores de PVAF continuos de material genético del sujeto en dichos "loci (informativos)" que son informativos para el primer progenitor en una categoría que corresponde al primer progenitor;
- determinar "loci (de una primera subcategoría)", es decir, loci en los que el primer alelo del primer progenitor comprende un genotipo que es idéntico al genotipo de los dos alelos del segundo progenitor y subcategorizar los valores de PVAF continuos en dichos "loci (de una primera subcategoría)" en una primera subcategoría;
- determinar "loci (de una segunda subcategoría)", es decir, loci en los que el segundo alelo del primer progenitor comprende un genotipo que es idéntico al genotipo de los dos alelos del segundo progenitor y subcategorizar los valores de PVAF continuos en dichos "loci (de una segunda subcategoría)" en una segunda subcategoría;
- determinar "loci (girados)", es decir, loci en los que dicho primer progenitor tiene un genotipo en fase particular;
- reflejar los valores de PVAF en dichos "loci (girados)" determinados alrededor del eje medio;
- segmentar dichos valores de PVAF subcategorizados tras la etapa de reflexión; y
- proporcionar los valores de PVAF segmentados para indicar una anomalía genética en el material genético del sujeto y/o la herencia del material genético del sujeto.

Los métodos de la invención pueden comprender una etapa de categorización y subcategorización. Como es evidente también a partir de la redacción de las realizaciones (especialmente debido al uso de la expresión "que comprende"), no es necesario que las etapas de categorización y subcategorización se lleven a cabo por separado o en un orden particular (esto mismo es aplicable a cualquiera de las etapas del método). En particular, la categorización y subcategorización se pueden llevar a cabo de manera simultánea. Como es evidente a partir de la descripción de la invención, preferiblemente la categorización y subcategorización se llevan a cabo basándose en la información del genotipo de los progenitores. Mediante el uso de los métodos descritos en la presente memoria, en una única etapa se pueden subcategorizar los valores de PVAF basándose en los datos del genotipo (en fase) de los

progenitores.

Por ejemplo, en otra realización particular, la presente invención proporciona un método que comprende:

- obtener valores de frecuencias de alelos de variantes polimórficas (PVAF) continuos del material genético del sujeto;
- 5 - obtener información del genotipo en fase de un primer progenitor e información del genotipo en fase o no en fase de un segundo progenitor;
- determinar loci de una primera subcategoría en los que el primer alelo del primer progenitor comprende un genotipo que es:
  - diferente del genotipo del segundo alelo del primer progenitor; y
  - 10 - idéntico al genotipo de los dos alelos (homocigotos) del segundo progenitor;
- subcategorizar los valores de PVAF continuos en dichos loci de una primera subcategoría en una primera subcategoría;
- determinar loci de una segunda subcategoría en los que el primer alelo del primer progenitor comprende un genotipo que es:
  - 15 - diferente del genotipo del segundo alelo del primer progenitor; y
  - diferente del genotipo de los dos alelos (homocigotos) del segundo progenitor;
- subcategorizar los valores de PVAF continuos en dichos loci en una segunda subcategoría;
- determinar los loci de giro en los que dicho primer progenitor tiene un genotipo en fase particular;
- reflejar los valores de PVAF en la primera y segunda subcategoría en loci de giro determinados alrededor del eje medio;
- 20 - segmentar dichos valores de PVAF subcategorizados tras la etapa de reflexión; y
- proporcionar los valores de PVAF segmentados para indicar una anomalía genética en el material genético del sujeto y/o la herencia del material genético del sujeto.

25 En dicho caso, las etapas de subcategorización, como se describen, categorizan de manera simultánea los valores de PVAF (es decir, conservan los valores de PVAF en los loci informativos para el primer progenitor) y subcategorizan los valores de PVAF (distribuyendo los valores de PVAF específicos en loci con combinaciones de alelos específicos). Así, basándose en el genotipo (en fase), los valores de PVAF se pueden asignar en una única etapa p.ej. las subcategorías P1, P2, M1 y M2.

30 Se debe indicar además que la etapa de reflexión se puede llevar a cabo antes o durante la etapa de categorización y/o subcategorización. En particular, la segmentación se lleva a cabo tras la etapa de reflexión (si está presente en el método).

Por ejemplo, el concepto según las realizaciones de la invención comprende preferiblemente: determinar la fase de el/los genotipo(s) de PV parental(es), preferiblemente mediante el uso de un genotipo de PV disponible obtenido de un familiar cercano, p.ej. un hermano o los abuelos, o el propio genotipo de la muestra de ADN exenta de células, de células individuales, de unas pocas células o multi-celular (tras WGA o no). En una etapa posterior preferida, se identifican los loci de PV informativos. En las realizaciones preferidas, se define que un locus es informativo cuando en un progenitor es heterocigoto y en el otro progenitor es homocigoto para el mismo locus de PV. En las realizaciones preferidas, los loci de PV informativos se categorizan en forma de categorías parentales informativas, tales como paternas o maternas informativas. Un locus informativo se define preferiblemente como paterno cuando el genotipo de PV del padre en el locus es heterocigoto y el genotipo de PV de la madre es homocigoto. De forma similar, se define un locus informativo como materno cuando el genotipo de PV de la madre es heterocigoto en el locus y el genotipo de PV del padre es homocigoto. En las realizaciones preferidas adicionales, los loci de PV informativos maternos y paternos, determinados como se describió anteriormente, preferiblemente se subcategorizan adicionalmente basándose en combinaciones de genotipos de PV informativos parentales específicos (también denominados en la presente memoria combinación de genotipos específicos del primer y segundo progenitor). Así, en la categoría paterna, si el genotipo de PV del padre es AB y el genotipo de PV de la madre es AA, o si el genotipo de PV del padre es BA y el genotipo de PV de la madre es BB, estos loci se marcan como P1. Y si el genotipo de PV del padre es AB y el genotipo de PV de la madre es BB, o si el genotipo de PV del padre es BA y el genotipo de PV de la madre es AA, estos se marcan como P2. De forma similar, en la categoría materna, si el genotipo de PV de la madre es AB y el genotipo de PV del padre es AA, o si el genotipo de PV de la madre es BA y el genotipo de PV del padre es BB, estos loci se marcan como M1. Y si el genotipo de PV de la

madre es AB y el genotipo de PV del padre es BB, o si el genotipo de PV de la madre es BA y el genotipo de PV del padre es AA, estos se marcan como M2.

Se debe indicar que las subcategorías P1, P2, M1 y M2 tienen el significado descrito en la presente invención, cuyo significado no necesariamente corresponde a P1, P2, M1 o M2 descrito en otra parte (p.ej. en el documento EP1951897).

Se debe indicar que, dada la simetría de un sujeto que tiene dos progenitores, las realizaciones que describen el uso de la información de genotipo/haplotipo materno son aplicables también a los equivalentes paternos, y viceversa. Por dicha razón, las referencias en la presente memoria al primer y segundo progenitor se refieren al padre y a la madre o a la madre y al padre, respectivamente y viceversa. Además, la persona experta conoce cómo aplicar las realizaciones descritas mediante el uso de la información de genotipo/haplotipo de dos progenitores a situaciones en las que solamente se usa la información de genotipo/haplotipo (en fase) de un único progenitor. Como simple ejemplo, si hay disponibles dos genotipos en fase (como se prefiere y, p.ej., como se muestra en la Fig. 17), los valores de PVAF se pueden categorizar en dos categorías (primer progenitor y segundo progenitor), y cada categoría se puede subcategorizar adicionalmente en dos subcategorías basándose en las combinaciones específicas de genotipos en fase de los progenitores. En esa situación, los métodos de la invención permiten crear un perfil genético de valores de PVAF segmentados que muestran las anomalías genéticas de los cromosomas hereditarios de ambos progenitores, y que indican la herencia/haplotipo de los cromosomas hereditarios de ambos progenitores. Sin embargo, no siempre es necesaria tal información detallada sobre los cromosomas hereditarios de ambos progenitores. Por ejemplo, si un padre porta un defecto genético en una región cromosómica particular, la identificación de qué región cromosómica paterna ha sido heredada por el embrión/feto/niño permite determinar si se ha transmitido el defecto genético. Además, existen situaciones en las que la información de genotipo (en fase) solamente está disponible a partir de un único progenitor. Por lo tanto, los métodos descritos en la presente memoria se pueden aplicar mediante el uso de solamente información genética de un único progenitor, o mediante el uso de solamente información del genotipo en fase de un primer progenitor e información del genotipo no en fase de un segundo progenitor. P.ej., si está disponible la información del genotipo en fase del primer progenitor, los valores de PVAF se pueden categorizar en la primera categoría parental. La subcategorización se puede basar en el genotipo en fase específico del primer progenitor y el genotipo (homocigoto) no en fase del segundo progenitor. Por lo tanto, los métodos de la presente invención permiten determinar la constitución genética de los cromosomas heredados del primer progenitor.

En una realización adicional, los datos de PVAF de ADN exento de células, de células individuales, de unas pocas células o multi-celular se categorizan en dos perfiles paternos y maternos diferentes según los loci de PV informativos determinados como se describió anteriormente. Además, en las realizaciones preferidas, los datos de PVAF de la muestra se reflejan alrededor del eje medio cuando cualquiera de los progenitores tiene asignaciones de PV BA heterocigotas. De manera más específica, si el padre tiene el genotipo BA, los valores de PVAF coincidentes de una muestra de ADN exenta de células, de células individuales, de unas pocas células o multi-celular (tras WGA o no) del perfil paterno se reflejan preferiblemente alrededor del eje medio. De forma similar, si la madre tiene el genotipo BA, los valores de PVAF coincidentes de una muestra de ADN exenta de células, de células individuales, de unas pocas células o multi-celular (tras WGA o no) del perfil materno se reflejan alrededor del eje medio.

En las realizaciones preferidas, los datos de PVAF categorizados de la muestra de la etapa previa se subcategorizan, preferiblemente en dos sub-perfiles por categoría parental (p.ej. cuatro subcategorías en total al analizar las categorías paternas y maternas) según los loci informativos determinados antes. De manera más específica, los datos de PVAF paternos de la muestra se subcategorizan preferiblemente en los sub-perfiles P1 y P2. De forma similar, los datos de PVAF maternos de la muestra se subcategorizan preferiblemente en los sub-perfiles M1 y M2. Cada uno de los cuatro sub-perfiles se segmenta preferiblemente en una realización preferida mediante el uso de funciones constantes por partes (PCF).

De manera beneficiosa respecto de la técnica anterior, los métodos de la presente invención posibilitan un análisis de alta calidad del material genético mediante el uso de una potencia de cálculo limitada. En particular, los métodos de la invención se pueden aplicar mediante el uso de solamente valores de PVAF que se han categorizado, aunque opcionalmente descartando los valores de PVAF que no se han categorizado (p.ej. los valores de PVAF que no corresponden a los loci informativos). Sin embargo, los métodos de la invención usan valores de PVAF continuos, que tienen un contenido de información mayor en comparación con las asignaciones de PV discretas. Los métodos de la invención proporcionan un equilibrio óptimo entre la reducción del almacenamiento de datos y la potencia de cálculo necesaria por una parte (reducción a valores de PVAF categorizados) y la conservación de un contenido de información elevado (uso de valores de PVAF continuos).

Opcionalmente, las realizaciones preferidas del método posibilitan que se puedan visualizar los perfiles parentales en dos gráficos. De manera más específica, mediante la representación de los sub-perfiles P1 y P2 en negro y gris, respectivamente, se genera el gráfico del perfil paterno; posteriormente, los P1 y P2 segmentados se superponen sobre P1 y P2 en el mismo gráfico. De forma similar, representando los sub-perfiles M1 y M2 en negro y gris, respectivamente, se genera el gráfico del perfil materno; posteriormente los M1 y M2 segmentados se superponen sobre M1 y M2 en el mismo gráfico. Los P1, P2, M1 y M2 segmentados se usan para la interpretación posterior de los diferentes distintivos para las diferentes arquitecturas alélicas, como se describe más adelante.

En otra realización particular, la característica de paridad inherente (y/o la característica de complementariedad) se usa para determinar la calidad de los valores de PVAF segmentados obtenidos, en particular para determinar la calidad de la segmentación. En una realización adicional, los métodos de la invención comprenden además corregir la longitud de un segmento en una primera subcategoría mediante el uso de la longitud de un segmento correspondiente en la segunda categoría. En otra realización, los métodos de la invención comprenden además determinar un índice de calidad, en el que el índice de calidad se determina comparando los segmentos de una primera y segunda subcategorías en una categoría parental. En una realización particular, la comparación de segmentos comprende comparar la longitud de los segmentos. En otra realización particular, la comparación de segmentos comprende comparar la localización de los extremos de los segmentos. En una realización adicional, se asigna una puntuación de calidad que indica una baja calidad si los segmentos de una primera y segunda subcategoría son sustancialmente diferentes. Se asigna una puntuación de calidad que indica una elevada calidad si los segmentos de una primera y segunda subcategoría son sustancialmente similares (es decir, la longitud y/o extremo de los segmentos en dos subcategorías diferentes son básicamente las mismas, o la diferencia es inferior a un valor umbral especificado). En otra realización, proporcionar los valores de PVAF segmentados comprende evaluar la longitud de los segmentos de valores de PVAF segmentados en al menos dos subcategorías de la categoría de un progenitor para indicar un sitio de recombinación homóloga o un punto de ruptura cromosómico.

En otra realización, la característica de complementariedad (y/o la característica de paridad) se usa para determinar la calidad de los valores de PVAF segmentados obtenidos, en particular para determinar la calidad de la segmentación. En una realización adicional, los métodos de la invención comprenden además corregir la distancia entre segmentos de diferentes subcategorías en una primera categoría mediante el uso de la distancia entre segmentos de diferentes subcategorías en la segunda categoría. En particular, corregir la distancia alélica paterna ( $d_{Pat}$ ) mediante el uso de la distancia alélica materna ( $d_{Mat}$ ), o viceversa. En otra realización, los métodos de la invención comprenden además determinar una puntuación de calidad, en la que la puntuación de calidad se determina comparando las distancias alélicas en dos categorías parentales. Se asigna una puntuación de calidad que indica una baja calidad si las distancias alélicas paternas y maternas son sustancialmente incompatibles (no complementarias). Se asigna una puntuación de calidad que indica una elevada calidad si las distancias alélicas paternas y maternas son sustancialmente compatibles (complementarias) (es decir, las distancias alélicas paternas y maternas son básicamente compatibles). En una realización particular, los métodos de la presente invención comprenden (además) determinar las distancias alélicas paternas de compatibilidad comparando la suma de las distancias alélicas del primer y segundo progenitor respecto de un valor esperado y/o un valor umbral. Más en particular, el valor esperado de la suma de las distancias alélicas paternas es  $PVAF_{max}$ . P.ej., si los valores de PVAF continuos se presentan en forma de números que oscilan de 0 a 1, el valor esperado es 1 (en particular,  $d_{Pat} + d_{Mat} = 1$ ). Así, en una realización preferida, la suma de la distancia alélica paterna y la distancia alélica materna para una región cromosómica se compara con 1. La suma que es cercana a 1 indica la compatibilidad de las distancias alélicas. La suma que es sustancialmente diferente de 1 indica la incompatibilidad. Evidentemente, la suma de las distancias alélicas paternas en una región cromosómica se puede comparar con un valor umbral alto y bajo. Por lo tanto, en una realización particular, si la suma de las distancias alélicas parentales en una región cromosómica es mayor que el valor umbral elevado o inferior que el valor umbral bajo, esto indica una puntuación de baja calidad. De forma similar, la distancia alélica en una categoría parental particular se puede corregir mediante el uso de la distancia alélica en la otra categoría parental. Más en particular, la distancia alélica en una categoría parental particular se puede corregir a 1 menos la distancia alélica en la otra categoría parental. P.ej.,  $d_{Pat}(corr) = 1 - d_{Mat}$  o viceversa. En otra realización particular, proporcionar los valores de PVAF segmentados comprende para una región cromosómica:

- determinar una primera distancia entre a) un valor de PVAF segmentado en una primera subcategoría de la categoría del primer progenitor en dicha región cromosómica y b) un valor de PVAF segmentado en una segunda subcategoría de la categoría del primer progenitor en dicha región cromosómica;

- determinar una segunda distancia entre a) un valor de PVAF segmentado en una primera subcategoría de la categoría del segundo progenitor en dicha región cromosómica y b) un valor de PVAF segmentado en una segunda subcategoría de la categoría del segundo progenitor en dicha región cromosómica; y

- comparar la primera distancia y la segunda distancia para indicar una anomalía del número de copias de dicha región cromosómica.

En una realización particular, la presente invención proporciona un informe que muestra los valores de PVAF segmentados obtenibles mediante los métodos de la invención. En otra realización, la presente invención proporciona un informe que muestra los valores de PVAF segmentados de una región cromosómica, en el que los valores de PVAF segmentados se han subcategorizado en al menos dos subcategorías, en el que la distancia entre los segmentos de las diferentes subcategorías indican el número de copias de dicha región cromosómica, y en el que los extremos de los segmentos indican un sitio de recombinación homóloga o punto de ruptura cromosómico en dicha región cromosómica. Preferiblemente, dicho informe muestra además la herencia de dicha región cromosómica. En otra realización preferida, la presente invención proporciona una estructura de haplotipo (o "cariomapa") de un cromosoma o región cromosómica, en particular de un cromosoma completo, más en particular de dos o más cromosomas. La estructura del haplotipo (o cariomapa) representa esquemáticamente dicho(s) cromosoma(s) o región(es) cromosómica(s), e indica qué regiones se han heredado de qué regiones cromosómicas

parentales.

En una realización particular, la presente invención proporciona un método asistido por ordenador para el análisis del material genético de un sujeto, y dicho método comprende:

- 5
- cargar en un sistema informático los valores de frecuencias de alelos de variantes polimórficas (PVAF) continuos del material genético de un sujeto;
  - cargar en un sistema informático la información del genotipo un primer y un segundo progenitor; y
  - mediante el sistema informático, llevar a cabo una o más de las etapas de análisis restantes del método de la invención como se describe en la presente memoria.

10 En una realización particular adicional, la presente invención proporciona un método asistido por ordenador para el análisis del material genético de un sujeto, y dicho método comprende:

- 15
- cargar en un sistema informático los valores de frecuencias de alelos de variantes polimórficas (PVAF) continuos del material genético de un sujeto;
  - cargar en un sistema informático la información del genotipo un primer y un segundo progenitor;
  - categorizar mediante el sistema informático los valores de PVAF continuos en una categoría que corresponde al primer progenitor basándose en la información del genotipo del primer progenitor;
  - segmentar mediante el sistema informático dichos valores de PVAF categorizados; y
  - proporcionar mediante el sistema informático los valores de PVAF segmentados para indicar una anomalía genética en el material genético del sujeto y/o la herencia del material genético del sujeto.

20 En otra realización particular, la presente invención proporciona un método asistido por ordenador para el análisis del material genético de un sujeto, y dicho método comprende:

- 25
- cargar en un sistema informático los valores de frecuencias de alelos de variantes polimórficas (PVAF) continuos del material genético de un sujeto;
  - cargar en un sistema informático la información del genotipo en fase de un primer progenitor y la información del genotipo en fase o no en fase de un segundo progenitor;
  - categorizar mediante el sistema informático los valores de PVAF continuos en una categoría que corresponde al primer progenitor basándose en la información del genotipo del primer y segundo progenitor;
  - subcategorizar mediante el sistema informático los valores de PVAF continuos de la categoría que corresponde al primer progenitor en subcategorías;
  - segmentar mediante el sistema informático dichos valores de PVAF subcategorizados; y

- 30
- proporcionar mediante el sistema informático los valores de PVAF segmentados para indicar una anomalía genética en el material genético del sujeto y/o la herencia del material genético del sujeto.

35 En una realización particular, cargar la información del genotipo en fase de un progenitor incluye a) cargar la información del genotipo no en fase del progenitor, b) cargar la información del genotipo de un familiar cercano, y c) determinar la fase de dichos datos del genotipo no en fase mediante un ordenador con el uso de los datos del genotipo no en fase del progenitor y la información del genotipo del familiar cercano, para obtener la información del genotipo en fase del progenitor.

En otra realización particular, la presente invención proporciona un método asistido por ordenador para el análisis del material genético de un sujeto, y dicho método comprende:

- 40
- cargar en un sistema informático los valores de frecuencias de alelos de variantes polimórficas (PVAF) continuos del material genético de un sujeto;
  - cargar en un sistema informático la información del genotipo en fase de un primer progenitor y la información del genotipo en fase de un segundo progenitor;
  - categorizar mediante el sistema informático los valores de PVAF continuos en una primera categoría que corresponde al primer progenitor y una segunda categoría que corresponde al segundo progenitor basándose en la información del genotipo del primer y segundo progenitor;
  - subcategorizar mediante el sistema informático los valores de PVAF continuos en la primera y segunda categorías en subcategorías;
- 45

- segmentar mediante el sistema informático dichos valores de PVAF subcategorizados; y

- determinar mediante el sistema informático los valores de PVAF segmentados para indicar una anomalía genética en el material genético del sujeto y/o la herencia del material genético del sujeto.

5 En una realización adicional de los métodos asistidos por ordenador anteriormente mencionados de la presente invención, la subcategorización de los valores de PVAF se combina con la reflexión de los valores de PVAF en loci de giro determinados alrededor del eje medio como se describe en la presente memoria.

