

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 768 344**

51 Int. Cl.:

C12Q 1/6827 (2008.01)

C12Q 1/6858 (2008.01)

C12Q 1/6883 (2008.01)

C12Q 1/6874 (2008.01)

G16B 30/00 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.06.2012 PCT/US2012/042668**

87 Fecha y número de publicación internacional: **20.12.2012 WO12174378**

96 Fecha de presentación y número de la solicitud europea: **15.06.2012 E 12801070 (9)**

97 Fecha y número de publicación de la concesión europea: **18.12.2019 EP 2721181**

54 Título: **Métodos y materiales para evaluar el desequilibrio alélico**

30 Prioridad:

17.06.2011 US 201161498418 P

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

22.06.2020

73 Titular/es:

MYRIAD GENETICS, INC. (100.0%)

320 Wakara Way

Salt Lake City, UT 84108, US

72 Inventor/es:

GUTIN, ALEXANDER;

TIMMS, KIRSTEN y

LANCHBURY, JERRY

74 Agente/Representante:

SÁEZ MAESO, Ana

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 768 344 T3

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

DESCRIPCIÓN

Métodos y materiales para evaluar el desequilibrio alélico

5 Campo de la invención

La invención en general se refiere al diagnóstico molecular, y en particular a un método y sistema para detectar el desequilibrio alélico en muestras de pacientes.

10 Antecedentes de la invención

En general, una comparación de secuencias presentes en el mismo locus en cada cromosoma (cada cromosoma autosómico para varones) de un par de cromosomas puede revelar si ese locus particular es homocigoto o heterocigoto dentro del genoma de una célula. Los loci polimórficos dentro del genoma humano son generalmente heterocigotos dentro de un individuo ya que el individuo típicamente recibe una copia del padre biológico y una copia de la madre biológica. En algunos casos, un locus polimórfico o una cadena de loci polimórficos dentro de un individuo es homocigota como un resultado de heredar copias idénticas de ambos padres biológicos. En otros casos, la homocigosidad resulta de una pérdida de heterocigosidad (LOH) de la línea germinal. Debido a que la información de LOH y del número de copias puede ser clínicamente útil, existe la necesidad de métodos mejorados para identificar los loci y las regiones de LOH en las muestras.

Jacobs y otros (2007) describen un análisis integrado de genotipo, LOH y número de copias para el ADN derivado de tejidos FFPE. El análisis se lleva a cabo mediante el uso de una micromatriz que contiene más de 500K SNP. El ordenador se usa para el cálculo

25 Breve resumen de la invención

La invención se define por las reivindicaciones adjuntas.

30 El análisis del número de copias (incluido el desequilibrio alélico y la LOH) de los tejidos tumorales se ha realizado tradicionalmente mediante el uso de matrices de polimorfismo de nucleótido simple (SNP). La calidad de los datos frecuentemente es muy variable y, especialmente para las muestras FFPE, tiende a ser pobre. Los inventores han desarrollado un método de análisis de número de copias de todo el genoma que produce datos de alta calidad a partir de todos los tipos de muestras que se basa en la captura en solución de fragmentos de ADN que abarcan los loci diana (por ejemplo, los SNP), seguido de una secuenciación paralela para identificar y cuantificar los alelos. Los datos resultantes permiten un análisis de alta calidad de LOH y del número de copias de la muestra.

En consecuencia, en un aspecto de la presente descripción se proporciona un método para detectar el estado de desequilibrio alélico en una pluralidad de loci genómicos en una muestra tumoral de un paciente con cáncer, que comprende las etapas de enriquecer una muestra de ADN genómico en moléculas de ADN cada una que comprende un locus de interés; secuenciar dichas moléculas de ADN para determinar el genotipo en cada locus; determinar para cada locus si existe un desequilibrio alélico.

45 En otro aspecto de la presente descripción, se proporciona un método para detectar el estado de LOH en una pluralidad de loci genómicos en una muestra tumoral de un paciente con cáncer, que comprende las etapas de enriquecer una muestra de ADN genómico en moléculas de ADN cada una que comprende un locus de interés; secuenciar dichas moléculas de ADN para determinar el genotipo en cada uno de tales loci; determinar para cada locus homocigoto si es homocigoto debido a LOH.

50 A menos que se defina de cualquier otra forma, todos los términos técnicos y científicos usados en la presente descripción tienen el mismo significado que el que entiende comúnmente el experto en la técnica a la cual pertenece la invención. Aunque los métodos y materiales similares o equivalentes a los descritos en la presente descripción pueden usarse en la práctica o prueba de la presente invención, los métodos y materiales adecuados se describen más abajo. En caso de conflicto, prevalecerá la presente descripción, incluidas las definiciones. Además, los materiales, métodos, y ejemplos son ilustrativos solamente y no pretenden ser limitantes.

Otras características y ventajas de la invención serán evidentes a partir de la siguiente descripción detallada y a partir de las reivindicaciones.

60 Breve descripción de las figuras

La Figura 1 es un gráfico que representa las dosificaciones de alelos de células de cáncer de mama de una paciente con cáncer de mama a lo largo del cromosoma 1 según se determinó mediante el uso de un matriz de SNP. La región cromosómica entre las flechas es una región de LOH que tiene aproximadamente 103 Mb de longitud.

65

La Figura 2 es un gráfico que representa las dosificaciones de alelos de células de cáncer de mama para el mismo paciente con cáncer de mama que en la Figura 1 a lo largo del cromosoma 1 según se determinó mediante el uso de secuenciación de alto rendimiento. La región cromosómica entre las flechas es una región de LOH que tiene aproximadamente 103 Mb de longitud.

La Figura 3 es un diagrama de un ejemplo de un dispositivo informático y un dispositivo informático móvil que puede usarse para implementar las técnicas descritas en la presente descripción.

Descripción detallada de la invención

La invención se define por las reivindicaciones adjuntas. Se ha descubierto sorprendentemente que determinar el desequilibrio alélico (por ejemplo, número de copias anormales, LOH) en muestras embebidas en parafina, fijadas con formalina ("FFPE") mediante el uso de secuenciación de regiones genómicas que comprenden los loci de interés (por ejemplo, los SNP) produce datos de calidad muy superior en comparación con los datos del número de copias y de desequilibrio alélico generados mediante micromatrices. Esta invención permite el análisis a gran escala (por ejemplo, genoma completo) del número de copias (por ejemplo, desequilibrio alélico) de muestras de calidad variable. En particular, permite la producción de datos de alta calidad a partir de ADN derivado de FFPE. Las plataformas actuales basadas en matrices no pueden producir datos de calidad suficiente a partir de este tipo de muestra.

En consecuencia, en un aspecto de la presente descripción, se proporciona un método para detectar el estado de desequilibrio alélico en una pluralidad de loci genómicos en una muestra tumoral de un paciente con cáncer, que comprende las etapas de enriquecer una muestra de ADN genómico en moléculas de ADN cada una que comprende un locus de interés; secuenciar dichas moléculas de ADN para determinar el genotipo en cada locus; determinar para cada locus si existe un desequilibrio alélico. "Locus" como se usa en la presente descripción tiene su significado habitual en la técnica. Como se usa en la presente descripción, "región" significa una pluralidad de loci sustancialmente adyacentes. A menos que se indique lo contrario o, a menos que el contexto indique claramente lo contrario, las declaraciones hechas sobre un locus generalmente se aplicarán a una región.

