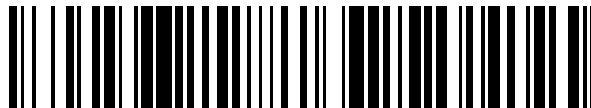


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 703 764**

51 Int. Cl.:

C12Q 1/6886 (2008.01)

C12Q 1/6809 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **13.12.2013 PCT/US2013/075015**

87 Fecha y número de publicación internacional: **19.06.2014 WO14093825**

96 Fecha de presentación y número de la solicitud europea: **13.12.2013 E 13863107 (2)**

97 Fecha y número de publicación de la concesión europea: **17.10.2018 EP 2931922**

54 Título: **Biomarcadores personalizados para el cáncer**

30 Prioridad:
14.12.2012 US 201261737626 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
12.03.2019

73 Titular/es:
**CHRONIX BIOMEDICAL (100.0%)
5941 Optical Court, Suite 203E
San Jose, California 95138, US**

72 Inventor/es:
**SCHÜTZ, EKKEHARD;
BECK, JULIA y
URNOVITZ, HOWARD**

74 Agente/Representante:
PONS ARIÑO, Ángel

ES 2 703 764 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

Biomarcadores personalizados para el cáncer

5 **Antecedentes de la invención**

Existe una necesidad de métodos para monitorizar la respuesta de un paciente con cáncer a una terapia. Esta invención aborda esa necesidad.

10 **Breve resumen de la invención**

La invención se basa, en parte, en el descubrimiento de que las inserciones de elementos repetitivos se producen en los tumores de pacientes con cáncer, y que dicha región de inserción puede usarse como un biomarcador individualizado del paciente en particular para monitorizar la respuesta del paciente a una terapia.

15 La invención proporciona métodos para la selección de un biomarcador personalizado para un paciente que tiene un tumor. Los métodos de la invención comprenden (a) la preparación de una colección de fragmentos de ADN a partir de muestras de ácidos nucleicos que comprenden el ADN obtenido del paciente a partir de cada una de las siguientes fuentes: células mononucleares de sangre periférica (PBMC), el tumor y ácidos nucleicos acelulares previos a la terapia, ácidos nucleicos que se obtienen antes de una terapia; (b) la secuenciación de los fragmentos de ADN de las colecciones para identificar los fragmentos de ADN que comprenden una región de inserción de un elemento repetitivo; y (c) la comparación de las regiones de inserción de un elemento repetitivo identificadas en la muestra de PBMC con las regiones de inserción de un elemento repetitivo identificadas en la muestra del tumor y/o en la muestra de los ácidos nucleicos acelulares para identificar las regiones de inserción de un elemento repetitivo que están presentes en el tumor, pero no en las PBMC, identificando así un biomarcador personalizado para el paciente. El biomarcador puede ser cualquier elemento repetitivo tal como una secuencia Alu, un elemento nuclear disperso corto (SINE), un elemento nuclear disperso largo (LINE) o un retrovirus endógeno (ERV).

30 En algunas realizaciones, la etapa (c) comprenderá adicionalmente la comparación de las regiones de inserción de un elemento repetitivo identificadas en la muestra de PBMC con las correspondientes regiones de la base de datos del genoma humano para identificar las regiones de inserción en la muestra de PBMC que no están presentes en la base de datos del genoma humano.

35 En una realización típica, se amplificará el ADN de una o más muestras (por ejemplo, mediante el uso de una PCR). En este caso, los métodos pueden comprender adicionalmente las etapas de la adición de secuencias adaptadoras que comprenden sitios de cebado universales a cada uno de los fragmentos de ADN y la amplificación de los fragmentos de ADN, antes de la secuenciación de los fragmentos. La amplificación mediante una PCR puede llevarse a cabo mediante el uso de un cebador específico del elemento repetitivo y un cebador universal, en cuyo caso solo se amplificarán los fragmentos que comprenden el elemento repetitivo. Por lo tanto, permitiendo la identificación de los fragmentos que comprenden las inserciones de un elemento repetitivo. Alternativamente, la etapa de amplificación puede llevarse a cabo mediante el uso de dos cebadores universales, en cuyo caso se amplificarán todos los fragmentos. Los fragmentos que comprenden las inserciones de un elemento repetitivo también pueden ser identificados y/o enriquecidos mediante el uso de sondas oligonucleotídicas específicas para el elemento repetitivo. Las ondas pueden usarse para extraer físicamente los fragmentos de ADN deseados que comprenden el elemento repetitivo de otros fragmentos de ADN de la colección, enriqueciendo así la colección en los fragmentos de ADN deseados. La etapa de enriquecimiento puede incluir, por ejemplo, la unión de complejos de hibridación que comprenden sondas biotiniladas a un soporte sólido (por ejemplo, a una microesfera magnética).

50 En algunas realizaciones, los métodos pueden comprender adicionalmente la etapa de la obtención de la muestra de ácidos nucleicos del paciente a partir de ácidos nucleicos acelulares posteriores a la terapia, los ácidos nucleicos que son obtenidos después de la terapia. En estas realizaciones, las regiones de un elemento repetitivo identificadas en la muestra de ácidos nucleicos acelulares previos a la terapia se comparan con la muestra de ácidos nucleicos acelulares posteriores a la terapia para identificar las regiones de inserción de un elemento repetitivo que no son detectadas o que están muy reducidas en la muestra de ácidos nucleicos acelulares posteriores a la terapia.

55 En un aspecto, la invención proporciona un método para la selección de un biomarcador personalizado para un paciente que tiene un tumor, método que comprende: (a) la realización de al menos una reacción de amplificación del ADN de las muestras de ácido nucleico que comprenden el ADN obtenido del paciente a partir de cada una de las siguientes fuentes: células mononucleares de sangre periférica (PBMC), el tumor, ácidos nucleicos acelulares previos a la terapia, los ácidos nucleicos que se obtienen antes de una terapia, y ácidos nucleicos acelulares posteriores a la terapia, los ácidos nucleicos que se obtienen después de la terapia, en el que la amplificación se lleva a cabo mediante el uso de un cebador de una secuencia Alu y un cebador universal; (b) la secuenciación del ADN obtenido en las reacciones de amplificación para determinar las regiones que comprenden una secuencia de inserción Alu con respecto a las correspondientes regiones de la base de datos del genoma humano; (c) la comparación de las regiones de inserción Alu identificadas en la etapa (b) a partir de la muestra de PBMC con las regiones de inserción Alu identificadas en la etapa (b) de la muestra del tumor para identificar las regiones de

inserción Alu que están presentes en el tumor, pero no en las PBMC; (d) la comparación de las regiones de inserción Alu identificadas en la etapa (c) con las regiones de inserción Alu identificadas en la muestra de ácidos nucleicos acelulares del paciente previa a la terapia, para identificar las regiones de inserción Alu que están presentes en la muestra del tumor y en la muestra de ácidos nucleicos acelulares previa a la terapia; y (e) la comparación de las regiones de inserción Alu identificadas en la etapa (d) con las regiones de inserción Alu identificadas en la muestra de ácidos nucleicos acelulares del paciente posterior a la terapia, para identificar las regiones de inserción Alu que no son detectadas en la muestra de ácidos nucleicos acelulares posterior a la terapia, identificando así un biomarcador personalizado para el paciente. En algunas realizaciones de los métodos de la invención, la terapia es una cirugía. En algunas realizaciones, la terapia es una radiación. En algunas realizaciones, la terapia es una quimioterapia.

En otro aspecto, la invención proporciona un método para la selección de un biomarcador personalizado para un paciente que tiene un tumor, método que comprende: (a) la realización de al menos una reacción de amplificación del ADN de la muestra de ácidos nucleicos que comprende el ADN obtenido de los ácidos nucleicos en circulación previa a la terapia y de la muestra de ácidos nucleicos que comprende el ADN obtenido de los ácidos nucleicos en circulación posterior a la terapia, en el que la amplificación se lleva a cabo mediante el uso de un cebador de una secuencia Alu y un cebador universal; (b) la secuenciación del ADN obtenido en las reacciones de amplificación para determinar las regiones que comprenden una secuencia de inserción Alu con respecto a un genoma humano normal y que pueden ser ubicadas de forma única en una región definida de un genoma humano normal; (c) la cuantificación del nivel de una región de inserción Alu identificada en la etapa (b) a partir de la muestra de ADN en circulación previa a la terapia en comparación con el nivel de las regiones de inserción Alu presente en los controles normales para identificar las regiones de inserciones de Alu que están presentes en el ADN acelular en circulación a unos niveles mayores que los controles normales; y (d) la comparación del nivel de las regiones de inserción Alu identificadas en la etapa (c) con el nivel de las regiones de inserción Alu en la muestra de ADN acelular posterior a la terapia para identificar las regiones de inserción Alu que no son detectadas a un nivel diferente del normal, identificando así un biomarcador personalizado para el paciente. En algunas realizaciones de los métodos de la invención, la terapia es una cirugía. En algunas realizaciones, la terapia es una radiación. En algunas realizaciones, la terapia es una quimioterapia.

En algunas realizaciones de los métodos de la invención, el método comprende adicionalmente la realización de una reacción de amplificación adicional del ADN de cada una de las muestras de ácidos nucleicos, en el que las segundas reacciones de amplificación comprenden un segundo cebador de una secuencia Alu y un segundo cebador universal. En algunas realizaciones, una de las reacciones de amplificación comprende un par de cebadores que comprende la SEQ ID NO: 5 y un cebador universal; o un par de cebadores que comprende la SEQ ID NO: 6 y un cebador universal. En algunas realizaciones, una de las reacciones de amplificación comprende un par de cebadores que comprende la SEQ ID NO: 5 y un cebador universal, y las reacciones de amplificación adicionales comprenden un par de cebadores que comprende la SEQ ID NO: 6 y un cebador universal.

Los biomarcadores personalizados pueden ser detectados para un paciente con cualquier tipo de cáncer. En algunas realizaciones, el paciente tiene cáncer de mama, cáncer colorrectal, cáncer de próstata, cáncer de pulmón, cáncer de hígado, cáncer de páncreas, cáncer de ovario, cáncer cervical, cáncer gástrico, cáncer epidermoide de cabeza y cuello, cáncer de vejiga o cáncer renal. En algunas realizaciones, el paciente tiene un sarcoma. En algunas realizaciones, el paciente tiene un linfoma o una leucemia. En algunas realizaciones, el paciente tiene un tumor sólido. En algunas realizaciones, el paciente tiene un tumor de mama.

En un aspecto adicional, la invención proporciona un método para la monitorización de la respuesta de un paciente que tiene un tumor a una terapia, método que comprende la realización de un método de la invención para la identificación de un biomarcador individualizado y la evaluación del nivel de dicho biomarcador en los ácidos nucleicos en circulación acelulares obtenidos a partir del paciente antes y después de la terapia. En algunas realizaciones, la terapia es una cirugía. En algunas realizaciones, la terapia es una radiación. En algunas realizaciones, la terapia es una quimioterapia.

Breve descripción de los dibujos

La Figura 1 muestra un diagrama que ilustra la evaluación de la región genómica del ADN acelular. Se muestran los valores de Z previos a la cirugía (círculos blancos) y las diferencias en los valores de Z previos y posteriores a la cirugía (rombos oscuros). Las barras representan las sumas de las lecturas por intervalo que entraban en los análisis. Los intervalos en los que los valores de Z previos a la cirugía < 3 fueron censurados a cero.

La Figura 2 muestra un diagrama de flujo que ilustra un algoritmo de filtrado usado para la identificación de las regiones que es probable que hereden las inserciones Alu específicas de las células tumorales.

La Figura 3A hasta E muestra las imágenes de un gel de agarosa que ilustran las bandas generadas mediante una PCR de una PCR específica de una inserción Alu de un tumor individual. En los paneles A hasta E, el lado derecho muestra el número normalizado de lecturas detectadas para la inserción Alu específica obtenida a partir de la secuenciación dirigida a la Alu. El lado izquierdo de los paneles A hasta E proporciona los resultados de las

respectivas PCR. Se diseñaron cebadores para amplificar la región de las inserciones Alu específicas; si la región está presente, se observa una segunda banda de mayor tamaño en el ADN de una muestra.

La Figura 4A-E ilustra los valores de Z para las lecturas del ADN en circulación y del ADN tumoral cartografiadas para los elementos repetitivos conocidos del genoma humano, separadas para la PCR de la ALU40 y la de la ALU 259.

La Figura 5 ilustra los resultados de dos pacientes que muestran una reducción diferencial en los biomarcadores de ADN acelular.

La Figura 6 muestra la ubicación de las lecturas L1 que tienen un par cartografiado de forma única con respecto a la secuencia LI-HS de la línea humana. Las cifras representan la posición en la L1-HS; las barras negras de la parte inferior representan la posición de las sondas de enriquecimiento (Tabla 4).

Descripción detallada de la invención

Según se usa en el presente documento, un "biomarcador" se refiere a una región del ADN que comprende un elemento repetitivo (por ejemplo, secuencias Alu, elementos nucleares dispersos cortos (SINE) y elementos nucleares dispersos largos (LINE) o retrovirus endógenos (ERV). Dicho biomarcador, en el contexto de esta invención, es particular de un paciente específico y no está presente en el ADN normal, por ejemplo, en el ADN de una muestra de sangre periférica del paciente, o no está presente en la correspondiente región de un genoma humano normal. Un ejemplo de un genoma humano normal de referencia es el genoma de *Homo sapiens* (humano), hg18/ build 36.1, versión del genoma divulgada en marzo de 2006.

El término "asignado inequívocamente" o "ubicado de forma única en una región definida de un genoma humano normal" en el contexto de esta invención se refiere a la determinación de que la lectura de una secuencia obtenida bien a partir del ADN acelular en circulación o de un tumor, o bien del ADN de las células mononucleares de sangre periférica, puede ser asignada a una región en particular del genoma de referencia. En la determinación de un biomarcador personalizado, se asigna una secuencia a una región en particular del genoma de referencia basándose en algoritmos de identidad bien conocidos, tales como el algoritmo BLAST, mediante el uso de unos parámetros de elevada rigurosidad, tales como $e < 0,0001$. Además, dicha secuencia no tiene una coincidencia adicional de igual ajuste en el genoma de referencia que se emplea. Por ejemplo, en las realizaciones relativas a pacientes humanos, el genoma de referencia es normalmente un genoma humano construido a partir de la base de datos del genoma humano.

El término "ADN acelular en circulación" según se usa en el presente documento significa moléculas libres de ADN de 25 nucleótidos o más que no están contenidas en ninguna célula intacta de la sangre humana, y que pueden ser obtenidas a partir del suero o del plasma humano.

El término "cebador" se refiere a un oligonucleótido que actúa como punto de inicio de la síntesis de ADN en unas condiciones en las que se induce la síntesis de un producto de extensión del complementario de una hebra de ácido nucleico, es decir, en presencia de cuatro trifosfatos de nucleósido diferentes y de un agente de polimerización (es decir, una polimerasa de ADN o una transcriptasa inversa) en un tampón apropiado a una temperatura adecuada. Un cebador es preferentemente un oligodesoxirribonucleótido monocatenario. El cebador incluye una "región de hibridación" exacta o sustancialmente complementaria de la secuencia objetivo, preferentemente de entre aproximadamente 15 y aproximadamente 35 nucleótidos de longitud. Un cebador oligonucleotídico puede consistir totalmente en la región de hibridación o puede contener características adicionales que permiten la detección, la inmovilización o la manipulación del producto amplificado, pero que no alteran la capacidad del cebador para que sirva como reactivo de inicio de la síntesis de ADN. Por ejemplo, puede incluirse una cola de una secuencia de ácidos nucleicos en el extremo 5' del cebador que hibride con un oligonucleótido de captura.

El término "sonda" se refiere a un oligonucleótido que hibrida selectivamente con un ácido nucleico objetivo en las condiciones adecuadas. Una sonda para la detección de las secuencias de biomarcadores descritas en el presente documento puede tener cualquier longitud, por ejemplo, entre 15-500 pb de longitud. Normalmente, en los ensayos basados en sondas, se prefieren las sondas de hibridación que tienen menos de 50 pb.

El término "secuencia objetivo" o "región objetivo" se refiere a una región de un ácido nucleico que va a ser analizada y que comprende la secuencia de interés.

Según se usa en el presente documento, los términos "ácido nucleico", "polinucleótido" y "oligonucleótido" se refieren a cebadores, a sondas y a fragmentos oligoméricos. Los términos no están limitados por la longitud y son genéricos para los polímeros lineales de polidesoxirribonucleótidos (que contienen 2-desoxi-D-ribosa), de polirribonucleótidos (que contienen D-ribosa) y para cualquier otro N-glicósido de una base de purina o de pirimidina, o de bases de purina o de pirimidina modificadas. Estos términos incluyen ADN bi- y monocatenario, así como ARN bi- y monocatenario. Los oligonucleótidos para su uso en la invención pueden usarse como cebadores y/o como sondas.

Un ácido nucleico, un polinucleótido o un oligonucleótido puede comprender enlaces o enlaces modificados fosfodiéster incluyendo, pero no se limitan a, fosfotriéster, fosforamido, siloxano, carbonato, carboximetil éster, acetamido, carbamato, tioéter, fosforamido con puente, metilén fosfonato con puente, fosforotioato, metilfosfonato, fosforoditioato, fosforotioato con puente o enlaces de sulfona, y combinaciones de dichos enlaces.

Un ácido nucleico, un polinucleótido o un oligonucleótido puede comprender las cinco bases naturales biológicas (adenina, guanina, timina, citosina y uracilo) y/u otras bases distintas a las cinco bases naturales biológicas. Estas bases pueden servir para diversos fines, por ejemplo, para estabilizar o desestabilizar la hibridación; para promover o inhibir la degradación de la sonda; o como puntos de fijación de fracciones detectables o de fracciones de inactivación. Por ejemplo, un polinucleótido de la invención puede contener una o más fracciones de bases modificadas, no convencionales o derivatizadas, incluyendo, pero no se limitan a, N6-metil-adenina, N6-terc-butil-bencil-adenina, imidazol, imidazoles sustituidos, 5-fluorouracilo, 5 bromouracilo, 5-clorouracilo, 5-yodouracilo, hipoxantina, xantina, 4-acetilcitosina, 5 (carboxihidroximetil)uracilo, 5 carboximetilaminometil-2-tiouridina, 5 carboximetilaminometiluracilo, dihidrouracilo, beta-D-galactosilqueosina, inosina, N6 isopenteniladenina, 1-metilguanina, 1-metilinosina, 2,2-dimetilguanina, 2-metiladenina, 2-metilguanina, 3-metilcitosina, 5-metilcitosina, N6-metiladenina, 7-metilguanina, 5-metilaminometiluracilo, 5-metoxiaminometil-2-tiouracilo, beta-D manosilqueosina, 5'-metoxycarboximetiluracilo, 5-metoxiuracilo, 2-metiltio-N6-isopenteniladenina, ácido uracil-5-oxiacético (v), wibutosina, pseudouracilo, queosina, 2 tiocitosina, 5-metil-2-tiouracilo, 2-tiouracilo, 4-tiouracilo, 5-metiluracilo, metil éster del ácido uracil-5-oxiacético, 3-(3-amino-3-N-2-carboxipropil) uracilo, (acp3)w, 2,6-diaminopurina, y 5-propinil pirimidina. Otros ejemplos de fracciones de bases modificadas, no convencionales o derivatizadas pueden encontrarse en las Patentes de EE.UU. nº 6.001.611; 5.955.589; 5.844.106; 5.789.562; 5.750.343; 5.728.525; y 5.679.785. Adicionalmente, un ácido nucleico, un polinucleótido o un oligonucleótido puede comprender una o más fracciones de azúcar modificadas que incluyen, pero no se limitan a, arabinosa, 2-fluoroarabinosa, xilulosa y una hexosa.

Las "secuencias repetitivas o elementos repetitivos" se refieren a elementos de ADN muy repetidos presentes en un genoma. Estas secuencias habitualmente se categorizan en familias de secuencias y se clasifican ampliamente como ADN repetitivo disperso (véase, por ejemplo, Jelinak y Schmid, Ann. Rev. Biochem. 51: 831-844, 1982; Hardman, Biochem J. 234: 1-11, 1986; y Vogt, Hum. Genet. 84: 301-306, 1990) o ADN repetido en tándem. Algunos elementos repetitivos incluyen ADN satélite, minisatélite y microsatélite. En los seres humanos, el ADN repetitivo disperso incluye secuencias *Alu*, elementos nucleares dispersos cortos (SINE) y elementos nucleares dispersos largos (LINE) y retrovirus endógenos (ERV). La categorización de los elementos repetitivos y de las familias de elementos repetitivos y sus secuencias consenso de referencia está definida en bases de datos públicas (por ejemplo, la rebase (versión 18.10) - Genetic Information Research Institute (Jurka et al., Cytogenet Genome Res 2005; 110: 462-7)).

El término "secuencia Alu" o "secuencia Alu repetida" se refiere a una secuencia que está categorizada como miembro de una familia de elementos dispersos cortos conservados de ADN genómico. Existen aproximadamente 1.000.000 de copias de ADN con repeticiones de Alu en el genoma humano (véase, por ejemplo, Rowald y Herrera (2000) Genetics 108: 57-72). Una Alu atípica es un dímero, formado por dos elementos de secuencias similares (los brazos izquierdo y derecho) que están separados por un conector corto rico en A. La mayoría de las secuencias Alu tiene una larga cola de poli-A de aproximadamente 20-100 bases. Las secuencias Alu se nombran según el sitio de la enzima de restricción Alu en la secuencia Alu consenso. Las propiedades y el origen molecular de las secuencias Alu repetitivas se describen, por ejemplo, en Mighell et al., FEBS Lett. 417: 1-5, 1997 y en Batzer et al., J Mol Evol. 42: 3-6, 1996. Como es conocido en la materia, puede haber ligeras variaciones en las secuencias Alu entre diferentes clones celulares representativos de un tipo celular en particular, o entre diferentes organismos representativos de la misma especie. Por lo tanto, según se usa en el presente documento, las "secuencias Alu consenso" se refieren a secuencias derivadas del alineamiento de múltiples secuencias de diferentes secuencias Alu obtenidas a partir de la misma especie o de la misma célula, en las que las secuencias Alu consenso comprenden la secuencia de nucleótidos en común o más común entre los clones representados.

En los métodos de las presentes invenciones, un "cebador Alu" se refiere a un oligonucleótido que hibrida con una secuencia Alu o una secuencia Alu consenso y que, junto con otro cebador, puede amplificar secuencias que contienen Alu en reacciones de amplificación.

Los elementos SVA (Sina, VNTR, Alu) son elementos móviles repetitivos del genoma humano. Un SVA típico está flanqueado por dos duplicaciones del sitio objetivo y comprende un número variable de repeticiones hexaméricas, una región con homología de la secuencia Alu antisentido, una región variable de repeticiones en tándem de SINE-R y una señal de poli-A (véase, por ejemplo, Ostertag, et al., Am J Hum Genet 73 (6): 1444-1451, 2003)

Un sitio de cebado "universal" es un sitio con el que hibridará un cebador universal. En general, "universal" se refiere al uso de un único cebador o de un conjunto de cebadores para una pluralidad de reacciones de amplificación. Por ejemplo, al usar un cebador universal para la secuenciación del ADN acelular en circulación, el ADN es fragmentado y adaptado para que contenga una secuencia de cebado universal para permitir la secuenciación y/o la amplificación del ADN en circulación mediante el uso de un único conjunto de.