*Haplotipado de ADN exento de células, de células individuales, de unas pocas células o multi-celular para cromosomas disómicos con una arquitectura alélica equilibrada*

10 En las realizaciones preferidas del método, dicho método puede comprender un genotipo multi-celular de un hermano, denominado hermano de referencia, que va a servir como germen para el haplotipado. Por lo tanto, el genotipo multi-celular del hermano de referencia se usa preferiblemente para determinar la fase de los genotipos multi-celulares de los progenitores, suponiendo que los cromosomas homólogos parentales que se transmiten al hermano de referencia son P1 y M1. Aplicando los principios anteriormente mencionados de un método según las realizaciones de la invención, con la PVAF obtenida de muestras de ADN exentas de células, de células individuales, 15 de unas pocas células o multi-celular, se pueden generar los perfiles de PVAF paternos y maternos. Los valores de PVAF de P1 y P2 segmentados en el perfil paterno de cada cromosoma revelan bloques de haplotipos paternos. De forma similar, los valores de PVAF de M1 y M2 segmentados en el perfil materno de cada cromosoma revelan bloques de haplotipos maternos. Los puntos de ruptura en los segmentos de cada perfil de PVAF representan los sitios de recombinación homóloga. En esta realización, si en el perfil paterno los valores de PVAF de P1 y P2 segmentados se localizan en/alrededor de 0 y 0,5, respectivamente, se transmiten los alelos paternos similares al hermano de referencia. Sin embargo, si los valores de PVAF de P1 y P2 segmentados se localizan en/alrededor de 0,5 y 1, respectivamente, se transmiten los alelos paternos diferentes al hermano de referencia. De forma similar, si en el perfil materno los valores de PVAF de M1 y M2 segmentados se localizan en/alrededor de 0 y 0,5, respectivamente, se transmiten los alelos maternos similares al hermano de referencia. Sin embargo, si los valores de PVAF de M1 y M2 segmentados se localizan en/alrededor de 0,5 y 1, respectivamente, se transmiten los alelos maternos diferentes al hermano de referencia (véase, p.ej., la Fig. 1). En otras palabras, los puntos de ruptura emparejados en los valores de PVAF de SNP de células individuales de M1 y M2 segmentados marcan los sitios de recombinación homóloga materna, mientras los de los valores de PVAF de células individuales de P1 y P2 segmentados localizan los entrecruzamientos genéticos paternos. Por lo tanto, los segmentos M1-M2 y P1-P2 resultantes indican los bloques de haplotipos hereditarios de los genotipos parentales en fase. 30

En las realizaciones adicionales de la invención, el método comprende el uso de genotipos de PV multi-celulares de los abuelos para determinar la fase de los genotipos de PV multi-celulares de los progenitores, suponiendo que el primer cromosoma homólogo parental es del abuelo y el segundo cromosoma homólogo es de la abuela. Mediante la aplicación de los principios anteriormente mencionados de las realizaciones del método según la presente invención sobre las frecuencias de alelos B de PV, obtenidas de muestras ADN exento de células, de células individuales, de unas pocas células o multi-celular, se generan los perfiles de PVAF paternos y maternos de la muestra. Los valores de PVAF de P1 y P2 segmentados en el perfil de PVAF paterno de cada cromosoma revelan bloques de haplotipos paternos. De forma similar, los valores de PVAF de M1 y M2 segmentados en el perfil de PVAF materno de cada cromosoma revelan bloques de haplotipos maternos. En esta realización, un patrón con valores de PVAF de P1 y P2 segmentados localizados en/alrededor de 0 y 0,5, respectivamente, implica la herencia del alelo del abuelo paterno. Sin embargo, un patrón con valores de PVAF de P1 y P2 segmentados localizados en/alrededor de 0,5 y 1, respectivamente, implica la herencia del alelo de la abuela paterna. De forma similar, un patrón con valores de PVAF de M1 y M2 segmentados localizados en/alrededor de 0 y 0,5, respectivamente, implica la herencia del alelo del abuelo materno. Sin embargo, un patrón con valores de PVAF de M1 y M2 segmentados localizados en/alrededor de 0,5 y 1, respectivamente, implica la herencia del alelo de la abuela materna. 45

*Haplotipado de ADN exento de células, de células individuales, de unas pocas células o multi-celular y determinación del perfil del número de copias de regiones genómicas con una arquitectura alélica desequilibrada*

50 El estado del número de copias específico de alelo/haplotipo así como su origen parental y mecanístico se puede determinar mediante el uso de un método según las realizaciones de la invención. Los desequilibrios alélicos en combinación con los valores del número de copias de ADN (relativos) coincidentes (p.ej. valores de logR) dan como resultado las realizaciones adicionales del método.

En las realizaciones preferidas del método, el método proporciona que se puede determinar que el estado de ploidía de un cromosoma es una monosomía de origen paterno, es decir, una copia paterna y ninguna copia materna, cuando los valores de logR en ese cromosoma son alrededor de -1, pero los valores de PVAF de P1 y P2 segmentados en el perfil de PVAF paterno se solapan y tienen valores de/alrededor de 0 o 1, mientras los valores de PVAF de M1 y M2 segmentados en el perfil de PVAF materno son de/alrededor de 0 y 1, respectivamente. En esta realización, los puntos de ruptura en el perfil de PVAF paterno representan los sitios de recombinación homóloga paterna. En esta realización, si se usa un genotipo del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados y solapantes de/alrededor de 0 55

implica la herencia de los mismos alelos paternos que el hermano de referencia. Sin embargo, un patrón con valores de PVAF de P1 y P2 segmentados y solapantes de/alrededor de 1 implica la herencia de los alelos paternos diferentes del hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados y solapantes de/alrededor de 0 implica la herencia del alelo del abuelo paterno. Sin embargo, un patrón con valores de PVAF de P1 y P2 segmentados y solapantes de/alrededor de 1 implica la herencia del alelo de la abuela materna. La Figura 2 ilustra cuatro perfiles de PVAF paternos probables de una monosomía con un evento hipotético de recombinación homóloga entre las cromátidas no hermanas, suponiendo que se conserva el alelo paterno.

Como se puede observar en la Figura 2, antes de la etapa de reflexión de los valores de PVAF sobre el eje medio, los valores de PVAF que pertenecen a las categorías P1 y P2 están alrededor de 0 y alrededor de 1. La segmentación de dichos datos generaría segmentos de P1 y P2 en 0 y segmentos de P1 y P2 en 1 en las cuatro situaciones presentadas en la Figura 2. Por lo tanto, p.ej. el evento de recombinación homóloga mostrado en las situaciones III y IV no sería visible a partir de los perfiles segmentados. En particular, se debe indicar que la Figura 2 es una representación esquemática de una situación hipotética, "ideal", en la que los valores de PVAF tienen el valor esperado de 1 o 0. Además, cuando se usan los datos de genotipo obtenidos de muestras que tienen una cantidad baja de material genético (tales como muestras de células individuales), los valores de PVAF contendrán ruido, y los patrones antes de la reflexión y segmentación serán aún menos claros. Los métodos de la presente invención permiten por primera vez identificar puntos de ruptura cromosómicos y/o sitios de recombinación homóloga en muestras aneuploides, especialmente cuando dichas muestras comprenden cantidades bajas de material genético, que provocan valores de PVAF bastante distorsionados y con ruido. Estas observaciones también son aplicables a los otros ejemplos descritos en la presente memoria.

En las realizaciones adicionales de un método según la presente invención, se puede determinar que el estado de ploidía de un cromosoma es una monosomía de origen materno, es decir, ninguna copia paterna y una copia materna, cuando los valores de logR en ese cromosoma son alrededor de -1, los valores de PVAF de M1 y M2 segmentados en el perfil de PVAF materno se solapan y tienen valores de/alrededor de 0 o 1, y los valores de PVAF de P1 y P2 segmentados en el perfil de PVAF paterno están separados y tienen valores de/alrededor de 0 y 1, respectivamente. En esta realización, los puntos de ruptura del perfil materno representan los sitios de recombinación homóloga. En esta realización, si se usa un genotipo del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de M1 y M2 segmentados y solapantes de/alrededor de 0 implica la herencia de los mismos alelos maternos que el hermano de referencia. Sin embargo, un patrón con valores de PVAF de M1 y M2 segmentados y solapantes de/alrededor de 1 implica la herencia de los alelos maternos diferentes del hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de M1 y M2 segmentados y solapantes de/alrededor de 0 implica la herencia del alelo del abuelo materno. Sin embargo, un patrón con valores de PVAF de M1 y M2 segmentados y solapantes de/alrededor de 1 implica la herencia del alelo de la abuela materna. La Figura 2 ilustra cuatro perfiles de PVAF maternos probables de una monosomía con un evento hipotético de recombinación homóloga entre las cromátidas no hermanas, suponiendo que se conserva el alelo materno.

En las realizaciones preferidas adicionales, se puede determinar que un cromosoma es una isodisomía uniparental (UPID) de origen paterno, es decir, dos copias paternas (iguales) y ninguna copia materna, cuando los valores de logR en ese cromosoma son alrededor de 0, los valores de PVAF de P1 y P2 segmentados en el perfil paterno son solapantes y tienen valores de/alrededor de 0 o 1, y los valores de PVAF de M1 y M2 segmentados en el perfil materno están separados y tienen valores de/alrededor de 0 y 1, respectivamente, para los mismos segmentos de PVAF de P1 y P2. En esta realización, los puntos de ruptura del perfil paterno representan los sitios de recombinación homóloga. En esta realización, si se usa un genotipo de PV del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados y solapantes de/alrededor de 0 implica la herencia de los mismos alelos paternos que el hermano de referencia. Sin embargo, un patrón con valores de PVAF de P1 y P2 segmentados y solapantes de/alrededor de 1 implica la herencia de los alelos maternos diferentes del hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados y solapantes de/alrededor de 0 implica la herencia del alelo del abuelo paterno. Sin embargo, un patrón con valores de PVAF de P1 y P2 segmentados y solapantes de/alrededor de 1 implica la herencia del alelo de la abuela paterna. La Figura 3 ilustra cuatro perfiles de PVAF paternos muy probables de un cromosoma con una anomalía UPID con un evento hipotético de recombinación homóloga entre las cromátidas no hermanas, suponiendo que solamente se conservan los cromosomas paternos.

En las realizaciones preferidas, se determina que un cromosoma es una isodisomía uniparental (UPID) de origen materno, es decir, dos copias maternas (iguales) y ninguna copia paterna, cuando los valores de logR en ese cromosoma son de/alrededor de 0, los valores de PVAF de M1 y M2 segmentados en el perfil de PVAF materno son solapantes y tienen valores de/alrededor de 0 o uno, mientras los valores de PVAF de P1 y P2 segmentados en el perfil de PVAF paterno están separados y tienen valores de/alrededor de 0 y 1, respectivamente, para los mismos segmentos de PVAF de M1 y M2. En esta realización, los puntos de ruptura del perfil materno representan los sitios de recombinación homóloga. En esta realización, si se usa un genotipo del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de M1 y M2 segmentados y

- solapantes de/alrededor de 0 implica la herencia de los mismos alelos maternos que el hermano de referencia. Sin embargo, un patrón con valores de PVAF de M1 y M2 segmentados y solapantes de/alrededor de 1 implica la herencia de los alelos maternos diferentes del hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de M1 y M2 segmentados y solapantes de/alrededor de 0 implica la herencia de los alelos del abuelo materno. Sin embargo, un patrón con valores de PVAF de M1 y M2 segmentados y solapantes de/alrededor de 1 implica la herencia de los alelos de la abuela materna. La Figura 3 ilustra cuatro perfiles de PVAF maternos muy probables de un cromosoma con una anomalía UPiD con un evento hipotético de recombinación homóloga entre las cromátidas no hermanas, suponiendo que solamente se conservan los cromosomas maternos.
- 5
- 10 Las realizaciones de un método según la invención proporcionan que se puede determinar que un cromosoma es una heterodisomía uniparental (UPhD) de origen paterno, es decir, dos copias paternas (distintas) (es decir, ambos homólogos paternos) y ninguna copia materna, cuando los valores de logR en ese cromosoma son de/alrededor de 0, los valores de PVAF de P1 y P2 segmentados en el perfil de PVAF paterno son solapantes y son de/alrededor de 0,5 en posición proximal al centrómero, y los valores de PVAF de P1 y P2 segmentados en el perfil paterno son solapantes y son de/alrededor de 0 o 1 en posición distal al centrómero si hay presente un sitio de recombinación homóloga. Además, los valores de PVAF de M1 y M2 segmentados en el perfil materno están separados y son de/alrededor de 0 y 1, respectivamente. La Figura 4 ilustra cuatro perfiles de PVAF paternos muy probables de un cromosoma con una anomalía UPhD con un evento hipotético de recombinación homóloga entre las cromátidas no hermanas, suponiendo que solamente se conservan los cromosomas paternos.
- 15
- 20 Las realizaciones de un método según la invención proporcionan que se determina que un cromosoma es una heterodisomía uniparental (UPhD) de origen materno, es decir, ninguna copia paterna y dos copias maternas (distintas) (es decir, ambos homólogos maternos), cuando los valores de logR en ese cromosoma son de/alrededor de 0, los valores de PVAF de M1 y M2 segmentados en el perfil materno son solapantes y son de/alrededor de 0,5 en posición proximal al centrómero, y los valores de PVAF de M1 y M2 segmentados en el perfil materno son solapantes y tienen valores de/alrededor de 0 o 1 en posición distal al centrómero si hay presente un sitio de recombinación homóloga. Además, los valores de PVAF de P1 y P2 segmentados en el perfil de PVAF paterno están separados y tienen valores de/alrededor de 0 y 1, respectivamente. La Figura 4 ilustra cuatro perfiles de PVAF maternos muy probables de un cromosoma con una anomalía UPhD con un evento hipotético de recombinación homóloga entre las cromátidas no hermanas, suponiendo que solamente se conservan los cromosomas maternos.
- 25
- 30 Las realizaciones de un método según la invención proporcionan que se puede determinar que el estado de ploidía de un cromosoma es una trisomía de origen completamente paterno, es decir, la presencia de tres cromosomas paternos idénticos y ninguna copia materna, cuando los valores de logR en ese cromosoma son alrededor de 0,58, los valores de PVAF de P1 y P2 segmentados en el perfil de PVAF paterno son solapantes y son de/alrededor de 0 o 1, y los valores de PVAF de M1 y M2 segmentados en el perfil de PVAF materno están separados en/alrededor de 0 y 1, respectivamente. En esta realización, los puntos de ruptura del perfil paterno representan los sitios de recombinación homóloga. En esta realización, si se usa un genotipo del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados y solapantes de/alrededor de 0 implica la herencia de los mismos alelos paternos que el hermano de referencia. Y un patrón con valores de PVAF de P1 y P2 segmentados y solapantes de/alrededor de 1 implica la herencia de los alelos paternos diferentes de los heredados por el hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados y solapantes de/alrededor de 0 implica la herencia del alelo del abuelo paterno. Y un patrón con valores de PVAF de P1 y P2 segmentados y solapantes de/alrededor de 1 implica la herencia del alelo de la abuela materna. En esta realización, los perfiles de PVAF parentales finales son similares a los perfiles de PVAF paternos de la monosomía paterna o los perfiles de PVAF de UPiD (véanse, p.ej., las Figs. 2 y 3). Sin embargo, los valores de logR segmentados son de/alrededor de 0,58.
- 35
- 40
- 45
- Las realizaciones de un método según la invención proporcionan que se puede determinar que el estado de ploidía de un cromosoma es una trisomía de origen completamente materno, es decir, la presencia de tres cromosomas maternos idénticos y ninguna copia paterna, cuando los valores de logR en ese cromosoma son alrededor de 0,58, los valores de PVAF de M1 y M2 segmentados en el perfil de PVAF materno son solapantes y son de/alrededor de 0 o 1, y los valores de PVAF de P1 y P2 segmentados en el perfil de PVAF paterno están separados en/alrededor de 0 y 1, respectivamente. En esta realización, los puntos de ruptura del perfil materno representan los sitios de recombinación homóloga. En esta realización, si se usa un genotipo del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de M1 y M2 segmentados y solapantes de/alrededor de 0 implica la herencia de los mismos alelos maternos que el hermano de referencia. Y un patrón con valores de PVAF de M1 y M2 segmentados y solapantes de/alrededor de 1 implica la herencia de los alelos maternos diferentes de los heredados por el hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de M1 y M2 segmentados y solapantes de/alrededor de 0 implica la herencia del alelo del abuelo materno. Y un patrón con valores de PVAF de M1 y M2 segmentados y solapantes de/alrededor de 1 implica la herencia del alelo de la abuela materna. En esta realización, los perfiles de PVAF maternos finales son similares a los perfiles de PVAF maternos de la monosomía materna o los perfiles de PVAF de UPiD (véanse, p.ej., las Figs. 2 y 3), ya que solamente se conservan las copias maternas. Sin embargo, los valores de logR segmentados son de/alrededor de
- 50
- 55
- 60

0,58.

Las realizaciones de un método según la invención proporcionan que se puede determinar que el estado de ploidía de un cromosoma es una trisomía paterna, es decir, la presencia de dos copias paternas y una copia materna, con origen meiótico I cuando los valores de logR en ese cromosoma son alrededor de 0,58, los valores de PVAF de P1 y P2 segmentados en el perfil paterno son respectivamente alrededor de 0,33 y 0,67 en posición proximal al centrómero y son alrededor de "0 y 0,33" o "0,67 y 1" en posición distal al centrómero después de un sitio de recombinación homóloga. En esta realización, los valores de PVAF de M1 y M2 segmentados en el perfil materno son respectivamente de/alrededor de 0 y 0,67 o de/alrededor de 0,33 y 1. En esta realización, los puntos de ruptura en los perfiles de PVAF paternos y maternos representan sitios de recombinación homóloga. En esta realización, si se usa un genotipo del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón de PVAF con valores de PVAF de M1 y M2 segmentados de/alrededor de 0 y 0,67, respectivamente, implica la herencia de los mismos alelos maternos que el hermano de referencia. Y un patrón con valores de PVAF de M1 y M2 segmentados de/alrededor de 0,33 y 1, respectivamente, implica la herencia de los alelos maternos diferentes de los heredados por el hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de M1 y M2 segmentados de/alrededor de 0 y 0,67 implica la herencia del alelo del abuelo materno. Y un patrón con valores de PVAF de M1 y M2 segmentados de/alrededor de 0,33 y 1 implica la herencia del alelo de la abuela materna. La Figura 5 ilustra cuatro perfiles de PVAF paternos muy probables de una trisomía meiótica I con un evento hipotético de recombinación homóloga entre las cromátidas no hermanas, suponiendo que se transmite un cromosoma paterno adicional.

De forma similar a los comentarios anteriores sobre la Figura 2, sin la reflexión de ciertos valores de PVAF (basándose en el genotipo parental), no se pueden visualizar los eventos de recombinación en las situaciones II y III. La presente invención permite identificar el estado de ploidía (trisomía paterna) así como el haplotipo mediante el uso de un único método. La identificación del haplotipo de muestras aneuploides, incluso cuando tales muestras comprenden cantidades bajas de material genético, es una mejora respecto de la técnica anterior. Esto puede ser importante, por ejemplo, en el diagnóstico preimplantacional cuando el embrión tiene una trisomía X. Como la trisomía X conlleva diferencias fenotípicas limitadas o inexistentes, todavía se puede implantar un embrión con trisomía X. El conocer de manera combinada el haplotipo de los cromosomas X puede proporcionar información valiosa sobre la transmisión de una región cromosómica que porta un trastorno genético asociado a X de uno de los progenitores al embrión.

Las realizaciones de un método según la invención proporcionan que se puede determinar que el estado de ploidía de un cromosoma es una trisomía materna, es decir, la presencia de dos copias maternas y una copia paterna, con un origen meiótico I cuando los valores de logR en ese cromosoma son de/alrededor de 0,58, los valores de PVAF de M1 y M2 segmentados en el perfil de PVAF materno son respectivamente de/alrededor de 0,33 y 0,67 en posición proximal al centrómero y son de/alrededor de "0 y 0,33" o "0,67 y 1" en posición distal al centrómero después de haberse dado un sitio de recombinación homóloga. En esta realización, los valores de PVAF de P1 y P2 segmentados en el perfil paterno son respectivamente de/alrededor de 0 y 0,67 o de/alrededor de 0,33 y 1. En esta realización, los puntos de ruptura en los perfiles paternos y maternos representan sitios de recombinación homóloga. En esta realización, si se usa un genotipo del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0 y 0,67, respectivamente, implica la herencia de los mismos alelos paternos que el hermano de referencia. Y un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0,33 y 1, respectivamente, implica la herencia de los alelos paternos diferentes de los heredados por el hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0 y 0,67 implica la herencia del alelo del abuelo paterno. Y un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0,33 y 1 implica la herencia del alelo de la abuela paterna. La Figura 5 ilustra cuatro perfiles de PVAF maternos muy probables de una trisomía meiótica I con un evento hipotético de recombinación homóloga entre las cromátidas no hermanas, suponiendo que se transmite un cromosoma materno adicional.

Las realizaciones de un método según la invención proporcionan que se puede determinar que el estado de ploidía de un cromosoma es una trisomía paterna, es decir, la presencia de dos copias paternas y una copia materna, con un origen meiótico II cuando los valores de logR en ese cromosoma son alrededor de 0,58, los valores de PVAF de P1 y P2 segmentados en el perfil paterno son de/alrededor de 0 y 0,33 o de/alrededor de 0,67 y 1 en posición proximal al centrómero y son de/alrededor de 0,33 y 0,67 en posición distal al centrómero después de que haya presente un sitio de recombinación homóloga. En esta realización, los valores de PVAF de M1 y M2 segmentados en el perfil materno son respectivamente de alrededor de 0 y 0,67 o de/alrededor de 0,33 y 1. En esta realización, los puntos de ruptura en los perfiles paternos y maternos representan sitios de recombinación homóloga. En esta realización, si se usa un genotipo del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de M1 y M2 segmentados de/alrededor de 0 y 0,67, respectivamente, implica la herencia de los mismos alelos maternos que el hermano de referencia. Sin embargo, un patrón con valores de PVAF de M1 y M2 segmentados de/alrededor de 0,33 y 1, respectivamente, implica la herencia de los alelos maternos diferentes de los heredados por el hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de

PVAF de M1 y M2 segmentados de/alrededor de 0 y 0,67 implica la herencia del abuelo materno. Sin embargo, un patrón con valores de PVAF de M1 y M2 segmentados de/alrededor de 0,33 y 1 implica la herencia de la abuela materna. La Figura 6 ilustra dos perfiles de PVAF paternos muy probables de una trisomía meiótica II con un evento hipotético de recombinación homóloga entre las cromátidas no hermanas, suponiendo que se transmite un cromosoma paterno adicional.

Las realizaciones de un método según la invención proporcionan que se puede determinar que el estado de ploidía de un cromosoma es una trisomía materna, es decir, la presencia de una copia paterna y dos copias maternas, de origen meiótico II cuando los valores de logR en ese cromosoma son de/alrededor de 0,58, los valores de PVAF de M1 y M2 segmentados en el perfil materno son de/alrededor de 0 y 0,33 o de/alrededor de 0,67 y 1 en posición proximal al centrómero, y son de/alrededor de 0,33 y 0,67 en posición distal al centrómero después de haberse dado un sitio de recombinación homóloga. En esta realización, los valores de PVAF de P1 y P2 segmentados en el perfil paterno son respectivamente de alrededor de 0 y 0,67 o de/alrededor de 0,33 y 1. En esta realización, los puntos de ruptura en los perfiles paternos y maternos representan sitios de recombinación homóloga. En esta realización, si se usa un genotipo del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0 y 0,67, respectivamente, implica la herencia de los mismos alelos paternos que el hermano de referencia. Sin embargo, un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0,33 y 1, respectivamente, implica la herencia de los alelos paternos diferentes de los heredados por el hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0 y 0,67 implica la herencia del alelo del abuelo paterno. Sin embargo, un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0,33 y 1 implica la herencia del alelo de la abuela paterna. La Figura 6 ilustra dos perfiles de PVAF maternos muy probables de una trisomía meiótica II con un evento hipotético de recombinación homóloga entre las cromátidas no hermanas, suponiendo que se transmite un cromosoma materno adicional.