Como se usa en la presente descripción, "desequilibrio alélico" significa cualquier caso en el que el número de copias somáticas difiere del número de copias de la línea germinal en un locus o región genómica. En algunas modalidades el desequilibrio alélico se expresa en términos de proporción de copias mayoritarias ("MCP"). La proporción de copias mayoritarias o MCP, como se usa en la presente descripción, significa la relación del número de copias alélicas mayoritarias con relación al número de copias alélicas mayoritarias + minoritarias, como sigue:

$$MCP = \frac{[\text{número de copias alélicas mayoritarias}]}{[\text{número de copias alélicas mayoritarias}] + [\text{número de copias alélicas minoritarias}]}$$

En algunas modalidades, un locus o región muestra un desequilibrio alélico si la MCP en tal locus o región es 0,51, 0,52, 0,53, 0,54, 0,55, 0,6, 0,65, 0,7, 0,75, 0,8, 0,85, 0,9, 0,95 o 1.

Un ejemplo de desequilibrio alélico es la pérdida de heterocigosidad ("LOH"), en la que un locus es heterocigoto en la línea germinal pero homocigoto en el tejido somático. En este sentido, la homocigosidad puede incluir la pérdida homocigótica (es decir, la delección) del locus en el tejido somático. Los diferentes tipos de LOH y desequilibrio alélico posibles se analizan con más detalle a continuación.

Por lo tanto, en algunas modalidades la presente descripción proporciona un método para detectar el estado de LOH en una pluralidad de loci genómicos en una muestra tumoral de un paciente con cáncer, que comprende enriquecer una muestra de ADN genómico en moléculas de ADN cada una que comprende un locus de interés; secuenciar dichas moléculas de ADN para determinar el genotipo en cada uno de tal locus; determinar para cada locus homocigoto si es homocigoto debido a LOH.

De acuerdo con la presente descripción, las técnicas de secuenciación de ácido nucleico pueden usarse para identificar los loci y/o regiones que tienen un desequilibrio alélico. Por ejemplo, el ADN genómico de una muestra de células (por ejemplo, una muestra de células cancerosas) puede extraerse y fragmentarse. Puede usarse cualquier método apropiado para extraer y fragmentar el ácido nucleico genómico, incluidos, sin limitación, los kits comerciales tales como QIAamp DNA Mini Kit (Qiagen), MagNA Pure DNA Isolation Kit (Roche Applied Science) y GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich). Una vez extraído y fragmentado, puede realizarse una secuenciación dirigida o no dirigida para determinar los genotipos de la muestra en los loci de interés. Por ejemplo, la secuenciación del genoma completo, el transcriptoma completo o el exoma completo puede realizarse para determinar genotipos en millones o incluso miles de millones de pares de bases (es decir, los pares de bases pueden ser "loci" a evaluar).

En algunos casos, la secuenciación dirigida de los loci polimórficos conocidos (por ejemplo, los SNP y las secuencias circundantes) puede realizarse como una alternativa al análisis de micromatrices. Por ejemplo, el ADN genómico puede enriquecerse en aquellos fragmentos que contienen un locus (por ejemplo, ubicación de los SNP) a analizar mediante el uso de kits diseñados para este propósito (por ejemplo, Agilent SureSelect, Illumina TruSeq Capture, Nimblegen SeqCap EZ Choice, Raindance Thunderstorm™). Por ejemplo, el ADN genómico que contiene los loci a analizar puede hibridarse

a fragmentos de ARN de captura biotinilados para formar complejos biotinilados de ARN/ADN genómico. Alternativamente, pueden utilizarse sondas de captura de ADN que dan como resultado la formación de híbridos de ADN biotinilado/ADN genómico. Pueden usar perlas magnéticas recubiertas con estreptavidina y una fuerza magnética para separar los complejos de ARN biotinilado/ADN genómico de los fragmentos de ADN genómico que no están presentes dentro de un complejo de ARN biotinilado/ADN genómico. Los complejos de ARN biotinilado/ADN genómico obtenidos pueden tratarse para eliminar el ARN capturado de las perlas magnéticas, dejando de este modo fragmentos de ADN genómico intactos que contienen un locus a analizar. Estos fragmentos de ADN genómico intactos que contienen los loci a analizar pueden amplificarse mediante el uso de, por ejemplo, técnicas de PCR. Alternativamente, puede emplearse una reacción de PCR multiplex para enriquecer los loci de interés. Los cebadores de PCR pueden diseñarse para flanquear los loci de interés y puede ejecutarse una reacción de PCR para amplificar secuencias que comprenden tales loci.

Los fragmentos de ADN genómico enriquecido pueden secuenciarse mediante el uso de cualquier técnica de secuenciación. Además de la secuenciación de Sanger, numerosas máquinas y estrategias de secuenciación adecuadas son bien conocidas en la técnica, incluidas, pero no se limitan a, las desarrolladas por Illumina (el Genome Analyzer; Bennett y otros (2005) *Pharmacogenomics*, 6:373-382; HiSeq; MiSeq); por Applied Biosystems, Inc. (el secuenciador SOLiD™; solid.appliedbiosystems.com); por Roche (por ejemplo, el secuenciador 454 GS FLX™; Margulies y otros (2005) *Nature*, 437:376-380; patentes núms. 6,274,320; 6,258,568; 6,210,891); por Helicos Biosciences (sistema Heliscope™, ver, por ejemplo, publicación de solicitud de patente de Estados Unidos núm. 2007/0070349); por Oxford Nanopore (por ejemplo, GridION™ y MinION™, ver, por ejemplo, la solicitud internacional núm. PCT/GB2009/001690, publicación núm. WO/2010/004273); y por otros.

Los resultados de secuenciación de los fragmentos de ADN genómico pueden usarse para identificar los loci que tienen un desequilibrio alélico. En algunos casos, puede realizarse un análisis del estado de desequilibrio alélico de los loci sobre una longitud de un cromosoma para determinar la longitud de las regiones de desequilibrio alélico. Por ejemplo, un tramo de ubicaciones de SNP que están separadas (por ejemplo, separadas de aproximadamente 25 kb a aproximadamente 100 kb) a lo largo de un cromosoma puede evaluarse mediante secuenciación, y los resultados de secuenciación pueden usarse para determinar no solo la presencia de una región de desequilibrio alélico (por ejemplo, homocigosidad somática) a lo largo de un cromosoma sino también la longitud de esa región de desequilibrio. Los resultados de secuenciación obtenidos pueden usarse para generar un gráfico que traza las dosificaciones de alelos a lo largo de un cromosoma. La dosificación de alelos d_i para SNP i puede calcularse a partir del número ajustado de sondas capturadas para dos alelos (A_i y B_i): $d_i = A_i / (A_i + B_i)$. Un ejemplo de tal gráfico se presenta en la Figura 2.