Según se usa en el presente documento, las formas singulares "un", "uno/a" y "el/la" incluyen los referentes plurales salvo que el contexto lo indique claramente de otro modo. Por lo tanto, por ejemplo, la referencia a "una molécula" incluye una pluralidad de dichas moléculas, y similares.

5 Introducción

La invención se basa, al menos en parte, en el descubrimiento de que las regiones cromosómicas que contienen secuencias repetitivas pueden ser identificadas en el ADN acelular en circulación en pacientes que tienen un tumor. Dichas secuencias proporcionan unos biomarcadores personalizados que pueden usarse para la monitorización de la respuesta de un paciente con cáncer a una terapia.

Consecuentemente, en un aspecto, la invención proporciona un método para el análisis de ácidos nucleicos acelulares en una muestra (de sangre, de suero o de plasma) de un paciente con cáncer que comprende la determinación de los biomarcadores personalizados para el paciente. En un aspecto adicional, la invención proporciona métodos de uso de dichos biomarcadores personalizados.

DetECCIÓN DE LOS BIOMARCADORES PERSONALIZADOS

Preparación de la colección

En la realización típica se prepara una colección de fragmentos de ADN obtenidos a partir de las muestras de ácidos nucleicos de las diversas fuentes del paciente. Estas muestras normalmente incluyen una muestra del tumor, una muestra de tejido normal, por ejemplo, de células mononucleares de sangre periférica; y muestras de ácidos nucleicos acelulares obtenidas antes de tratamiento con un régimen terapéutico y después del tratamiento con el régimen terapéutico. Para las muestras de ácidos nucleicos acelulares, las muestras se obtienen a partir de la sangre, por ejemplo, de suero o de plasma. Los ácidos nucleicos pueden aislarse a partir del suero o del plasma mediante el uso de técnicas bien conocidas, véanse, por ejemplo, las secciones de ejemplos. En el contexto de la actual invención, las secuencias de ácidos nucleicos que se analizan son secuencias de ADN.

Las colecciones se preparan a partir del ADN aislado a partir de las muestras mediante el uso de técnicas conocidas. En el caso del ADN aislado a partir de células, el ADN puede ser fragmentado para asegurar un tamaño medio adecuado para el análisis subsiguiente (por ejemplo, aproximadamente 200 pb). Normalmente, los fragmentos de ADN son modificados para que contengan secuencias adaptadoras con sitios de unión del cebador (universal) para permitir la amplificación de la colección. Los adaptadores y los códigos de barras (véase a continuación) para su uso en las técnicas de propagación de colecciones de biología molecular son bien conocidos por los expertos en la materia. Un cebador universal se une a dicha secuencia adaptadora, de forma que la amplificación universal de las secuencias objetivo ligadas al adaptador.

Los cebadores universales y los sitios de unión del cebador son bien conocidos en la materia. Estos se usan a menudo en la construcción de colecciones para producir fragmentos de ADN con extremos conocidos que pueden usarse para un procesamiento adicional, por ejemplo, una secuenciación. Los cebadores y los adaptadores universales están disponibles en el mercado. Algunos ejemplos son los empleados en la secuenciación masiva paralela mediante el uso de las plataformas de secuenciación 454/Roche (por ejemplo, Titanium), Lifetechnologies (SOLiD o Personal Genome Machine) o Illumina (HiSeq).

Identificación del ADN que comprende los elementos repetitivos presentes en las muestras de un paciente

Los fragmentos de ADN que comprenden los elementos repetitivos pueden ser identificados mediante el uso de diversas técnicas bien conocidas por los expertos en la materia. Por ejemplo, puede realizarse una PCR de las colecciones para identificar los fragmentos presentes en la colección que comprenden secuencias repetitivas. Por ejemplo, la amplificación puede llevarse a cabo mediante el uso de un cebador que hibride con el elemento repetitivo (por ejemplo, una secuencia Alu) como uno de los cebadores de amplificación, y un cebador universal como el otro cebador. Alternativamente, los fragmentos de ADN deseados pueden ser identificados mediante el uso de sondas oligonucleotídicas específicas para la secuencia repetitiva objetivo. En algunas realizaciones, las sondas están marcadas (por ejemplo, con biotina) para permitir el subsiguiente enriquecimiento en los fragmentos deseados mediante el uso de un soporte sólido (por ejemplo, microesferas magnéticas). En estas realizaciones, los fragmentos enriquecidos de ADN pueden ser identificados mediante el uso de cebadores universales.

La preparación de cebadores y sondas específicas para cualquier elemento repetitivo en particular está ampliamente en la pericia de los expertos habituales en la materia. Las secuencias de numerosos elementos repetitivos son bien conocidas en la materia (véase, por ejemplo, la rebase (versión 18.10) - Genetic Information Research Institute (Jurka et al., Cytogenet Genome Res 2005; 110: 462-7)). Por ejemplo, las secuencias Alu son bien conocidas en la materia. Las secuencias Alu individuales del genoma humano son divergentes entre sí únicamente una media de entre el 15 y el 20 %. Los cebadores de la PCR basados en secuencias Alu, por ejemplo, una secuencia Alu consenso, pueden seleccionarse basándose en las secuencias Alu conocidas. Algunos ejemplos de secuencias Alu consenso se proporcionan en Batzer *et al*, 1996, *supra*; y en Bennet et al., Genome Res. 18: 1875-1883, 2008. En

las realizaciones típicas, puede ser beneficiosa la realización de dos reacciones de amplificación para cada colección para identificar las regiones de inserción enriquecidas en Alu. Una reacción emplea un cebador Alu inverso y un apropiado cebador universal directo, y la segunda reacción emplea un cebador Alu directo y un apropiado cebador universal inverso. Por ejemplo, los cebadores usados para el enriquecimiento en Alu pueden ser diseñados sobre la base de un alineamiento multiseuencia mostrado en Batzer. Este alineamiento multiseuencia es un alineamiento de la subfamilia de secuencias consenso AluJ, AluSx, AluSq, AluSp, AluSc, AluY, AluYa5, AluYa8 y AluYb8. Un cebador inverso puede ser seleccionado sobre la base de los pares de bases conservados en la región 5' de las secuencias alineadas, por ejemplo, los residuos 22-43 del alineamiento multiseuencia. Un cebador directo puede ser seleccionado sobre la base de los pares de bases conservados en la región 3' de las secuencias alineadas, por ejemplo, la secuencia entre los residuos 259 y 283 del mismo alineamiento multiseuencia. Ambos cebadores están diseñados para amplificar todas las subfamilias indicadas. En algunas realizaciones, los cebadores de la PCR están basados en subfamilias de Alu que son activas, es decir, capaces de una transposición. Por lo tanto, por ejemplo, puede usarse un alineamiento de la subfamilia de secuencias consenso AluS o AluY para identificar las regiones conservadas en las subfamilias deseadas. Como se entiende en la materia, no es necesario que la secuencia del cebador concuerde exactamente con la secuencia Alu consenso, siempre que se puedan amplificar las subfamilias Alu de interés. Las secuencias típicas del cebador tienen al menos 20 nucleótidos, preferentemente 25, 30, 35, 40, 45 o 50, o más nucleótidos, de longitud. En algunas realizaciones, un cebador de Alu para su uso en la invención comprende una secuencia de nucleótidos de la SEQ ID NO: 5 o 6, o tiene una identidad de al menos el 75 %, normalmente una identidad de al menos el 80 %, el 85 %, el 90 %, el 95 % o más, con la SEQ ID NO: 5 o la SEQ ID NO: 6.

De forma análoga, la información de la secuencia de los SINE, los LINE y los ERV está fácilmente disponible. Como se ha mencionado anteriormente, la categorización de los elementos repetitivos y de las familias de elementos repetitivos y de sus secuencias consenso de referencia está definida en las bases de datos públicas (por ejemplo, rebase (version 18.10) - Genetic Information Research Institute (Jurka et al., Cytogenet Genome Res 2005; 110:462-7)).

Después, los productos amplificados son secuenciados, y las secuencias evaluadas para identificar las regiones que tienen una inserción de un elemento repetitivo. Como se entiende en la materia, los productos amplificados pueden ser sometidos a manipulaciones, por ejemplo, la adición de códigos de barras moleculares, para facilitar los procedimientos de secuenciación.

Secuenciación del ADN

La detección de los biomarcadores individuales según la invención comprende la secuenciación del ADN aislado a partir de diversas muestras del paciente, por ejemplo, los ácidos nucleicos acelulares presentes en la sangre, del ADN aislado a partir del tumor de un paciente, etc. Dicha secuenciación puede llevarse a cabo mediante el uso de las metodologías de secuenciación conocidas, por ejemplo, los sistemas de secuenciación Illumina, Lifetechnologies, y Roche 454. En las realizaciones típicas, se secuencia el ADN acelular, o el ADN aislado a partir del tumor de un paciente, o de tejidos no tumorales, por ejemplo, de células mononucleares de sangre periférica, a partir de un paciente mediante el uso de un método de secuenciación a gran escala que proporciona la capacidad para obtener información de la secuencia a partir de muchas lecturas. Dichas plataformas de secuenciación incluyen las comercializadas por Roche 454 Life Sciences (GS systems), Illumina (por ejemplo, HiSeq, MiSeq) y Lifetechnologies (por ejemplo, SOLiD systems).

La plataforma de secuenciación Roche 454 Life Sciences implica el uso de una PCR en emulsión y la inmovilización de los fragmentos de ADN en microesferas. La incorporación de nucleótidos durante la síntesis es detectada mediante la medición de la luz que se genera cuando se incorpora un nucleótido.

La tecnología Illumina implica la unión de un ADN genómico fragmentado aleatoriamente a una superficie plana ópticamente transparente. Los fragmentos de ADN unidos son extendidos y amplificados en puente para crear una celda de flujo de secuenciación de densidad ultra alta con agregados que contienen copias del mismo molde. Estos moldes son secuenciados mediante el uso de una tecnología de secuenciación mediante una síntesis que emplea terminadores reversibles con colorantes fluorescentes eliminables.

También pueden usarse métodos que emplean una secuenciación mediante hibridación. Dichos métodos usados, por ejemplo, en la tecnología Lifetechnologies SOUD4+, usan un conjunto de todos los posibles oligonucleótidos de una longitud fija, marcados según la secuencia. Los oligonucleótidos son apareados y ligados; la ligación preferente mediante una ligasa de ADN de las secuencias coincidentes da como resultado una señal informativa del nucleótido en esa posición.

La secuencia puede ser determinada mediante el uso de cualquier otro método de secuenciación del ADN incluyendo, por ejemplo, los métodos que usan una tecnología de semiconductor para la detección de los nucleótidos que son incorporados en un cebador extendido mediante la medición de los cambios que se producen en la corriente cuando se incorpora un nucleótido (véanse, por ejemplo, las Publicaciones de Solicitud de Patente de EE.UU. nº 20090127589 y 20100035252). Otras técnicas incluyen una secuenciación directa con exonucleasa

exenta de marcadores, en la que los nucleótidos escindidos del ácido nucleico se detectan haciéndolos pasar a través de un nanoporo (Oxford Nanopore) (Clark et al., Nature Nanotechnology 4: 265 - 270, 2009); y la tecnología de secuenciación del ADN Single Molecule Real Time (SMRT™) (Pacific Biosciences), que es una técnica de secuenciación mediante síntesis.

5

Detección de las regiones de inserción del biomarcador personal

Para cada muestra de ADN, el ADN es secuenciado y evaluado para analizar la presencia de secuencias de inserción de elementos repetitivos. En la identificación de las regiones de inserción de un elemento repetitivo que pueden usarse como biomarcadores personalizados, todas las secuencias que se cartografían en una ventana de aproximadamente 250 pares de bases en las proximidades del elemento repetitivo (por ejemplo, elementos Alu, SVA o LINE) anotadas en el genoma de referencia normalmente se extraen a partir de la consideración como una secuencia de inserción candidata mediante el uso de aplicaciones tales como RepeatMasker. Para el resto de las secuencias, se identifican las regiones de inserción que comprenden secuencias repetitivas. El experto comprende que las regiones de inserciones pueden ser definidas mediante el uso de varios criterios. En una realización, las regiones de inserción se definen como aquellos fragmentos que tienen unas secuencias procedentes de la reacción de amplificación empleando el cebador inverso Alu que está junto con una ventana de 200 pares de bases de una secuencia procedente del cebador Alu directo de la PCR.

10

15

20

25

Después se emplean varios procedimientos de filtrado para determinar aquellas regiones de inserción que son biomarcadores tumorales para el paciente en particular. En una realización, unas regiones de inserción Alu es un biomarcador individualizado cuando tiene las siguientes propiedades: está presente en una muestra del tumor de un paciente, pero no en las muestras normales, y está presente en una muestra de ácidos nucleicos acelulares del paciente previa a la terapia, pero no en una muestra de ácidos nucleicos acelulares del paciente posterior a la terapia.

30

35

40

En algunas realizaciones se lleva a cabo una comparación directa de varias muestras del paciente. Como ejemplos, pueden llevarse a cabo las siguientes comparaciones para identificar los biomarcadores. Las regiones de inserción de un elemento repetitivo identificadas en el ADN de una muestra del tumor del paciente se comparan con las regiones de inserción identificadas en el ADN de los controles normales. La muestra normal puede ser una muestra no tumoral del tumor del paciente y/o una o más muestras de individuos normales (es decir, individuos a los que no se les ha diagnosticado un cáncer). En algunas realizaciones, la muestra normal es de tejido no tumoral del paciente, por ejemplo, de células mononucleares de sangre periférica. Aquellas regiones de inserción de un elemento repetitivo que están presentes en una muestra del tumor, pero no en el ADN normal, también se comparan con aquellas presentes en las muestras de ADN acelular previas a la terapia obtenidas de los pacientes. Las regiones de inserción de un elemento repetitivo que están presentes tanto en la muestra del tumor como en las muestras previas a la terapia se comparan con las regiones de inserción contenidas en una muestra de ADN acelular posterior a la terapia obtenida del paciente. Aquellas regiones de inserción que están presentes tanto en la muestra del tumor como en la muestra de ADN acelular previa a la terapia se comparan después con las regiones de inserción que son detectadas en una muestra de ADN acelular posterior a la terapia. Aquellas regiones de inserción en están presentes en la muestra posterior a la terapia son biomarcadores para ese paciente individual. El experto en la materia comprende que no es necesario que estas comparaciones se realicen en ningún orden en particular.

45

50

55

Las comparaciones pueden llevarse a cabo de una forma semicuantitativa o cuantitativa. Por lo tanto, por ejemplo, en una metodología semicuantitativa, la presencia o la ausencia de las regiones de inserción en una muestra se identifica simplemente como positiva o negativa cuando se determina un biomarcador. Alternativamente, pueden emplearse procedimientos más cuantitativos para identificar los biomarcadores. En algunas realizaciones, la comparación entre las regiones de inserción detectadas en la muestra del tumor y las regiones de inserción detectadas en las muestras normales, de ese paciente y de controles normales independientes, puede ser cuantitativa. Por ejemplo, los métodos de identificación de un biomarcador pueden comprender una etapa de comparación en la que únicamente se consideran aquellas regiones como potenciales biomarcadores cuando el recuento normalizado de las PBMC difiere en al menos 5 veces en comparación con el ADN tumoral. El experto entiende que pueden usarse factores distintos a "5 veces", por ejemplo, entre 2 y 4 veces, o 6 veces o mayor, en dicho análisis cuantitativo. También puede aplicarse el mismo tipo de comparación cuantitativa para evaluar las regiones de inserción posteriores a la terapia.

60

Otro aspecto de una metodología cuantitativa es el establecimiento de límites para las comparaciones entre el tumor y las regiones de inserción normales de un elemento repetitivo y las regiones de inserción del tumor con las regiones de inserción previas a la terapia, en la que una cobertura analítica mínima por región encontrada puede considerarse un criterio. Por ejemplo, se consideran las inserciones en una posición genómica definida que solo se observan en una muestra del tumor, pero no en una normal, en controles no cancerosos, si dicha inserción se encuentra en la secuenciación en al menos tres casos. Este número puede ser establecido en cualquier número > 0. Puede emplearse cualquier combinación de medio de filtrado cuantitativo y cualitativo.

65

En la evaluación de las muestras de ADN acelular, el experto comprende que las secuencias de las muestras previas a la terapia y posteriores a la terapia normalmente se comparan con las de un conjunto de controles

normales. Para cualquier comparación, los resultados obtenidos mediante un método definido pueden compararse con los resultados obtenidos mediante la misma metodología metodológica, pero mediante el uso de una muestra o de conjuntos de muestras diferentes. Dichas muestras para la comparación pueden proceder de individuos de control, por ejemplo, una muestra de plasma o de suero de individuos de control normales. Dicho grupo de individuos sin cáncer normalmente se considera un "grupo de referencia". Las muestras obtenidas a partir de dicho grupo de referencia pueden usarse como las muestras de referencia definidas para las comparaciones transversales. En los casos en los que haya disponibles células somáticas normales de un individuo, dichas células pueden definirse como una referencia para una posterior comparación intra-individual. El uso de una muestra "temporizada" (por ejemplo, previa a la terapia que puede compararse con una muestra posterior a la terapia como referencia) permite una comparación intra-individual (longitudinal).

Una vez identificada una región de inserción que puede servir como biomarcador personalizado, la región es secuenciada para facilitar el uso del marcador como se describe a continuación.

15 **Uso de los biomarcadores personalizados**

Un biomarcador personalizado identificado según la invención puede usarse para evaluar la respuesta de ese paciente en particular a una terapia contra el cáncer. Dicha evaluación puede llevarse a cabo, por ejemplo, mediante el uso de una reacción de amplificación y/o de hibridación de un ácido nucleico para detectar los niveles del biomarcador personalizado en los ácidos nucleicos acelulares presentes en una muestra de sangre del paciente.

Con objeto de detectar la presencia del biomarcador personalizado, se obtiene una muestra de sangre del paciente antes y/o después del tratamiento con una terapia contra el cáncer. Después se analiza el suero o el plasma de la muestra de sangre para evaluar la presencia del biomarcador personalizado.

Las técnicas de detección para la evaluación de la presencia de un biomarcador personalizado en los ácidos nucleicos implican procedimientos bien conocidos en el campo de la genética molecular. En algunas realizaciones típicas, la detección de un biomarcador personalizado implica la amplificación de los ácidos nucleicos. Una amplia guía para la realización de dichas técnicas se proporciona en la materia. Algunas referencias ejemplares incluyen manuales tales como PCR Technology: Principles and Applications for DNA Amplification (ed. H. A. Erlich, Freeman Press, NY, N. Y., 1992); PCR Protocols: A Guide to Methods and Applications (eds. Innis, et al., Academic Press, San Diego, Calif., 1990); Current Protocols in Molecular Biology, Ausubel, 1994-1999, incluyendo las actualizaciones complementarias a lo largo de abril de 2004; Sambrook & Russell, Molecular Cloning, A Laboratory Manual (3ª Ed, 2001).

Aunque los métodos pueden emplear etapas de una PCR, también pueden usarse otros protocolos de amplificación. Algunos métodos de amplificación incluyen una reacción en cadena de la ligasa (véase, por ejemplo, Wu & Wallace, Genomics 4: 560-569,1988); un ensayo de desplazamiento de hebra (véase, por ejemplo, Walker et al., Proc. Natl. Acad. Sci. EE.UU. 89: 392-396,1992; la Patente de EE.UU. nº 5.455.166); y diversos sistemas de amplificación basados en la transcripción, incluyendo los métodos descritos en las Patentes de EE.UU. nº 5.437.990; 5.409.818; y 5.399.491; el sistema de amplificación de la transcripción (TAS) (Kwoh et al., Proc. Natl. Acad. Sci. EE.UU. 86: 1173-1177, 1989); y la replicación de secuenciada autosostenida (3SR) (Guatelli et al., Proc. Natl. Acad. Sci. EE.UU. 87: 1874-1878, 1990; el documento WO 92/08800). Alternativamente, pueden usarse métodos para la amplificación de la sonda hasta unos niveles detectables, tales como la amplificación de la Q β -replicasa (Kramer & Lizardi, Nature 339: 401-402,1989; Lomeli et al., Clin. Chem. 35: 1826-1831,1989). Una revisión de los métodos de amplificación conocidos la proporcionan, por ejemplo, Abramson y Myers en Current Opinion in Biotechnology 4: 41-47, 1993.

En algunas realizaciones se emplean cebadores oligonucleotídicos que amplifican el biomarcador de interés. La presencia del biomarcador puede ser determinada, por ejemplo, mediante la detección de la longitud del biomarcador en el ácido nucleico acelular en comparación con un control. En otras realizaciones, también puede usarse una sonda para detectar la presencia del biomarcador.

Los oligonucleótidos que se emplean como cebadores y/o sondas para la detección de los biomarcadores pueden seleccionarse mediante el uso de métodos bien conocidos en la materia. Por ejemplo, pueden diseñarse cebadores de una PCR mediante el uso de las técnicas informáticas de diseño de cebadores convencionales conocidas por los individuos expertos en la materia. Las variables consideradas durante el diseño de los cebadores de una PCR pueden incluir la longitud del cebador, el contenido en pares de GC, la temperatura de fusión y el tamaño del ácido nucleico objetivo amplificado por el par de cebadores.

En una realización, el biomarcador es identificado mediante una hibridación en unas condiciones de hibridación específicas de la secuencia con una sonda que se dirige a la región del biomarcador (por ejemplo, se dirige a alguna porción asignada inequívocamente del biomarcador objetivo).

Los formatos de hibridación adecuados son bien conocidos en la materia, incluyendo, pero no se limitan a, ensayos en fase de solución, en fase sólida, en formatos de matriz de oligonucleótidos, en fase mixta o de hibridación *in situ*. En las hibridaciones en fase de solución (o líquida), tanto el ácido nucleico objetivo como la sonda o los cebadores

son libres de interactuar en la mezcla de reacción. También se han desarrollado técnicas tales como sistemas de PCR en tiempo real que permiten el análisis, por ejemplo, la cuantificación, de los productos amplificados durante una reacción de PCR. En este tipo de reacción, la hibridación con una sonda oligonucleotídica específica se produce durante el programa de amplificación para identificar la presencia de un ácido nucleico objetivo. La hibridación de sondas oligonucleotídicas asegura la mayor especificidad debido a la transición entre dos estados controlados termodinámicamente. Algunos ejemplos de estos formatos de ensayos son las sondas de hibridación de transferencia de energía por resonancia de fluorescencia, las balizas moleculares, los escorpiones moleculares y las sondas de hibridación de exonucleasa (por ejemplo, revisados en Bustin, J. Mol. Endocrin. 25: 169-93, 2000).

Puede inmovilizarse tanto la sonda como la secuencia objetivo. Por ejemplo, la sonda puede comprender un marcador (por ejemplo, biotina) que permite que la sonda y la secuencia objetivo sean capturadas sobre un soporte sólido (por ejemplo, una microesfera magnética).