Las realizaciones de un método según la invención proporcionan que se puede determinar que el estado de ploidía de un cromosoma es una trisomía paterna, es decir, la presencia de dos copias paternas y una copia materna, con un origen mitótico cuando los valores de logR en ese cromosoma son de alrededor de 0,58, los valores de PVAF de P1 y P2 segmentados en el perfil paterno son de/alrededor de 0 y 0,33 o de/alrededor de 0,67 y 1, respectivamente. En esta realización, los valores de PVAF de M1 y M2 segmentados en el perfil materno son respectivamente de/alrededor de 0 y 0,67 o de/alrededor de 0,33 y 1. En esta realización, los puntos de ruptura en los perfiles paternos y maternos representan sitios de recombinación homóloga. En esta realización, si se usa un genotipo del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0 y 0,33, respectivamente, implica la herencia y duplicación de los mismos alelos paternos que los heredados por el hermano de referencia. Sin embargo, un patrón con valores de PVAF de P1 y P2 segmentados de alrededor de 0,67 y 1, respectivamente, implica la herencia y duplicación de los alelos paternos diferentes de los heredados por el hermano de referencia. Además, en el perfil materno un patrón con valores de PVAF de M1 y M2 segmentados de/alrededor de 0 y 0,67, respectivamente, implica la herencia de los mismos alelos maternos que el hermano de referencia. Sin embargo, un patrón con valores de PVAF de M1 y M2 segmentados de/alrededor de 0,33 y 1, respectivamente, implica la herencia de los alelos maternos diferentes de los heredados por el hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados de alrededor de 0 y 0,33 implica la herencia de los alelos del abuelo materno. Sin embargo, un patrón con valores de PVAF de P1 y P2 segmentados de alrededor de 0,67 y 1 implica la herencia de los alelos de la abuela paterna. Además, en el perfil materno, un patrón con valores de PVAF de M1 y M2 segmentados de alrededor de 0 y 0,67 implica la herencia de los alelos del abuelo materno. Sin embargo, un patrón con valores de PVAF de M1 y M2 segmentados de alrededor de 0,33 y 1 implica la herencia de los alelos de la abuela materna. La Figura 7 ilustra dos perfiles de PVAF paternos muy probables de trisomía mitótica con un evento hipotético de recombinación homóloga entre las cromátidas no hermanas, suponiendo que se duplica el cromosoma paterno.

Las realizaciones de un método según la invención proporcionan que se puede determinar que el estado de ploidía de un cromosoma es una trisomía materna, es decir, la presencia de una copia paterna y dos copias maternas, con un origen mitótico cuando los valores de logR en ese cromosoma son de/alrededor de 0,58, los valores de PVAF de M1 y M2 segmentados en el perfil de PVAF materno son de alrededor de 0 y 0,33 o de alrededor de 0,67 y 1, respectivamente. En esta realización, los valores de PVAF de P1 y P2 segmentados en el perfil de PVAF paterno son respectivamente de/alrededor de 0 y 0,67 o de/alrededor de 0,33 y 1. En esta realización, los puntos de ruptura en los perfiles de PVAF paternos y maternos representan sitios de recombinación homóloga. En esta realización, si se usa un genotipo del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de M1 y M2 segmentados de/alrededor de 0 y 0,33, respectivamente, implica la herencia y duplicación de los mismos alelos maternos que los portados por el hermano de referencia. Sin embargo, un patrón con valores de PVAF de M1 y M2 segmentados de/alrededor de 0,67 y 1, respectivamente, implica la herencia de los alelos maternos diferentes de los heredados por el hermano de referencia. Además, en el perfil de PVAF paterno un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0 y 0,67, respectivamente, implica la herencia de los mismos alelos paternos que el hermano de referencia. Sin embargo, un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0,33 y 1, respectivamente, implica la herencia de los alelos

paternos diferentes de los heredados por el hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de M1 y M2 segmentados de alrededor de 0 y 0,33 implica la herencia de los alelos del abuelo materno. Sin embargo, un patrón con valores de PVAF de M1 y M2 segmentados de alrededor de 0,67 y 1 implica la herencia de los alelos de la abuela materna. Además, en el perfil paterno, un patrón con valores de PVAF de P1 y P2 segmentados de alrededor de 0 y 0,67 implica la herencia de los alelos del abuelo paterno. Sin embargo, un patrón con valores de PVAF de P1 y P2 segmentados de alrededor de 0,33 y 1 implica la herencia de los alelos de la abuela paterna. La Figura 7 ilustra dos perfiles de PVAF maternos muy probables de trisomía mitótica con un evento hipotético de recombinación homóloga entre las cromátidas no hermanas, suponiendo que se duplica el cromosoma materno.

Las realizaciones de un método según la invención proporcionan que se puede determinar que el estado de ploidía de un cromosoma es tetrasomía, es decir, la presencia de dos copias paternas y dos copias maternas, cuando los valores de logR en ese cromosoma son de alrededor de 1. En esta realización, los valores de PVAF de P1 y P2 segmentados en el perfil paterno son respectivamente de/alrededor de 0 y 0,5 o de/alrededor de 0,5 y 1. De forma similar, los valores de PVAF de M1 y M2 segmentados en el perfil materno son respectivamente de/alrededor de 0 y 0,5 o de/alrededor de 0,5 y 1. En esta realización, los puntos de ruptura en los perfiles paternos y maternos representan sitios de recombinación homóloga. En esta realización, si se usa un genotipo del hermano de referencia para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0 y 0,5, respectivamente, implica la herencia de los mismos alelos paternos que los portados por el hermano de referencia. Y un patrón con valores de PVAF de P1 y P2 segmentados de/alrededor de 0,5 y 1, respectivamente, implica la herencia de los alelos paternos diferentes de los heredados por el hermano de referencia. De forma similar, un patrón con valores de PVAF de M1 y M2 segmentados de alrededor de 0 y 0,5, respectivamente, implica la herencia de los mismos alelos maternos que los portados por el hermano de referencia. Y un patrón con valores de PVAF de M1 y M2 segmentados de alrededor de 0,5 y 1, respectivamente, implica la herencia de los alelos maternos diferentes de los heredados por el hermano de referencia. En esta realización, si se usan los genotipos de los abuelos para la determinación de la fase de los genotipos parentales, un patrón con valores de PVAF de P1 y P2 segmentados de alrededor de 0 y 0,5, respectivamente, implica la herencia del alelo del abuelo paterno. Y un patrón con valores de PVAF de P1 y P2 segmentados de alrededor de 0,5 y 1, respectivamente, implica la herencia del alelo de la abuela paterna. De forma similar, un patrón con valores de PVAF de M1 y M2 segmentados de alrededor de 0 y 0,5, respectivamente, implica la herencia del alelo del abuelo materno. Y un patrón con valores de PVAF de M1 y M2 segmentados de alrededor de 0,5 y 1, respectivamente, implica la herencia del alelo de la abuela materna. En esta realización, los perfiles de PVAF parentales finales son similares a los perfiles de PVAF parentales de un cromosoma disómico normal. Sin embargo, los valores de logR segmentados son de/alrededor de 1.

En ciertas realizaciones, se puede determinar el estado de ploidía y su origen mediante el uso simplemente de datos de tres personas, es decir, los progenitores y un descendiente. En esta realización, el genotipo del descendiente se usa respecto de los genotipos parentales en fase. Mediante la aplicación de los principios anteriormente mencionados (ii a ix) en las frecuencias del alelo B de PV obtenidas a partir de ADN exento de células, de células individuales, de unas pocas células o multi-celular del embrión/feto/descendencia/tumor, se puede revelar el origen parental y mecánico de una anomalía cromosómica del número de copias o una anomalía sin cambios en el número de copias.

#### *Haplotipo multi-celular y determinación del perfil del número de copias para detectar una arquitectura en mosaico en una masa de células*

Las realizaciones de un método según la invención proporcionan que se puede determinar que un cromosoma es una monosomía en mosaico de origen materno. En esta realización, una masa de células puede comprender, por ejemplo, dos poblaciones de células: (1) un  $p\%$  de células anormales, es decir, células con una copia materna y sin copia paterna, y (2)  $(100 - p)\%$  células normales, es decir, células con una copia materna y una copia paterna. En esta realización, los valores de logR en ese cromosoma son de 0 a -1 (dependiendo del grado de mosaicismo,  $p\%$ ), los valores de PVAF de M1 y M2 segmentados en el perfil de PVAF materno se aproximan entre sí y tienen una distancia total de  $<0,5$ , mientras los valores de PVAF de P1 y P2 segmentados en el perfil de PVAF paterno se separan entre sí y tienen una distancia total de  $>0,5$ , para los mismos segmentos de PVAF de M1 y M2. En esta realización, la distancia entre M1 y M2 y entre P1 y P2, así como los valores de logR segmentados, reflejan el grado de mosaicismo,  $p\%$ . En esta realización, los puntos de ruptura representan preferiblemente los sitios de recombinación homóloga.

Las realizaciones de un método según la invención proporcionan que se puede determinar que un cromosoma es una monosomía en mosaico de origen paterno. En esta realización, una masa de células puede comprender, por ejemplo, dos poblaciones de células: (1) un  $p\%$  de células anormales, es decir, células con una copia paterna y sin copia materna, y (2)  $(100 - p)\%$  células normales, es decir, células con una copia materna y una copia paterna. En esta realización, los valores de logR en ese cromosoma son de 0 a -1 (dependiendo del grado de mosaicismo,  $p\%$ ), los valores de PVAF de P1 y P2 segmentados en el perfil de PVAF materno se aproximan entre sí y por tanto tienen una distancia total de  $<0,5$ , mientras los valores de PVAF de M1 y M2 segmentados en el perfil de PVAF paterno se separan entre sí y tienen una distancia total de  $>0,5$ , para los mismos segmentos de PVAF de P1 y P2. En esta realización, la distancia entre M1 y M2 y entre P1 y P2, así como los valores de logR segmentados, reflejan el grado

de mosaicismo,  $\rho\%$ . En esta realización, los puntos de ruptura representan los sitios de recombinación homóloga.

Las realizaciones de un método según la invención proporcionan que se puede determinar que un cromosoma es una disomía uniparental de origen materno. En esta realización, una masa de células puede comprender, por ejemplo, dos poblaciones de células: (1) un  $\rho\%$  de células UPD anormales, es decir, células con dos copias maternas idénticas y sin copia paterna, y (2)  $(100 - \rho)\%$  células normales, es decir, células con una copia materna y una copia paterna. En esta realización, los valores de logR en ese cromosoma son de/alrededor de 0, los valores de PVAf de M1 y M2 segmentados en el perfil de PVAf materno se aproximan entre sí y tienen una distancia total de  $<0,5$ , mientras los valores de PVAf de P1 y P2 segmentados en el perfil de PVAf paterno se separan entre sí y tienen una distancia total de  $>0,5$ , para los mismos segmentos de PVAf de M1 y M2. En esta realización, la distancia entre M1 y M2 y la distancia entre P1 y P2 refleja el grado de mosaicismo,  $\rho\%$ . En esta realización, los puntos de ruptura representan los sitios de recombinación homóloga. La Figura 8 ilustra una UPD en mosaico con un grado de mosaicismo  $\rho=50\%$ .

Las realizaciones de un método según la invención proporcionan que se puede determinar que un cromosoma es una disomía uniparental de origen paterno. En esta realización, una masa de células puede comprender dos poblaciones: (1) un  $\rho\%$  de células UPD anormales, es decir, células con dos copias paternas idénticas y sin copia materna, y (2)  $(100 - \rho)\%$  células normales, es decir, células con una copia materna y una copia paterna. En esta realización, los valores de logR en ese cromosoma son de/alrededor de 0, los valores de PVAf de P1 y P2 segmentados en el perfil de PVAf paterno se aproximan entre sí y tienen una distancia total de  $<0,5$ , mientras los valores de PVAf de M1 y M2 segmentados en el perfil de PVAf materno se separan entre sí y tienen una distancia total de  $>0,5$ , para los mismos segmentos de PVAf de M1 y M2. En esta realización, la distancia entre M1 y M2 y la distancia entre P1 y P2 refleja el grado de mosaicismo,  $\rho\%$ . En esta realización, los puntos de ruptura representan los sitios de recombinación homóloga. La Figura 8 ilustra una UPD en mosaico con un grado de mosaicismo  $\rho=50\%$ .

Las realizaciones de un método según la invención proporcionan que se puede determinar que un cromosoma es una trisomía en mosaico de origen materno. En esta realización, una masa de células puede comprender, por ejemplo, dos poblaciones: (1) un  $\rho\%$  de células con trisomía anormales, es decir, células con dos copias maternas y una copia paterna, y (2)  $(100 - \rho)\%$  células normales, es decir, células con una copia materna y una copia paterna. En esta realización, los valores de logR en ese cromosoma son entre 0 y 0,58, los valores de PVAf de M1 y M2 segmentados en el perfil de PVAf materno se aproximan entre sí y tienen una distancia total de  $<0,5$ , mientras los valores de PVAf de P1 y P2 segmentados en el perfil de PVAf paterno se separan entre sí y tienen una distancia total de  $>0,5$ , para los mismos segmentos de PVAf de M1 y M2. En esta realización, los valores de logR segmentados, la distancia entre M1 y M2 y la distancia entre P1 y P2 reflejan el grado de mosaicismo,  $\rho\%$ . En esta realización, los puntos de ruptura representan los sitios de recombinación homóloga.

Las realizaciones de un método según la invención proporcionan que se puede determinar que un cromosoma es una trisomía en mosaico de origen paterno. En esta realización, una masa de células puede comprender, por ejemplo, dos poblaciones: (1) un  $\rho\%$  de células con trisomía anormales, es decir, células con dos copias paternas y una copia materna, y (2)  $(100 - \rho)\%$  células normales, es decir, células con una copia materna y una copia paterna. En esta realización, los valores de logR en ese cromosoma son entre 0 y 0,58, los valores de PVAf de P1 y P2 segmentados en el perfil de PVAf paterno se aproximan entre sí y tienen una distancia total de  $<0,5$ , mientras los valores de PVAf de M1 y M2 segmentados en el perfil de PVAf materno se separan entre sí y tienen una distancia total de  $>0,5$ , para los mismos segmentos de PVAf de P1 y P2. En esta realización, los valores de logR segmentados, la distancia entre M1 y M2 y la distancia entre P1 y P2 reflejan el grado de mosaicismo,  $\rho\%$ . En esta realización, los puntos de ruptura representan los sitios de recombinación homóloga.

#### *Procedimiento de normalización opcional*

En las realizaciones preferidas adicionales, en las que se usan muestras de ADN obtenidas de células con inestabilidad cromosómica, tales como, p.ej., muestras de ADN obtenidas de embriones en etapa de segmentación humanos o células cancerosas, los valores del número de copias de ADN (relativos) (p.ej. valores de logR) se normalizan preferiblemente para el análisis adecuado del número de copias. En esta realización, el proceso de normalización comprende preferiblemente la determinación de los cromosomas disómicos normales más probables para la corrección del valor de logR. Los valores del progenitor de origen específicos de alelo se pueden determinar como se describió en Voet et al en "*Breakage-fusion-bridge cycles leading to inv dup del occur in human cleavage stage embryos*", Human mutation 32, 783 (2011). Tras el cálculo de la puntuación parental específica de cromosoma, se calculan las proporciones relativas parentales y se usan para la normalización preliminar de los valores del número de copias de ADN (relativos) (p.ej. valores de logR). En esta realización, se pueden calcular las puntuaciones paternas y maternas,  $PS_k$  y  $MS_k$ , respectivamente, para cada cromosoma k, dados los valores del progenitor de origen específico de alelo:

$$PS_k = \frac{\sum_j P_{k,j}}{\sum_j S_{k,j}}$$

$$MS_k = \frac{\sum_j M_{k,j}}{\sum_j S_{k,j}}$$

5 en la que  $P_{k,j}$  y  $M_{k,j}$  representan los valores del progenitor de origen paterno y materno para el locus  $j$  en el cromosoma  $k$ , respectivamente; y  $S_{k,j}$  es una asignación de PV en el cromosoma  $k$  que es informativa para el análisis del progenitor de origen, que incluye las asignaciones de PV parentales que no son ambas heterocigotas u homocigotas idénticas. En esta realización, el cálculo de la proporción relativa parental posterior puede comprender:

$$Pat_k = \frac{PS_k}{PS_k + MS_k}$$

$$Mat_k = \frac{MS_k}{PS_k + MS_k}$$

10 en la que  $Pat_k$  y  $Mat_k$  son las proporciones relativas paternas y maternas, respectivamente, del cromosoma  $k$ . En una realización preferida, los valores del número de copias de ADN (relativos) en bruto (p.ej. valores de logR) se suavizan mediante el uso de una ventana deslizante. Dichos valores del número de copias de ADN (relativos) suavizados/en bruto (p.ej. valores de logR) se corrigen preferiblemente con respecto al sesgo de % de contenido de GC mediante un ajuste de regresión local y los valores del número de copias de ADN (relativos) (p.ej. valores de logR) se normalizan de manera preliminar mediante el uso de una media (truncada) de los cromosomas disómicos probablemente normales determinados mediante los criterios de puntuación parental anteriormente mencionados. En esta realización, la integrar dichos perfiles de los valores del número de copias de ADN (relativos) preliminares (p.ej. valores de logR) con valores de PVAf haplotipados / en fase y segmentados, se hace una selección final de cromosomas y se usa para la corrección de la media (truncada) consiguiente de dichos valores del número de copias de ADN (relativos) preliminares (p.ej. valores de logR). Además, los valores del número de copias de ADN (relativos) normalizados (p.ej. valores de logR) que se integran posteriormente con valores de PVAf de células individuales haplotipados / en fase y segmentados se pueden usar para asignar anomalías del ADN. Por ejemplo, para loci nulisómicos, monosómicos, disómicos, disómicos uniparentales y trisómicos se pueden esperar patrones típicos en los valores de PVAf haplotipados / en fase y segmentados.

25 Las Figs. 13a-13c ilustran un diagrama de flujo del canal computacional, que comprende un módulo basado en un método según las realizaciones de la invención. Dicho canal computacional proporciona de manera ventajosa el haplotipado de células individuales y la asignación de las variantes patológicas asociadas. Además de los módulos para el control de calidad, genotipado, haplotipado bimodal y visualización de los datos de matrices de SNP, el módulo que constituye un método según las realizaciones de la invención se puede equipar adicionalmente con un módulo de análisis supervisado del número de copias que permite no solamente la identificación de desequilibrios cromosómicos y su origen parental, sino también la detección de anomalías del ADN sin cambios en el número de copias, como las disomías uniparentales (UPDs). Este módulo además puede permitir descubrir la naturaleza meiótica o mitótica de las anomalías cromosómicas.

*Transformación opcional de valores de PVAf*

En las realizaciones opcionales adicionales, los valores de PVAf continuos en las categorías paternas y maternas se pueden transformar en haplotipos paternos y maternos discretos.

35 En esta realización, debido a la paridad y las características de complementariedad (véanse las definiciones) de la presente invención, se puede reconstruir el haplotipo paterno de la muestra de ADN mediante el uso de todos los valores de PVAf genéticos de todo el genoma o una selección de los mismos. Por lo cual, dicha transformación puede comprender:

- i. Calcular de manera iterativa la distancia entre cada valor de PVAf de P1 en bruto con el valor de PVAf de P2 vecino más cercano;
- 40 ii. Asignar las distancias calculadas a los loci de P1 correspondientes;
- iii. Sumar las distancias asignadas a los valores de P1 en bruto correspondientes, concatenar el último al P2 y ordenar toda la matriz basándose en su posición física, y los valores resultantes se denominan P1 ajustado (adj-P1);
- 45 iv. Calcular de manera iterativa la distancia entre cada valor de PVAf de P2 en bruto con el valor de PVAf de P1 vecino más cercano;

- v. Asignar las distancias calculadas a los loci de P2 correspondientes;
- vi. Restar las distancias asignadas de los valores de P2 en bruto correspondientes, concatenar el último al P1 y ordenar toda la matriz basándose en su posición física, y los valores resultantes se denominan P2 ajustado (adj-P2);
- 5 vii. Omitir todos los valores menores de un cierto umbral (p.ej. 0,8) de los adj-P1 y adj-P2 resultantes;
- viii. Omitir todos los valores mayores de un cierto umbral (p.ej. 0,2) de los adj-P1 y adj-P2 resultantes;
- ix. Reasignar todos los valores restantes menores de 0,5 en la categoría paterna a 1 (es decir, el mismo bloque de haplotipos que se transmite al familiar cercano);
- 10 x. Reasignar todos los valores restantes mayores de 0,5 en la categoría paterna a 2 (es decir, un bloque de haplotipos distinto del que se transmite al familiar cercano);

En esta realización, debido a la paridad y las características de complementariedad de la presente invención, se puede reconstruir el haplotipo materno de la muestra de ADN mediante el uso de todos los valores de PVAF genéticos de todo el genoma o una selección de los mismos. Por lo cual, dicha transformación puede comprender:

- 15 i. Calcular de manera iterativa la distancia entre cada valor de PVAF de M1 en bruto con el valor de PVAF de M2 vecino más cercano;
- ii. Asignar las distancias calculadas a los loci de M1 correspondientes;
- iii. Sumar las distancias asignadas a los valores de M1 en bruto correspondientes, concatenar el último al M2 y ordenar toda la matriz basándose en su posición física, y los valores resultantes se denominan M1 ajustado (adj-M1);
- 20 iv. Calcular de manera iterativa la distancia entre cada valor de PVAF de M2 en bruto con el valor de PVAF de M1 vecino más cercano;
- v. Asignar las distancias calculadas a los loci de M2 correspondientes;
- vi. Restar las distancias asignadas de los valores de M2 en bruto correspondientes, concatenar el último al M1 y ordenar toda la matriz basándose en su posición física, y los valores resultantes se denominan M2 ajustado (adj-M2);
- 25 vii. Omitir todos los valores menores de un cierto umbral (p.ej. 0,8) de los adj-M1 y adj-M2 resultantes;
- viii. Omitir todos los valores mayores de un cierto umbral (p.ej. 0,2) de los adj-M1 y adj-M2 resultantes;
- ix. Reasignar todos los valores restantes menores de 0,5 en la categoría materna a 1 (es decir, el mismo bloque de haplotipos que se transmite al familiar cercano);
- 30 x. Reasignar todos los valores restantes mayores de 0,5 en la categoría materna a 2 (es decir, un bloque de haplotipos distinto del que se transmite al familiar cercano);

La Figura 15 ilustra la exactitud de la presente invención en la detección del sitio de recombinación homóloga (sitio HR).

- 35 Se debe entender que esta invención no se limita a las características particulares de los medios y/o las etapas del proceso de los métodos descritos, ya que tales medios y métodos pueden variar. También se debe entender que la terminología usada en la presente memoria tiene el propósito de describir realizaciones particulares únicamente, y no pretende ser limitante. Se debe indicar que, tal como se usa en la memoria descriptiva y las reivindicaciones adjuntas, las formas singulares "un/una", "uno" y "el/la" incluyen las referencias singulares y/o plurales a menos que el contexto lo dicte claramente de otra manera. También se debe entender que las formas plurales incluyen las referencias singulares y/o plurales, a menos que el contexto lo dicte claramente de otra manera. Además se debe
- 40 entender que, en caso de que se proporcionen intervalos de parámetros que están delimitados por valores numéricos, se considera que el intervalo incluye estos valores limitativos.