Una vez que se ha determinado el genotipo de la muestra (por ejemplo, homocigosidad) para una pluralidad de loci (por ejemplo, los SNP), pueden usarse técnicas comunes para identificar los loci y las regiones de desequilibrio alélico debido a un cambio somático (por ejemplo, LOH). Una forma de determinar si el desequilibrio se debe al cambio somático es comparar el genotipo somático con la línea germinal. Por ejemplo, el genotipo para una pluralidad de loci (por ejemplo, los SNP) puede determinarse tanto en una muestra de línea germinal (por ejemplo, sangre) como en una muestra somática (por ejemplo, tumor). Los genotipos para cada muestra pueden compararse (típicamente por ordenador) para determinar si el genoma de la célula de la línea germinal era, por ejemplo, heterocigoto y si el genoma de la célula somática era, por ejemplo, homocigoto. Tales loci son loci de LOH y las regiones de tales loci son regiones de LOH.

Las técnicas computacionales también pueden usarse para determinar si el desequilibrio alélico es somático (por ejemplo, debido a LOH). Tales técnicas son particularmente útiles cuando una muestra de línea germinal no está disponible para análisis y comparación. Por ejemplo, los algoritmos tales como los descritos en otra parte pueden usarse para detectar regiones de desequilibrio alélico mediante el uso de la información a partir de matrices de SNP (Nannya y otros, *Cancer Res.*, 65:6071-6079 (2005)). Típicamente estos algoritmos no tienen en cuenta explícitamente la contaminación de muestras tumorales con tejido benigno. Cf. solicitud internacional núm. PCT/US2011/026098 de Abkevich y otros; Goransson y otros, *PLoS One* (2009) 4(6):e6057. Frecuentemente esta contaminación es lo suficientemente alta como para dificultar la detección de regiones de desequilibrio alélico. Los métodos analíticos mejorados de acuerdo con la presente invención para identificar un desequilibrio alélico, incluso a pesar de la contaminación, incluyen los incorporados en productos de software informático como se describe a continuación.

Lo siguiente es un ejemplo. Si la relación observada (por ejemplo, MCP) de las señales de dos alelos, A y B, es de dos a uno, hay dos posibilidades. La primera posibilidad es que las células cancerosas tienen LOH con delección del alelo B en una muestra con un 50 % de contaminación con células normales. La segunda posibilidad es que no hay LOH pero el alelo A está duplicado en una muestra sin contaminación con células normales. Puede implementarse un algoritmo como un programa informático como se describe en la presente descripción para reconstruir regiones de LOH en función de los datos de genotipo (por ejemplo, genotipo de SNP). Un punto del algoritmo es reconstruir primero los números de copias específicas de alelo (ASCN) en cada locus (por ejemplo, los SNP). Los ASCN son los números de copias de ambos alelos paternos y maternos. Una región de LOH se determina entonces como un tramo de los SNP con uno de los ASCN (paternos o maternos) que es cero. El algoritmo puede basarse en maximizar una función de probabilidad y puede ser conceptualmente parecido a un algoritmo descrito previamente diseñado para reconstruir el número total de copias (en lugar de ASCN) en cada locus (por ejemplo, los SNP). Ver *la* solicitud internacional núm. PCT/US2011/026098 (publicación núm. WO/2011/106541). La función de probabilidad puede maximizarse sobre ASCN de todos los loci, el nivel de contaminación con tejido benigno, el número total de copias promedio en todo el genoma y el nivel de ruido

específico de la muestra. Los datos de entrada para el algoritmo pueden incluir o consistir en (1) intensidades de señal normalizadas específicas de muestra para ambos alelos de cada locus y (2) conjunto de parámetros específicos de ensayo (específicos para diferentes matrices de SNP y para un enfoque basado en secuencias) definido en función del análisis de gran número de muestras con perfiles de ASCN conocidos.

5

En algunos casos, puede usarse un proceso de selección para seleccionar los loci (por ejemplo, loci de SNP) a evaluar mediante el uso de un ensayo configurado para identificar los loci que tienen un desequilibrio alélico (por ejemplo, ensayos basados en matriz de SNP y ensayos basados en secuenciación). Por ejemplo, cualquier ubicación de SNP humano puede seleccionarse para su inclusión en un ensayo basado en matriz de SNP o un ensayo basado en secuenciación configurado para identificar los loci que tienen un desequilibrio alélico dentro del genoma de las células. En algunos casos, 0,5, 1,0, 1,5, 2,0, 2,5 millones o más loci (por ejemplo, ubicaciones de SNP) presentes dentro del genoma humano pueden evaluarse para identificar los loci que (a) no están presentes en el cromosoma Y, (b) no son loci mitocondriales, (c) tienen una frecuencia de alelos minoritarios de al menos aproximadamente 5 % en la población de interés (por ejemplo, caucásicos), (d) tienen una frecuencia de alelos minoritarios de al menos aproximadamente 1 % en tres poblaciones distintas a la población de interés (por ejemplo, chino, japonés y yoruba), y/o (e) no tienen una desviación significativa del equilibrio de Hardy-Weinberg en ninguna de estas poblaciones. En algunos casos, pueden seleccionarse más de 100 000, 150 000 o 200 000 loci humanos que cumplen con los criterios (a) a (e). De los loci humanos que cumplen con los criterios (a) a (e), puede seleccionarse un grupo de loci (por ejemplo, máximo 2 500, 5 000, 7 500, 10 000, 20 000, 30 000, 40 000, 50 000, 75 000, 100 000, 150 000 o 200 000 loci) de manera que los loci tengan un alto grado de frecuencia de alelos en la población de interés, abarquen el genoma humano de una manera un tanto espaciada (por ejemplo, al menos un locus de interés cada 5 kb, 10 kb, 25 kb, 50 kb, 75 kb, 100 kb, 150 kb, 200 kb, 300 kb, 400 kb, 500 kb o más), y no estén en desequilibrio de ligamiento con otro locus seleccionado en ninguna de las poblaciones usadas para el análisis. En algunos casos, pueden seleccionarse alrededor de 40, 50, 60, 70, 80, 90, 100, 110, 120, 130 mil o más loci para cumplir con cada uno de estos criterios e incluirlos en un ensayo configurado para identificar regiones de desequilibrio alélico en un genoma humano. Por ejemplo, pueden seleccionarse entre aproximadamente 70 000 y aproximadamente 90 000 (por ejemplo, aproximadamente 80 000) SNP para el análisis con un ensayo basado en matriz de SNP, y entre aproximadamente 45 000 y aproximadamente 55 000 (por ejemplo, aproximadamente 54 000) SNP pueden seleccionarse para el análisis con un ensayo basado en secuenciación.

En consecuencia, en un aspecto de la presente descripción, se proporciona un método para detectar el estado de desequilibrio alélico en una pluralidad de loci genómicos en una muestra de un paciente, que comprende las etapas de enriquecer una muestra de ADN genómico en moléculas de ADN cada una que comprende un locus de interés; secuenciar dichas moléculas de ADN para determinar el genotipo en cada locus; determinar para cada locus si tiene un desequilibrio alélico.