En un formato en el que el objetivo está inmovilizado, el ADN objetivo amplificado es inmovilizado sobre un soporte sólido, y el complejo objetivo es incubado con la sonda en unas condiciones de hibridación adecuadas, la sonda no hibridada se retira mediante un lavado en unas condiciones adecuadamente rigurosas, y se monitoriza el soporte sólido para evaluar la presencia de la sonda unida. En los formatos en los que las sondas están inmovilizadas sobre un soporte sólido, el ADN objetivo normalmente se marca, habitualmente durante la amplificación. La sonda inmovilizada es incubada con el ADN objetivo amplificado en unas condiciones de hibridación adecuadas, el ADN objetivo no hibridado se extrae mediante un lavado en unas condiciones adecuadamente rigurosas, y el soporte sólido/sonda se monitoriza para evaluar la presencia del ADN objetivo unido.

En algunas realizaciones se inmovilizan múltiples sondas, por ejemplo, que se dirigen a diferentes biomarcadores objetivo para ese paciente, sobre un soporte sólido, y el ADN acelular de un paciente se analiza mediante el uso simultáneo de las múltiples sondas. Algunos ejemplos de dichas matrices de ácidos nucleicos se describen en el documento WO 95/11995.

En un método alternativo sin sondas, el ácido nucleico amplificado correspondiente a un ácido nucleico objetivo se lleva a cabo mediante el uso de cebadores de ácidos nucleicos de la región cromosómica, y se detecta mediante la monitorización del aumento en la cantidad total de ADN bicatenario en la mezcla de reacción, se describe, por ejemplo, en la Patente de EE.UU. nº 5.994.056; y en las Publicaciones de Patente Europea nº 487.218 y 512.334. La detección del ADN bicatenario objetivo se basa en el aumento en la fluorescencia de varios colorantes de unión al ADN, por ejemplo, SYBR Green, que aparece cuando se unen al ADN bicatenario.

Los oligonucleótidos pueden ser preparados mediante cualquier método adecuado, habitualmente mediante una síntesis química, y también pueden ser adquiridos a través de fuentes comerciales. Los oligonucleótidos pueden incluir enlaces fosfodiéster modificados (por ejemplo, fosforotioato, metilfosfonatos, fosfoamidato o boranofosfato) o pueden usarse enlaces distintos a un derivado del ácido fosforoso en un oligonucleótido para prevenir la escisión en un sitio seleccionado. Además, el uso de azúcares modificados en 2'-amino tiende a favorecer el desplazamiento sobre la digestión del oligonucleótido cuando hibrida con un ácido nucleico que también es el molde para la síntesis de una nueva hebra de ácido nucleico.

En algunas realizaciones, puede determinarse el nivel del biomarcador en circulación en el paciente con respecto a un valor índice, tal como la cantidad de biomarcador que es identificada en los controles normales.

Dispositivos y kits

En un aspecto adicional, la divulgación proporciona kits útiles para la identificación de uno o más biomarcadores individualizados en el ácido nucleico acelular en circulación de un paciente. En un aspecto de la divulgación, el kit proporciona al menos un cebador Alu y un cebador universal y/o anotadores para ligar un sitio de unión del cebador universal al ADN acelular fragmentado en circulación del paciente. Dichos reactivos pueden usarse para la identificación de biomarcadores individuales.

Determinación de un biomarcador personalizado del cáncer

En el contexto de la invención, la "detección" o la "identificación" o "identificar la presencia" o "detectar la presencia" de un personalizado en una muestra de ácidos nucleicos acelulares en circulación de un paciente se refiere a la determinación de cualquier nivel del biomarcador en la muestra de ácidos nucleicos en circulación del paciente. Una vez que se ha detectado el biomarcador, puede usarse la información para ayudar en la evaluación de la respuesta de un paciente a una terapia. Por ejemplo, la presencia de un biomarcador individualizado en el ADN acelular en circulación en el paciente después del tratamiento con un protocolo terapéutico en particular puede indicar que el paciente no ha respondido completamente al protocolo. Consecuentemente, la información puede usarse para ayudar en la determinación de los ajustes del protocolo terapéutico y/o para determinar si debería emplearse una terapia alternativa.

La información obtenida a partir del análisis del biomarcador puede almacenarse en una forma legible por

ordenador. Dicho sistema informático comprende normalmente unos subsistemas principales tales como un procesador, un sistema de memoria (normalmente RAM), un controlador de entrada/salida (I/O), un dispositivo externo tal como una pantalla a través de un adaptador de pantalla, puertos en serie, un teclado, una unidad de disco fija a través de una interfaz de almacenamiento y una unidad de disco floppy operativa para recibir un disco floppy, y un dispositivo de CD-ROM (o de DVD-ROM) operativo para recibir un CD-ROM. Pueden conectarse otros muchos dispositivos, tales como una interfaz de red conectada a través de un puerto en serie.

El sistema informático también puede estar conectado a una red, que comprende una pluralidad de dispositivos de computación conectados a través de una conexión de datos, tal como un cable Ethernet (coaxial o 10BaseT), una línea telefónica, una línea de ISDN, una red inalámbrica, de fibra óptica u otro medio de transmisión de señales adecuado, mediante lo cual al menos un dispositivo de red (por ejemplo, un ordenador, una matriz de disco, etc.) comprende un patrón de dominios magnéticos (por ejemplo, un disco magnético) y/o dominios de carga (por ejemplo, una matriz de celdas DRAM) que componen un patrón de bits que codifica los datos adquiridos a partir de un ensayo de la invención.

El sistema informático puede comprender un código para la interpretación de los resultados de un estudio para la determinación de los biomarcadores personalizados o para la evaluación de la presencia de uno o más de los biomarcadores personalizados identificados según la invención para ayudar en el pronóstico. Por lo tanto, en un ejemplo de realización, los resultados del análisis del biomarcador son proporcionados a un ordenador en el que un procesador central ejecuta un programa informático para la evaluación de uno o más de los biomarcadores.

La invención puede comprender el uso de un sistema informático, tal como el que se ha descrito anteriormente, que comprende: (1) un ordenador; (2) un patrón de bits almacenado que codifica los resultados de las pruebas del biomarcador obtenidos mediante los métodos de la invención, que pueden ser almacenados en el ordenador; (3) y, opcionalmente, (4) un programa para la evaluación de un biomarcador.

La invención puede comprender la generación de un informe basándose en la detección de uno o más biomarcadores personalizados para el paciente.

Por lo tanto, la presente invención puede comprender el uso de un sistema relacionado con los anteriores métodos de la invención. En una realización, la invención comprende el uso de un sistema para el análisis del ADN acelular en circulación, que comprende: (1) un analizador de muestras para la ejecución del método de análisis del ADN acelular en circulación en la sangre, el suero o el plasma de un paciente según se describe en las diversas realizaciones anteriores; (2) un sistema informático para recibir y analizar automáticamente los datos obtenidos en la etapa (1) para proporcionar un valor de la prueba que representa el estado (la presencia o la ausencia o la cantidad, es decir, la concentración o el número de copias) de un biomarcador personalizado para el paciente.

La función de análisis basados en ordenador puede ser implementada en cualquier lenguaje y/o navegadores adecuados. Por ejemplo, puede ser implementada en un lenguaje C y preferentemente mediante el uso de lenguajes de programación de alto nivel orientados al objeto, tales como Visual Basic, SmallTalk, C++, y similares. La aplicación puede escribirse para que se ajuste a entornos tales como el entorno Microsoft Windows™, incluyendo Windows™ 98, Windows™ 2000, Windows™ NT, y similares. Además, la aplicación también puede escribirse para el entorno Macintosh™, SUN™, UNIX o LINUX. Además, las etapas funcionales también pueden ser implementadas mediante el uso de un lenguaje de programación universal o independiente de plataforma. Algunos ejemplos de dichos lenguajes de programación multiplataforma incluyen, pero no se limitan a, el lenguaje de etiquetado de hipertexto (HTML), JAVA™, JavaScript™, el lenguaje de programación Flash, la interfaz de entrada común/lenguaje de consulta estructurada (CGI/SQL), el lenguaje práctico de extracción de informe (PERL), AppleScript™ y otros lenguajes *script* de sistema, el lenguaje de programación/lenguaje de consulta estructurada (PL/SQL), y similares. Pueden usarse navegadores con Java™- o JavaScript™- activado, tales como HotJava™, Microsoft™ Explorer™ o Netscape™. Cuando se usa el contenido activo de las páginas web, pueden incluir miniaplicaciones de Java™ controles de ActiveX™ u otras tecnologías de contenido activo.

La función de análisis también puede materializarse en productos de programas informáticos y usarse en los sistemas descritos anteriormente o en otros sistemas basados en un ordenador o en internet. Consecuentemente, otro aspecto de la presente divulgación se refiere a un producto de programa informático que comprende un medio usable por ordenador que tiene códigos de programa o instrucciones legibles por ordenador materializados en el mismo para permitir que un procesador lleve a cabo el análisis y la correlación de las funciones como se ha descrito anteriormente. Estas instrucciones del programa informático pueden ser cargadas en un ordenador o en otro aparato programable para producir una máquina, tal que las instrucciones que se ejecutan en el ordenador o en otro aparato programable crean un medio para implementar las funciones con las etapas descritas anteriormente. Estas instrucciones del programa informático también pueden ser almacenadas en una memoria o un medio legible por ordenador que puede dirigir un ordenador u otro aparato programable para que funcione de una forma en particular, de forma que las instrucciones almacenadas en la memoria o el medio legible por ordenador produzcan un artículo de elaboración que incluye un medio de instrucción que implementa el análisis. Las instrucciones del programa informático también pueden ser cargadas en un ordenador o en otro aparato programable para provocar una serie de etapas operativas que se van a llevar a cabo en el ordenador o en el otro aparato programable para producir un

proceso implementado por ordenador tal que las instrucciones que se ejecutan en el ordenador o en el otro aparato programable proporciona las etapas para la implementación de las funciones o etapas descritas anteriormente.

5 Los siguientes ejemplos se proporcionan únicamente como ilustración, y no como limitación. Los expertos en la materia reconocerán fácilmente una diversidad de parámetros no críticos que podrían cambiarse o modificarse para producir esencialmente unos resultados similares.

Ejemplos

10 Identificación de los biomarcadores individualizados para el cáncer

Ejemplo 1

15 Este ejemplo describe la identificación de biomarcadores Alu individualizados para pacientes con cáncer de mama.

Secuenciación

20 Se prepararon colecciones de secuencias a partir del ADN genómico nuclear de células mononucleares de sangre periférica y tumorales (PBMC). El ADN genómico extraído (100 ng) se sometió a un esfuerzo de cizalla por ultrasonidos hasta un tamaño de aproximadamente 200 pb. Se repararon los extremos de los fragmentos (50 ng) mediante el uso de 0,4 unidades/μl de un fragmento Klenow (USB, Affymetrics) y de 0,2 unidades/μl de cinasa (USB, Affymetrics) durante 30 min a 37 °C. La reacción se purificó mediante el uso de microesferas AMPure (Agencourt). Se fijaron salientes individuales de A a los fragmentos con los extremos reparados mediante una incubación con 0,4 unidades/μl de un fragmento Klenow exo-(NEB) durante 30 min a 37 °C. La reacción se purificó mediante el uso de microesferas AMPure (Agencourt). Se ligaron adaptadores a los fragmentos con las colas de A. Los adaptadores eran dos oligonucleótidos de ADN monocatenario:

P9 gatcggaagagcatcgcaaggcaagcaagctaaactgaaccgggttctcattagcg (SEQ ID NO: 1)

30 P5 aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatc*t (SEQ ID NO: 2)

El oligonucleótido P9 tiene un grupo fosfato en 5 prima. El oligo P5 tiene un enlace internucleótido de fosforotioato entre los dos nucleótidos en 3 prima, que está indicado con un asterisco. Ambos oligonucleótidos se aparearon para formar un adaptador en Y parcialmente bicatenario.

35 Los adaptadores en Y se ligaron al fragmento A con las colas de A mediante una incubación con 1,2 unidades/μl de ligasa (USB, Affymetrics) durante 30 min a la temperatura ambiente. Las concentraciones del adaptador en la reacción eran de 30x la cantidad molar de los extremos de los fragmentos. La reacción se purificó mediante el uso de la solución SureClean Plus (Biolina).

40 Los fragmentos ligados se amplificaron en 12 ciclos de PCR mediante el uso de los cebadores:

P9-PCR-Primer cgctaatgaggaacccgggttcagtttagcttgcttgccttgcgatgctcttccgatc (SEQ ID NO: 3)

45 P5-Universal aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatct (SEQ ID NO: 4)

50 Las condiciones de la reacción de PCR eran 200 μmol/l de cada dNTP, 0,02 unidades/μl de la polimerasa de ADN Phusion High-Fidelity, 0,5 μmol/l de cada cebador en 1x de tampón Phusion HF. Las condiciones de ciclado eran: a 98 °C durante 30 s de desnaturalización inicial, 12 ciclos a 98 °C durante 10 s, a 62 °C durante 30 s, a 70 °C durante 30 s y a 70 °C durante 5 min de elongación final. La reacción se purificó mediante el uso del kit GenElute PCR Clean-Up (Sigma-Aldrich).

La colección de ADN amplificado (10 ng) se sometió a una PCR de enriquecimiento en Alu. Los cebadores usados en esta PCR eran:

55 Alu40.REV gtgactggagttcagacgtgtgctcttccgatctcccaagtgtctgggattacagg (SEQ ID NO: 5)

Alu259.FOR gtgactggagttcagacgtgtgctcttccgatcggcracagagcgaracygtctca (SEQ ID NO: 6)

60 P5-Universal aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatct (SEQ ID NO: 4)

65 Para cada colección se llevaron a cabo dos PCR mediante el uso de cualquiera de los cebadores Alu40.REV y P5-Universal o Alu259.FOR y P5-Universal. Las condiciones de la reacción de PCR eran 200 μmol/l de cada dNTP, 0,02 unidades/μl de la polimerasa de ADN Phusion High-Fidelity, 1 μmol/l de cada cebador en 1x de tampón Phusion HF. Las condiciones de ciclado eran: a 98 °C durante 30 s de desnaturalización inicial, 12 ciclos a 98 °C durante 10 s, a 68 °C durante 30 s, a 70 °C durante 30 s y a 70 °C durante 5 min de elongación final. Las reacciones se purificaron mediante el uso del kit innuPrep PCRpure (Analytik Jena).

Para cada muestra y cada PCR de enriquecimiento en Alu, se sometieron 200 ng de los productos de la PCR a 4 ciclos de PCR mediante el uso de unos cebadores de fusión que portan una secuencia de índice individual de 7 pb. Esta secuencia sirve como código de barras molecular para permitir la asignación de cada lectura de la respectiva muestra y PCR después de la secuenciación. La parte en 3' de los cebadores de índice se aparea con la secuencia subrayada del cebador Alu40.REV y Alu259.FOR. La parte en 5' de este cebador añade el adaptador de secuenciación full Illumina a los fragmentos. El siguientes un ejemplo de una secuencia de un cebador de índice: caagcagaagacggcatacgagat-7bpindex-gtgactggagttcagacgtgtgctcttccgatc. (SEQ ID NOs: 7 y 8)

La preparación de las colecciones para los fragmentos de ADN acelular extraídos a partir del suero de los pacientes se llevó a cabo según se describe para el ADN tumoral y de las PBMC, pero se omitió la etapa de fragmentación. La cantidad inicial del ADN acelular era de 10 ng. Los fragmentos de colecciones enriquecidos de los pacientes y de las diferentes muestras se agruparon y se secuenciaron con un sistema Illumina HiSeq o Illumina HiScan según los protocolos de los fabricantes. La secuenciación se llevó a cabo en forma de una secuenciación de lectura única durante 51 ciclos, seguidos de 8 ciclos de secuenciación de índice.

Análisis bioinformáticos

Las lecturas de secuencia se alinearon con el genoma humano (HG18, NCBI36, marzo de 2006) mediante el uso del programa informático CASAVA v1.8 (Illumina) con la parametrización por defecto. Mediante el uso de la versión de repeticiones enmascaradas del genoma humano (HG18, NCBI36, marzo de 2006), todas las lecturas de secuencia cartografiadas en una ventana de ± 250 pb alrededor de un elemento Alu o SVA anotadas en el genoma de referencia.

Del resto de las lecturas, se construyeron regiones de inserción para cada muestra y cada espécimen. Las regiones de inserción se definieron mediante los siguientes criterios: al menos una lectura procedente de la PCR del Alu40.REV y al menos al menos una lectura procedente de la PCR del Alu259.FOR están conjuntamente en una ventana de 200 pb. La región de "inicio" está definida por la posición inicial de la primera lectura. La región de "detención" está definida por la posición final de la última lectura. La posición de inicio, la posición de detención y el número de lecturas fueron registrados para cada región de inserción de la respectiva muestra y espécimen.

Detección de las supuestas regiones de inserción específicas del tumor

Para cada muestra se compararon las regiones de inserción detectadas en el tumor frente a todas las regiones de inserción detectadas en los genomas de las PBMC de todos los pacientes ($n = 17$) y de controles normales independientes ($n = 8$). Se extrajeron todas las regiones de inserción del tumor que se solapaban (± 200 pb) con las regiones detectadas en el ADN de las PBMC. Para cada paciente, se compararon el resto de las regiones de inserción del tumor frente a las regiones de inserción detectadas en el ADN sérico previo la cirugía del respectivo paciente. Se conservaron todas las regiones de inserción del tumor que se solapaban (± 200 pb) con una región de inserción detectada en el suero previo la cirugía. Se extrajeron todas las regiones del tumor no solapantes. Para cada paciente se compararon el resto de las regiones de inserción del tumor frente a las regiones de inserción detectadas en el ADN sérico posterior a la cirugía del respectivo paciente. Se extrajeron todas las regiones de inserción del tumor que se solapaban (± 200 pb) con una región de inserción detectada en el suero posterior a la cirugía. Se conservaron todas las regiones del tumor no solapantes.

En una metodología alternativa, en lugar de filtrar de la forma cualitativa mencionada anteriormente (si/no), se diseñó una metodología cuantitativa, por ejemplo, mediante el uso de un filtro para la comparación. Por ejemplo, únicamente se conservan aquellas regiones en las que el recuento normalizado de las PBMC difiere en al menos 5 veces en comparación con el ADN tumoral. Otros factores que pueden usarse son entre 2 y 4 veces, o 6 veces o mayor. Puede aplicarse la misma comparación cuantitativa al filtrado descrito en el filtrado de las regiones de inserción posteriores a la cirugía. Otro aspecto de una metodología cuantitativa es el establecimiento de límites para las comparaciones entre las regiones de inserción Alu tumorales con respecto a las normales, y las regiones de inserciones tumorales Alu con las regiones de inserciones Alu previas a la terapia, en las que un recuento de lectura mínimo o región encontrada puede ser considerado como criterio. Puede emplearse cualquier combinación de medio de filtrado cuantitativo y cualitativo.

Las lecturas de secuencia generadas a partir del ADN sérico previo a la cirugía y posterior a la cirugía, y de 24 individuos adicionales aparentemente sanos, se sometieron a un análisis de la profundidad de cobertura. Los datos se usaron bien como tales, o bien se omitieron las lecturas que fueron cartografiadas a una distancia no mayor de 200 pb contiguos a una ALU anotada.

Las lecturas de secuencia cartografiadas se contaron en unas ventanas de 250 kbp a lo largo del genoma. Los recuentos de la lectura se normalizaron a recuentos por 10 millones de lecturas.

Definición de las regiones con una desviación significativa con respecto a los controles normales

Usando los recuentos de la ventana de lectura de las 24 muestras de control sanas, se calcularon los límites de

confianza superior e inferior mediante el uso de una metodología de "dejando uno fuera". En resumen, se calculó el valor de Z para cada intervalo con el resto de los 23 normales como base. Después se calculó la desviación media y típica para cada intervalo mediante el uso de los 24 puntos de datos resultantes (valores de Z) para cada ventana.

5 Los recuentos de la lectura de ventana de los especímenes previos a la cirugía se seleccionaron como potenciales regiones diagnósticas de un desequilibrio en el número de copias si estaban fuera de los límites de confianza superior o inferior. La desviación de cada recuento de lectura de ventana para cada espécimen previo y posterior a la cirugía con respecto a la media de los recuentos de lectura de ventana de los controles sanos se expresaron como el valor de Z (base: 24 normales).

10

Comparación de los desequilibrios en el número de copias previo a la cirugía frente a los posteriores a la cirugía

Por paciente, se examinaron las regiones genómicas con un desequilibrio significativo en el número de copias en la muestra sérica previa la cirugía, para evaluar los desequilibrios significativos en la muestra sérica posterior a la cirugía. Para esto, se restó la diferencia entre el valor de Z de las ventanas de la muestra posterior la cirugía del valor de Z de las ventanas de la muestra previa la cirugía. Una diferencia en el valor de Z muy positiva o negativa es indicativa de una región desequilibrada que está relacionada con la presencia del tumor, debido a que indica una elevada desviación con respecto a los controles normales en el ADN sérico previo la cirugía, pero una baja desviación con respecto al control normal en el ADN sérico posterior a la cirugía.

20

Estimación de las inserciones en, o en las proximidades de, los elementos repetitivos (ER)

Debido a que una porción del ~ 20 % de las lecturas secuencia no pudo ser cartografiada de forma única con el genoma humano, se llevó a cabo una alineación de la base de datos de las repeticiones humanas conocidas (base de datos Genetic Information Research Institute (GIRI)). Se contaron las lecturas resultantes cartografiadas de forma única para cada entrada de la base de datos y se normalizaron mediante el uso del número total de lecturas cartografiadas. Después, los recuentos de lectura normalizados se filtraron para aquellos que eran diferentes en el tumor en comparación con el ADN genómico de las PBMC, y además, diferentes con respecto a una comparación del CNA sérico previo y posterior a la cirugía. Para la comparación de los valores del tumor, los valores son convertidos después en valores de Z mediante el uso de los valores medios de las PBMC y la desviación típica con respecto a los valores tumorales. Para los valores séricos, el respectivo grupo de muestra sirvió como base (por ejemplo, todos los datos previos a la cirugía para los resultados previos a la cirugía, todos los datos posteriores a la cirugía para los resultados posteriores a la cirugía).

25

30

Resultado de la profundidad del análisis de cobertura

La Figura 1 proporciona unas gráficas que muestran los resultados para cinco 5 pacientes. Las gráficas fueron generadas mediante a) con un filtrado Alu de (No Alus) y b) sin. Las cifras normalizadas de las lecturas de secuencia obtenidas a partir del ADN acelular en circulación que se alineaban con la base de datos genómica humana HG18 se sumaron en unos intervalos de ventana de 250 kbp. Los resultados se convirtieron en valores de Z según se describe. Para cada uno de dichos intervalos, los valores previos a la cirugía < 3 fueron censurados a cero. Los círculos representan el valor de Z previo a la cirugía, los rombos la diferencia entre el valor de Z previo a la cirugía y los valores de Z posteriores a la cirugía. Las barras muestran la suma de recuentos que se usó para calcular los valores de Z para la comparación.