**REIVINDICACIONES**

1. Un método implementado por ordenador para el análisis del material genético de un sujeto, y dicho método comprende:
- 5 - obtener valores de frecuencias de alelos de variantes polimórficas (PVAF) continuos del material genético del sujeto;
- obtener información del genotipo en fase de un primer progenitor e información del genotipo en fase o no en fase de un segundo progenitor;
- categorizar los valores de PVAF continuos en una primera categoría que corresponde al primer progenitor basándose en la información del genotipo del primer y segundo progenitor;
- 10 - subcategorizar los valores de PVAF continuos de la primera categoría que corresponde al primer progenitor en subcategorías;
- segmentar dichos valores de PVAF subcategorizados; y
- proporcionar los valores de PVAF segmentados para indicar una anomalía genética en el material genético del sujeto y/o la herencia del material genético del sujeto.
- 15 2. El método de la reivindicación 1, en el que la información del genotipo del segundo progenitor es información del genotipo en fase;
- que comprende además
- categorizar los valores de PVAF continuos en una segunda categoría que corresponde al segundo progenitor basándose en la información del genotipo del primer y segundo progenitor; y
- 20 - subcategorizar los valores de PVAF continuos en la primera y segunda categorías en subcategorías.
3. El método de cualquiera de las reivindicaciones previas, que comprende además la reflexión de los valores de PVAF obtenidos respecto del eje medio antes de la segmentación.
4. El método de la reivindicación 1 o 2, que comprende además la reflexión de los valores de PVAF obtenidos, en una posición que corresponde a un genotipo parental en fase específico, respecto del eje medio antes de la segmentación.
- 25 5. El método de cualquiera de las reivindicaciones previas, que comprende además obtener valores de cantidad de ADN, tales como valores de logR o valores de recuentos de lecturas, y normalizar dichos valores de cantidad de ADN basándose en dichos valores de PVAF segmentados.
6. El método de cualquiera de las reivindicaciones previas, en el que los valores de PVAF continuos se han determinado mediante el uso de una muestra que comprende una cantidad baja de material genético de dicho sujeto, tal como una muestra que comprende solamente una o unas pocas células de dicho sujeto, o una muestra de plasma obtenida de una madre embarazada con dicho sujeto.
- 30 7. El método de cualquiera de las reivindicaciones previas, en el que la anomalía genética indicada por los valores de PVAF segmentados comprende una anomalía cromosómica numérica o estructural, o mosaicismo.
- 35 8. El método de cualquiera de las reivindicaciones previas, en el que los valores de PVAF segmentados indican el origen meiótico o mitótico de una anomalía genética en el material genético del sujeto.
9. El método de cualquiera de las reivindicaciones previas, en el que los valores de PVAF segmentados indican el haplotipo del material genético del sujeto.
- 40 10. El método de cualquiera de las reivindicaciones previas, en el que los valores de PVAF segmentados indican el número de copias de un cromosoma o región cromosómica en el material genético del sujeto y/o un sitio de recombinación homóloga o un punto de ruptura cromosómica.
11. El método de la reivindicación 1 o 2, en el que proporcionar los valores de PVAF segmentados comprende evaluar la longitud de segmentos de valores de PVAF segmentados en al menos dos subcategorías de la categoría de un progenitor para indicar un sitio de recombinación homóloga o un punto de ruptura cromosómica.
- 45 12. El método de la reivindicación 2, en el que proporcionar los valores de PVAF segmentados comprende para una región cromosómica:
- determinar una primera distancia entre a) un valor de PVAF segmentado en una primera subcategoría de la categoría del primer progenitor en dicha región cromosómica y b) un valor de PVAF segmentado en una segunda

subcategoría de la categoría del primer progenitor en dicha región cromosómica;

- determinar una segunda distancia entre a) un valor de PVAf segmentado en una primera subcategoría de la categoría del segundo progenitor en dicha región cromosómica y b) un valor de PVAf segmentado en una segunda subcategoría de la categoría del segundo progenitor en dicha región cromosómica; y

- 5 - comparar la primera distancia y la segunda distancia para indicar una anomalía del número de copias de dicha región cromosómica.

13. El método de cualquiera de las reivindicaciones previas, en el que categorizar los valores de PVAf continuos en una categoría que corresponde al primer progenitor comprende:

- 10 - determinar los loci que son informativos para el primer progenitor mediante el uso de la información del genotipo del primer y segundo progenitor; y

- categorizar los valores de PVAf continuos de material genético del sujeto en dichos loci que son informativos para el primer progenitor en una categoría que corresponde al primer progenitor.

14. El método de la reivindicación 1, en el que subcategorizar los valores de PVAf continuos de la categoría que corresponde al primer progenitor comprende:

- 15 - determinar los loci con una combinación de genotipo específico del primer y segundo progenitor;

- subcategorizar los valores de PVAf continuos de material genético del sujeto en dichos loci con una combinación de genotipos específica del primer y segundo progenitor.

15. Un producto de programa informático que es capaz, cuando se ejecuta en un motor de procesamiento, de llevar a cabo el método de cualquiera de las reivindicaciones previas.

- 20 16. Un medio de almacenamiento legible por ordenador no transitorio que almacena el producto de programa informático de la reivindicación 15.