35

En otro aspecto de la presente descripción, se proporciona un método para detectar el estado de LOH en una pluralidad de loci genómicos en una muestra de un paciente, que comprende las etapas de enriquecer una muestra de ADN genómico en moléculas de ADN cada una que comprende un locus de interés; secuenciar dichas moléculas de ADN para determinar el genotipo en cada locus; determinar para cada locus homocigoto si es homocigoto debido a LOH.

40

En otro aspecto de la presente descripción, se proporciona un método para detectar el estado del número de copias en una pluralidad de loci genómicos en una muestra de un paciente, que comprende las etapas de enriquecer una muestra de ADN genómico en moléculas de ADN cada una que comprende un locus de interés; secuenciar dichas moléculas de ADN; y cuantificar cada alelo en cada locus para determinar su número de copias.

45

En algunas modalidades se evalúan al menos 10, 50, 100, 1 000, 10 000, 50 000, 55 000, 75 000, 100 000, 150 000, 200 000, 300 000, 400 000, 500 000, 750 000, 1 000 000, 2 000 000 o más loci. En algunas modalidades estos loci están separados uniformemente a lo largo del genoma. Como se usa en la presente descripción, los loci están "separados uniformemente a lo largo del genoma" cuando la diferencia porcentual entre la distancia_{AB} entre dos loci A y B y la distancia_{CD} entre cualquiera de otros dos loci C y D (es decir, $100 \cdot (\text{distancia}_{AB} - \text{distancia}_{CD}) / \text{distancia}_{AB}$ o $100 \cdot (\text{distancia}_{AB} - \text{distancia}_{CD}) / \text{distancia}_{CD}$) es menor o igual a 50 %, 40 %, 30 %, 20 %, 15 %, 10 %, 9 %, 8 %, 7 %, 6 %, 5 %, 4 %, 3 %, 2 %, o 1 %. Tal diferencia porcentual se denomina en la presente descripción la "separación genómica" de los loci. En algunas modalidades la muestra es una muestra de tejido FFPE. En algunas modalidades la muestra es una muestra tumoral del paciente.

55

Otro aspecto de la descripción proporciona un sistema para determinar el estado de desequilibrio alélico en una pluralidad de loci en una muestra que comprende: un analizador de muestras para (1) enriquecer una muestra de ADN genómico en moléculas de ADN cada una que comprende un locus de interés y (2) secuenciar dichas moléculas de ADN para producir una pluralidad de señales cuantitativas sobre cada locus; un programa informático para analizar dicha pluralidad de señales cuantitativas para determinar si cada locus tiene un desequilibrio alélico.

60

Otro aspecto de la descripción proporciona un sistema para determinar el estado de LOH en una pluralidad de loci en una muestra que comprende: un analizador de muestras para (1) enriquecer una muestra de ADN genómico en moléculas de ADN cada una que comprende un locus de interés y (2) secuenciar dichas moléculas de ADN para producir una pluralidad de señales cuantitativas sobre cada locus; un programa informático para analizar dicha pluralidad de señales cuantitativas

65

para determinar si cada locus es homocigoto en la muestra; y un programa informático para determinar para cada locus homocigoto si es homocigoto debido a LOH.

Otro aspecto de la descripción proporciona un sistema para detectar el estado del número de copias en una pluralidad de loci genómicos en una muestra de un paciente que comprende: un analizador de muestras para (1) enriquecer una muestra de ADN genómico en moléculas de ADN cada una que comprende un locus de interés y (2) secuenciar dichas moléculas de ADN para producir una pluralidad de señales cuantitativas sobre cada locus; y un programa informático para analizar dicha pluralidad de señales cuantitativas para cuantificar cada alelo en cada locus para determinar su número de copias.

En algunas modalidades de los sistemas de la descripción, un analizador de muestras enriquece la muestra de ADN de interés y secuencia ese ADN. En otras modalidades dos o más analizadores de muestras realizan estas funciones. En algunas modalidades, un programa de software analiza la pluralidad de señales cuantitativas para determinar si cada locus es homocigoto en la muestra y también determina para cada locus homocigoto si es homocigoto debido a LOH.

La figura 3 es un diagrama de un ejemplo de un dispositivo informático 1400 y un dispositivo informático móvil 1450, que puede usarse con las técnicas descritas en la presente descripción. El dispositivo informático 1400 está destinado a representar varias formas de ordenadores digitales, tales como ordenadores portátiles, ordenadores de escritorio, estaciones de trabajo, asistentes digitales personales, servidores, servidores blade, unidades centrales de procesamiento y otros ordenadores apropiados. El dispositivo informático 1450 está destinado a representar diversas formas de dispositivos móviles, tales como asistentes digitales personales, teléfonos celulares, teléfonos inteligentes y otros dispositivos informáticos similares. Los componentes que se muestran aquí, sus conexiones y relaciones, y sus funciones, están destinados a ser solo ilustrativos, y no están destinados a limitar las implementaciones de las invenciones descritas y/o reivindicadas en este documento.

El dispositivo informático 1400 incluye un procesador 1402, memoria 1404, un dispositivo de almacenamiento 1406, una interfaz de alta velocidad 1408 que se conecta a la memoria 1404 y los puertos de expansión de alta velocidad 1410, y una interfaz de baja velocidad 1415 que se conecta al bus de baja velocidad 1414 y al dispositivo de almacenamiento 1406. Cada uno de los componentes 1402, 1404, 1406, 1408, 1410 y 1415, están interconectados mediante diversos buses y pueden montarse en una placa base común o de otras maneras según sea apropiado. El procesador 1402 puede procesar instrucciones para la ejecución dentro del dispositivo informático 1400, incluidas las instrucciones almacenadas en la memoria 1404 o en el dispositivo de almacenamiento 1406 para mostrar información gráfica para un GUI en un dispositivo externo de entrada/salida, tal como una pantalla 1416 acoplada a una interfaz de alta velocidad 1408. En otras implementaciones, pueden usarse múltiples procesadores y/o múltiples buses, según sea apropiado, junto con múltiples memorias y tipos de memoria. Además, pueden conectarse múltiples dispositivos informáticos 1400, donde cada dispositivo proporciona partes de las operaciones necesarias (por ejemplo, como un banco de servidores, un grupo de servidores blade o un sistema multiprocesador).

La memoria 1404 almacena información dentro del dispositivo informático 1400. En una implementación, la memoria 1404 es una unidad o unidades de memoria volátil. En otra implementación, la memoria 1404 es una unidad o unidades de memoria no volátiles. La memoria 1404 también puede ser otra forma de medio legible por ordenador, tal como un disco magnético u óptico.

El dispositivo de almacenamiento 1406 es capaz de proporcionar almacenamiento masivo para el dispositivo informático 1400. En una implementación, el dispositivo de almacenamiento 1406 puede ser o contener un medio legible por ordenador, tal como un dispositivo de disquete, un dispositivo de disco duro, un dispositivo de disco óptico o un dispositivo de cinta, una memoria flash u otro dispositivo de memoria de estado sólido similar, o una disposición de dispositivos, incluidos los dispositivos en una red de área de almacenamiento u otras configuraciones. Un producto de programa informático puede incorporarse de manera tangible en un portador de información. El producto de programa informático también puede contener instrucciones que, cuando se ejecutan, realizan uno o más métodos, tales como los descritos en la presente descripción. El portador de información es un medio legible por ordenador o máquina, tal como la memoria 1404, el dispositivo de almacenamiento 1406, la memoria en el procesador 1402 o una señal propagada.