45

Resultados de la detección de las supuestas regiones de inserción específicas del tumor

Las regiones de inserción que quedan después de los procedimientos de filtrado eran los supuestos sitios de inserción Alu específicos del tumor. Los resultados se proporcionan en la Tabla 1 y proceden del algoritmo de la Figura 2.

50

Las regiones de inserción detectadas en el filtrado cuantitativo se proporcionan en la Tabla 2. Estos son los resultados del filtrado cuantitativo de pacientes seleccionados, y se usan únicamente como un ejemplo. Esto no representa todas las regiones detectadas.

55

Resultados de la validación de la PCR

Se construyeron cebadores que abarcan algunos de los supuestos sitios de inserción Alu indicados en la Tabla 1 a partir del genoma de referencia genómico humano (HG 18, NCBI36, marzo de 2006). Los cebadores se probaron con ADN tumoral y con ADN de las PBMC del respectivo paciente. El molde de ADN usado para la PCR era, en algunos casos, el material que se sometió a una amplificación previa mediante el uso bien del kit Whole-Genome-Amplification 1 (WGA1, Sigma-Aldrich) o bien del kit Whole-Genome-Amplification 4 (WGA4, Sigma-Aldrich). Una lista de los cebadores y de la respectiva fuente del molde se proporciona en la Tabla 3. Las imágenes del gel de agarosa de los productos amplificados mediante la PCR se muestran en la Figura 3. La Figura 3 muestra los análisis del gel de cinco pacientes representativos, mostrándose los resultados del algoritmo de selección en la Figura 2. En el lado derecho de cada panel (A hasta E) se proporcionan los recuentos normalizados del sitio de inserción

60

65

específico, que está ubicado a ± 200 bp en la base de datos de la HG 18 (según se describe en [0082]). En el panel izquierdo se muestran las imágenes del gel de las PCR que fueron diseñadas para amplificar las inserciones en particular.

- 5 La Figura 4A-E muestra los valores de Z para la lectura del ADN en circulación y del ADN tumoral cartografiado con elementos repetitivos conocidos del genoma humano, separados para la PCR de la ALU40 y de la ALU 259.

La Figura 5 muestra los resultados de dos pacientes que mostraron una reducción diferencial en los biomarcadores del ADN acelular. Estos sitios de inserción Alu diferenciales fueron seleccionados a partir del procedimiento de selección cuantitativo ilustrado anteriormente. El Paciente 1 de la Figura 5A mostró la presencia de un sitio de polimorfismo de longitud se especifica tumoral localizado en una región única del cromosoma 18 en la muestra tumoral y en la previa a la cirugía. Estos fragmentos únicos no fueron detectados en la muestra de ADN acelular de las PBMC ni en la posterior a la cirugía. Se detectaron sitios de polimorfismos de longitud adicionales en regiones únicas de los cromosomas 1 y 5 y dos regiones únicas diferentes en el cromosoma 6 en la muestra de ADN acelular tumoral y en la previa a la cirugía, pero no en las PBMC. Sin embargo, estos 4 biomarcadores únicos todavía podían ser detectados en la muestra posterior a la cirugía. La reducción diferencial en los sitios de polimorfismos de longitud del ADN acelular también podía observarse en el Paciente 2 (Figura 5B). El Paciente 2 mostró una reducción completa en las muestras posteriores a la cirugía del ADN acelular en estos sitios únicos del cromosoma 1 y del cromosoma 13, pero únicamente una reducción parcial en las muestras posteriores a la cirugía del ADN acelular en otros nueve polimorfismos diferenciales únicos.

Ejemplo 2

Este ejemplo describe la identificación de biomarcadores individualizados L1-LINE para pacientes con cáncer colorrectal.

Preparación de la colección

Se extrajo el ADN nuclear de ($n = 12$) cánceres colorrectales y de los glóbulos blancos sanguíneos coincidentes mediante el uso del kit DNeasy Blood and Tissue (QIAGEN). El ADN extraído se sometió a un esfuerzo de cizalla por ultrasonidos hasta un tamaño medio de 200 pb mediante el uso de aparato de ultrasonidos focalizado Covaris S220 (Covaris) y microTubes AFA (Covaris). Se repararon los extremos del ADN fragmentado (100 ng), se añadieron colas de A y se ligaron a adaptadores de secuenciación Illumina mediante el uso del kit NEBNext Ultra Library Preparación (New England Biolabs) según las instrucciones del fabricante.

Enriquecimiento

El enriquecimiento en el fragmento L1 se llevó a cabo mediante el uso de oligonucleótidos biotinilados específicos para el L1 (Tabla 4) y 125 ng de ADN de la colección. Las mezclas de hibridación contenían 4 pmol de cada sonda L1, 5 μ g de ADN de esperma de salmón sometido a un esfuerzo de cizalla, 1.000 pmol de oligos de bloqueo contra los adaptadores específicos de la colección. Después de una desnaturalización inicial a 95° durante 10 min, se llevó a cabo la hibridación a 47 °C durante 72 h. La captura de las sondas biotiniladas junto con los fragmentos hibridados se llevó a cabo mediante el uso del kit TargetSeq Hybridization and Wash (Lifetechnologies) según las instrucciones del fabricante. Las secuencias capturadas se recuperaron a partir de las microesferas magnéticas mediante una desnacionalización con calor. Los fragmentos enriquecidos fueron amplificados durante 7 ciclos de PCR mediante el uso de cebadores específicos para los adaptadores de secuenciación ligados.

Secuenciación

La secuenciación de los extremos apareados se llevó a cabo con un Paired Illumina HiSeq 2000 produciendo una secuencia de 50 pb para cada extremo de los fragmentos.

Cartografiado y filtrado

Las secuencias obtenidas se cartografiaron frente a una colección de referencia que consiste en todas las secuencias consenso L1 recogidas en la GIRI RepBase versión 18.10. En una etapa posterior, se seleccionaron todos los pares de lecturas que tenían un cartografiado de secuencia y una secuencia no cartografiada en la base de datos L1. La lectura que no cartografiaba la L1 fue posteriormente cartografiada frente al genoma de referencia humano HG19 (CRCh37, febrero de 2009). Este segundo cartografiado produjo el grupo de pares de lectura "L1-uniqueHG19" que contiene todos los pares que portan una secuencia L1 en un lado y una secuencia no repetitiva única en el otro lado.

Las secuencias de cartografiado únicas se filtraron mediante la extracción de todas las secuencias que se cartografiaban en una ventana de 200 pb contigua a un L1 anotado en el genoma de referencia, debido a que esos elementos L1 son comunes a la población.

Todas las lecturas restantes con unas posiciones de cartografiado solapantes fueron agrupadas en regiones de lectura. Estas regiones de lectura fueron extendidas con las lecturas que fueron cartografiadas en 200 pb secuencia arriba o secuencia bajo.

- 5 Después, las regiones de lectura definidas en una muestra tumoral se compararon con la región de lectura definida por las muestras de linfocitos. Se extrajeron todas las regiones de lectura tumorales residentes en 50 pb secuencia arriba o secuencia abajo de una región de lectura de un linfocito, debido a que dichas lecturas identifican elementos L1 que están presentes en el genoma de la línea germinal y no son inserciones somáticas *de novo*.
- 10 Todas las lecturas duplicadas (duplicados de la PCR), definidas por compartir exactamente las mismas posiciones de cartografiado, se extrajeron de las regiones de lectura tumorales.

15 Finalmente, se seleccionaron todas las regiones que contienen al menos dos lecturas como la demarcación de una inserción somática L1 *de novo* en el genoma tumoral. Dichas regiones son indicativas de una inserción L1 específica tumoral, debido a que (i) contienen al menos 2 lecturas de la clase "L1-uniqueHG19" que (ii) no son contiguas a una L1 anotada (población L1) y (iii) no son contiguas a una L1 encontrada en los genomas de los linfocitos secuenciados (línea germinal L1).

20 *Resultados:*

La etapa de enriquecimiento mostró un enriquecimiento positivo del 60 % (entre el 44 % y el 76 %) de las lecturas en las que al menos una de las lecturas emparejadas podía ser alineada con las secuencias de referencia de L1. Además, se identificaron 1.500 lecturas en las que una lectura se alineaba de forma única con una secuencia no L1 de la base de datos humana HG19 y no estaba anotado que estuviera en las proximidades de una L1 identificada en la base de datos.

Cuando se alinearon los respectivos pares con la secuencia correspondiente a la línea L1-HS, pudo mostrarse una clara presentación de las regiones, usada para el enriquecimiento (Tabla 4), según se muestra en la Figura 6.

30 La Tabla 5 muestra las regiones cromosómicas descubiertas portadoras de una reinserción LI-HS que únicamente está presentes en el tumor, con la posición de la reinserción (HG19) y la posición en la LI-HS. Cuando se consideran únicamente aquellas inserciones genómicas que están comprobadas mediante al menos dos pares de lecturas no idénticos cartografiables de forma única, la distribución de éstas era variable.

35 Tabla 1

Paciente	Región	Recuento normalizado tumoral	Recuento normalizado presérico
1	chr1: 225668824-225669088	3,7	0,2
1	chr12: 24409844-24409970	10,2	10,5
1	chr13: 57722844-57722974	5,9	9,1
1	chr14: 42810280-42810458	8,3	12,4
1	chr19: 58728072-58728275	8,6	5,9
1	chr2: 21825477-21825648	13,6	2,4
1	chr2: 176436453-176436623	28,1	11,1
1	chr3: 16771014-16771238	7,7	5,9
1	chr8: 143078294-143078459	21,9	20,9
2	chr1: 84210120-84210310	6,0	53,5
2	chr12: 7515352-7515583	57,4	16,0
2	chr12: 125812263-125812481	21,8	38,6
2	chr4: 180935360-180935572	18,4	17,8
2	chr9: 112366790-112366988	10,9	43,7
3	chr1: 1639582-1639655	21,2	4,6
3	chr1: 4188245-4188411	11,8	17,1
3	chr1: 8400021-8400166	2,5	3,8
3	chr15: 95386890-95386999	2,5	12,7
3	chr3: 156523259-156523435	8,3	10,1
4	chr1: 42744531-42744674	21,3	8,3
4	chr1: 75476043-75476177	5,3	8,6

ES 2 703 764 T3

Paciente	Región	Recuento normalizado tumoral	Recuento normalizado presérico
4	chr10: 42569993-42570152	1,2	10,1
4	chr11: 37501169-37501278	24,2	5,4
4	chr12: 30311840-30311991	12,1	0,9
4	chr14: 83527163-83527342	44,6	9,7
4	chr16: 69712044-69712214	13,0	13,3
4	chr2: 195725724-195725913	31,6	9,0
4	chr2: 212553961-212554174	44,3	110,0
4	chr3: 105796302-105796413	1,8	0,9
4	chr4: 81355631-81355798	33,9	7,4
4	chr4: 110357587-110357711	1,8	4,3
4	chr5: 91066686-91066828	57,3	32,9
4	chr7: 54903338-54903491	1,5	2,3
4	chrX: 5791690-5791836	61,7	16,7
5	chr1: 43981427-43981622	1,3	4,5
5	chr13: 34083152-34083273	1,3	5,6
5	chr2: 159718375-159718655	14,3	16,2
5	chr21: 9733224-9733404	10,2	9,2
6	chr2: 189381538-189381696	7,4	12,1
7	chr1: 179854863-179855092	29,9	1,2
7	chr13: 106742744-106742969	1,1	5,2
7	chr2: 185709935-185710118	5,8	3,7
7	chr5: 42661495-42661726	30,6	81,9
8	chr1: 48745030-48745205	3,8	8,1
8	chr10: 82375975-82376135	2,5	3,5
8	chr15: 22308807-22308935	3,6	0,7
8	chr2: 186829894-186830068	10,4	2,4
8	chr4: 126197098-126197290	3,0	2,1
8	chr4: 167680779-167680958	2,7	1,2
8	chr5: 22454057-22454246	7,7	17,5
8	chr5: 39974140-39974294	8,8	4,3
8	chr5: 100087972-100088160	14,0	0,3
8	chr7: 13615684-13615825	3,8	8,0
8	chr7: 53325677-53325895	4,7	1,0
8	chr7: 70306861-70307058	11,5	10,0
8	chr9: 104533909-104534105	19,5	1,6
8	chr9: 104533909-104534079	19,2	1,6
8	chrX: 93248625-93248818	6,6	4,5
10	chr1: 243242339-243242484	29,3	6,4
10	chr10: 131942955-131943134	5,6	3,4
10	chr13: 90100261-90100403	2,5	1,1
10	chr16: 34560792-34560945	7,6	12,6
10	chr2: 196460079-196460192	2,0	9,6
10	chr3: 39551193-39551311	3,4	4,9
10	chr5: 37290632-37290828	5,1	9,4
10	chr6: 88754697-88754859	14,6	8,1
10	chr7: 66008688-66008802	1,1	1,5
10	chr7: 79070027-79070178	7,3	1,1

ES 2 703 764 T3

Paciente	Región	Recuento normalizado tumoral	Recuento normalizado presérico
10	chr9: 101889046-101889180	11,0	4,1
20	chrX: 118996698-118996791	2,0	0,9
12	chr1: 34015349-34015535	12,5	14,2
12	chr12: 75835772-75835995	79,1	8,2
12	chr22: 33566325-33566494	37,4	5,2
12	chr22: 33566363-33566494	37,0	5,2
13	chr1: 1639596-1639647	4,7	0,5
13	chr1 85528031-85528246	54,4	34,1
13	chr12: 24409804-24409963	24,0	4,8
13	chr12: 30311833-30312024	4,1	14,0
13	chr13: 31133521-31133722	29,4	8,7
13	chr13: 45993298-45993556	1,4	1,6
13	chr15: 65399680-65399845	2,4	1,6
13	chr7: 34286418-34286606	7,8	2,6
15	chr18: 26245362-26245606	41,4	2,8
16	chr1: 104871623-104871854	52,7	15,7
16	chr1: 160497612-160497806	50,6	29,6
16	chr1: 211908140-211908291	52,1	1,1
16	chr1: 224008418-224008652	1,5	0,8
16	chr10: 71673503-71673740	10,4	4,7
16	chr10: 78166563-78166804	10,4	11,9
16	chr11: 5834094-5834295	5,5	4,4
16	chr12: 79004761-79004964	35,0	36,0
16	chr12: 84141777-84141935	8,9	2,1
16	chr13: 88382677-88382881	11,7	16,7
16	chr14: 38578161-38578387	20,8	18,8
16	chr14: 60568102-60568269	7,4	3,0
16	chr15: 21681844-21682154	36,2	16,5
16	chr17: 27923192-27923393	17,5	7,4
16	chr18: 72971240-72971441	49,7	8,5
16	chr19: 22756096-22756273	19,3	2,8
16	chr19: 35541770-35541973	3,4	3,0
16	chr19: 36660194-36660423	22,1	3,6
16	chr2: 35143314-35143471	2,5	0,4
16	chr2: 84631424-84631656	86,5	13,3
16	chr2: 106026857-106027104	16,3	4,7
16	chr2: 137936742-137936888	17,8	10,2
16	chr2: 167109411-167109623	7,1	12,7
16	chr2: 216279158-216279366	12,0	47,2
16	chr2: 216279195-216279366	5,8	46,6
16	chr2: 229962184-229962346	26,1	29,4
16	chr2: 236356980-236357204	8,9	8,7
16	chr20: 8134573-8134740	3,7	5,9
16	chr20: 51803208-51803323	3,7	20,5
16	chr3: 74888046-74888192	8,9	2,1
16	chr3: 128978403-128978648	77,6	66,9
16	chr3: 136882961-136883146	11,0	4,9

ES 2 703 764 T3

Paciente	Región	Recuento normalizado tumoral	Recuento normalizado presérico
16	chr3: 167384636-167384854	11,3	35,2
16	chr4: 12257674-12257901	8,9	1,1
16	chr4: 12257674-12257899	8,3	1,1
16	chr4: 18081638-18081865	7,4	12,1
16	chr4: 129081215-129081401	20,5	9,1
16	chr5: 1493804-1493894	4,0	1,9
16	chr5: 10475757-10475924	46,6	13,1
16	chr5: 59655387-59655550	2,1	4,2
16	chr5: 95356487-95356755	54,6	10,8
16	chr5: 103230987-103231140	6,1	4,7
16	chr5: 129721638-129721818	13,8	3,0
16	chr5: 168710716-168710926	23,9	28,4
16	chr6: 5293628-5293816	3,1	8,3
16	chr6: 110557420-110557574	37,7	47,9
16	chr6: 119435653-119435855	15,0	7,6
16	chr6: 119435653-119435819	7,1	7,6
16	chr6: 123599662-123599856	14,4	3,0
16	chr7: 55554702-55554866	13,8	15,2
16	chr7: 61259199-61259372	28,8	99,1
16	chr7: 95550567-95550810	22,7	3,0
16	chr7: 102304106-102304342	22,7	20,8
16	chr7: 141355925-141356108	4,6	3,4
16	chr8: 59661474-59661654	37,4	101,9
16	chr8: 82587476-82587688	64,1	25,0
16	chr9: 26167010-26167139	10,1	17,4
16	chr9: 26167025-26167139	5,8	17,4
16	chr9: 107120758-107120970	6,7	4,2
16	chrX: 35176776-35176937	13,2	17,6
16	chrX: 93248594-93248891	4,6	13,1
17	chr1: 4188229-4188422	61,0	19,6
17	chr12: 79004801-79004893	4,7	4,1
17	chr16: 74212566-74212745	71,8	10,3
17	chr4: 41778429-41778642	81,4	65,0
17	chr5: 18955266-18955494	9,3	4,3
17	chr6: 53275296-53275532	9,6	2,1
17	chr7: 116753272-116753466	14,3	3,2
17	chr8: 3034589-3034766	2,3	10,7
17	chr8: 17987284-17987515	18,1	3,9
17	chrX: 86482191-86482398	7,9	1,8

Tabla 2

4x PBMC < Tu & 2x Post < Pre

Paciente	Región	Recuento normalizado tumoral	Recuento normalizado presérico de PBMC	Recuento normalizado presérico	Recuento normalizado postsérico
5	chr4: 180935376-180935564	23,23367202	0	23,96991786	6,13522487
6	chr10: 113995673-113995829	1,598943214		37,17017141	11,1092118

ES 2 703 764 T3

4x PBMC < Tu & 2x Post < Pre

Paciente	Región	Recuento normalizado tumoral	Recuento normalizado presérico de PBMC	Recuento normalizado presérico	Recuento normalizado postsérico
6	chr14: 38230428-38230622	32,9382302	0,24 / 0,88	123,1571679	44,8408184
6	chr14: 87657285-87657463	11,1926025		24,53231313	6,46354139
6	chr9: 14867350-14867546	8,634293354		58,48106969	10,9072261
6	chr9: 89135762-89135898	12,79154571		93,9166331	29,2879219
9	chr18: 20355821-20355995	15,67390694		35,50994404	13,0295755
9	chr3: 6611827-6611963	4,528017561		197,5921426	41,5875492
9	chr5: 35456477-35456658	4,179708518		35,50994404	7,31798076
9	chr8: 130024849-130025036	14,62897981		101,5192265	33,0201571
9	chr8: 132446963-132447159	20,55023355		120,9080917	27,8440243
14	chr10: 72275346-72275501	22,84380142	0,13575	23,52753959	15,614494
15	chr1: 69863579-69863750	282,3202084	1,033791	23,99745741	12,1546129
15	chr5: 55547575-55547662	49,98456148		47,58817825	18,0068339
15	chr6: 120421873-120421996	298,9817289		61,82395807	11,0291858

4x PBMC < Tu & 4x Post < Pre

6	chr9:14867350-14867546	8,634293354	0	58,48106969	10,9072261
7	chr1: 212977517-212977735	25,15932599		11,67053095	1,28400063
7	chr15: 84025866-84026064	9,844953648		7,141668192	1,07000053
7	chr17: 32996024-32996267	64,9037685		59,39777691	17,3340085
7	chr2: 115550251-115550458	8,751069909		12,88984015	2,14000105
7	chr3: 164908141-164908357	7,292558258		29,61179494	4,49400221
7	chr5: 98633992-98634189	15,31437234		69,32643757	11,7700058
7	chr6: 128375720-128375918	9,844953648		25,60549327	3,21000158
7	chr7: 53325693-53325892	4,740162868		19,85732131	5,35000263

4x PBMC < Tu & 4x Post < Pre

7	chr9: 102742619-102742818	9,844953648		24,21199704	3,21000158
11	chM5: 25852912-25853071	11,05459787 0		29,29110425	5,24337296
12	chr12: 95944094-95944285	22,97521656		44,16152383	5,5623034
12	chr17: 47404627-47404853	9,345850802		58,94864041	20,0242922
15	chr6: 120421873-120421996	298,9817289	3,205797	61,82395807	11,0291858

4 x PBMC < Tu & 10x Post < Pre

7	chr1: 212977517-212977735	25,15932599		11,67053095	1,28400063
12	chr12: 95944094-95944285	22,97521656		44,16152383	5,5623034

Tabla 3

Nombre del cebador	Secuencia del cebador	Fuente del molde ADN tumoral	Fuente del molde ADN de las PBMC
16-chr17: 27923338-27923393.F	TTGTATTACTTCTTGACCACATTAAA	WGA1	WGA1
16-chr17: 27923338-27923393.R	CGAGGGGGAGTAAAAGTGCT	WGA1	WGA1
16-chr14: 38578285-38578387.F	AAATGCACCTCTTCTCTCCA	WGA1	WGA1
16-chr14: 38578285-38578387.R	ATGCTGAGTTTTGGACTCAA	WGA1	WGA1
16-chr3: 74888046-74888097.F	AAACATCTTTAATCCCCAGCA	WGA1	WGA1
16-chr3: 74888046-74888097.R	TTCCTTAGATCAGGCACATTTTT	WGA1	WGA1
5-chr2: 159718375-159718655. F	GCCACAGGTCTTTGAAGAGG	WGA1	WGA1
5-chr2: 159718375-159718655. R	GGAACCTGCACACCCTGTAT	WGA1	WGA1
10-chr13: 90100261-90100403. F	CAAATCTCAAGAGTTTATGATGTACC	WGA1	WGA4

ES 2 703 764 T3

Nombre del cebador	Secuencia del cebador	Fuente del molde ADN tumoral	Fuente del molde ADN de las PBMC
10-chr13: 90100261-90100403.R	CCTAGAGAGATTGGGGGCTAA	WGA1	WGA4
10-chr2: 196460079-196460192. F	GCACATTTTAAAAAGCTTGCAG	WGA1	WGA4
10-chr2: 196460079-196460192. R	CGAATTTTCTATGTAAGTAAACA	WGA1	WGA4
10-chr10: 131942955-131943134. F	GAGAGGGCTTCTGAGATACC	WGA1	WGA4
10-chr10: 131942955-131943134.R	TCCACCATCGGAGAATAAGG	WGA1	WGA4
12-chr22: 33566363-33566494.F	AGCCACATGACCGCCTTCT	WGA1	WGA1
12-chr22: 33566363-33566494.R	TGTTTGACCAGGACCCTGAC	WGA1	WGA1