Figura 1

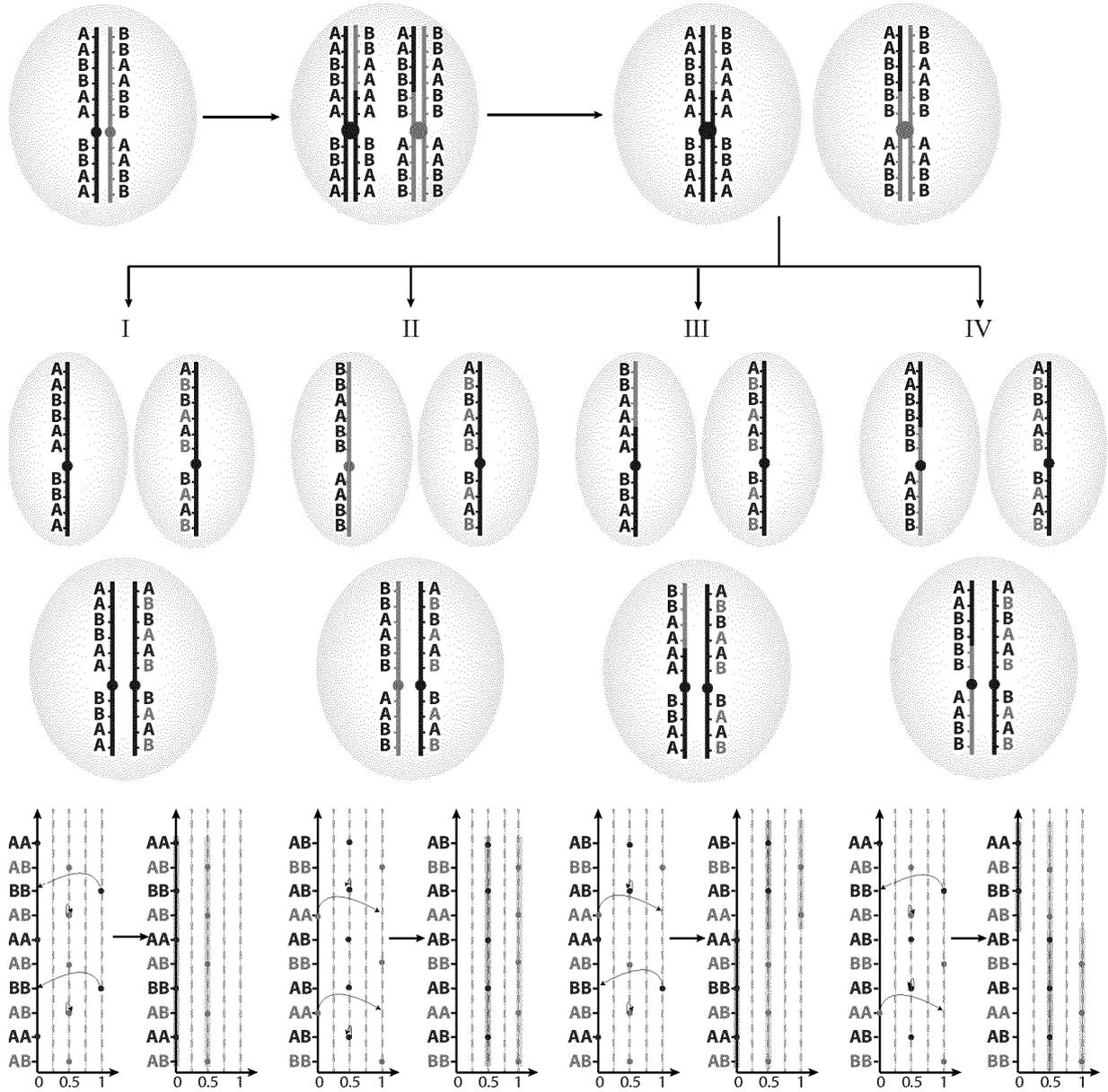


Figura 2

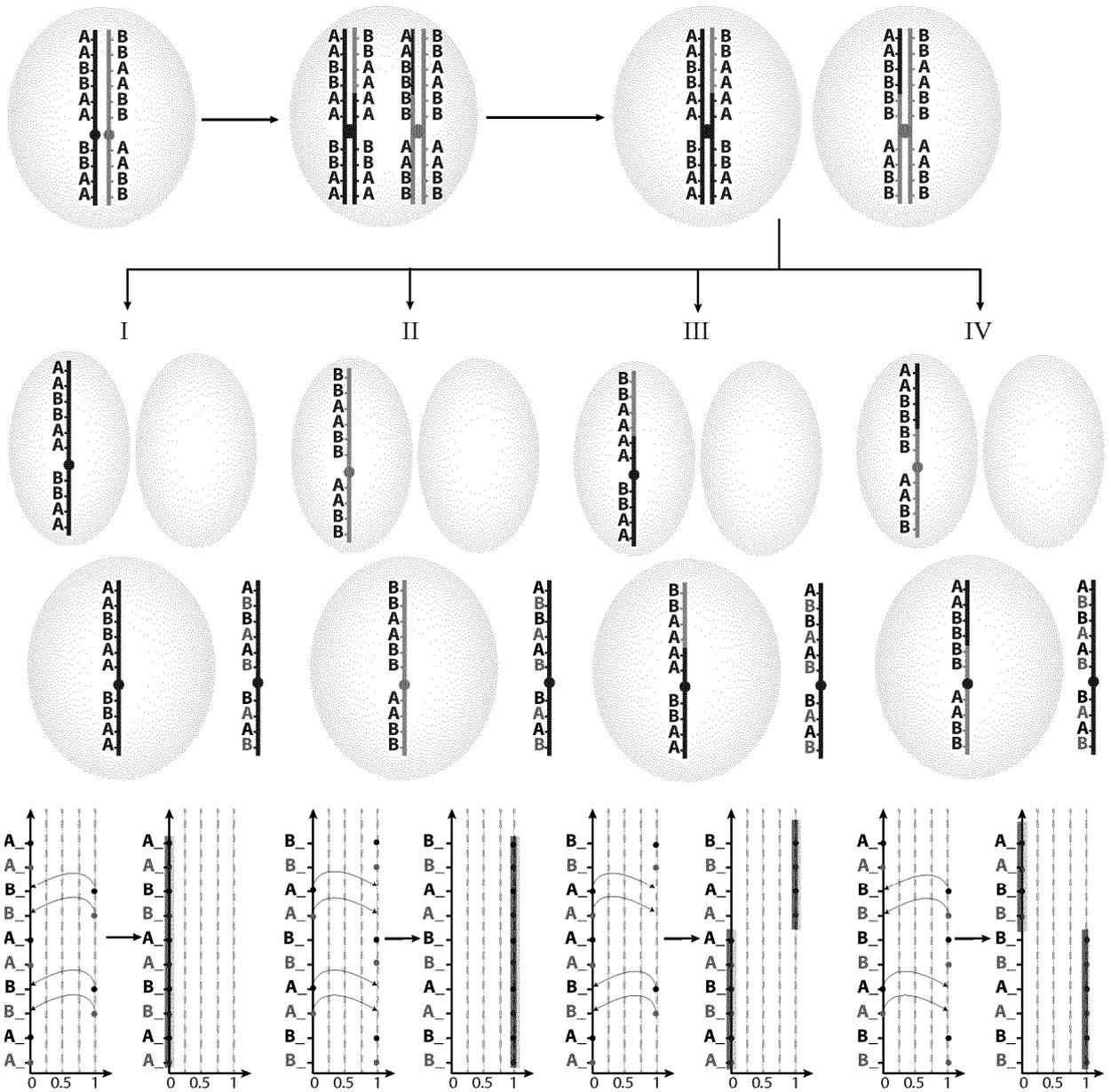


Figura 3

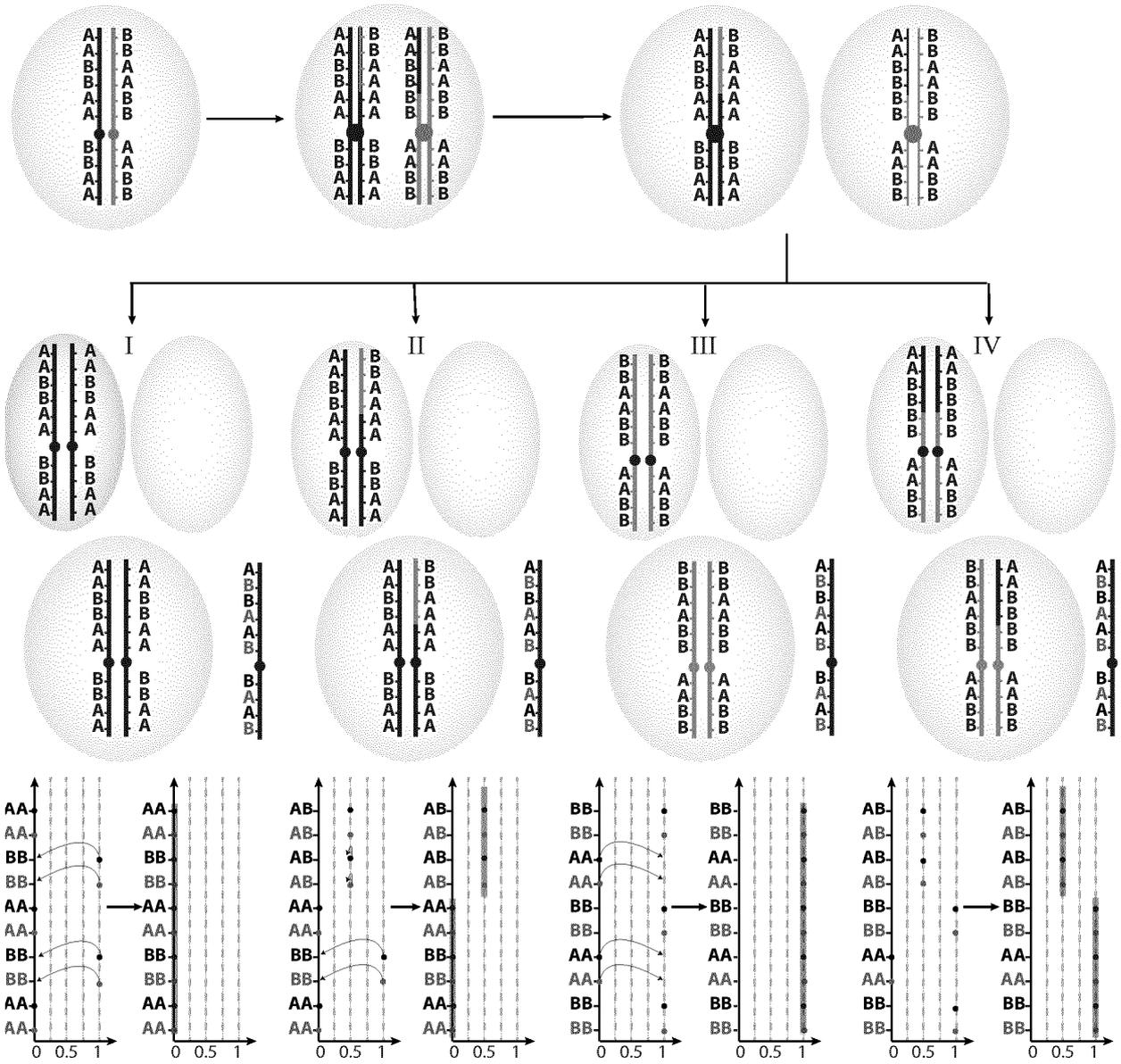


Figura 4

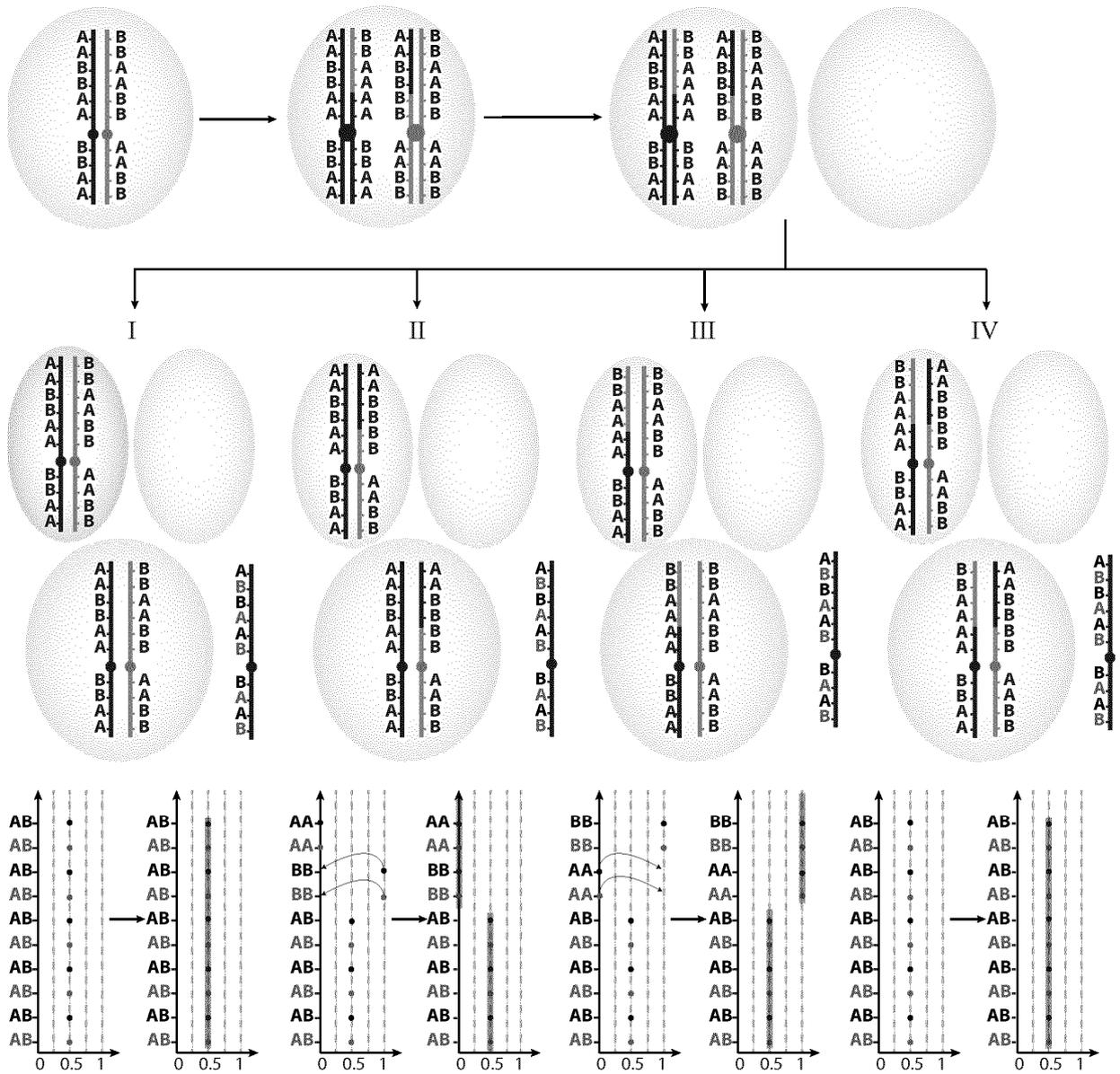


Figura 5

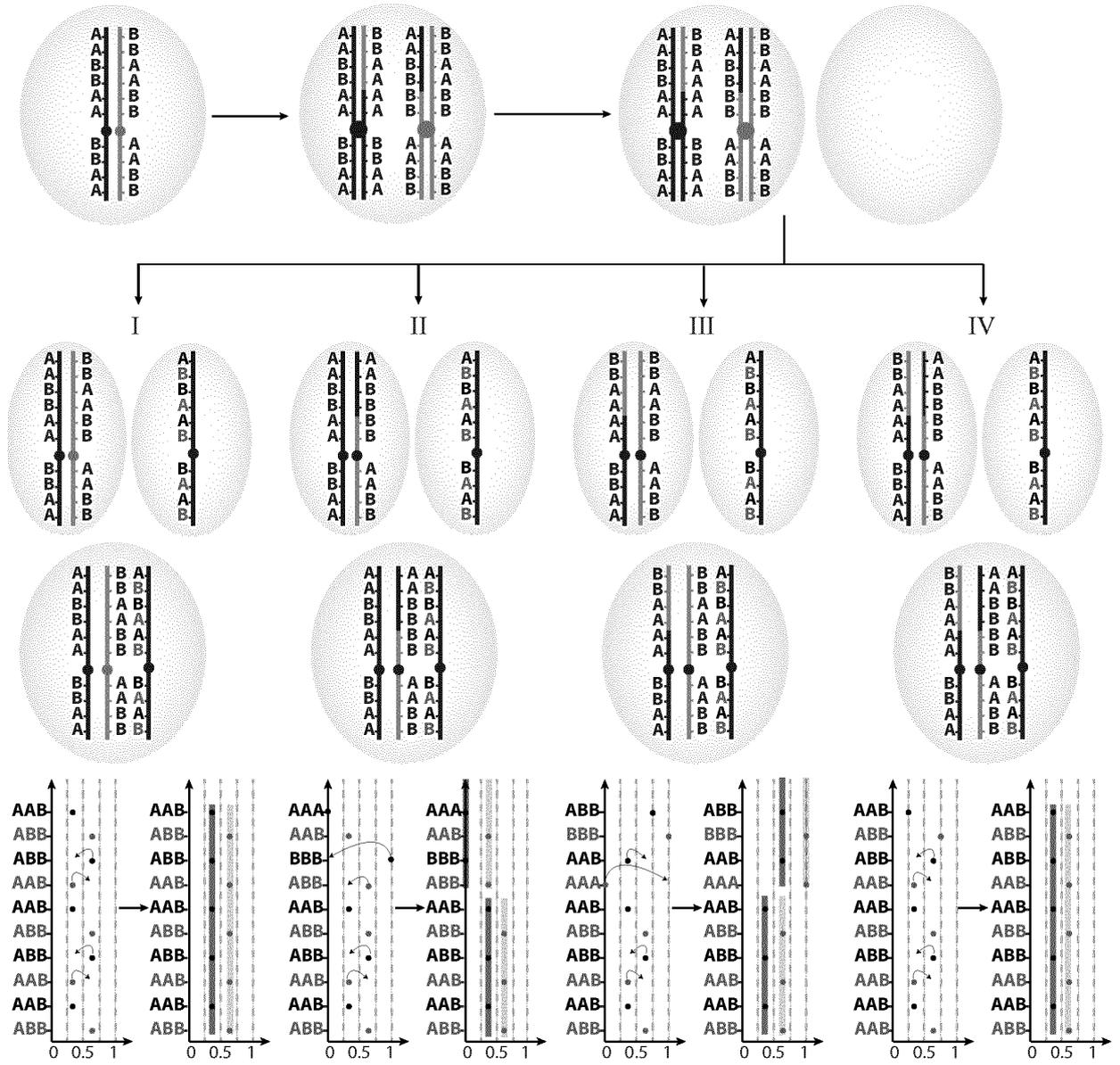


Figura 6

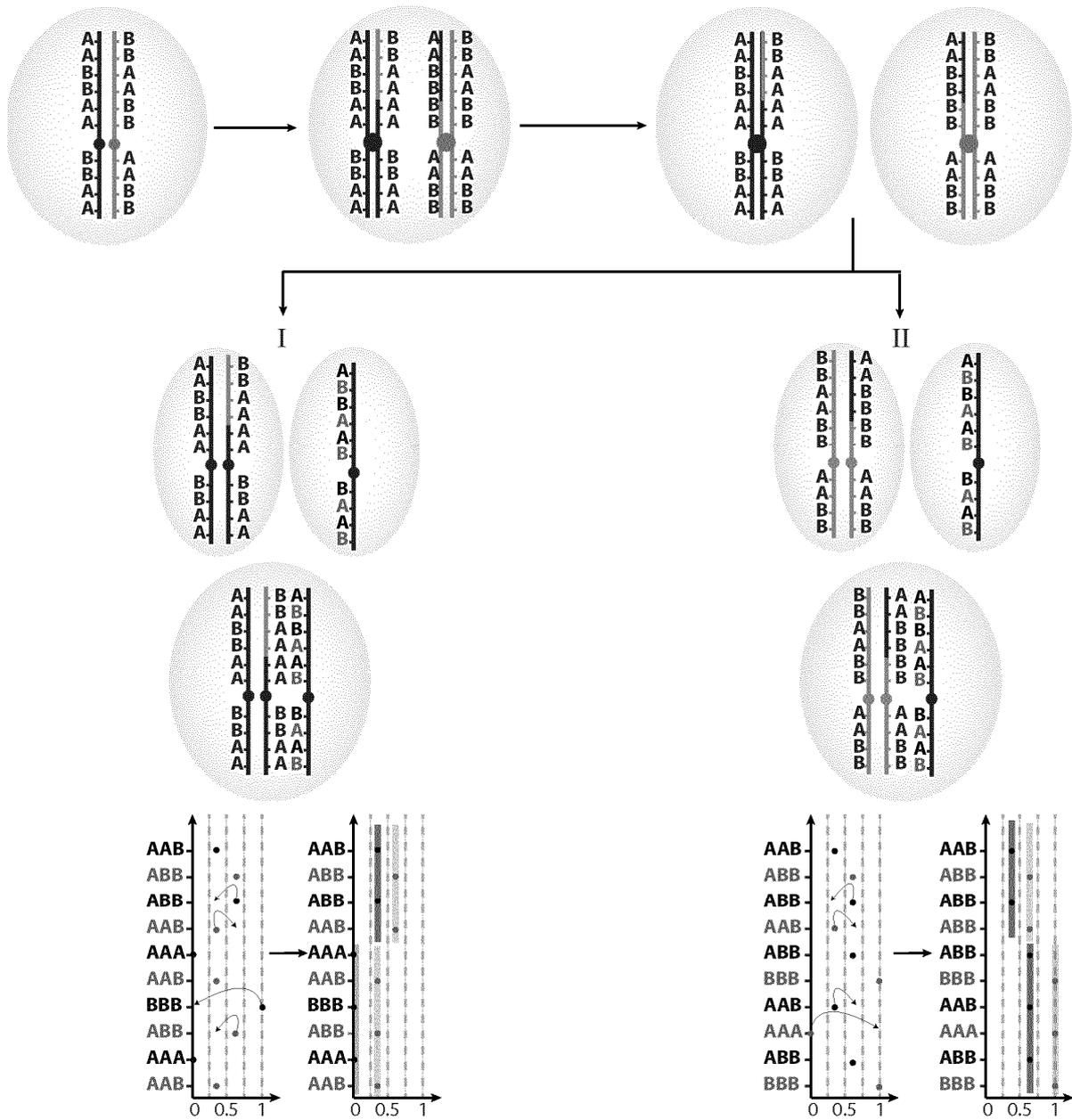


Figura 7

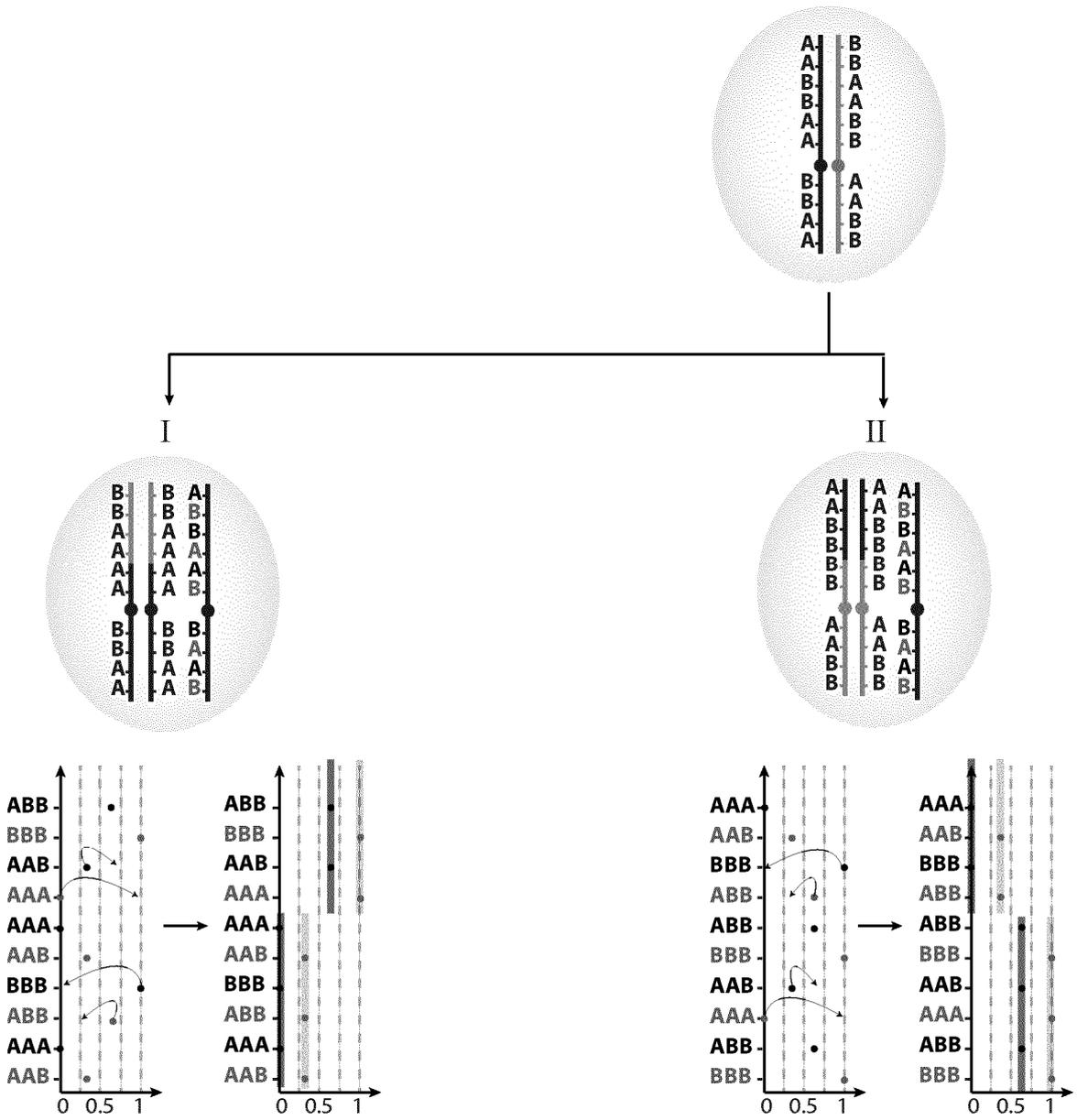


Figura 8a

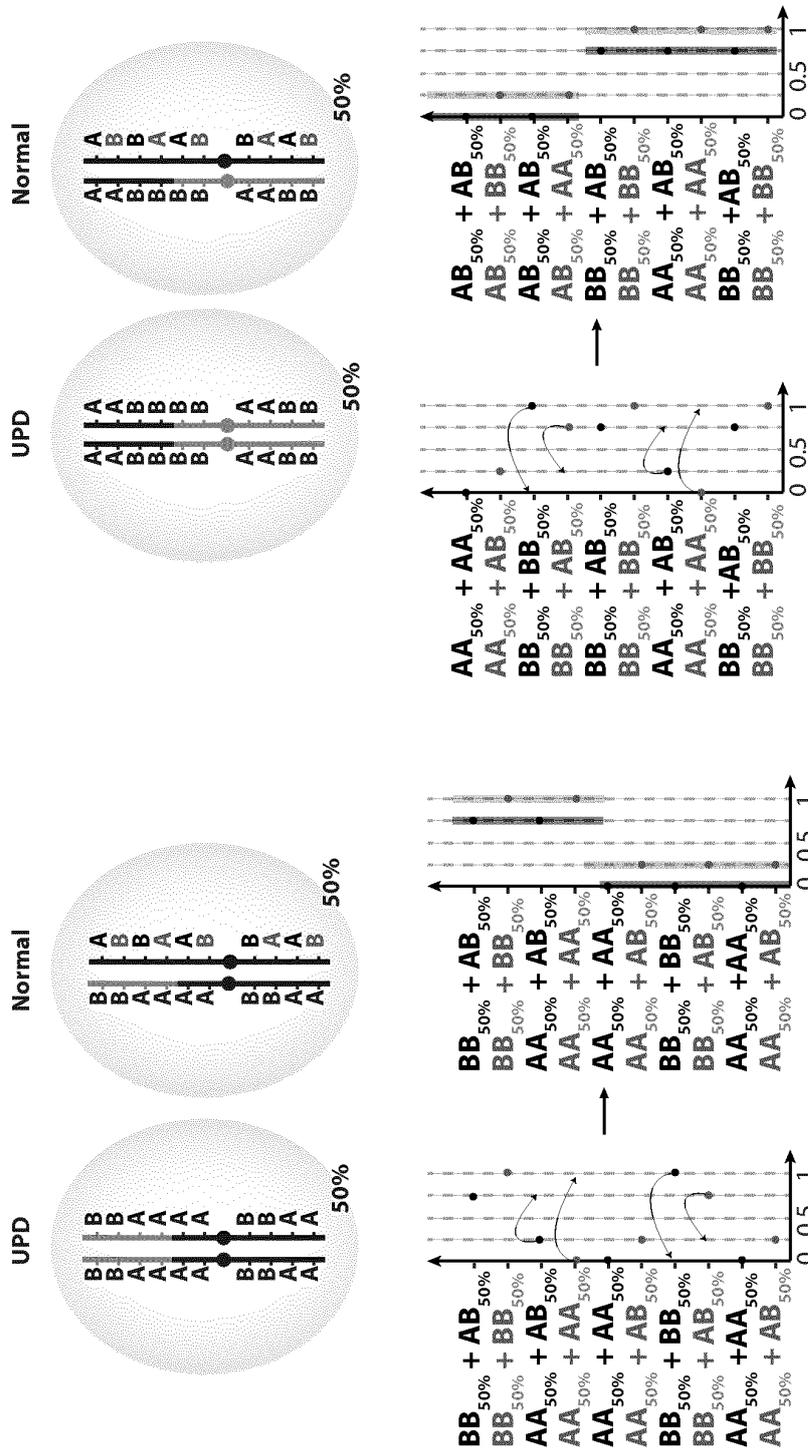


Figura 8b

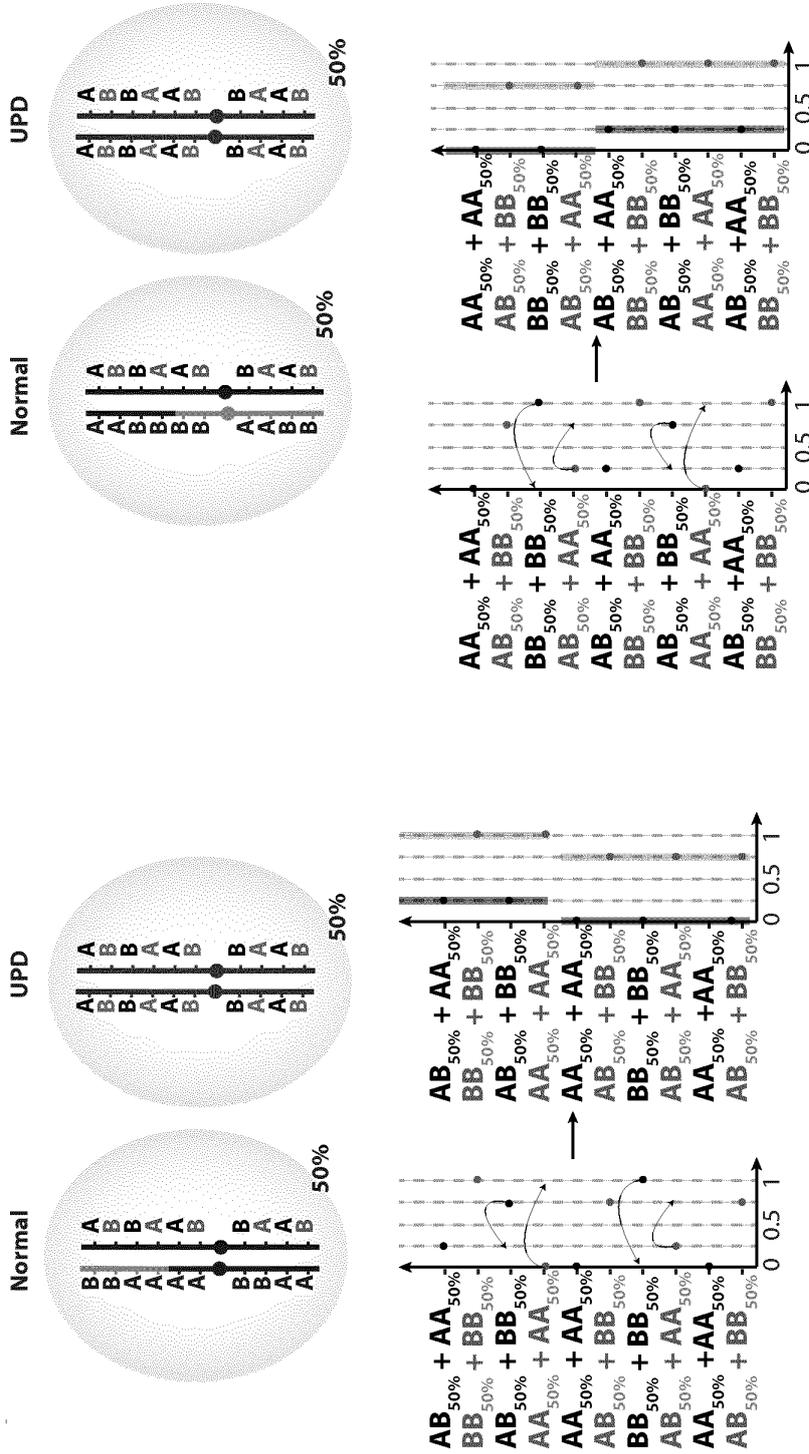