El controlador de alta velocidad 1408 gestiona las operaciones intensivas de ancho de banda para el dispositivo informático 1400, mientras que el controlador de baja velocidad 1415 gestiona las operaciones intensivas de menor ancho de banda. Tal asignación de funciones es solo ilustrativa. En una implementación, el controlador de alta velocidad 1408 está acoplado a la memoria 1404, la pantalla 1416 (por ejemplo, a través de un procesador o acelerador gráfico) y a puertos de expansión de alta velocidad 1410, que pueden aceptar varias tarjetas de expansión (no mostradas). En la implementación, el controlador de baja velocidad 1415 está acoplado al dispositivo de almacenamiento 1406 y al puerto de expansión de baja velocidad 1414. El puerto de expansión de baja velocidad, que puede incluir varios puertos de comunicación (por ejemplo, USB, Bluetooth, Ethernet o Ethernet inalámbrico) puede estar acoplado a uno o más dispositivos de entrada/salida, tal como un teclado, un dispositivo de puntero, un escáner, un lector óptico, un detector de señal fluorescente o un dispositivo de red tal como un conmutador o enrutador, por ejemplo, a través de un adaptador de red.

El dispositivo informático 1400 puede implementarse en varias formas diferentes, como se muestra en la figura. Por ejemplo, puede implementarse como un servidor estándar 1420, o múltiples veces en un grupo de tales servidores. También puede implementarse como parte de un sistema de servidores en bastidor 1424. Además, puede implementarse en un ordenador personal tal como un ordenador portátil 1422. Alternativamente, los componentes del dispositivo informático 1400 pueden combinarse con otros componentes en un dispositivo móvil (no mostrado), tal como el dispositivo 1450. Cada uno de tales dispositivos puede contener uno o más de dispositivo informático 1400, 1450, y un sistema completo puede estar compuesto de múltiples dispositivos informáticos 1400, 1450 que se comunican entre sí.

El dispositivo informático 1450 incluye un procesador 1452, memoria 1464, un dispositivo de entrada/salida tal como una pantalla 1454, una interfaz de comunicación 1466 y un transceptor 1468, entre otros componentes (por ejemplo, un escáner, un lector óptico, un detector de señal fluorescente). El dispositivo 1450 también puede proporcionarse con un dispositivo de almacenamiento, tal como un micro disco duro u otro dispositivo, para proporcionar almacenamiento adicional. Cada uno de los componentes 1450, 1452, 1464, 1454, 1466 y 1468, están interconectados mediante el uso de varios buses, y varios de los componentes pueden montarse en una placa base común o de otras maneras según sea apropiado.

El procesador 1452 puede ejecutar instrucciones dentro del dispositivo informático 1450, incluidas las instrucciones almacenadas en la memoria 1464. El procesador puede implementarse como un conjunto de circuitos integrados de chips que incluyen procesadores analógicos y digitales separados y múltiples. El procesador puede proporcionar, por ejemplo, la coordinación de los otros componentes del dispositivo 1450, tal como el control de las interfaces del usuario, las aplicaciones ejecutadas por el dispositivo 1450 y la comunicación inalámbrica por el dispositivo 1450.

El procesador 1452 puede comunicarse con un usuario a través de la interfaz de control 1458 y la interfaz de pantalla 1456 acoplada a una pantalla 1454. La pantalla 1454 puede ser, por ejemplo, una pantalla TFT LCD (pantalla de cristal líquido de transistores de película fina) o una pantalla OLED (diodo emisor de luz orgánica) u otra tecnología de pantalla apropiada. La interfaz de pantalla 1456 puede comprender circuitos apropiados para conducir la pantalla 1454 a presentar información gráfica y otra información a un usuario. La interfaz de control 1458 puede recibir comandos de un usuario y convertirlos para enviarlos al procesador 1452. Además, puede proporcionarse una interfaz externa 1462 en comunicación con el procesador 1452, para permitir la comunicación entre áreas cercanas del dispositivo 1450 con otros dispositivos. La interfaz externa 1462 puede proporcionar, por ejemplo, comunicación por cable en algunas implementaciones, o comunicación inalámbrica en otras implementaciones, y también pueden usarse múltiples interfaces.

La memoria 1464 almacena información dentro del dispositivo informático 1450. La memoria 1464 puede implementarse como uno o más de un medio o medios legibles por ordenador, una unidad o unidades de memoria volátil, o una unidad o unidades de memoria no volátil. La memoria de expansión 1474 también puede proporcionarse y conectarse al dispositivo 1450 a través de la interfaz de expansión 1472, que puede incluir, por ejemplo, una interfaz de tarjeta SIMM (módulo de memoria en línea individual). Tal memoria de expansión 1474 puede proporcionar espacio de almacenamiento extra para el dispositivo 1450, o también puede almacenar aplicaciones u otra información para el dispositivo 1450. Por ejemplo, la memoria de expansión 1474 puede incluir instrucciones para llevar a cabo o complementar los procesos descritos en la presente descripción, y también puede incluir información segura. Por lo tanto, por ejemplo, la memoria de expansión 1474 puede proporcionarse como un módulo de seguridad para el dispositivo 1450, y puede programarse con instrucciones que permitan el uso seguro del dispositivo 1450. Además, pueden proporcionarse aplicaciones seguras a través de las tarjetas SIMM, junto con información adicional, tal como colocar información de identificación en la tarjeta SIMM de manera no pirateable.

La memoria puede incluir, por ejemplo, memoria flash y/o memoria NVRAM, como se describe más adelante. En una implementación, un producto de programa informático se incorpora tangiblemente en un portador de información. El producto de programa informático contiene instrucciones que, cuando se ejecutan, realizan uno o más métodos, tales como los descritos en la presente descripción. El portador de información es un medio legible por ordenador o máquina, tal como la memoria 1464, la memoria de expansión 1474, la memoria en el procesador 1452 o una señal propagada que puede ser recibida, por ejemplo, a través del transceptor 1468 o la interfaz externa 1462.

El dispositivo 1450 puede comunicarse de forma inalámbrica a través de la interfaz de comunicación 1466, que puede incluir circuitos de procesamiento de señal digital cuando sea necesario. La interfaz de comunicación 1466 puede proporcionar comunicaciones bajo varios modos o protocolos, tales como llamadas de voz GSM, SMS, EMS o mensajería MMS, CDMA, TDMA, PDC, WCDMA, CDMA2000 o GPRS, entre otros. Tal comunicación puede ocurrir, por ejemplo, a través del transceptor de radiofrecuencia 1468. Además, puede ocurrir una comunicación de corto alcance, tal como mediante el uso de un Bluetooth, WiFi u otro transceptor (no se muestra). Además, el módulo receptor GPS (Sistema de posicionamiento global) 1470 puede proporcionar datos inalámbricos adicionales de navegación y ubicación al dispositivo 1450, que pueden usarse según sea apropiado por las aplicaciones que se ejecutan en el dispositivo 1450.