Tabla 4

Nombre	Secuencia
Probe4000 SEQ ID NO: 9	CTGGCACAAGACAGGGATGCCCTCTCTCACCGCTCCTATTCAACATAGTGTGGAAGTTC
Probe4117 SEQ ID NO: 10	CCCCATCGTCTCAGCCCAAATCTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCTGAGG
Probe4500 SEQ ID NO: 11	CTGGAGGCATCACACTACCTGACTTCAAACATACTACAAGGCTACAGTAACCAAAACAG
Probe4759 SEQ ID NO: 12	GCAATGGGGAAAGGATTCCCTATTTAATAAATGGTGCTGGGAAAAGTGGCTAGCCATATG
Probe5000 SEQ ID NO: 13	GGCGTGGGCAAGGACTTCATGTCCAAAACACAAAAGCAATGGCAACAAAAGCCAAAATT
Probe5182 SEQ ID NO: 14	AAACAACCCCATCAAAAAGTGGGCGAAGGACATGAACAGACACTTCTCAAAGAAGACAT
Probe5500 SEQ ID NO: 15	CACTGTTGGTGGGACTGTAAACTAGTTCAACCATTGTGGAAGTCAGTGTGGCGATTCTC
Probe5699 SEQ ID NO: 16	GATGAGTTCATATCCTTTGTAGGACATGGATGAAATTGGAAACCATCATTCTCAGTAAAC
Probe6000 SEQ ID NO: 17	GGGTGCAGCGCACCAGCATGGCACATGTATACATATGTAACCTGCACAATGTGCAC

5

Tabla 5

Tumor	Región	Lecturas en la región	Posiciones de cartografiado L1	Posiciones de cartografiado HG19 chr22: 37806100 (-), chr22:
344	chr22: 37806031-37806149	3	5266.5892.5892	37806033 (+), chr22: 37806032 (+) chr20: 62902257 (+), chr20:
344	chr20: 62902155-62902421	3	4843.5199.5199	62902156 (+), chr20: 62902372 (+) chr12: 16502810 (-), chr12:
344	chr12: 16502625-16502869	2	5936.5826	16502820 (-) chr3: 76827020 (+), chr3: 76827022
	chr3: 76827019-76827093		5609.5628.5647.5557	(+), chr3: 76827025 (+),
318		4		chr3: 76827044 (+) chr13: 42440358 (+), chr13:
317	chr13: 42440357-42440719	3	5349.5804.5970	42440567 (-), chr13: 42440670 (-) chr13: 108859662 (+), chr13:
315	chr 13: 108859513-108859718	2	4765.4801	108859669 (+) chr5: 88629450 (+), chr5: 88629499
119	chr5: 88629449-88629608	2	5956.5193	(-) chr7: 129888059 (+), chr7:
119	chr7: 129888058-129888246	3	5675.5650.5659	129888063 (+), chr7: 129888085 (+)
119	chr15: 58669421-58669609	3	5352.5810.5810	chr15: 58669422 (+), chr15: 58669558 (-), chr15: 58669559 (-)
119	chrX: 38768199-38768309	3	3903.3857.3847	chrX: 38768200 (+), chrX: 38768246 (+), chrX: 38768260 (+)

REIVINDICACIONES

1. Un método para la selección de un biomarcador personalizado para un paciente que tiene un tumor, método que comprende:
 - (a) la preparación de una colección de fragmentos de ADN a partir de muestras de ácidos nucleicos que comprenden el ADN obtenido del paciente a partir de cada una de las siguientes fuentes: células mononucleares de sangre periférica (PBMC), el tumor y los ácidos nucleicos acelulares previos a la terapia, ácidos nucleicos que se obtienen antes de una terapia;
 - (b) la secuenciación de los fragmentos de ADN de las colecciones para identificar los fragmentos de ADN que comprenden una región de inserción de un elemento repetitivo; y
 - (c) la comparación de las regiones de inserción de un elemento repetitivo identificadas en la muestra de PBMC con las regiones de inserción de un elemento repetitivo identificadas en la muestra del tumor y/o en la muestra de los ácidos nucleicos acelulares para identificar las regiones de inserción de un elemento repetitivo que están presentes en el tumor, pero no en las PBMC, identificando así un biomarcador personalizado para el paciente.
2. El método de la reivindicación 1, en el que el elemento repetitivo es un elemento Alu o un elemento LINE.
3. El método de una cualquiera de las reivindicaciones anteriores, en el que la etapa (c) comprende adicionalmente la comparación de las regiones de inserción de un elemento repetitivo identificadas en la muestra de PBMC con las correspondientes regiones en la base de datos del genoma humano para identificar las regiones de inserción en la muestra de PBMC que no están presentes en la base de datos del genoma humano.
4. El método de una cualquiera de las reivindicaciones anteriores, que comprende adicionalmente las etapas de añadir secuencias adaptadoras que comprenden sitios de cebado universales a cada uno de los fragmentos de ADN y la amplificación de los fragmentos de ADN, antes de la etapa (b).
5. El método de la reivindicación 4, en el que la etapa de amplificación se lleva a cabo mediante el uso de:
 - (i) un cebador específico para el elemento repetitivo y un cebador universal; o
 - (ii) dos cebadores universales.
6. El método de una cualquiera de las reivindicaciones anteriores, que comprende adicionalmente la etapa de enriquecimiento en los fragmentos de ADN en al menos una de las colecciones para los fragmentos de ADN que comprenden una región de inserción de un elemento repetitivo, opcionalmente en el que la etapa de enriquecimiento se lleva a cabo mediante el uso de sondas oligonucleotídicas que hibridan específicamente con el elemento repetitivo.
7. El método de una cualquiera de las reivindicaciones anteriores, que comprende adicionalmente la etapa de la obtención de una muestra de ácidos nucleicos del paciente a partir de los ácidos nucleicos acelulares posteriores a la terapia, ácidos nucleicos que se obtienen después de la terapia, que opcionalmente comprende adicionalmente la etapa de la comparación de las regiones de un elemento repetitivo identificadas en la muestra de ácidos nucleicos acelulares previa a la terapia con la muestra de ácidos nucleicos acelulares posterior a la terapia para identificar las regiones de inserción de un elemento repetitivo que no son detectadas en la muestra de ácidos nucleicos acelulares posterior a la terapia.
8. Un método para la selección de un biomarcador personalizado para un paciente que tiene un tumor, método que comprende:
 - (a) la realización de al menos una reacción de amplificación con el ADN de las muestras de ácidos nucleicos que comprenden el ADN obtenido del paciente a partir de cada una de las siguientes fuentes: células mononucleares de sangre periférica (PBMC), el tumor, los ácidos nucleicos acelulares previos a la terapia, ácidos nucleicos que se obtienen antes de una terapia, y los ácidos nucleicos acelulares posteriores a la terapia, ácidos nucleicos que se obtienen después de la terapia, en el que la amplificación se lleva a cabo mediante el uso de un cebador de una secuencia Alu y de un cebador universal;
 - (b) la secuenciación del ADN obtenido a partir de las reacciones de amplificación para determinar las regiones que comprenden una secuencia de inserción Alu con respecto a las correspondientes regiones en la base de datos del genoma humano;
 - (c) la comparación de las regiones de inserción Alu identificadas en la etapa (b) de la muestra de PBMC con las regiones de inserción Alu identificadas en la etapa (b) de la muestra del tumor, para identificar las regiones de inserción Alu que están presentes en el tumor, pero no en las PBMC;
 - (d) la comparación de las regiones de inserción Alu identificadas en la etapa (c) con las regiones de inserción Alu identificadas en la muestra de ácidos nucleicos acelulares del paciente previa a la terapia, para identificar las regiones de inserción Alu que están presentes en la muestra del tumor y en la muestra de ácidos nucleicos acelulares previa a la terapia; y
 - (e) la comparación de las regiones de inserción Alu identificadas en la etapa (d) con las regiones de inserción Alu

identificadas en la muestra de ácidos nucleicos acelulares del paciente posterior a la terapia, para identificar las regiones de inserción Alu que no se detectan en la muestra de ácidos nucleicos acelulares posterior a la terapia, identificando así un biomarcador personalizado para el paciente.

- 5 9. Un método para la selección de un biomarcador personalizado para un paciente que tiene un tumor, método que comprende:
- 10 (a) la realización de al menos una reacción de amplificación con el ADN de una muestra de ácidos nucleicos que comprende el ADN obtenido de los ácidos nucleicos en circulación previa a la terapia y una muestra de ácidos nucleicos que comprende el ADN obtenido de los ácidos nucleicos en circulación posterior a la terapia, en el que la amplificación se lleva a cabo mediante el uso de un cebador de una secuencia Alu y de un cebador universal;
- 15 (b) la secuenciación del ADN obtenido a partir de las reacciones de amplificación para determinar las regiones que comprenden una secuencia de inserción Alu con respecto a un genoma humano normal y que puede ser posicionada de forma única en una región definida de un genoma humano normal;
- 20 (c) la cuantificación del nivel de una región de inserción Alu identificada en la etapa (b) a partir de la muestra de ADN en circulación previa a la terapia en comparación con el nivel de las regiones de inserción Alu presentes en los controles normales, para identificar las regiones de inserciones Alu que están presentes en el ADN acelular en circulación a unos niveles mayores que en los controles normales; y
- (d) la comparación del nivel de las regiones de inserción Alu identificadas en la etapa (c) con el nivel de las regiones de inserción Alu en la muestra de ADN acelular posterior a la terapia, para identificar las regiones de inserción Alu que no son detectadas a un nivel diferente del normal, identificando así un biomarcador personalizado para el paciente.
- 25 10. El método de la reivindicación 8 o la reivindicación 9, que comprende adicionalmente la realización de una reacción de amplificación adicional con el ADN de cada una de las muestras de ácidos nucleicos en el que las segundas reacciones de amplificación que comprenden un segundo cebador de una secuencia Alu y un segundo cebador universal.
- 30 11. El método de la reivindicación 10, en el que una de las reacciones de amplificación comprende:
- (i) un par de cebadores que comprende la SEQ ID NO: 5 y un cebador universal; o
- (ii) un par de cebadores que comprende la SEQ ID NO: 6 y un cebador universal.
- 35 12. El método de la reivindicación 10, en el que una de las reacciones de amplificación comprende un par de cebadores que comprende la SEQ ID NO: 5 y un cebador universal, y las reacciones de amplificación adicionales comprenden un par de cebadores que comprende la SEQ ID NO: 6 y un cebador universal.
- 40 13. El método de una cualquiera de las reivindicaciones anteriores, en el que la terapia es una cirugía.
14. El método de una cualquiera de las reivindicaciones anteriores, en el que el paciente tiene un tumor de mama.
- 45 15. Un método de monitorización de la respuesta de un paciente que tiene un tumor a una terapia, método que comprende la realización de un método según una cualquiera de las reivindicaciones anteriores para la identificación de un biomarcador individualizado y la evaluación del nivel de dicho biomarcador en los ácidos nucleicos en circulación acelulares obtenidos del paciente antes y después de la terapia.

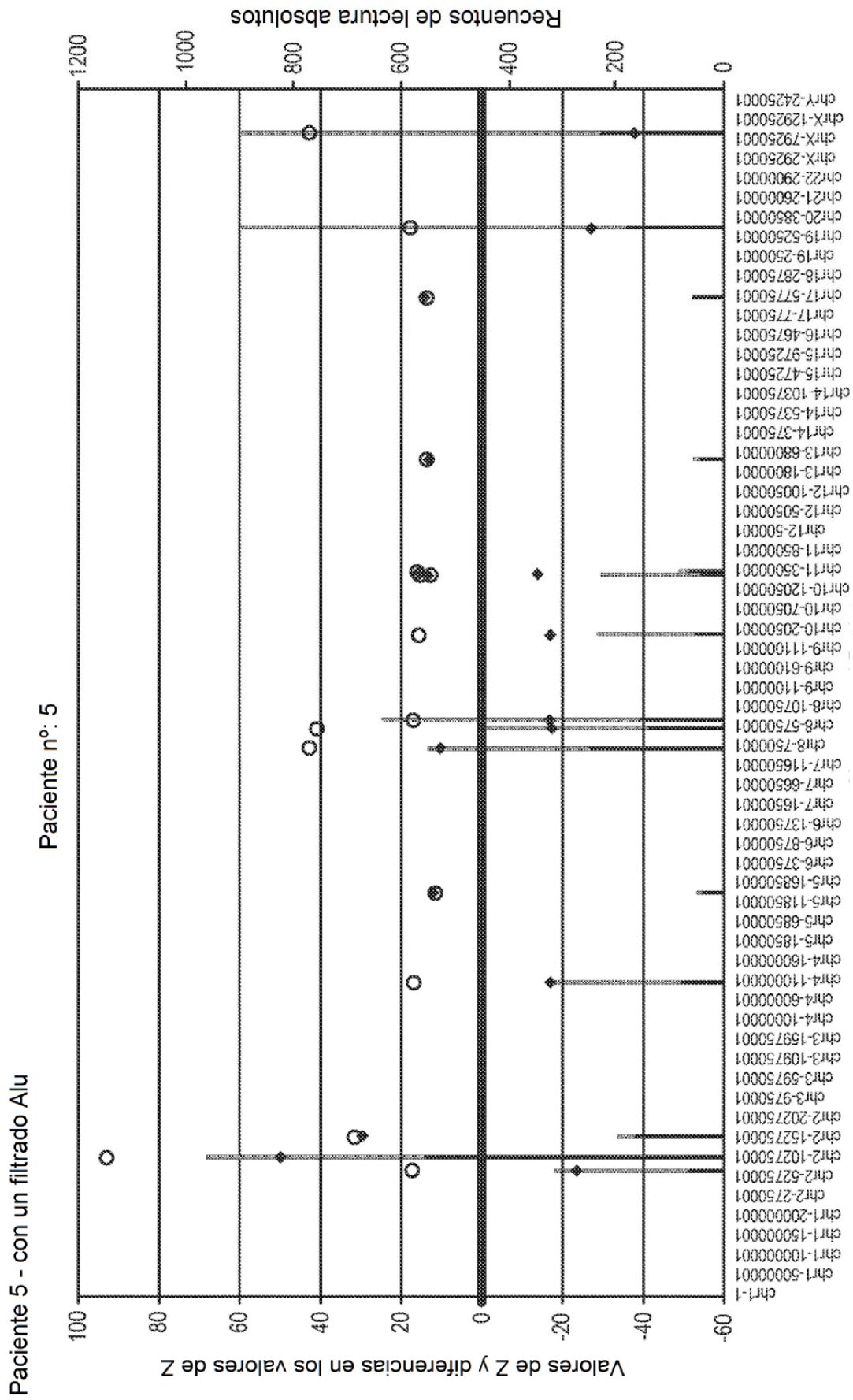


FIG. 1

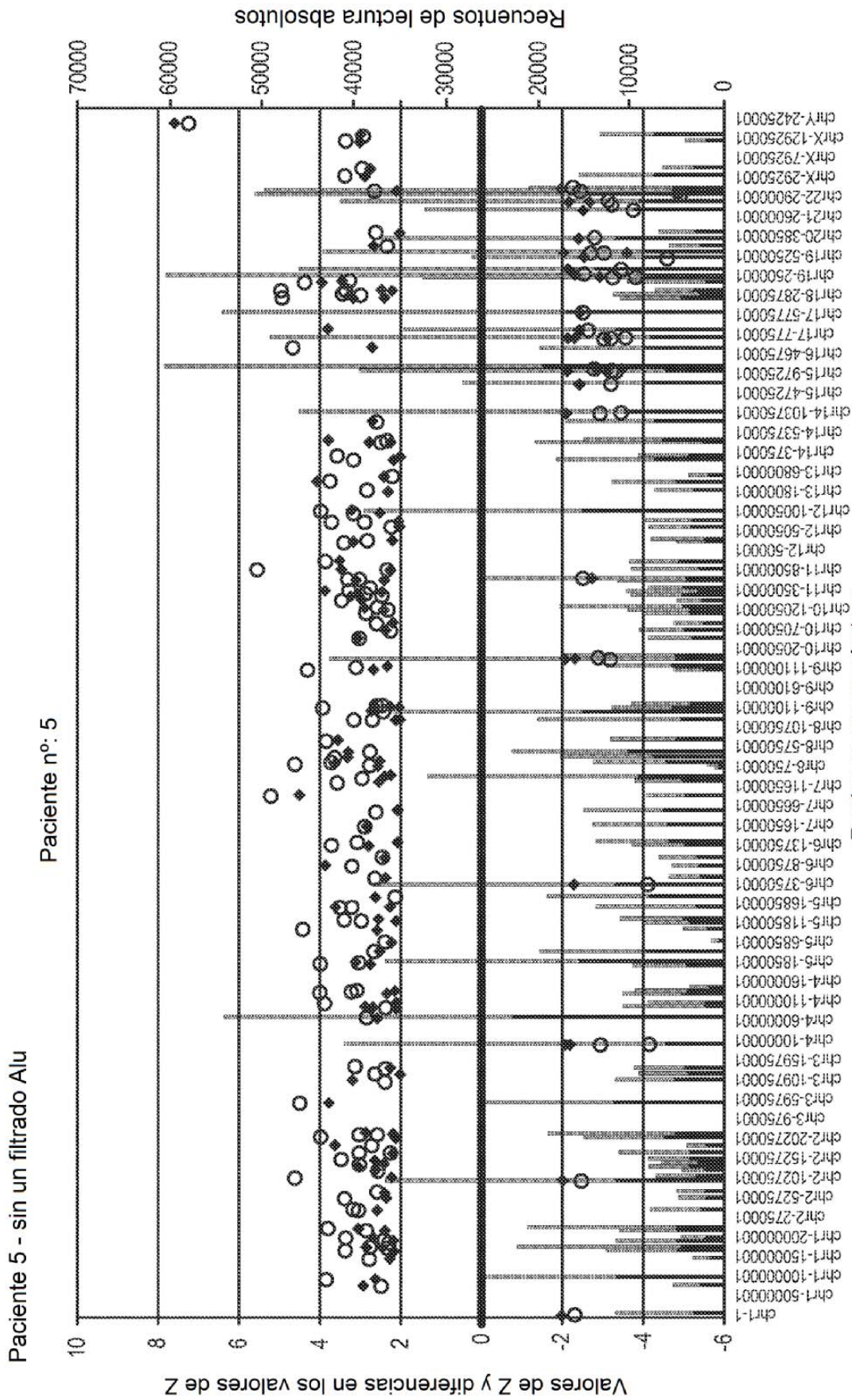
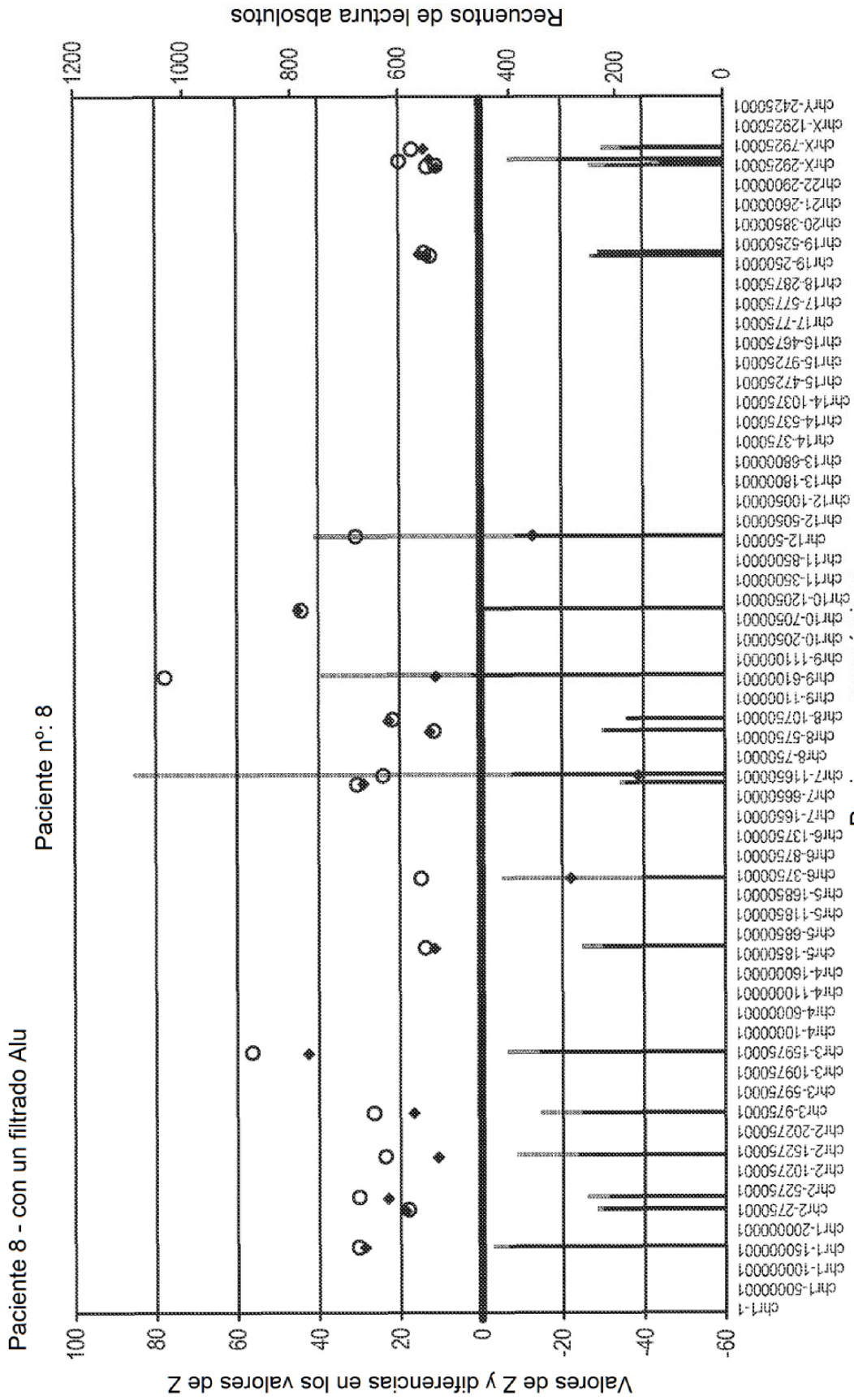


FIG. 1
(continuación)