Figura 9a

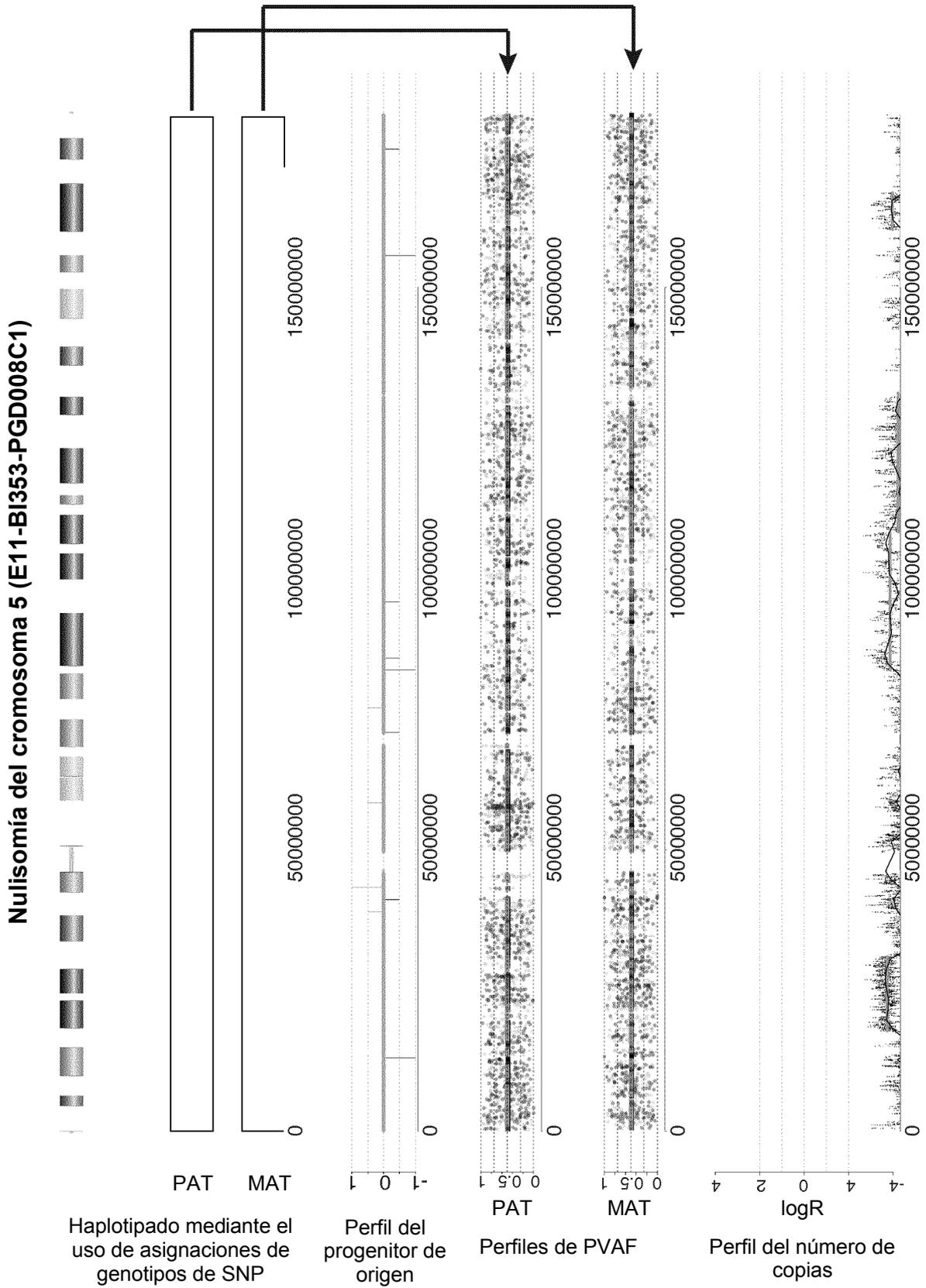


Figura 9b

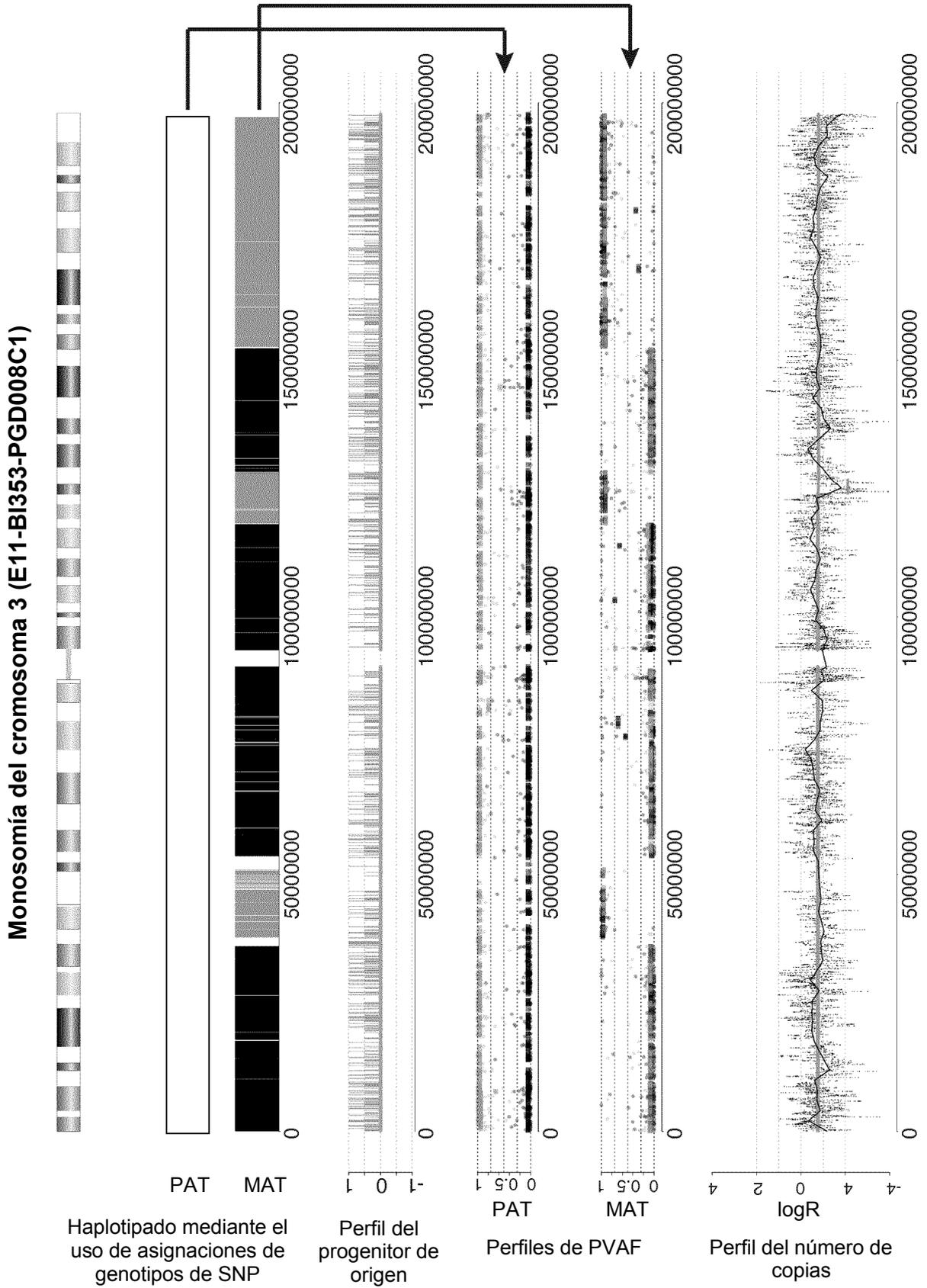


Figura 9c

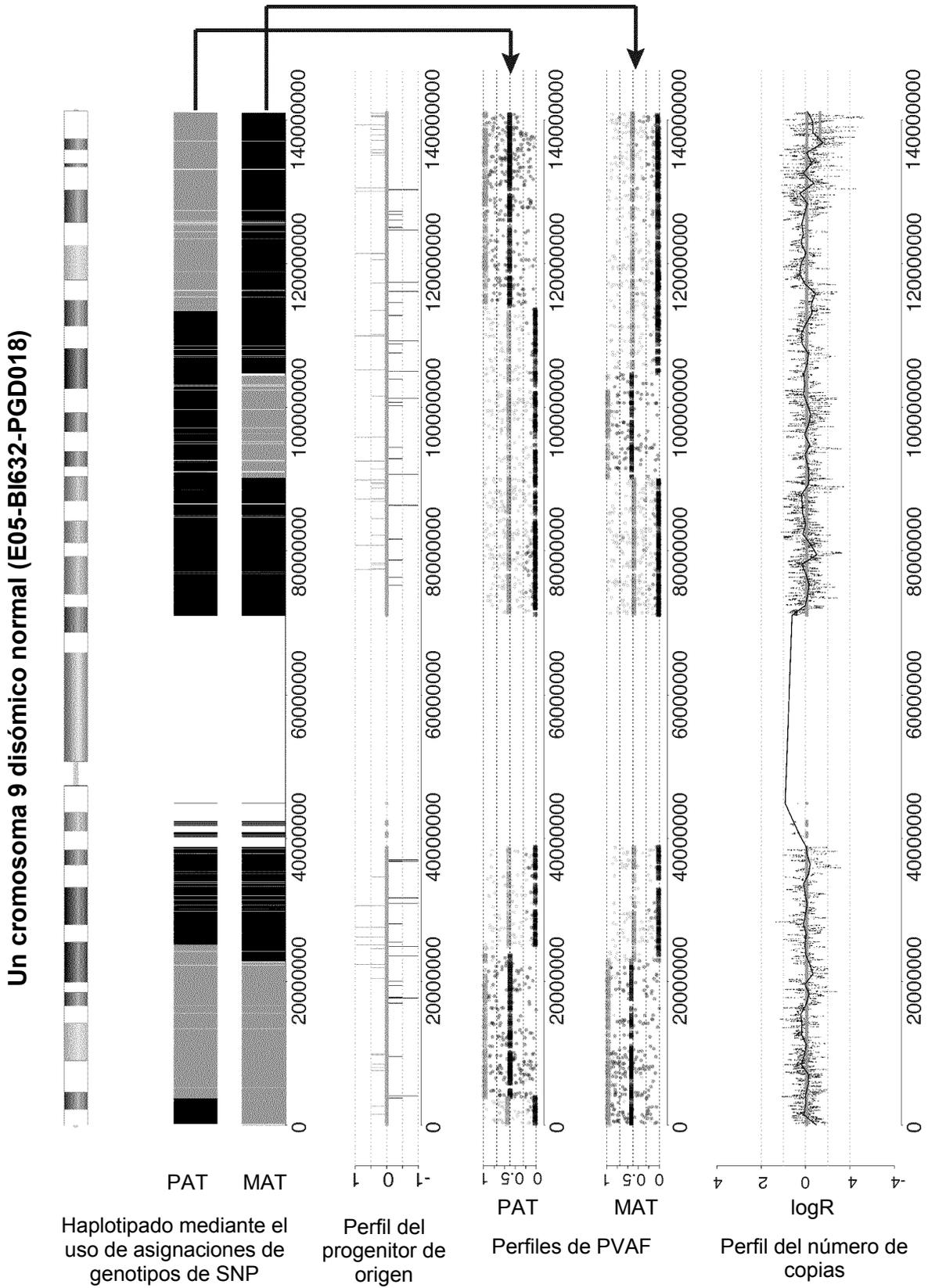


Figura 9d

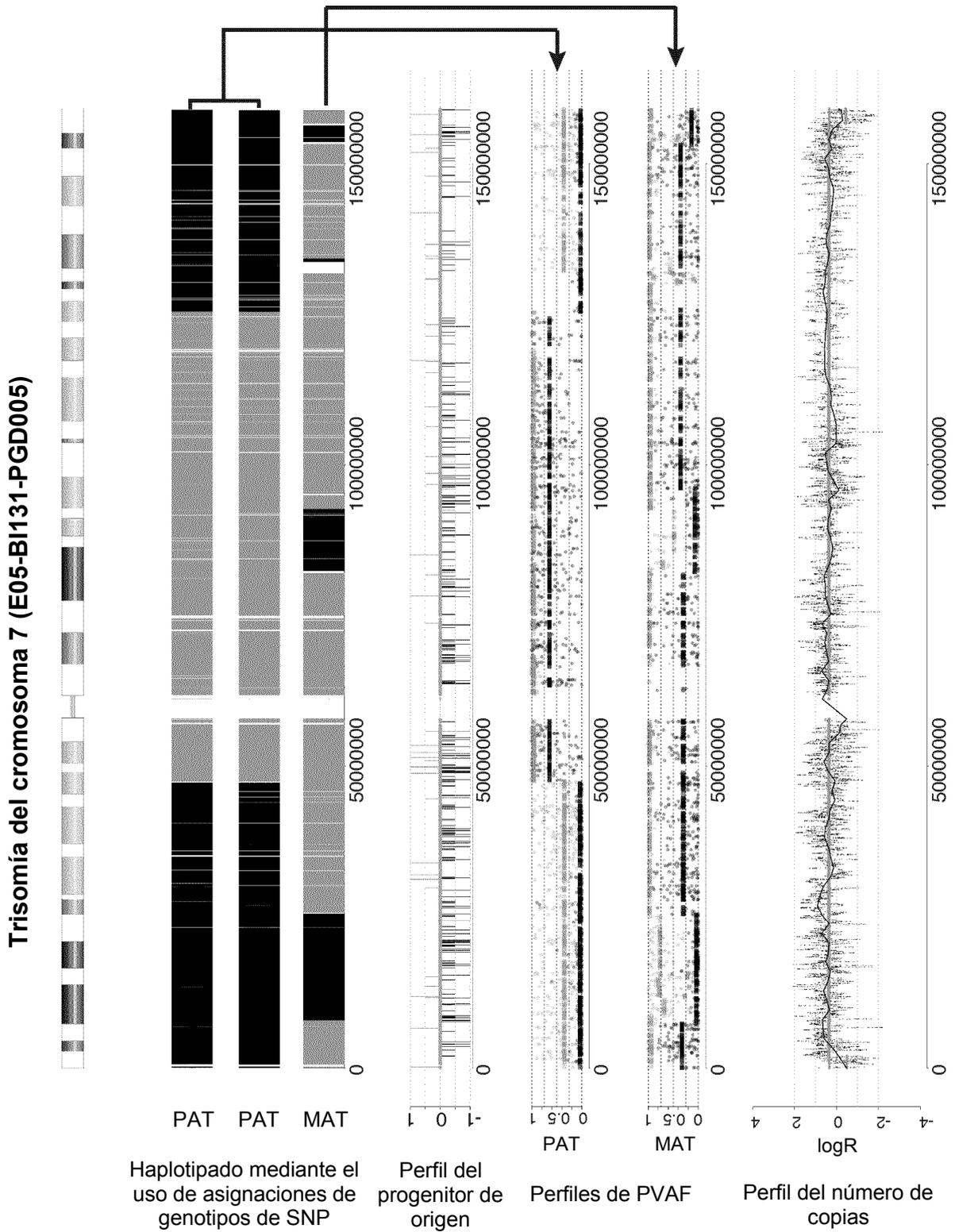


Figura 10a

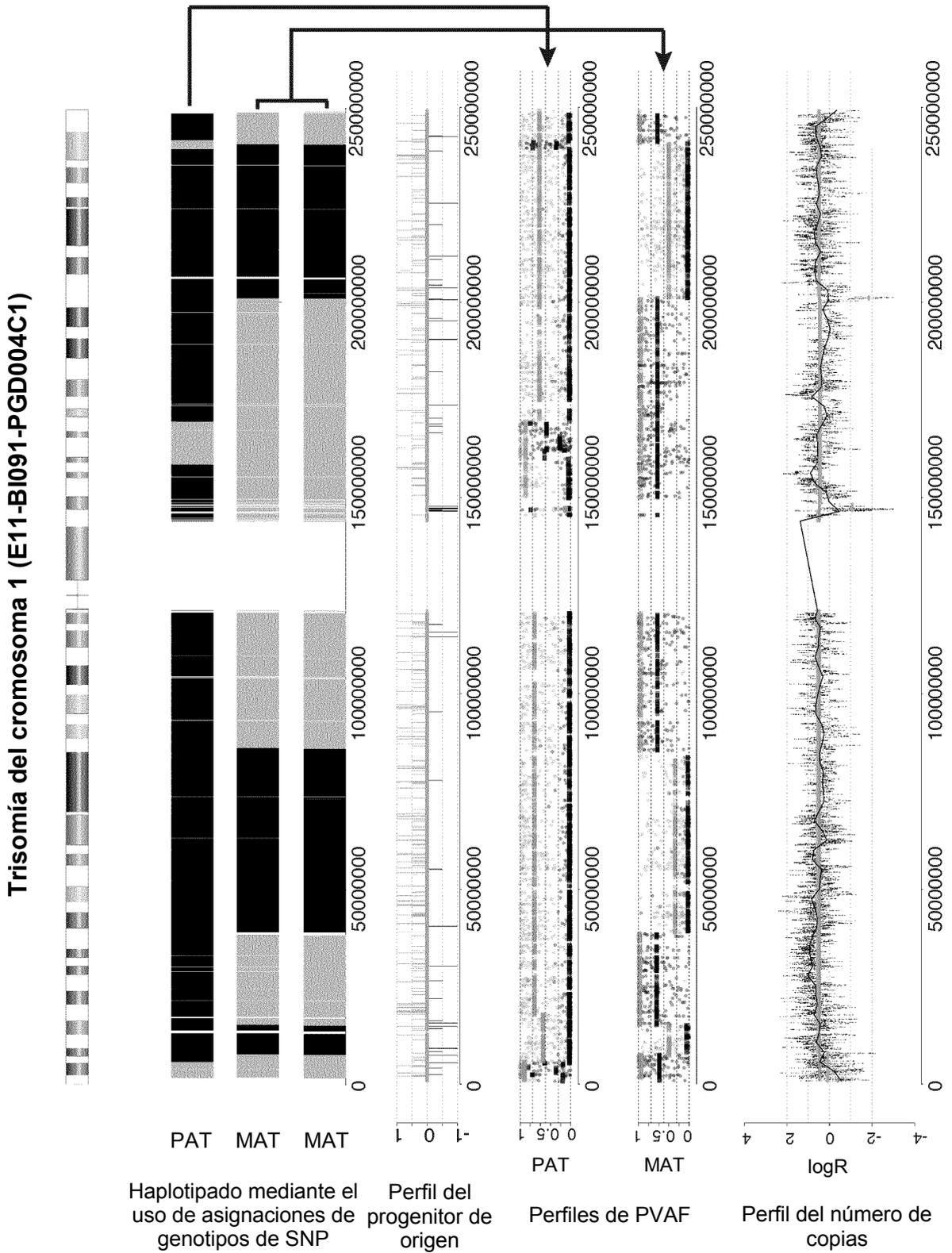


Figura 10b

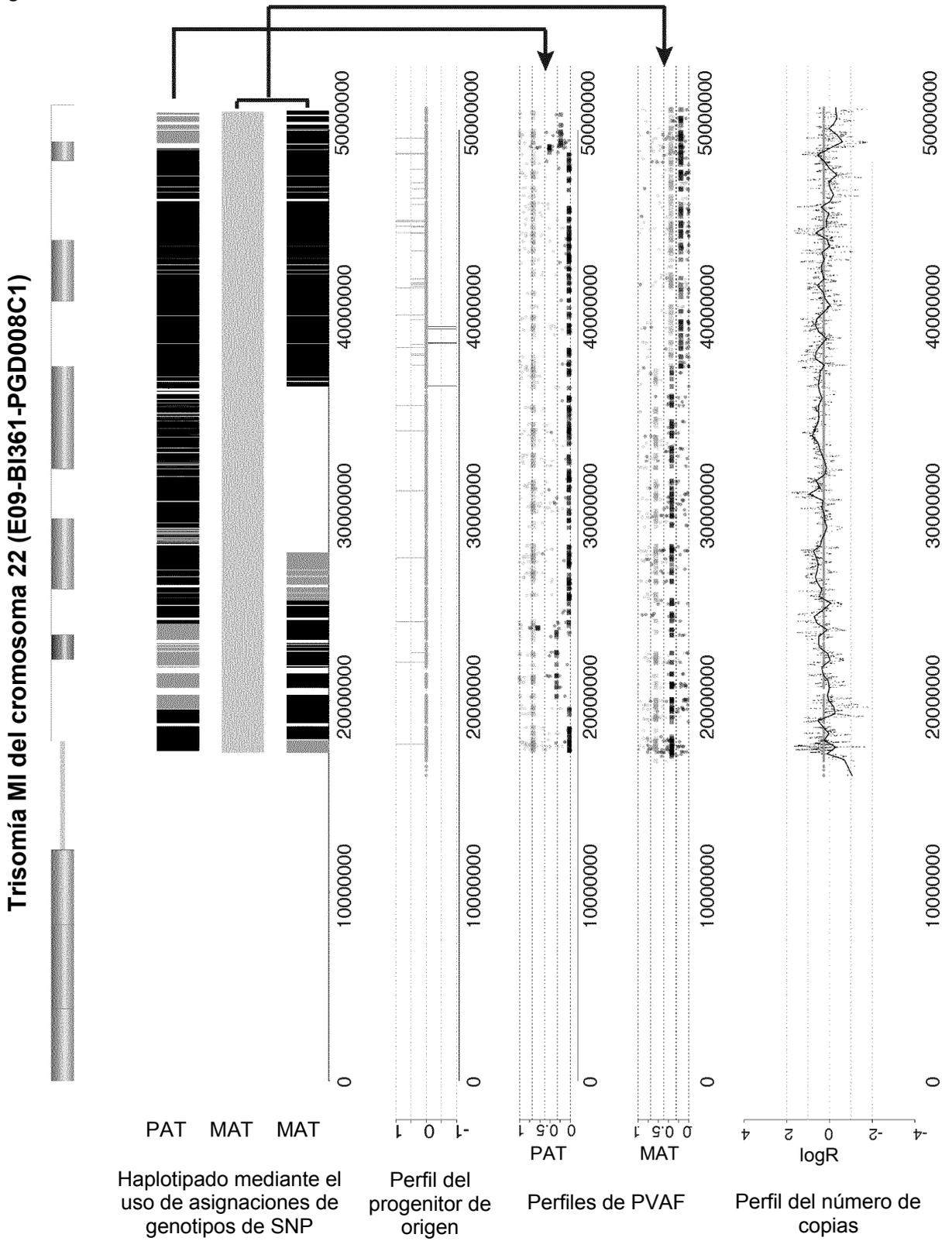


Figura 11a

Monosomía paterna del cromosoma 22

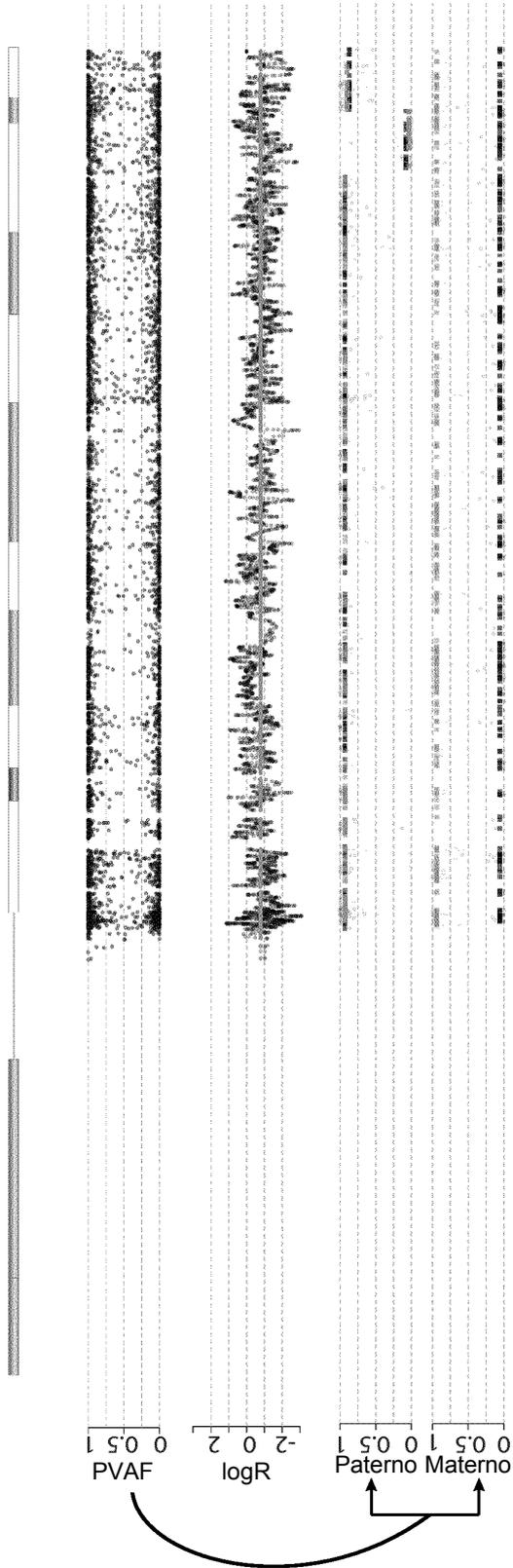


Figura 11b

Un cromosoma 10 disómico normal