El dispositivo 1450 también puede comunicarse de forma audible mediante el uso de un códec de audio 1460, que puede recibir información hablada de un usuario y convertirla en una información digital que puede usarse. El códec de audio 1460 también puede generar un sonido audible para un usuario, tal como a través de un altavoz, por ejemplo, en un microteléfono del dispositivo 1450. Tal sonido puede incluir sonido de llamadas telefónicas de voz, puede incluir sonido

grabado (por ejemplo, mensajes de voz, archivos de música, etc.) y también puede incluir sonido generado por aplicaciones que operan en el dispositivo 1450.

5 El dispositivo informático 1450 puede implementarse en varias formas diferentes, como se muestra en la figura. Por ejemplo, puede implementarse como un teléfono celular 1480. También puede implementarse como parte de un teléfono inteligente 1482, asistente digital personal u otro dispositivo móvil similar.

10 Pueden realizarse diversas implementaciones de los sistemas y técnicas descritos en la presente descripción en circuitos electrónicos digitales, circuitos integrados, ASIC especialmente diseñados (circuitos integrados específicos de aplicación), hardware de ordenador, firmware, software y/o combinaciones de estos. Estas diversas implementaciones pueden incluir la implementación en uno o más programas de ordenador que pueden ejecutarse y/o interpretarse en un sistema programable que incluye al menos un procesador programable, que puede tener un propósito especial o general, acoplado para recibir datos e instrucciones, y para transmitir datos e instrucciones a, un sistema de almacenamiento, al menos un dispositivo de entrada y al menos un dispositivo de salida.

15 Estos programas de ordenador (también conocidos como programas, software, aplicaciones de software o código) incluyen instrucciones de máquina para un procesador programable, y pueden implementarse en un procedimiento de alto nivel y/o lenguaje de programación orientado a objetos y/o lenguaje ensamblador/de máquina. Como se usa en la presente descripción, los términos "medio legible por máquina" y "medio legible por ordenador" se refieren a cualquier producto de programa informático, aparato y/o dispositivo (por ejemplo, discos magnéticos, discos ópticos, memoria y dispositivos lógicos programables (PLD)) usado para proporcionar instrucciones de máquina y/o datos a un procesador programable incluido un medio legible por máquina que recibe instrucciones de máquina como una señal legible por máquina. El término "señal legible por máquina" se refiere a cualquier señal usada para proporcionar instrucciones de máquina y/o datos a un procesador programable.

20 Para proporcionar interacción con un usuario, los sistemas y técnicas descritos en la presente descripción pueden implementarse en un ordenador que tenga un dispositivo de visualización (por ejemplo, un CRT (tubo de rayos catódicos) o un monitor LCD (pantalla de cristal líquido) para mostrar información al usuario y un teclado y un dispositivo de puntero (por ejemplo, un ratón o una bola rastreadora) mediante el cual el usuario puede proporcionar la entrada al ordenador. Otros tipos de dispositivos pueden usarse para proporcionar la interacción con un usuario también, por ejemplo, la retroalimentación proporcionada al usuario puede ser cualquier forma de retroalimentación sensorial (por ejemplo, retroalimentación visual, retroalimentación auditiva, y/o retroalimentación táctil), y la entrada del usuario puede ser recibida en cualquier forma, incluida la entrada acústica, hablada o táctil.

25 Los sistemas y técnicas descritos en la presente descripción pueden implementarse en un sistema informático que incluye un componente back end (por ejemplo, un servidor de datos), o que incluye un componente de middleware (por ejemplo, un servidor de aplicaciones), o que incluye un componente front end (por ejemplo, un ordenador del cliente que tiene una interfaz del usuario gráfica y/o un navegador a través del cual un usuario puede interactuar con una implementación de los sistemas y técnicas descritas en la presente descripción), o cualquier combinación de tales componentes back end, middleware, o front end. Los componentes del sistema pueden interconectarse mediante cualquier forma o medio de comunicación de datos digitales (por ejemplo, una red de comunicación). Los ejemplos de redes de comunicación incluyen una red de área local ("LAN"), y una red de área amplia ("WAN"), y el Internet.

30 El sistema informático puede incluir clientes y servidores. Un cliente y un servidor generalmente están alejados entre sí y generalmente interactúan a través de una red de comunicación. La relación entre cliente y servidor surge en virtud de los programas informáticos que se ejecutan en los ordenadores respectivos y que tienen una relación cliente-servidor entre sí.

35 En algunos casos, un sistema proporcionado en la presente descripción puede configurarse para incluir uno o más analizadores de muestras. Un analizador de muestras puede configurarse para producir una pluralidad de señales sobre el ADN genómico de una célula cancerosa. Por ejemplo, un analizador de muestras puede producir señales que pueden interpretarse de una manera que identifique el estado de desequilibrio alélico de los loci a lo largo de un cromosoma. En algunos casos, un analizador de muestras puede configurarse para llevar a cabo una o más etapas de un ensayo basado en secuenciación y puede configurarse para producir y/o capturar señales de tales ensayos. En algunos casos, un sistema informático proporcionado en la presente descripción puede configurarse para incluir un dispositivo informático. En tales casos, el dispositivo informático puede configurarse para recibir señales de un analizador de muestras.

40 El dispositivo informático puede incluir instrucciones ejecutables por ordenador o un programa informático (por ejemplo, software) que contenga instrucciones ejecutables por ordenador para llevar a cabo uno o más de los métodos o etapas descritos en la presente descripción. En algunos casos, tales instrucciones ejecutables por ordenador pueden dar una instrucción a un dispositivo informático para que analice las señales de un analizador de muestras, de otro dispositivo informático o de un ensayo basado en secuenciación. El análisis de tales señales puede llevarse a cabo para determinar genotipos, el desequilibrio alélico en ciertos loci, las regiones de desequilibrio alélico, el número de regiones de desequilibrio alélico, para determinar el tamaño de las regiones de desequilibrio alélico, para determinar el número de regiones de desequilibrio alélico que tienen un tamaño o intervalo de tamaños particular, o para determinar una combinación de estos elementos.

5 En algunos casos, un sistema proporcionado en la presente descripción puede incluir instrucciones ejecutables por ordenador o un programa informático (por ejemplo, software) que contiene instrucciones ejecutables por ordenador para formatear una salida que proporciona una indicación sobre el número de copias, el desequilibrio alélico, la LOH o una combinación de estos elementos.

10 En algunos casos, un sistema proporcionado en la presente descripción puede incluir un dispositivo de preprocesamiento configurado para procesar una muestra (por ejemplo, células cancerosas) de manera que pueda realizarse un ensayo basado en secuenciación. Los ejemplos de dispositivos de preprocesamiento incluyen, sin limitación, dispositivos configurados para enriquecer poblaciones celulares en células cancerosas en lugar de células no cancerosas, dispositivos configurados para lisar células y/o extraer ácido nucleico genómico, y dispositivos configurados para enriquecer una muestra en fragmentos de ADN genómico particulares.