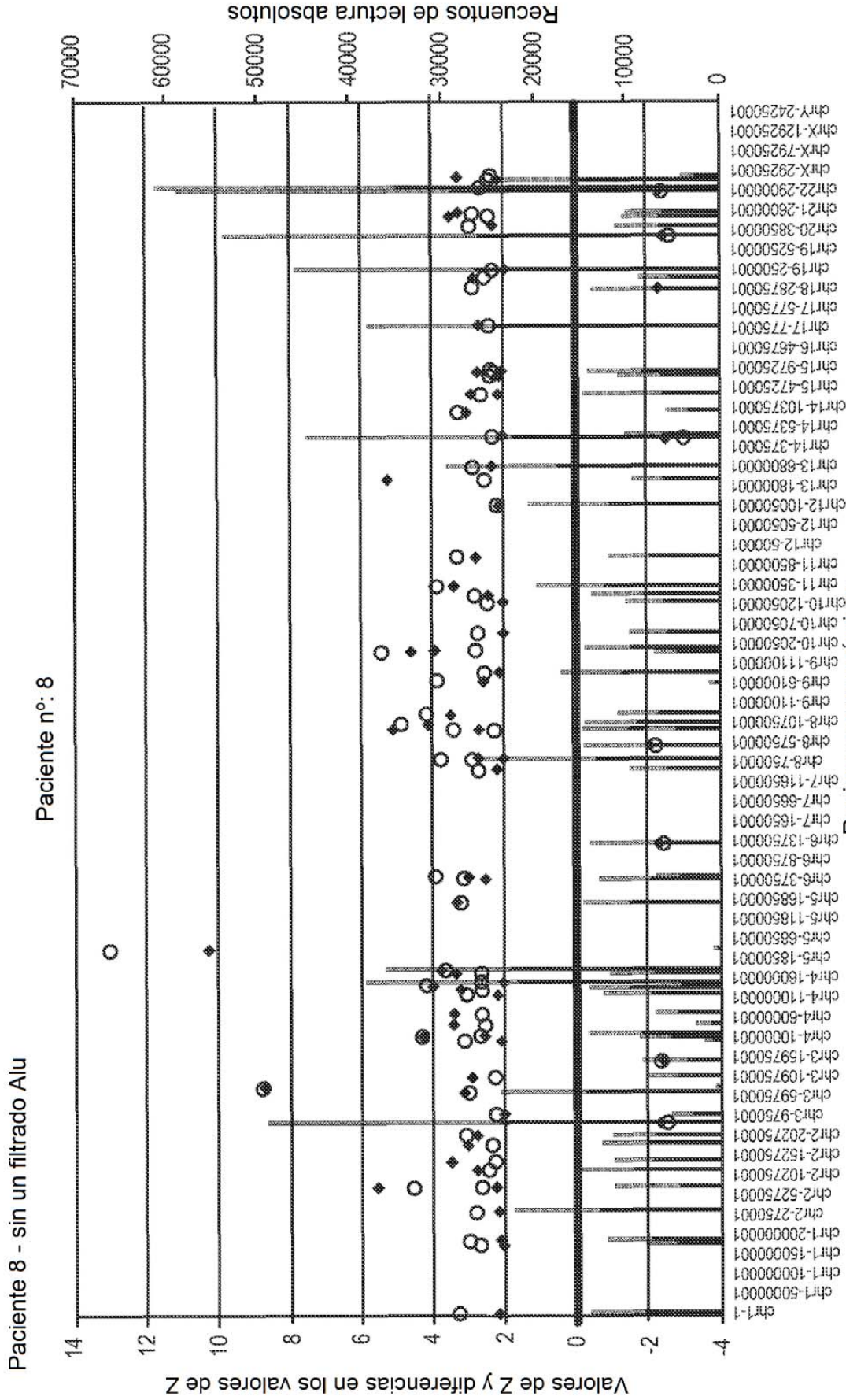


FIG. 1
(continuación)

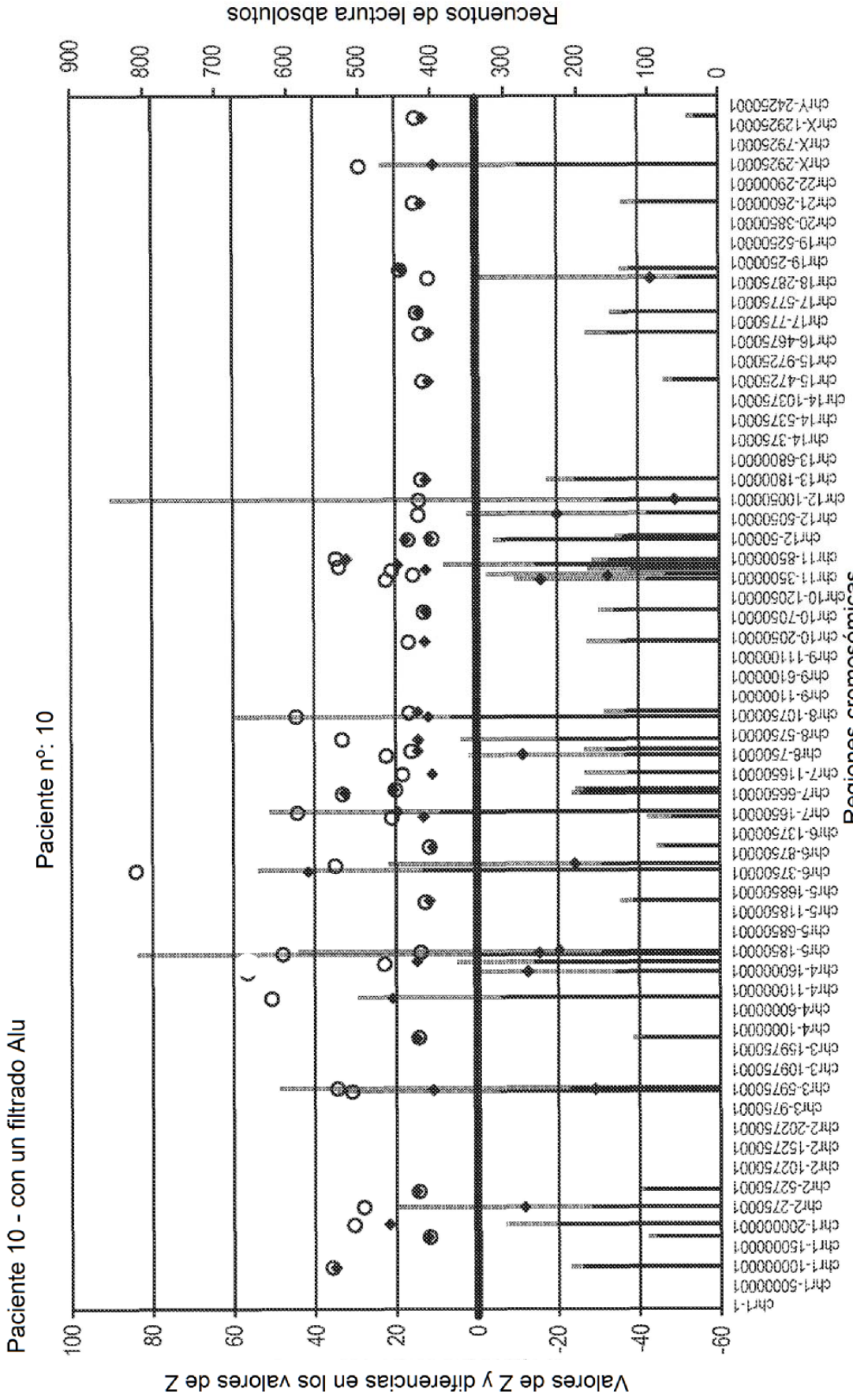
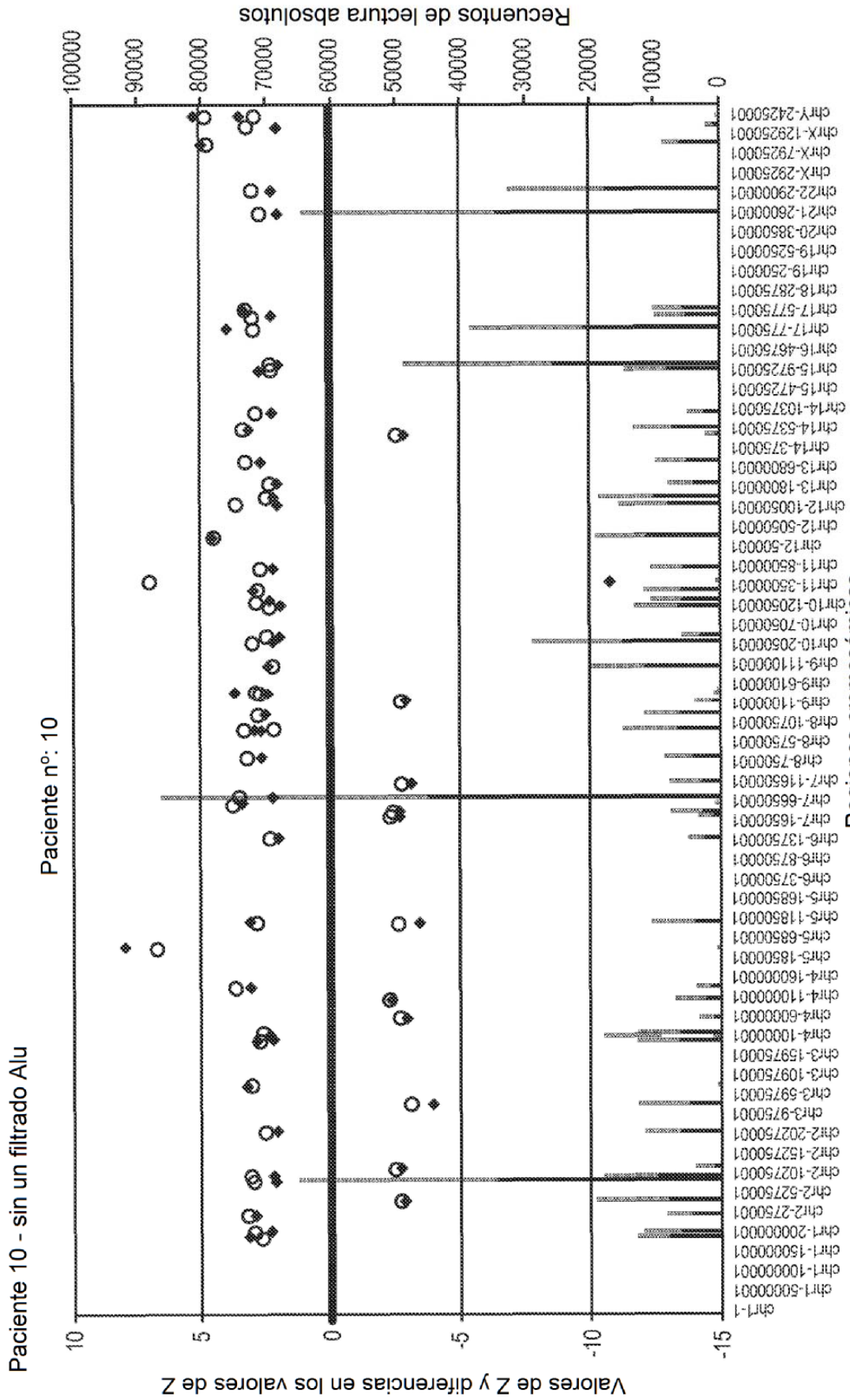
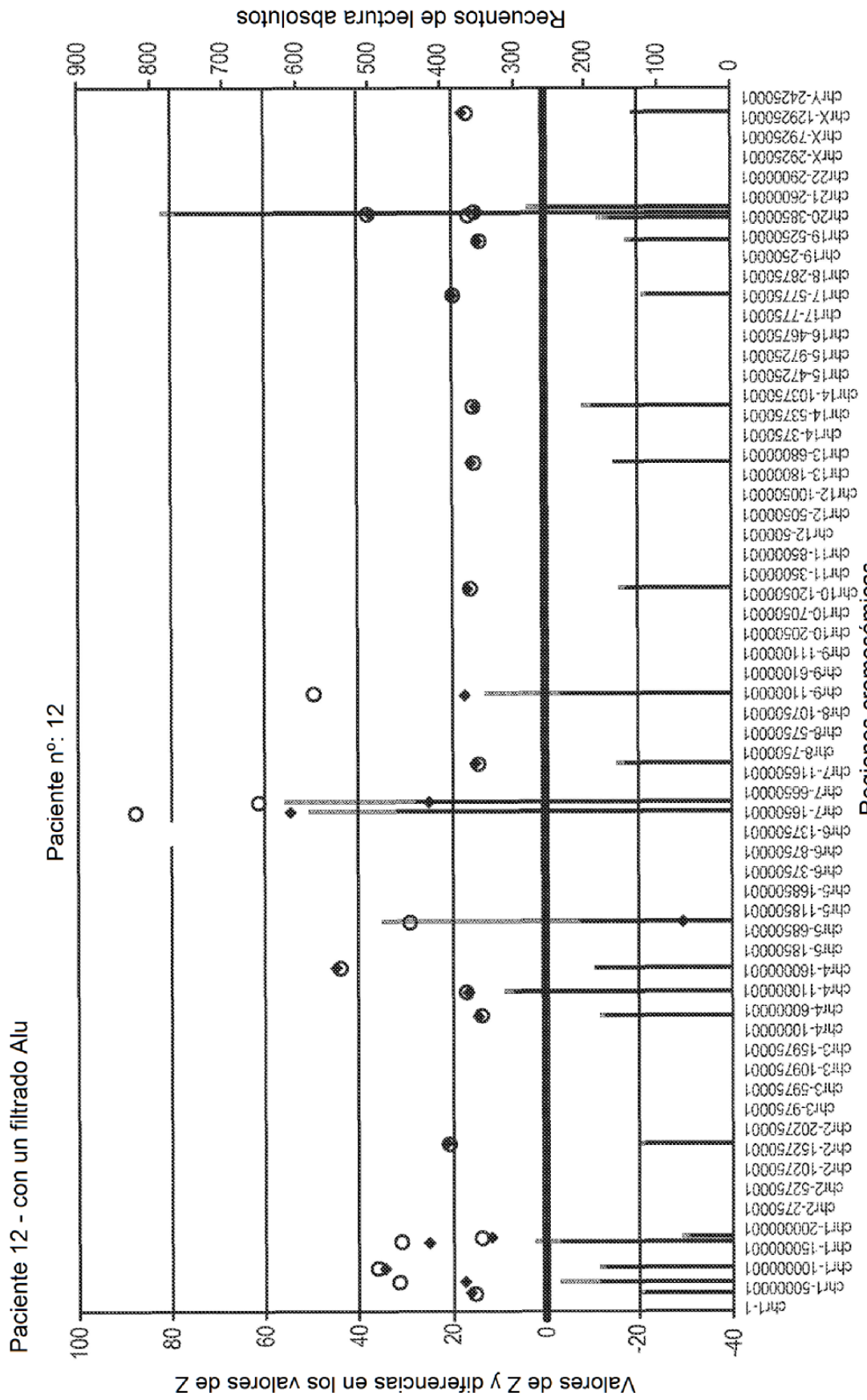


FIG. 1
(continuación)





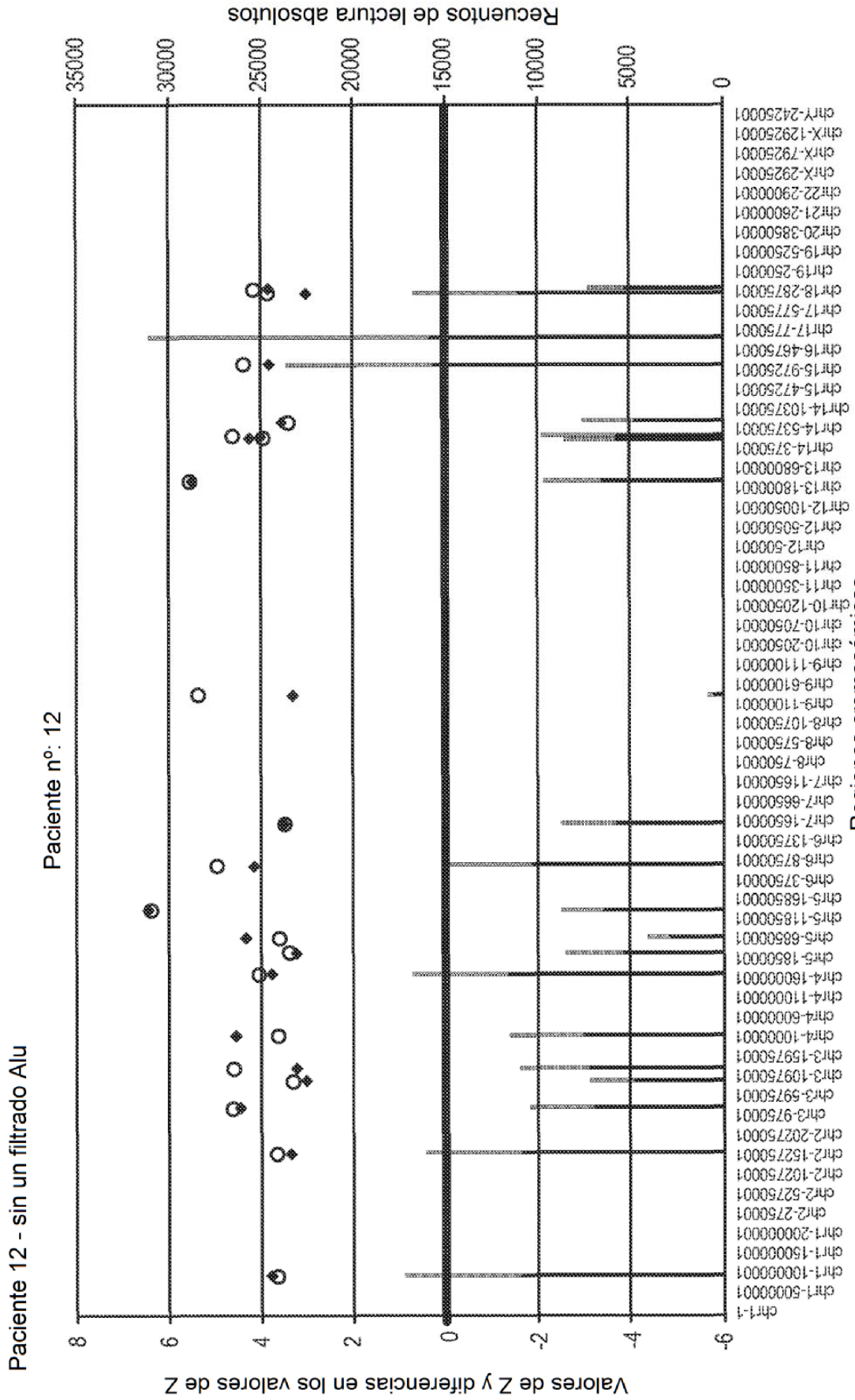
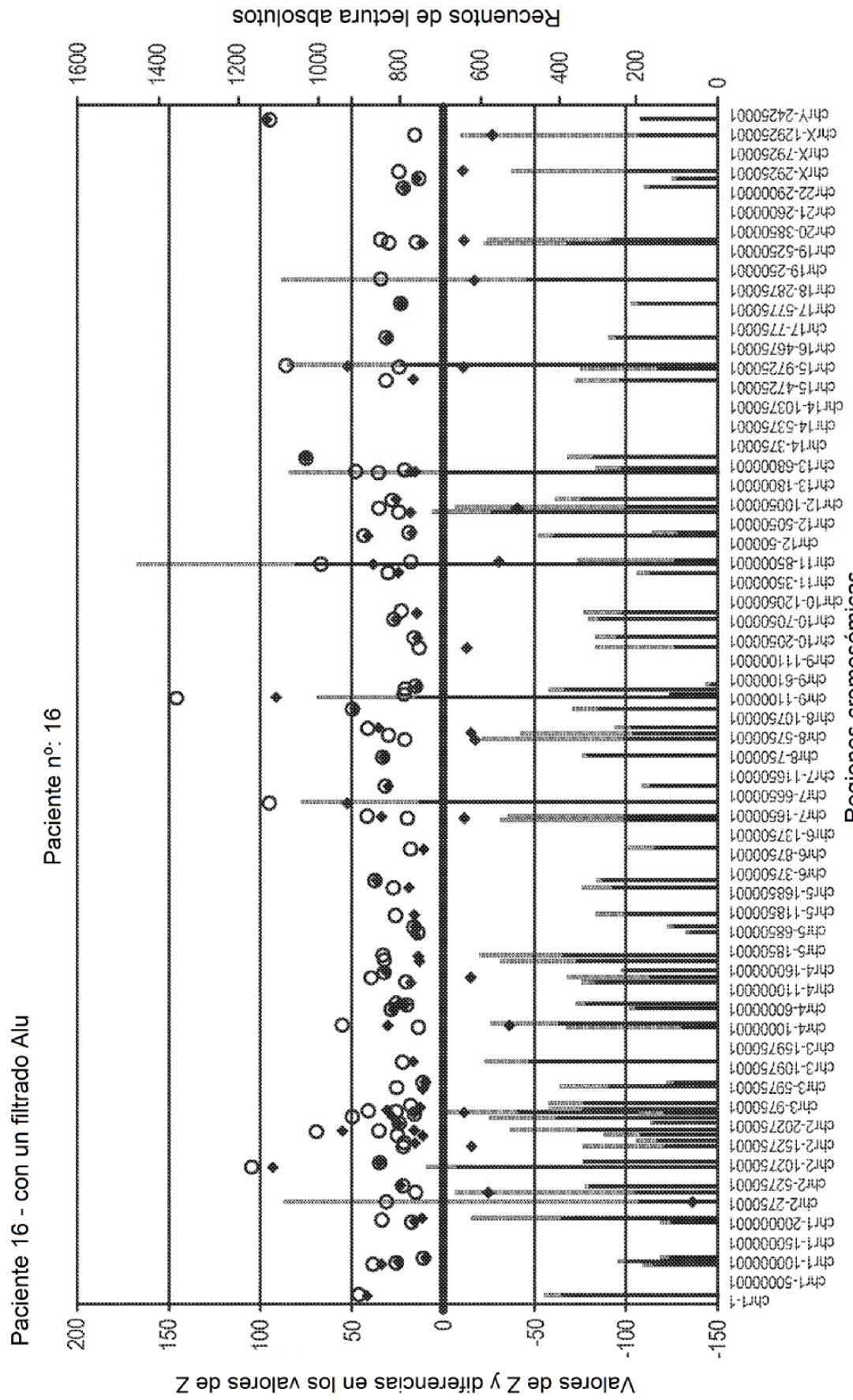
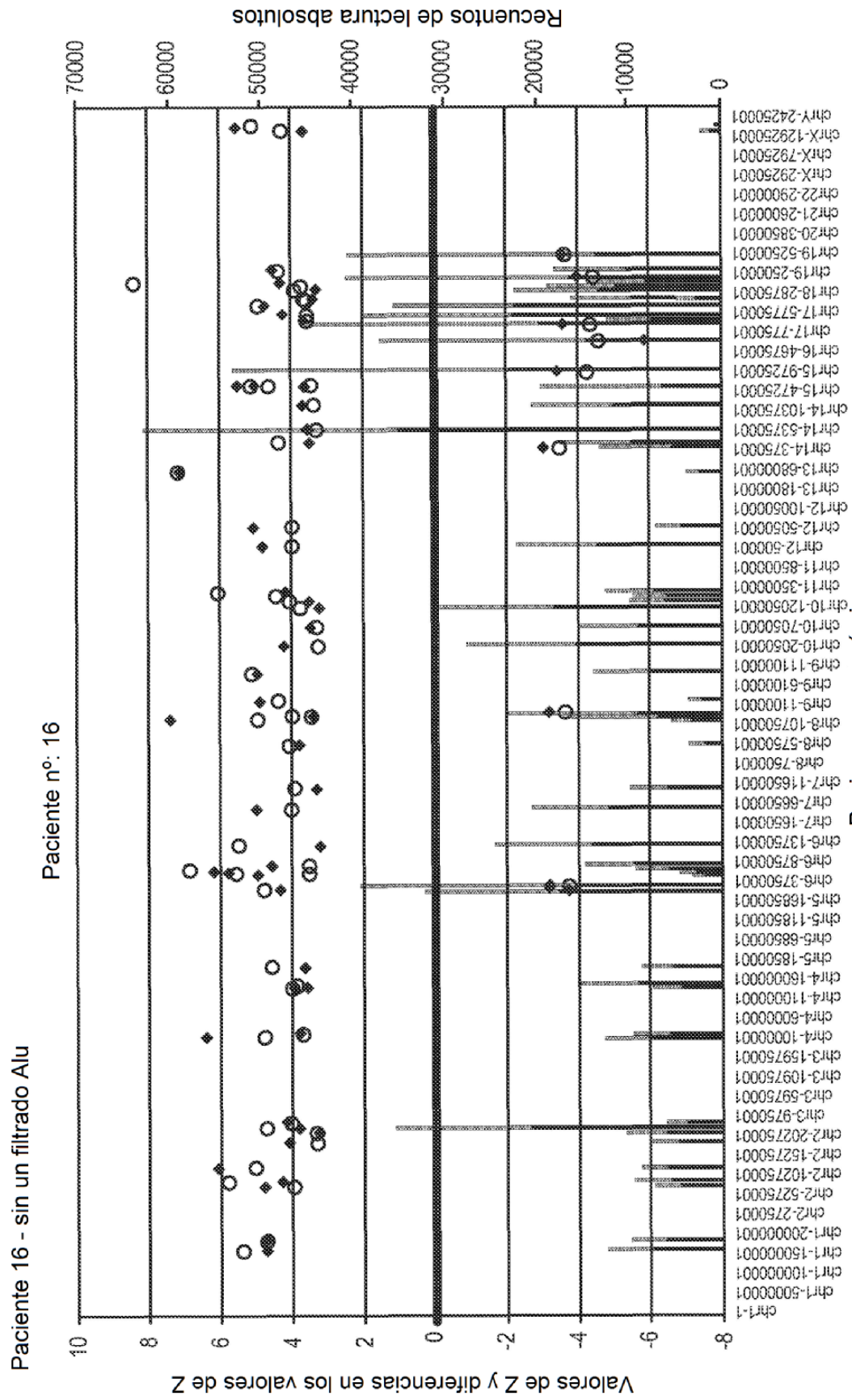