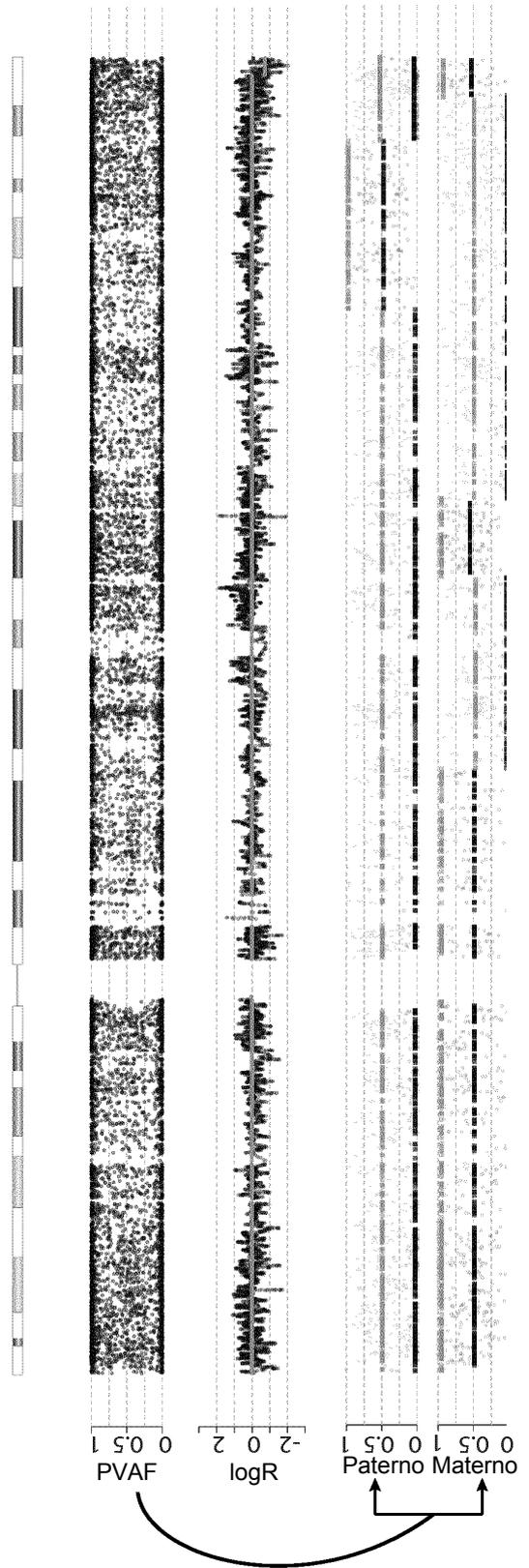


Figura 11c

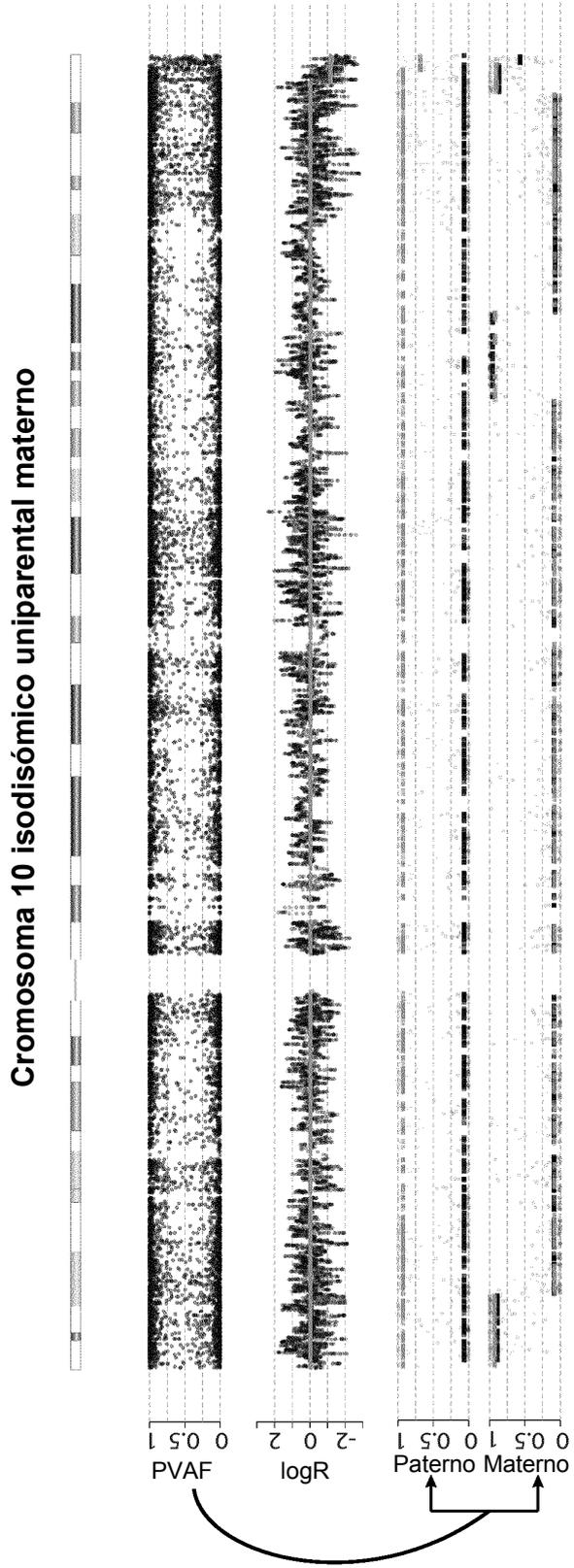


Figura 11d

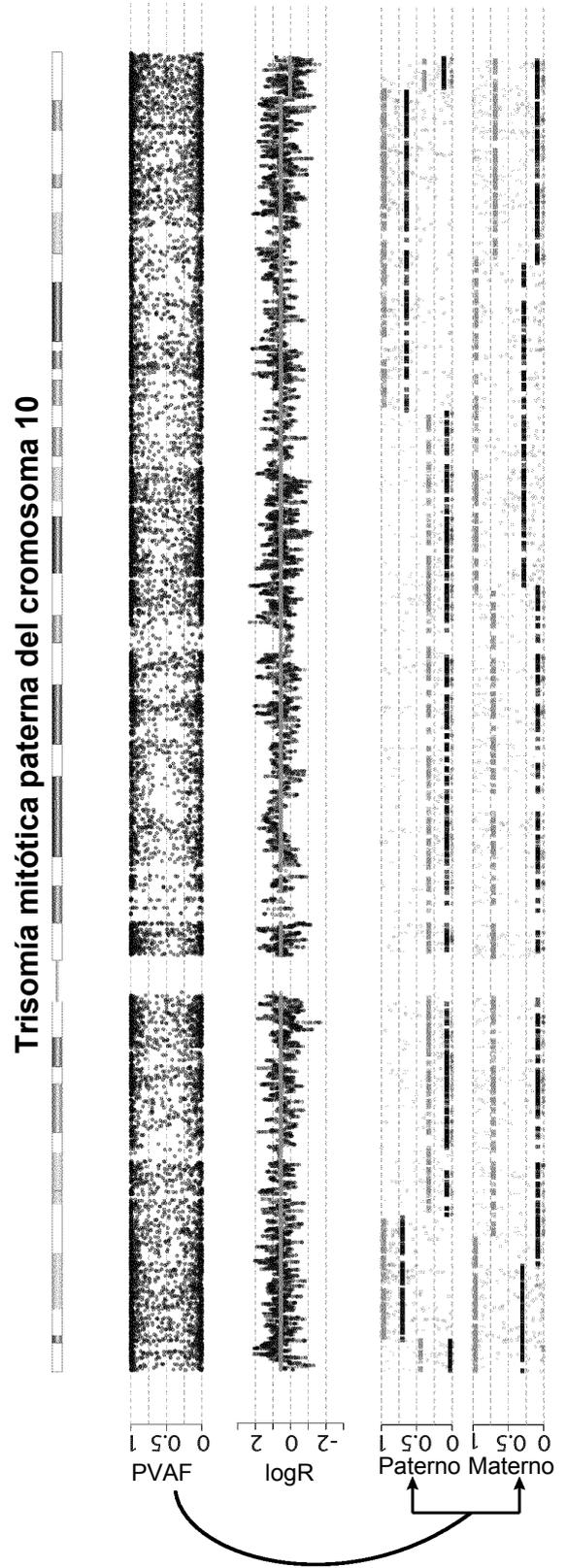


Figura 11e

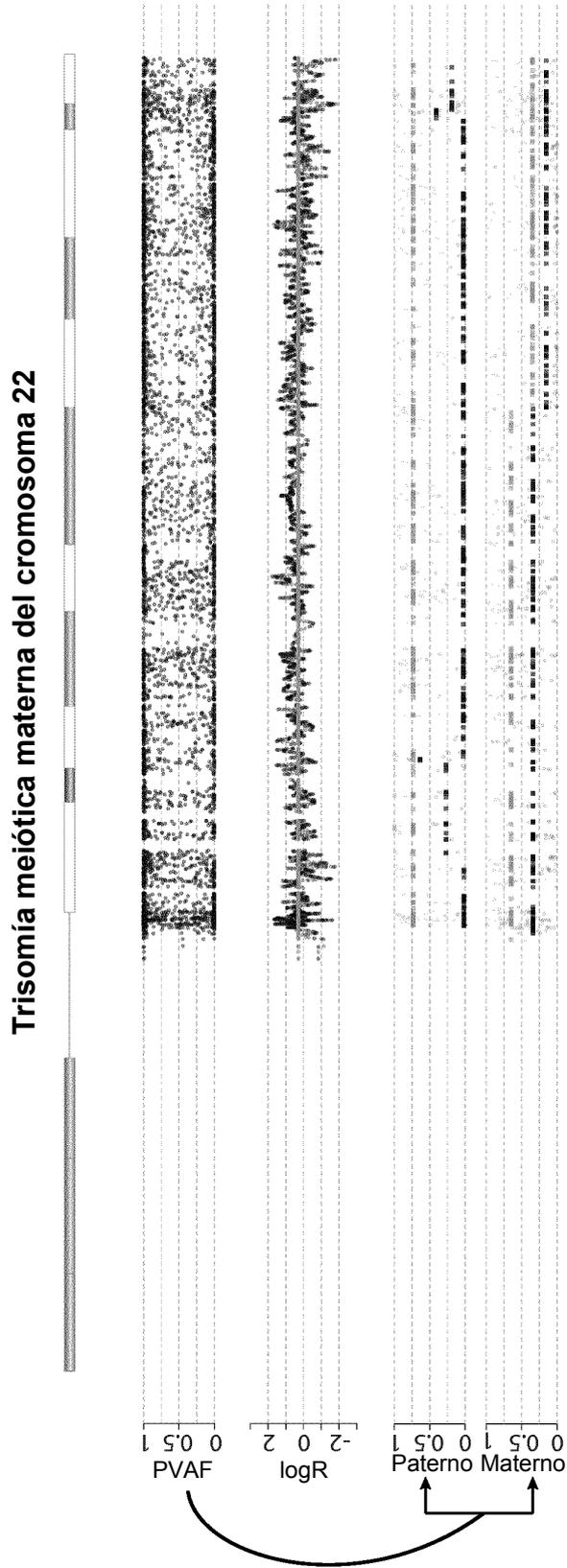


Figura 12a

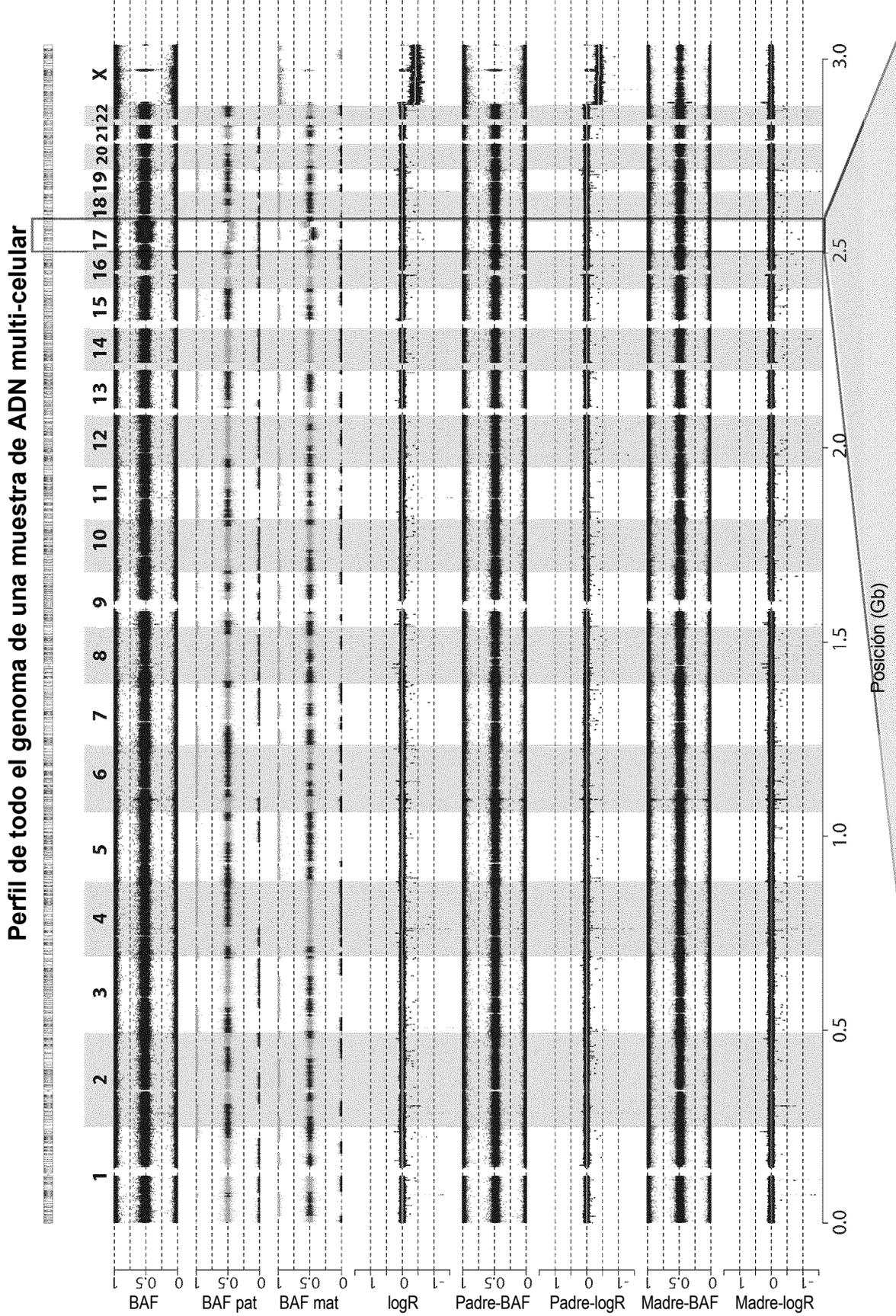


Figura 12b

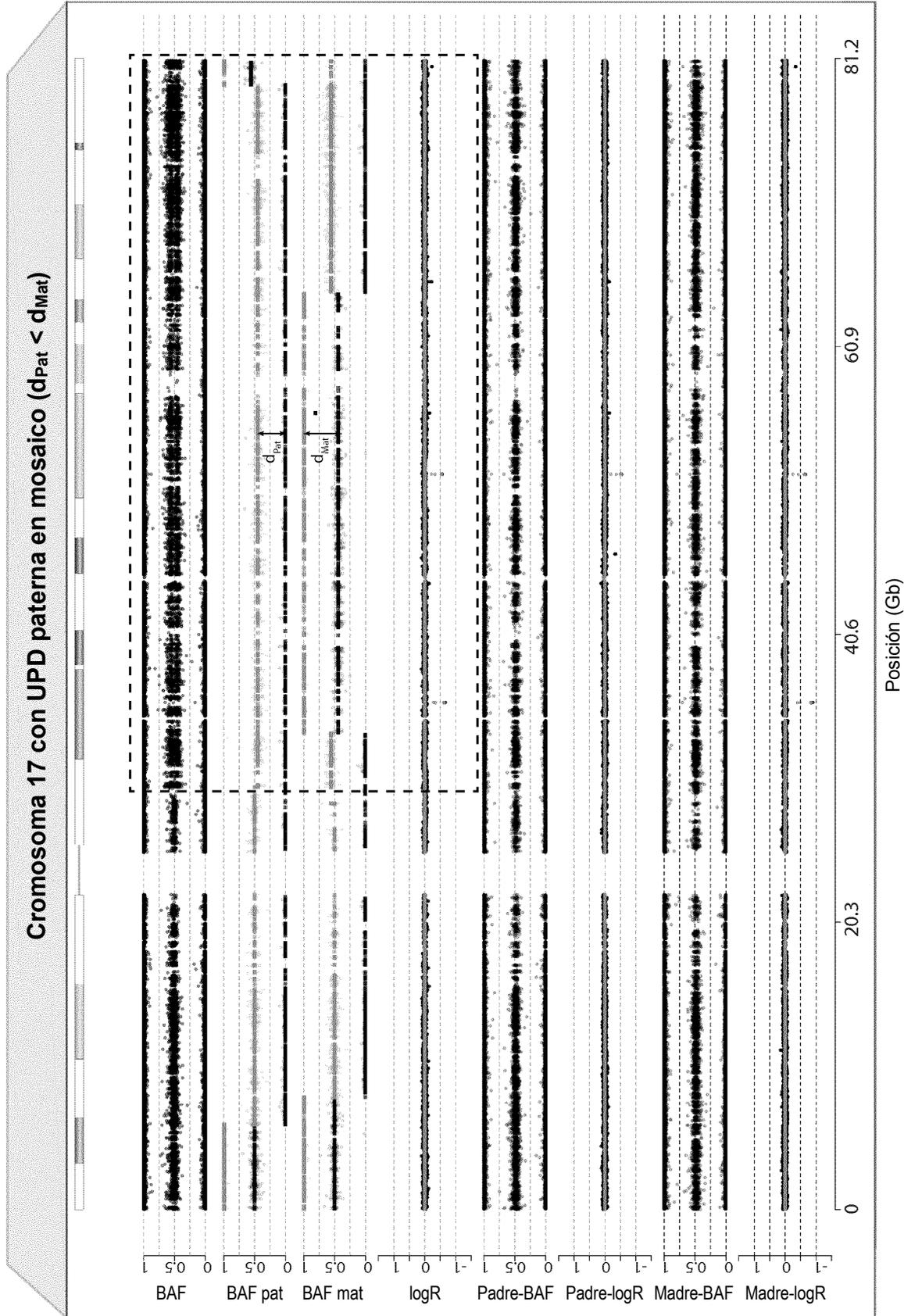


Figura 13

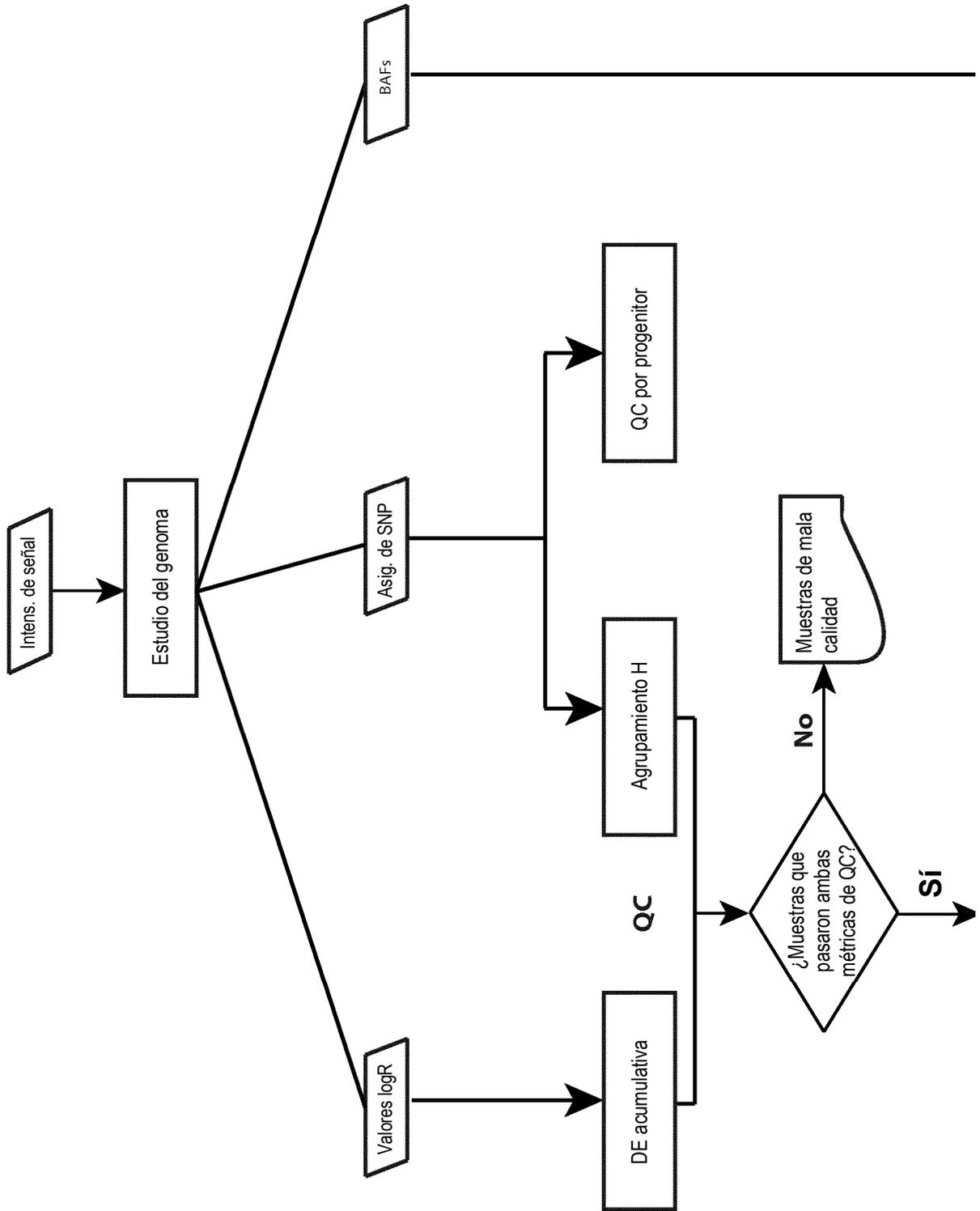


Figura 13 (continuación)

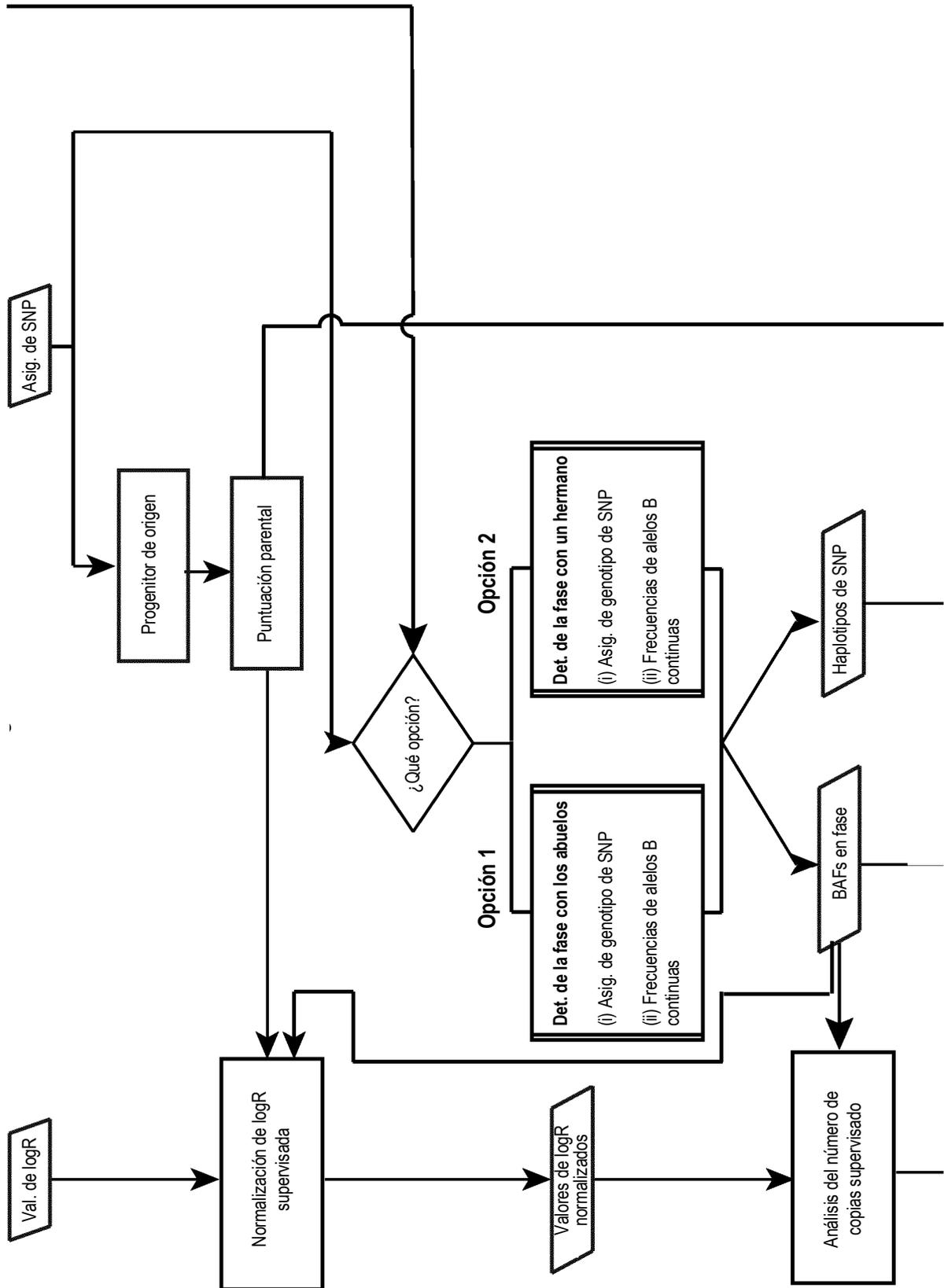


Figura 13 (continuación)