15 **EJEMPLOS**

El proceso descrito aquí utilizó un sistema de captura Agilent SureSelect seguido de la secuenciación en Illumina HiSeq, sin embargo, podría usarse cualquier método de captura en solución o basado en soporte sólido y la plataforma de secuenciación en paralelo de alto rendimiento.

20 El proceso de selección de diseño inicial utilizó los ~2,5 millones de SNP en la matriz de SNP Illumina Omni2.5. Se eligió esta lista de los SNP porque actualmente es la lista más grande de los SNP a partir de la cual existe información de genotipado disponible para múltiples grupos poblacionales diferentes. Todas las ubicaciones de 2 448 785 SNP se introdujeron en el asistente de Agilent eArray Sure Select Target Enrichment para lecturas largas de extremo único mediante el uso de las configuraciones predeterminadas. 1 353 042 pasaron los criterios de selección y tuvieron cebos diseñados.

25 Después, se seleccionaron 110 000 SNP con altas frecuencias de alelos minoritarios y que abarcan el genoma uniformemente. En la selección, se descartaron los SNP en desequilibrio de ligamiento fuerte y los SNP con una fuerte desviación del equilibrio de Hardy-Weinberg.

30 Se construyeron dos diseños de bibliotecas preliminares compuestos por 55 000 sondas, cada una que se dirige a 55 000 ubicaciones de SNP diferentes. Las pruebas se llevaron a cabo mediante el uso de una muestra de ADN normal de alta calidad para verificar la captura uniforme de ambos alelos de cada SNP. Además, se capturaron 4 muestras FFPE y se usaron para seleccionar las sondas con el rendimiento más óptimo. Se buscaron sondas que mostraran una captura robusta e incluso una profundidad de secuencia sin representación excesiva o insuficiente de las lecturas de secuencia en la biblioteca de secuenciación final.

35 El diseño final de bibliotecas de sondas de captura estaba compuesto por las 55 000 sondas óptimas identificadas mediante el uso de los diseños de captura preliminares.

40 Los resultados de la medición del número de copias y LOH mediante el uso de la técnica de secuenciación anterior se muestran en la Figura 2 (donde la Figura 1 muestra el análisis de micromatrices en tejido fresco congelado como una comparación).

45 Aunque la invención anterior se ha descrito en algún detalle a modo de ilustración y ejemplo para propósitos de claridad y comprensión, resultará evidente que ciertos cambios y modificaciones pueden ponerse en práctica dentro del alcance de las reivindicaciones adjuntas.

REIVINDICACIONES

- 5 1. Un método *in vitro* para detectar el estado de pérdida de heterocigosidad (LOH) en una pluralidad de loci genómicos de SNP en una muestra de tejido embebida en parafina, fijada en formalina de un paciente, que comprende:
 - 5 enriquecer una muestra de ADN genómico en moléculas de ADN cada una que comprende un locus de SNP de interés;
 - secuenciar dichas moléculas de ADN para determinar el genotipo en cada locus de SNP;
 - determinar para cada locus de SNP homocigoto si es homocigoto debido a LOH.
- 10 2. Uso de un sistema para determinar el estado de LOH en una pluralidad de loci genómicos de SNP en una muestra de tejido embebida en parafina, fijada en formalina, en donde el sistema comprende:
 - un analizador de muestras para
 - 15 (1) enriquecer una muestra de ADN genómico en moléculas de ADN cada una que comprende un locus de SNP de interés y
 - (2) secuenciar dichas moléculas de ADN para producir una pluralidad de señales cuantitativas sobre cada locus de SNP;
 - un medio de programa informático para analizar dicha pluralidad de señales cuantitativas para determinar el genotipo de cada locus de SNP en la muestra; y
 - un medio informático para determinar para cada locus de SNP homocigoto si es homocigoto debido a LOH.
- 20 3. El método de conformidad con la reivindicación 1 o el uso del sistema de conformidad con la reivindicación 2, en donde dicha pluralidad de loci de SNP genómicos comprende al menos 10, 50, 100, 1 000, 10 000, 50 000, 55 000, 75 000, 100 000, 150 000, 200 000, 300 000, 400 000, 500 000, 750 000, 1 000 000 o 2 000 000 o más loci.
- 25 4. El método de conformidad con la reivindicación 1 o el uso del sistema de conformidad con la reivindicación 2, en donde se determina que un locus de SNP tiene un desequilibrio alélico si la proporción de copias mayoritarias en el locus de SNP es 0.51 o mayor.
- 30 5. El método de conformidad con la reivindicación 1 o el uso del sistema de conformidad con la reivindicación 2, en donde dichos loci genómicos de SNP están espaciados de uniformemente a lo largo del genoma.
- 35 6. El método de conformidad con la reivindicación 5, en donde la diferencia porcentual entre la distancia entre dos cualesquiera de dicha pluralidad de loci genómicos de SNP y la distancia entre cualesquiera otros dos de dicha pluralidad de loci de SNP es menor o igual a 50 %, 40 %, 30 %, 20 %, 15 %, 10 %, 9 %, 8 %, 7 %, 6 %, 5 %, 4 %, 3 %, 2 % o 1 %.
7. El método de conformidad con la reivindicación 1 o el uso del sistema de conformidad con la reivindicación 2, en donde dicha muestra es una muestra tumoral extraída del paciente.
- 40 8. El método de conformidad con la reivindicación 7, en donde dicha muestra tumoral contiene al menos 5 %, 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 99 % o más de contaminación con células no tumorales.