FIG. 1
(continuación)





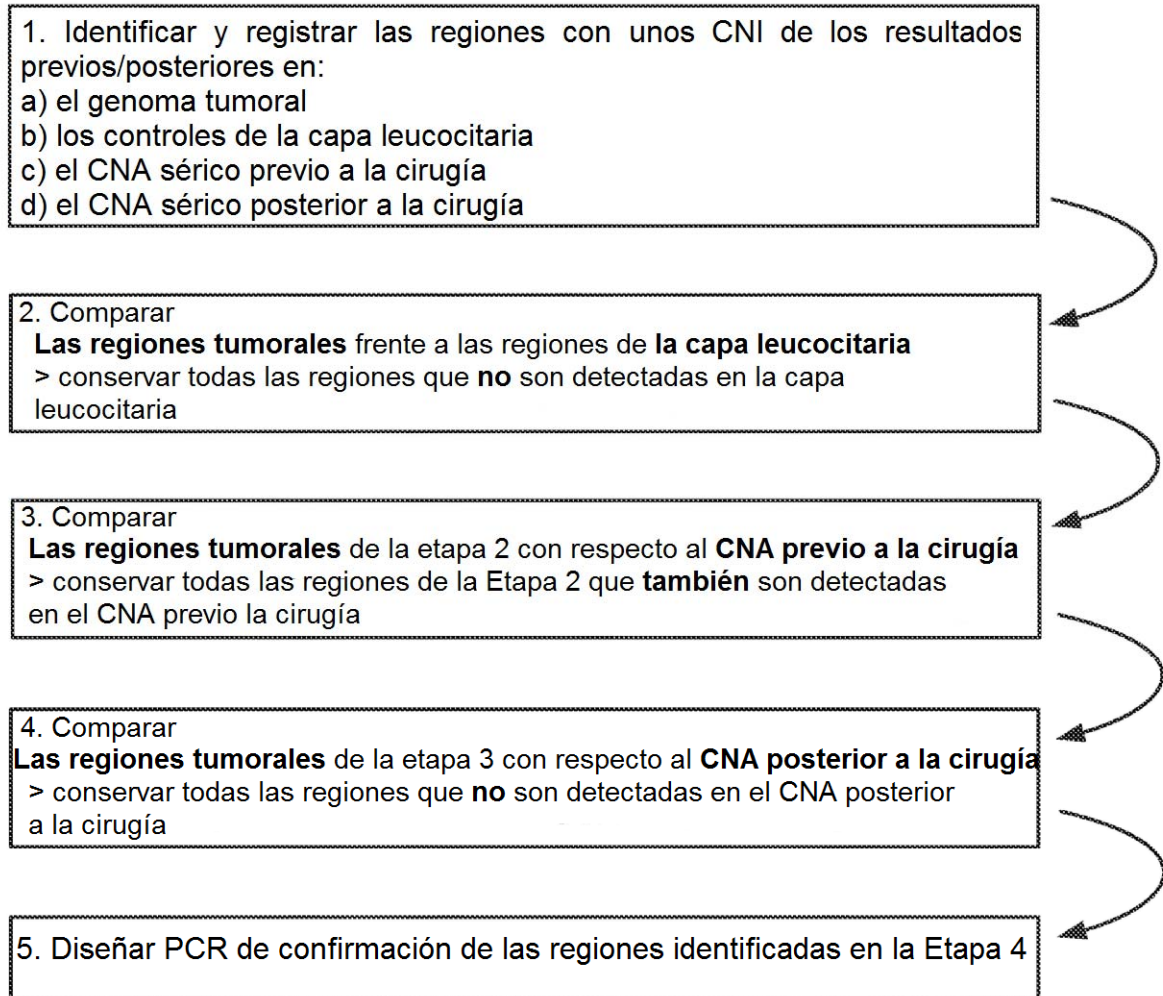


FIG. 2

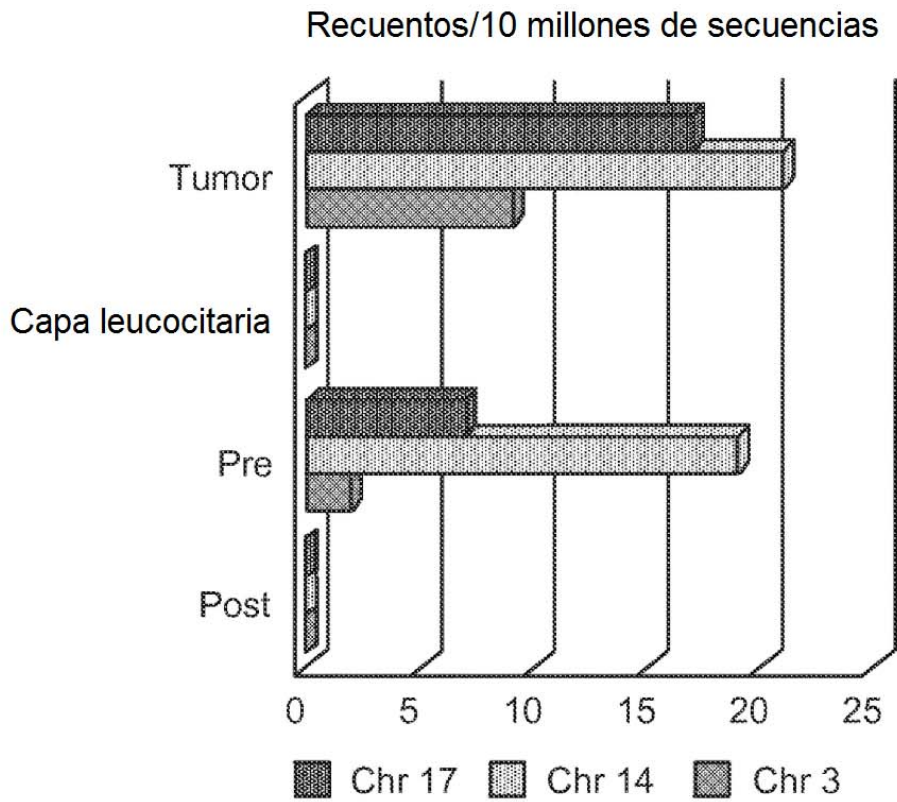
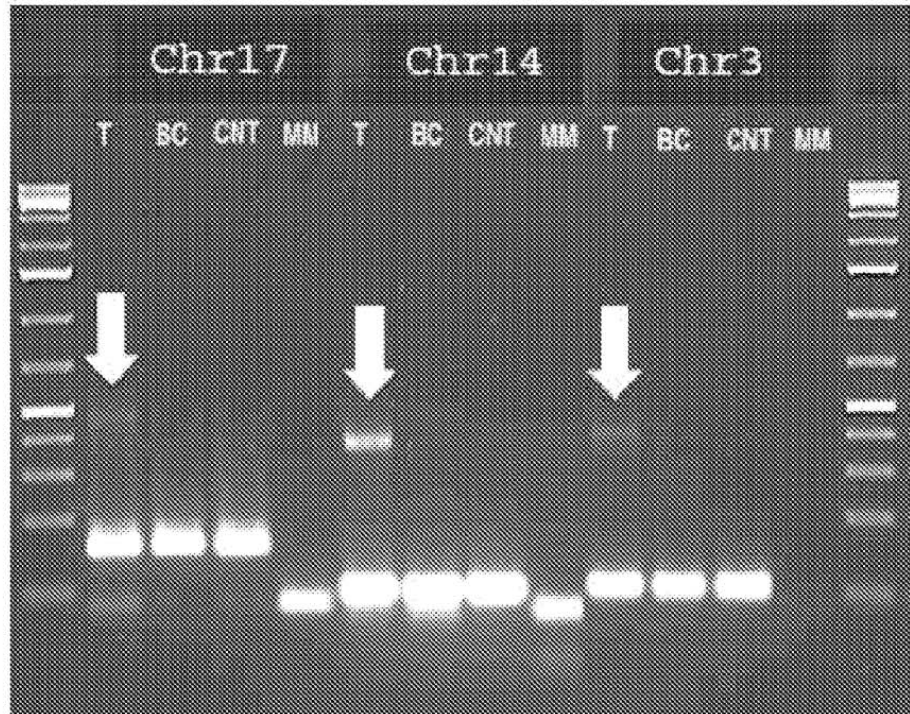


FIG. 3A

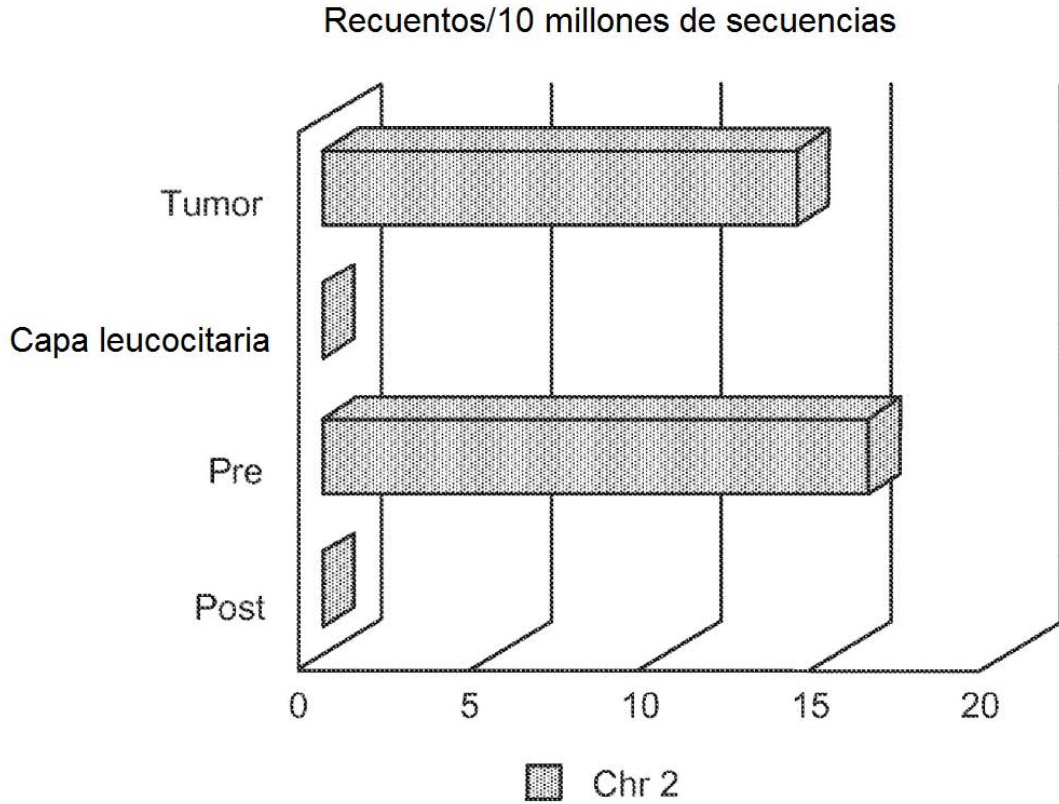
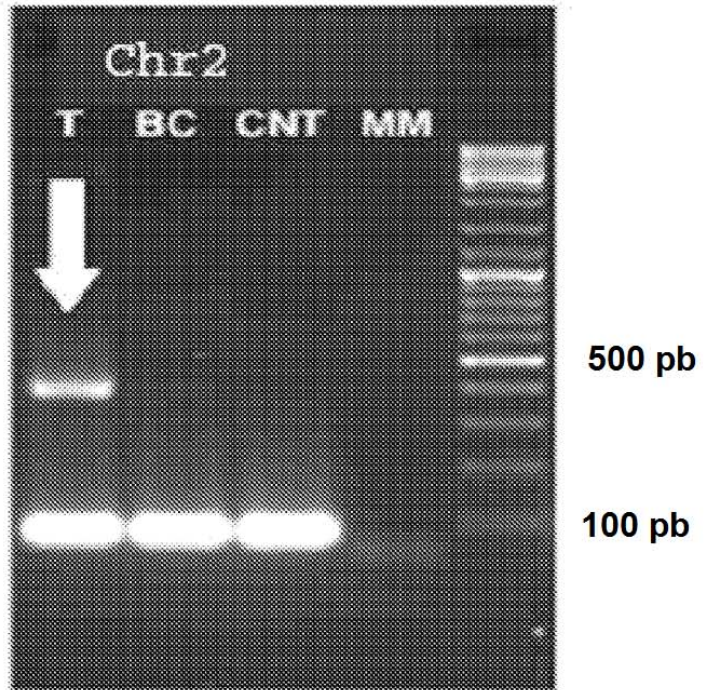


FIG. 3B

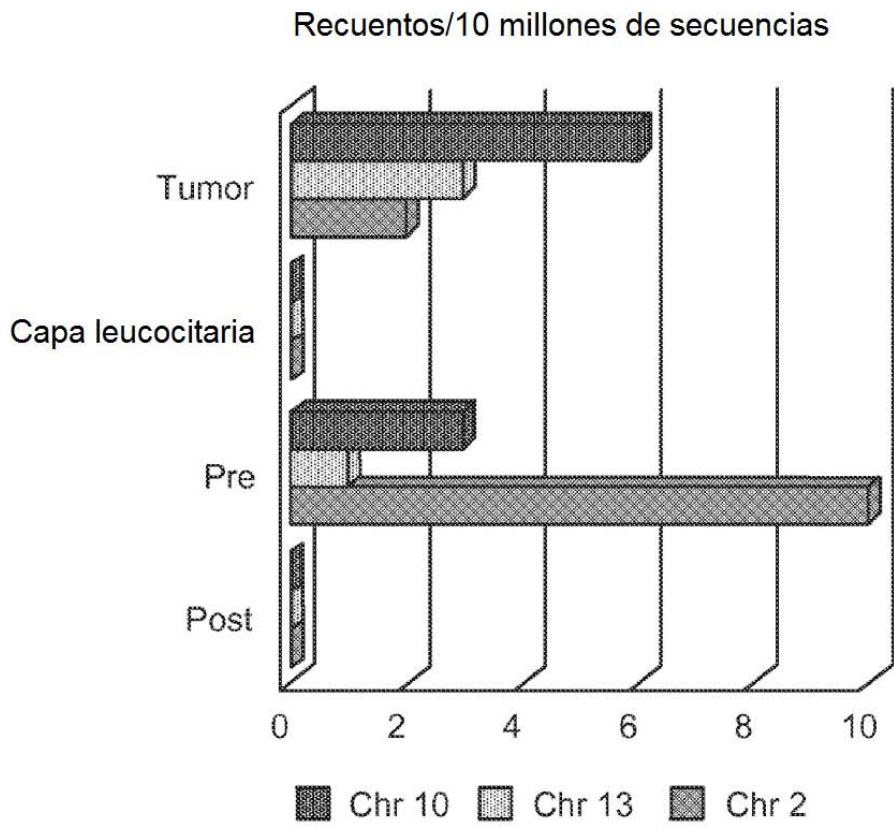
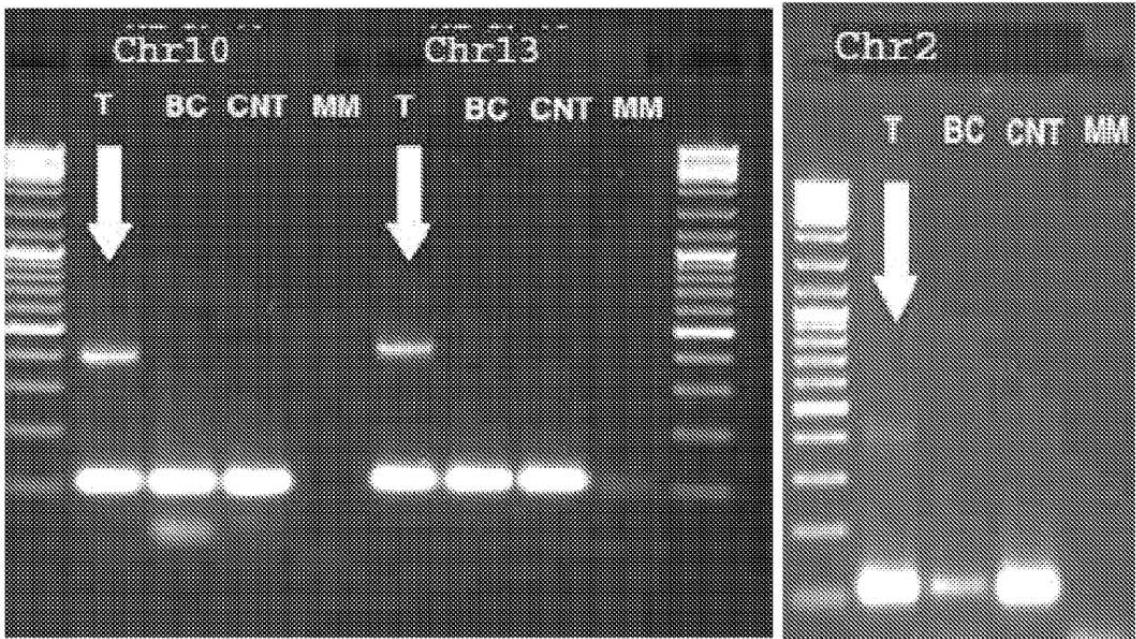