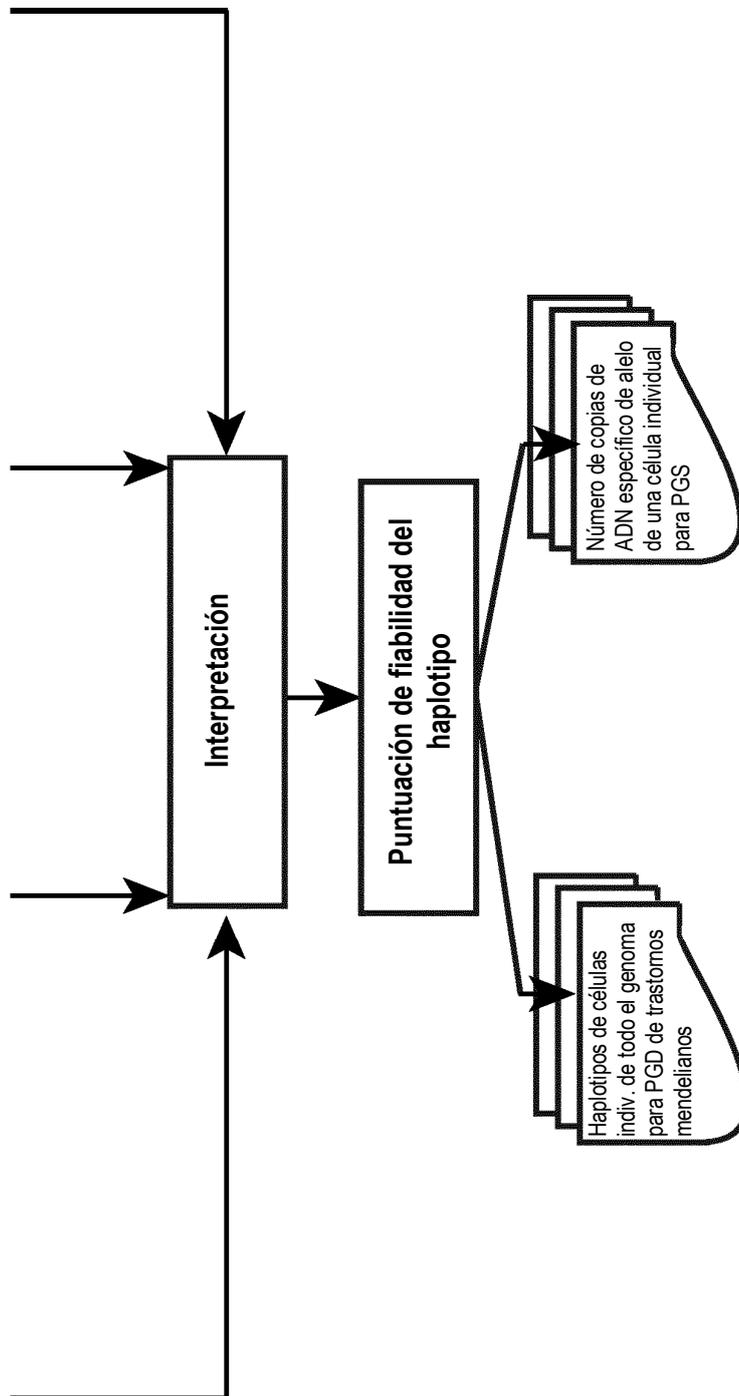


Figura 14a

Haplotipado de datos de secuenciación de exomas humanos

Datos de secuenciación de exomas

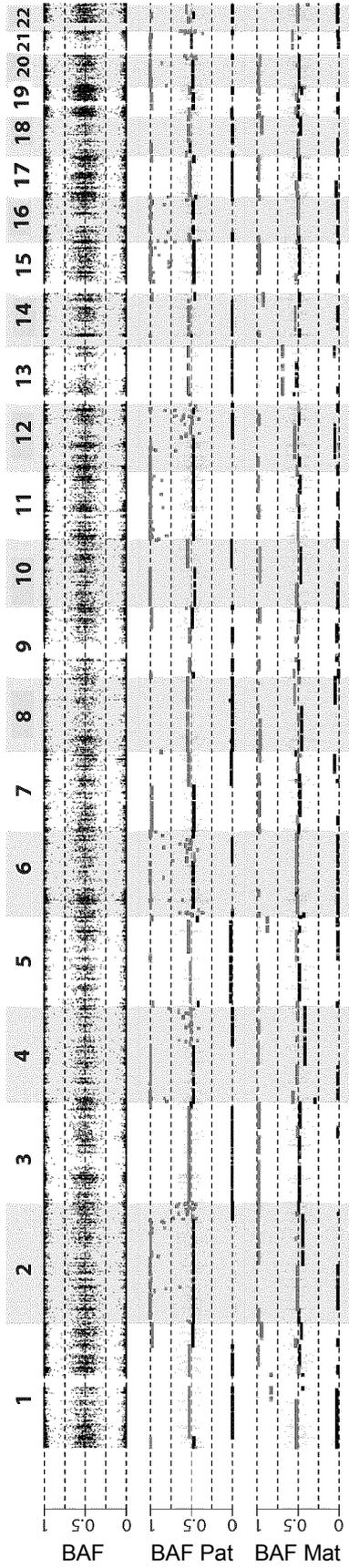


Figura 14b

Datos de matrices de SNP

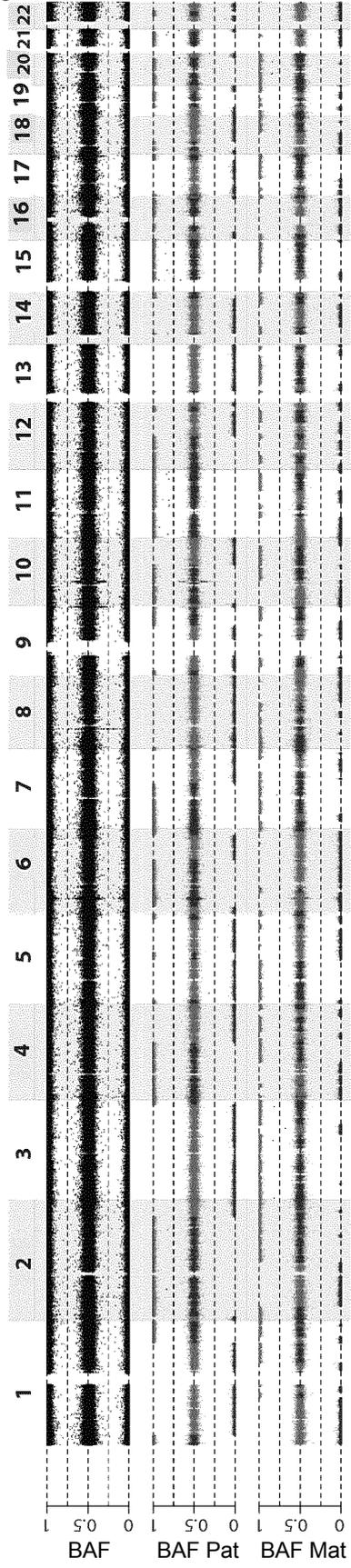


Figura 15a

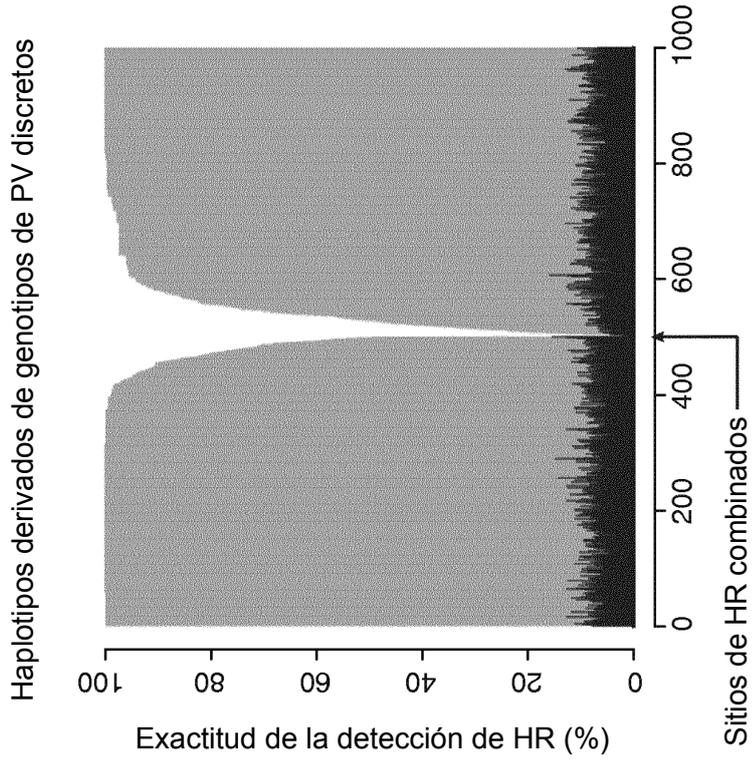


Figura 15b

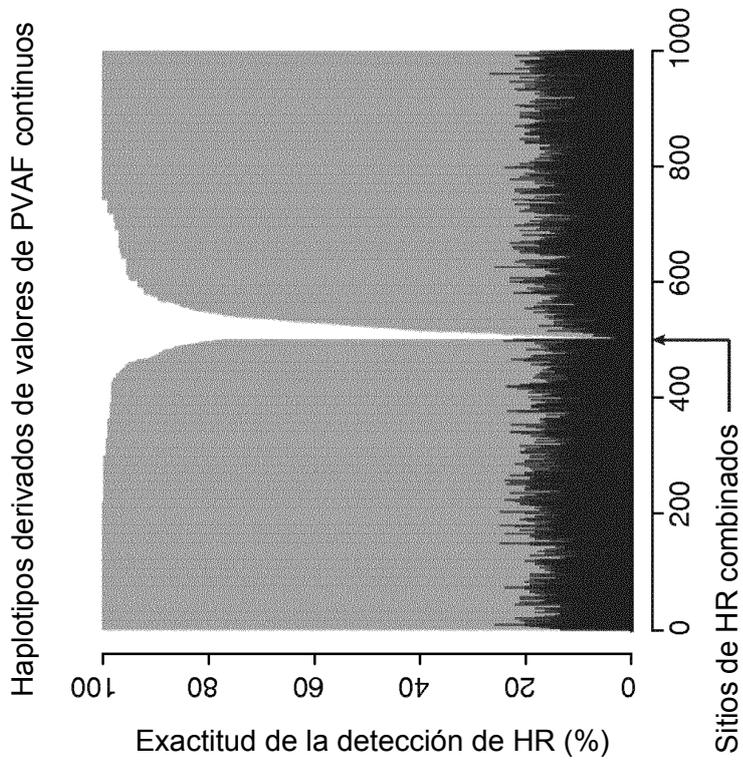


Figura 16a

**Cromosoma X (perfiles de PVAf)**

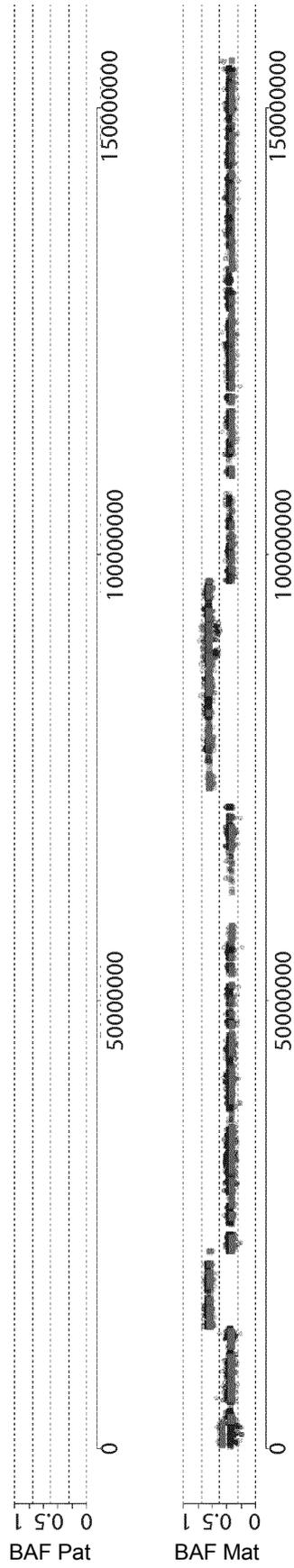
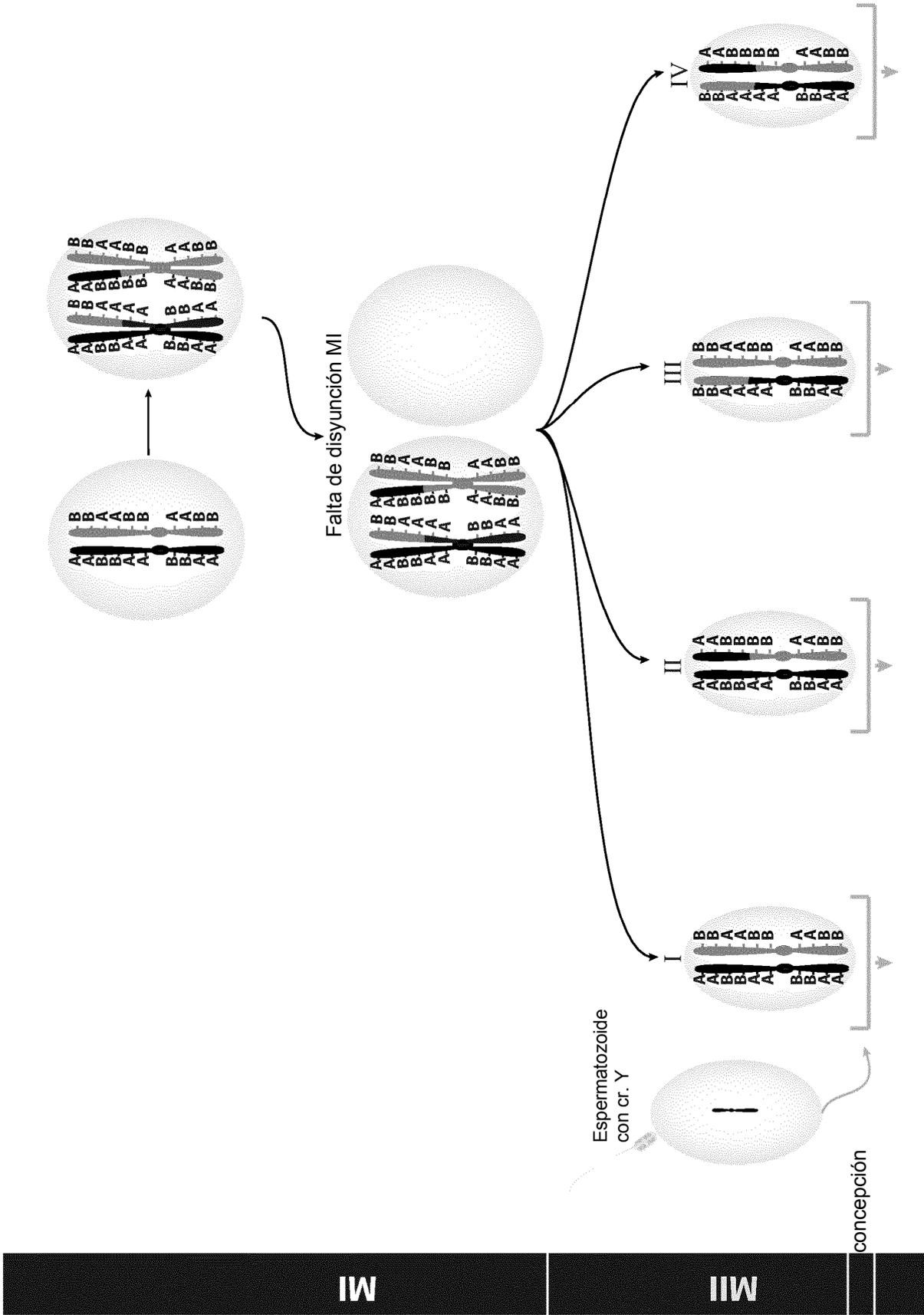


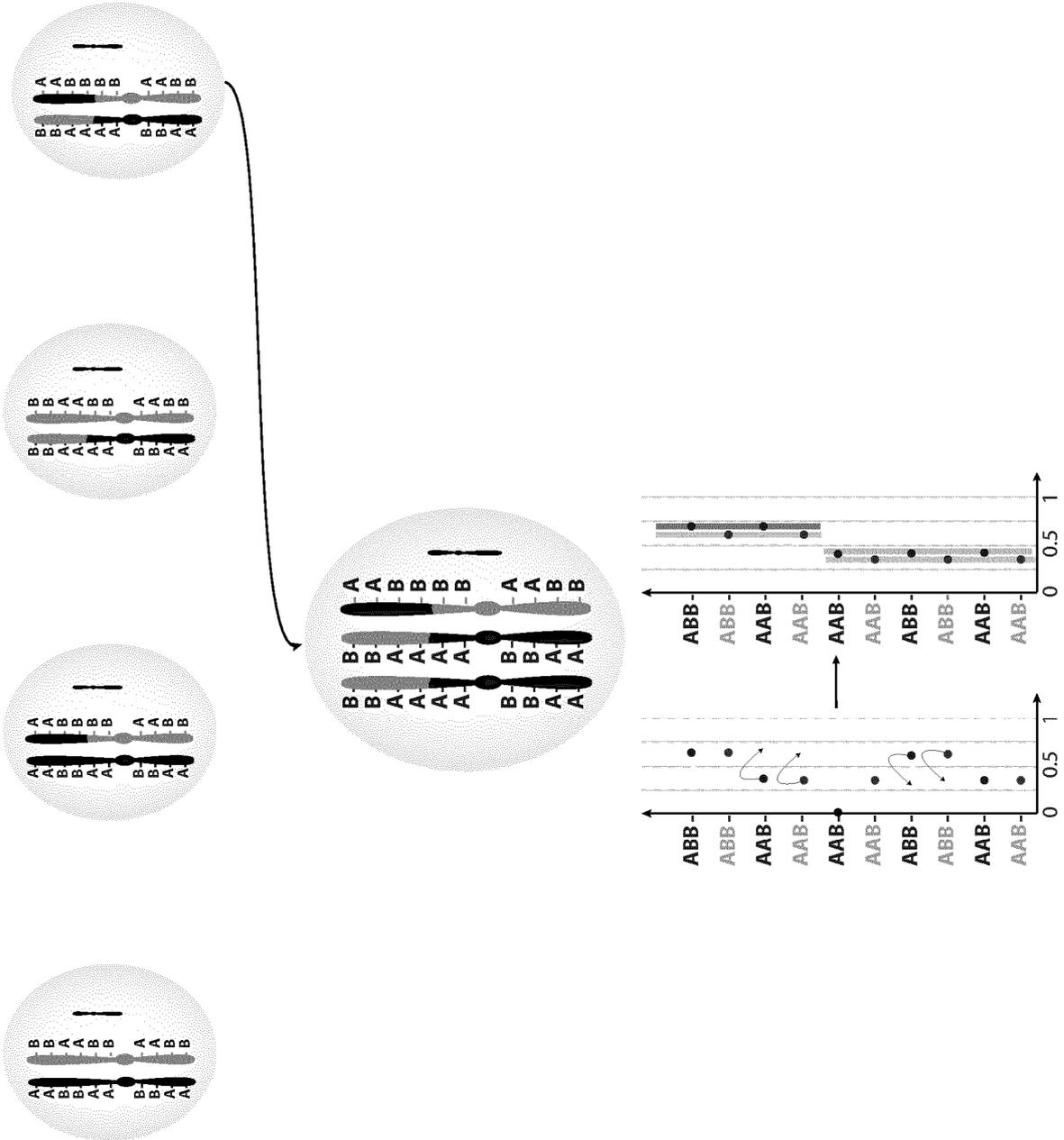
Figura 16b



MI

MI

Figura 16b (continuación)



Cigoto	Error mitótico	Distintivo de PVAF para Cr. X
--------	----------------	-------------------------------

Figura 17

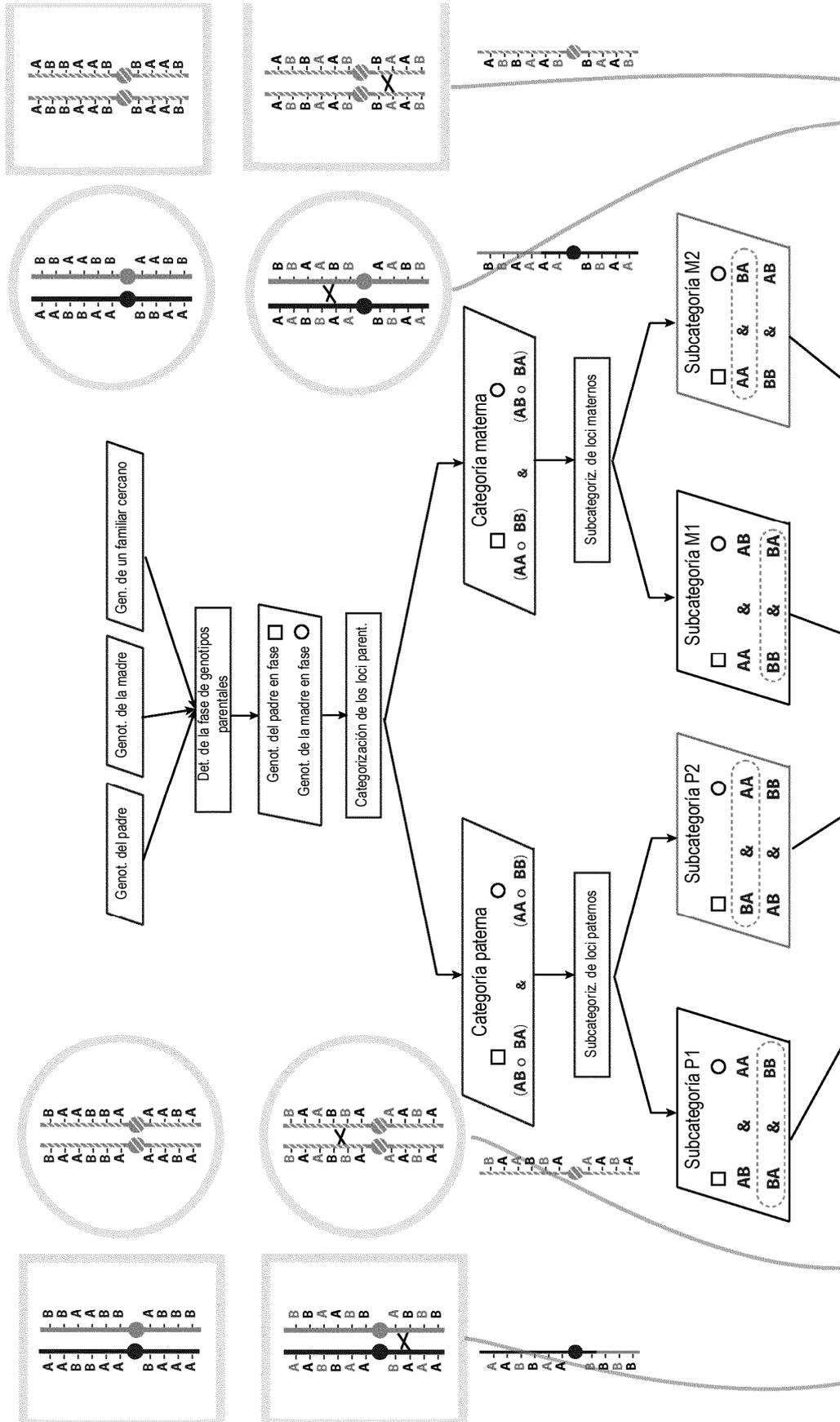


Figura 17 (continuación)