Figura 1

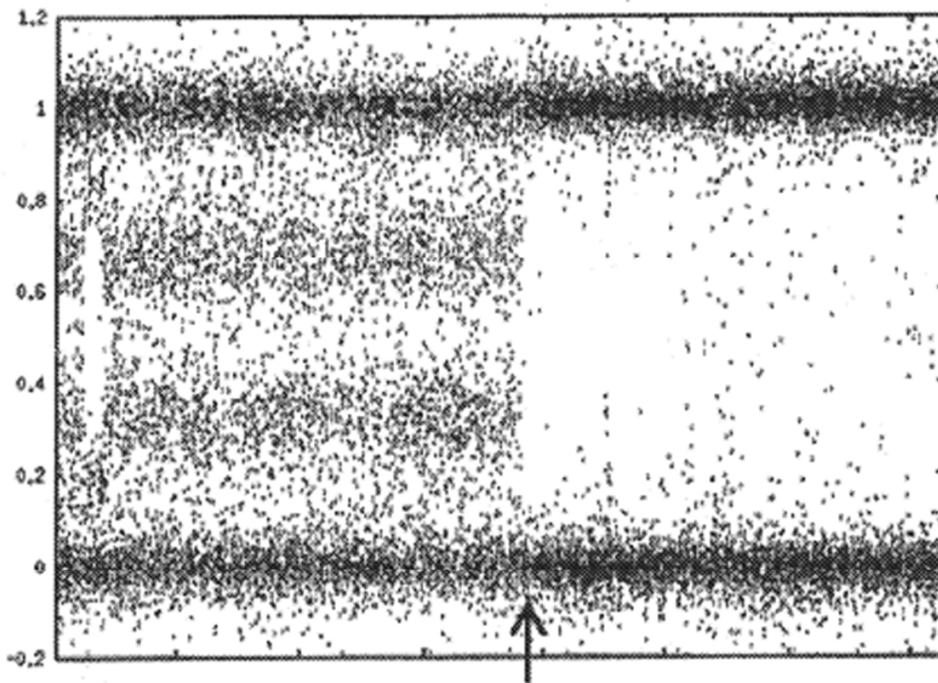
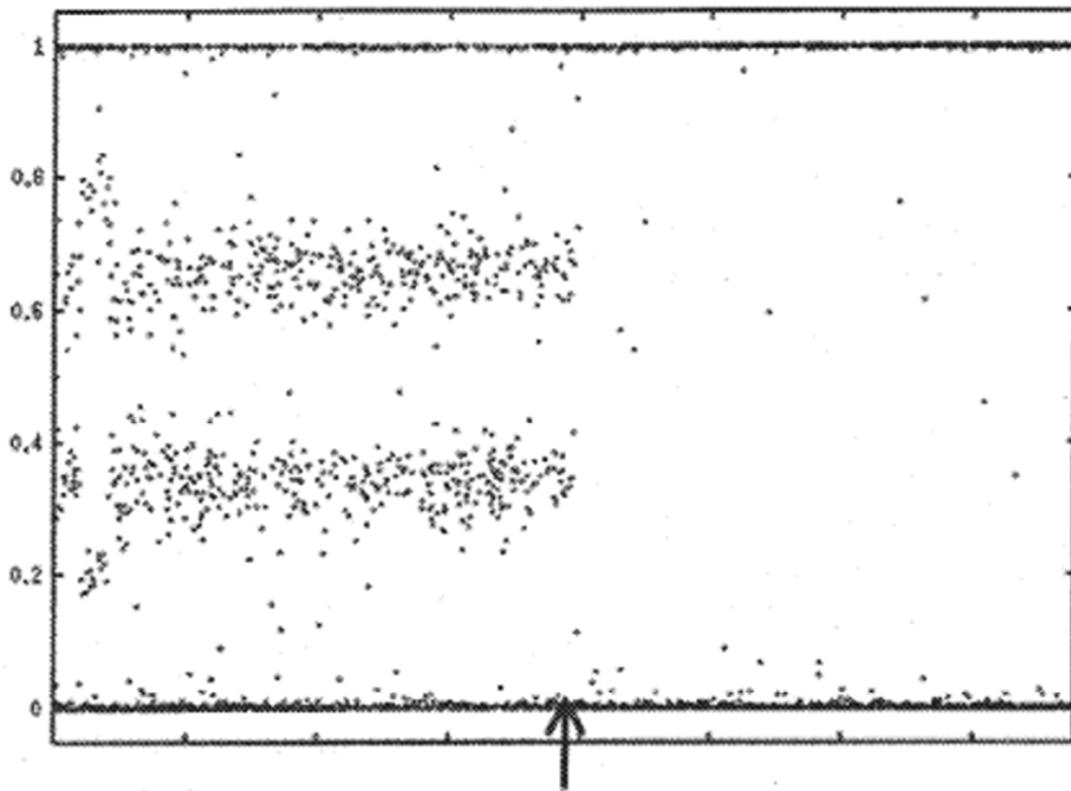


Figura 2



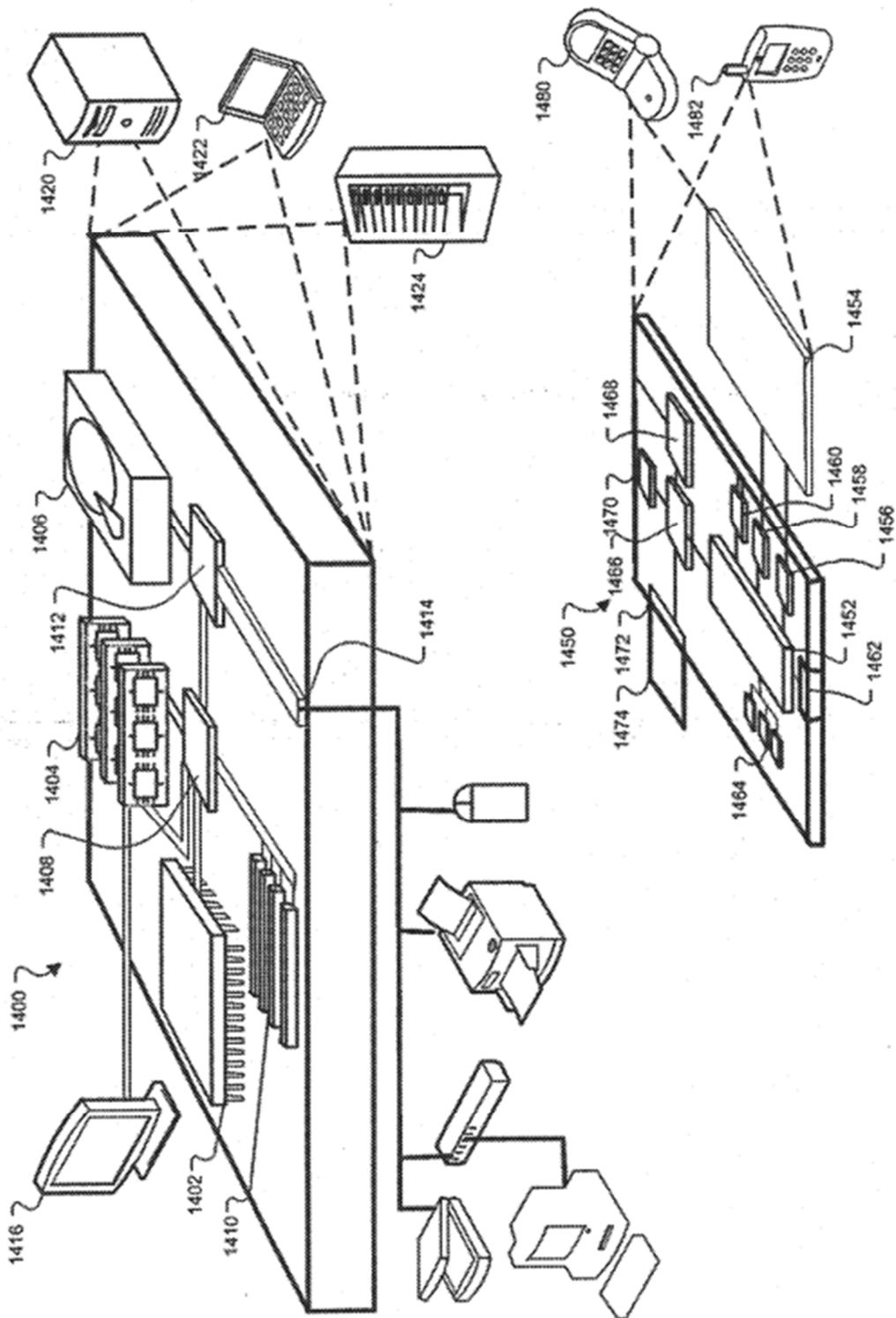


Figura 3