FIG. 3C

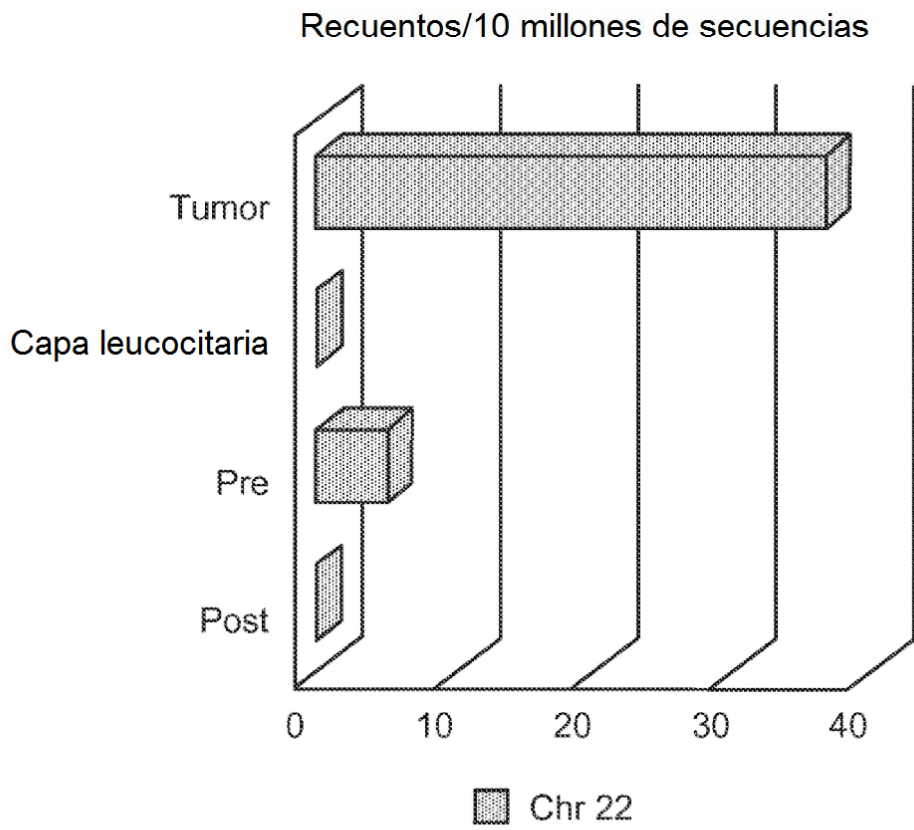
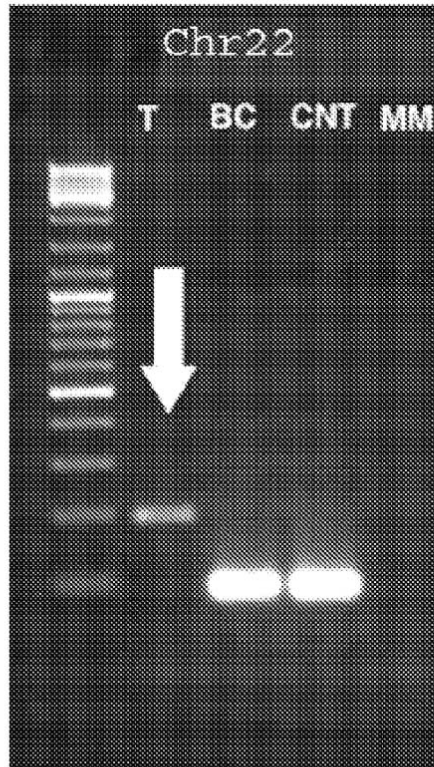
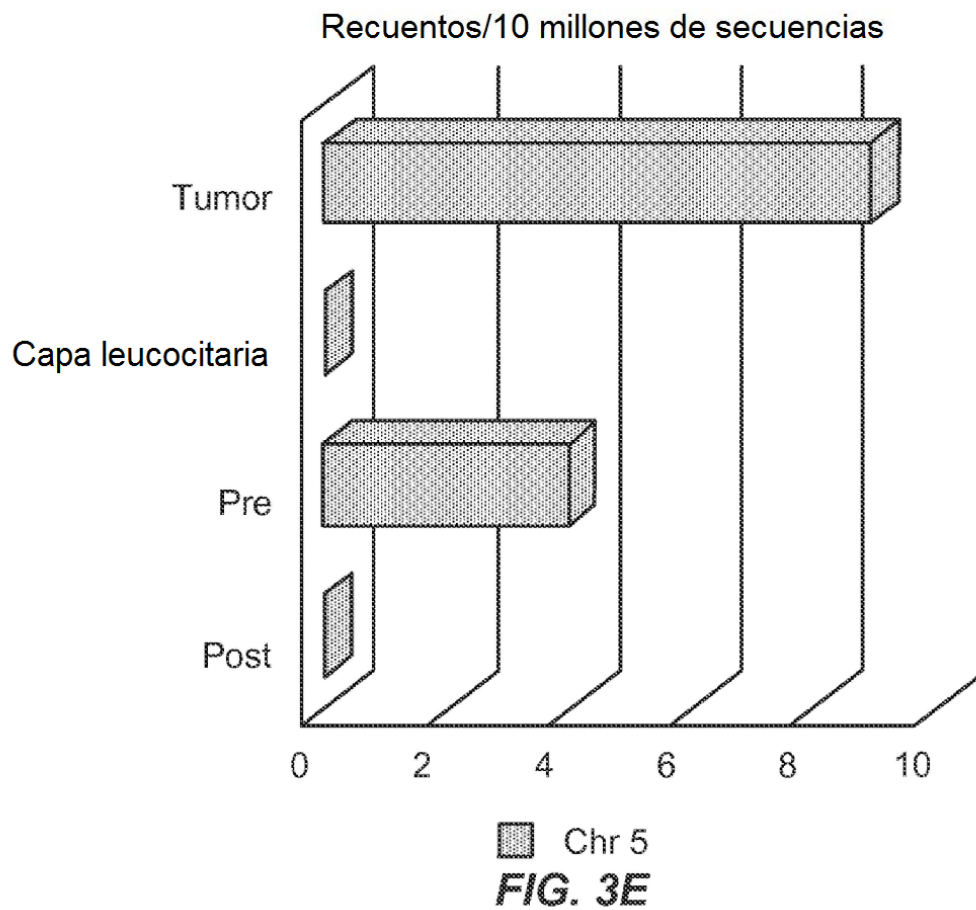
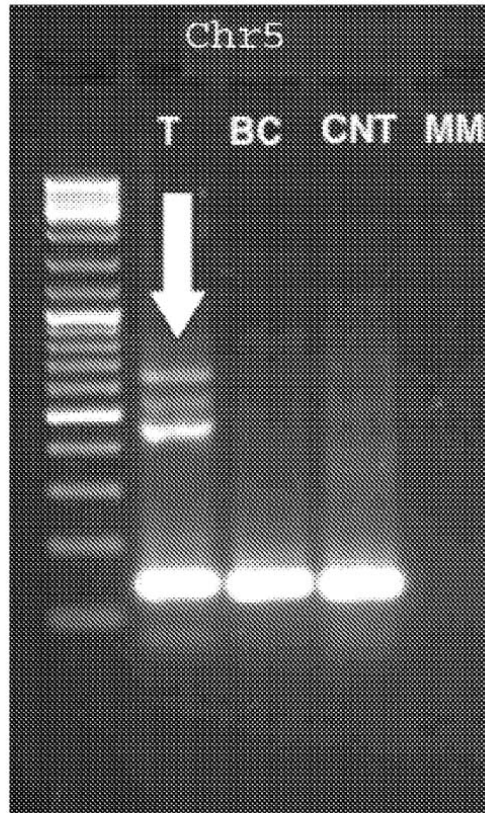


FIG. 3D



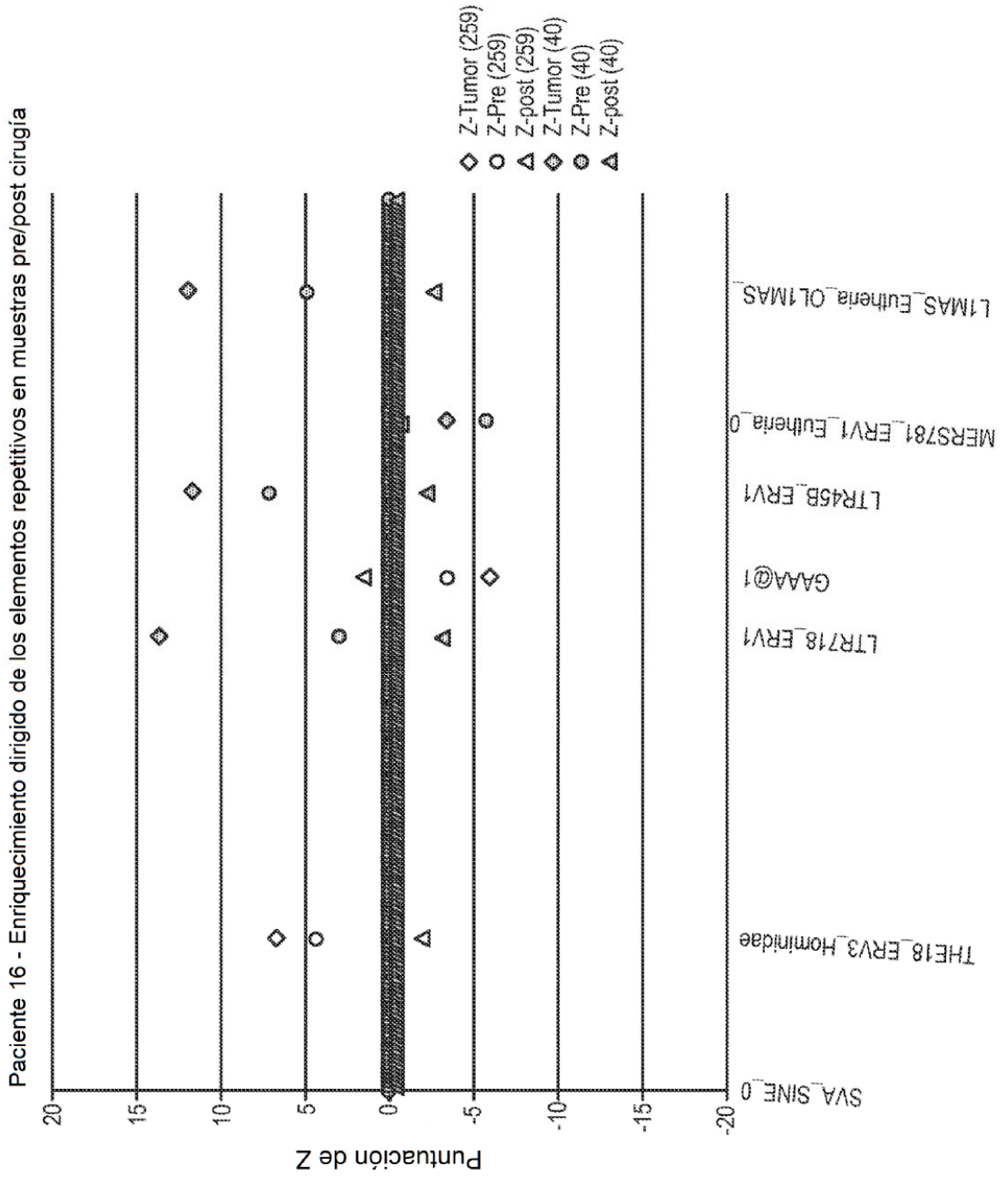


FIG. 4A

Paciente 16 - Enriquecimiento dirigido de los elementos repetitivos en muestras pre/post cirugía

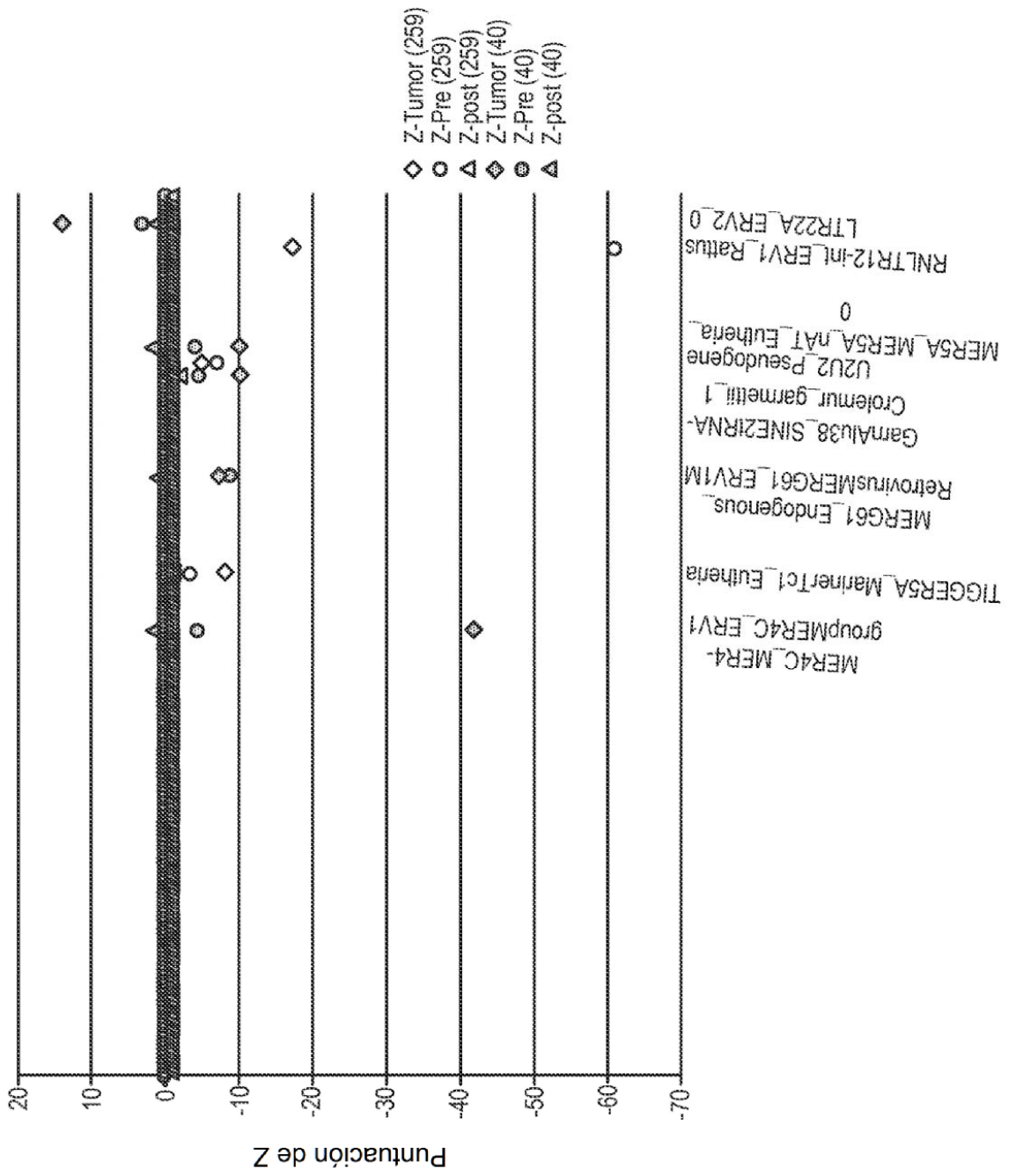


FIG. 4B

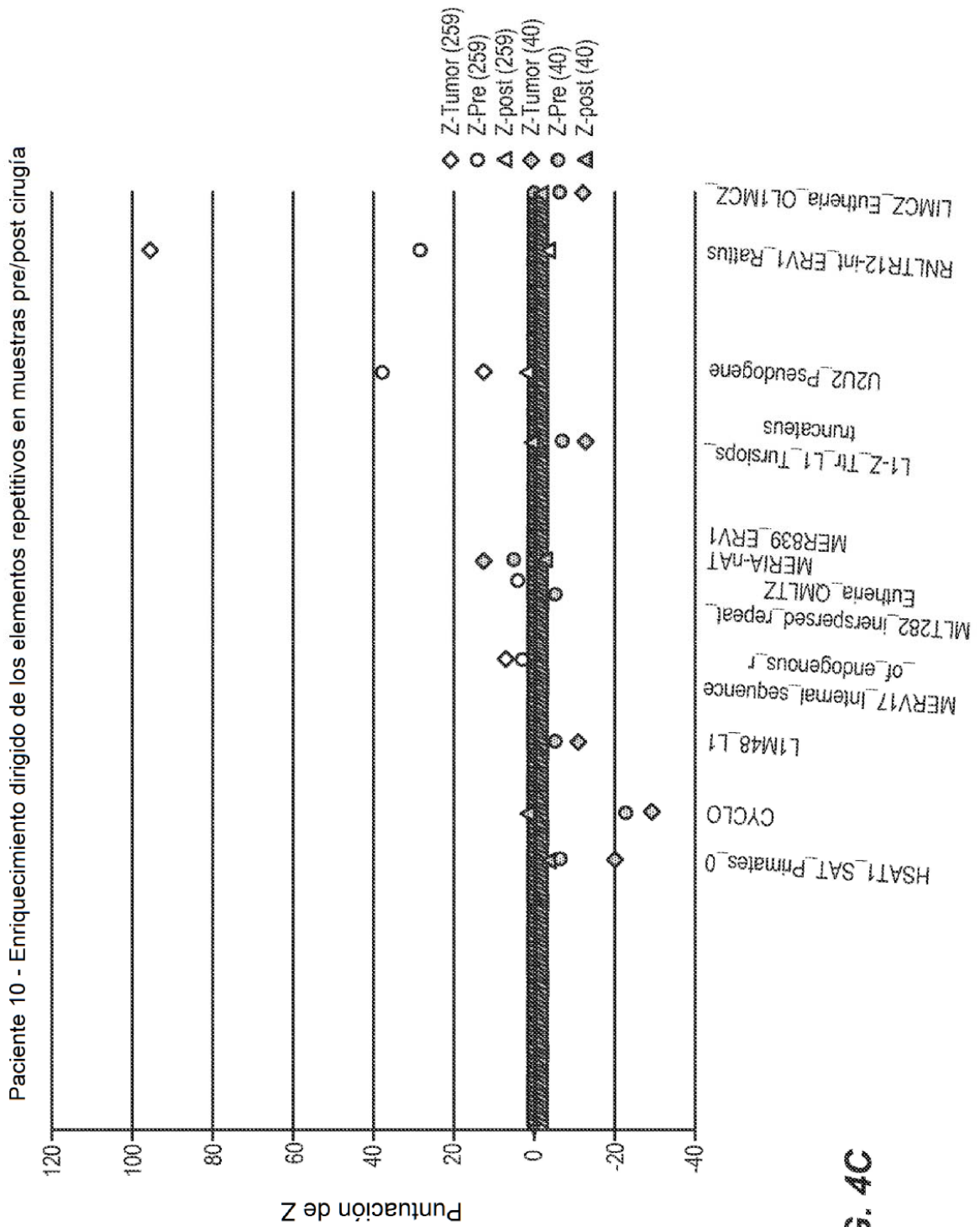


FIG. 4C

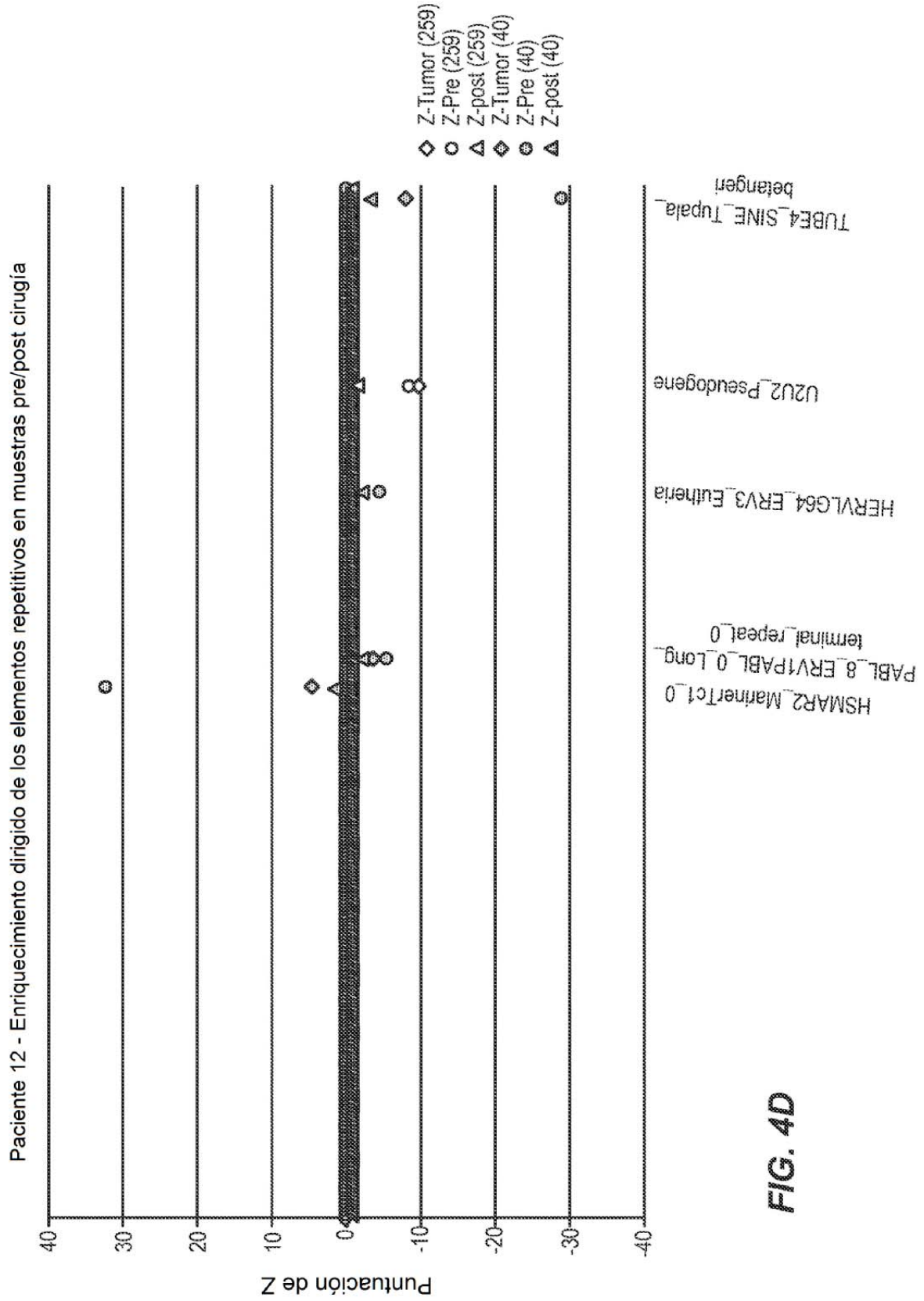


FIG. 4D

Paciente 8 - Enriquecimiento dirigido de los elementos repetitivos en muestras pre/post cirugía

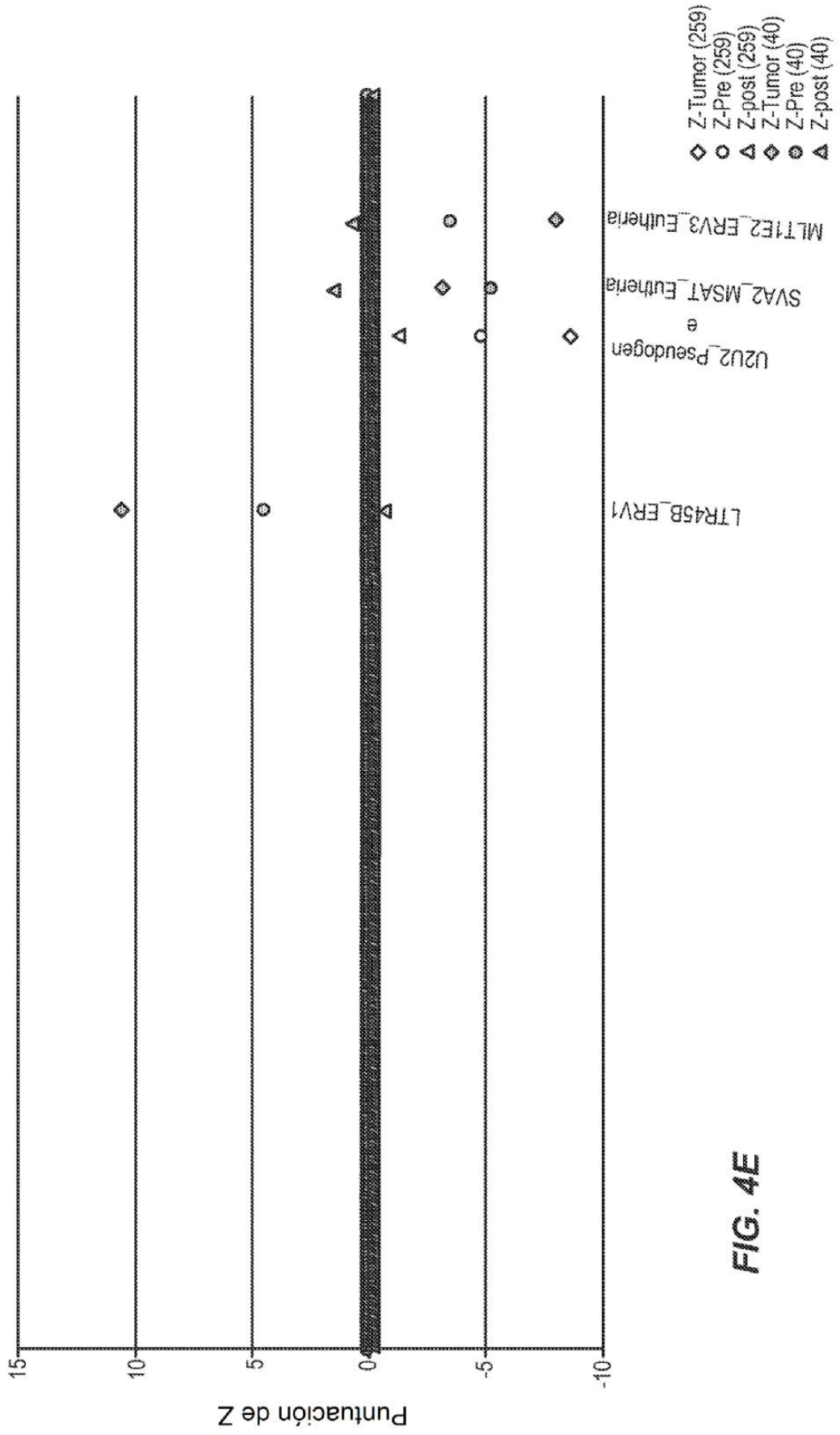


FIG. 4E

Sumario: Paciente 7

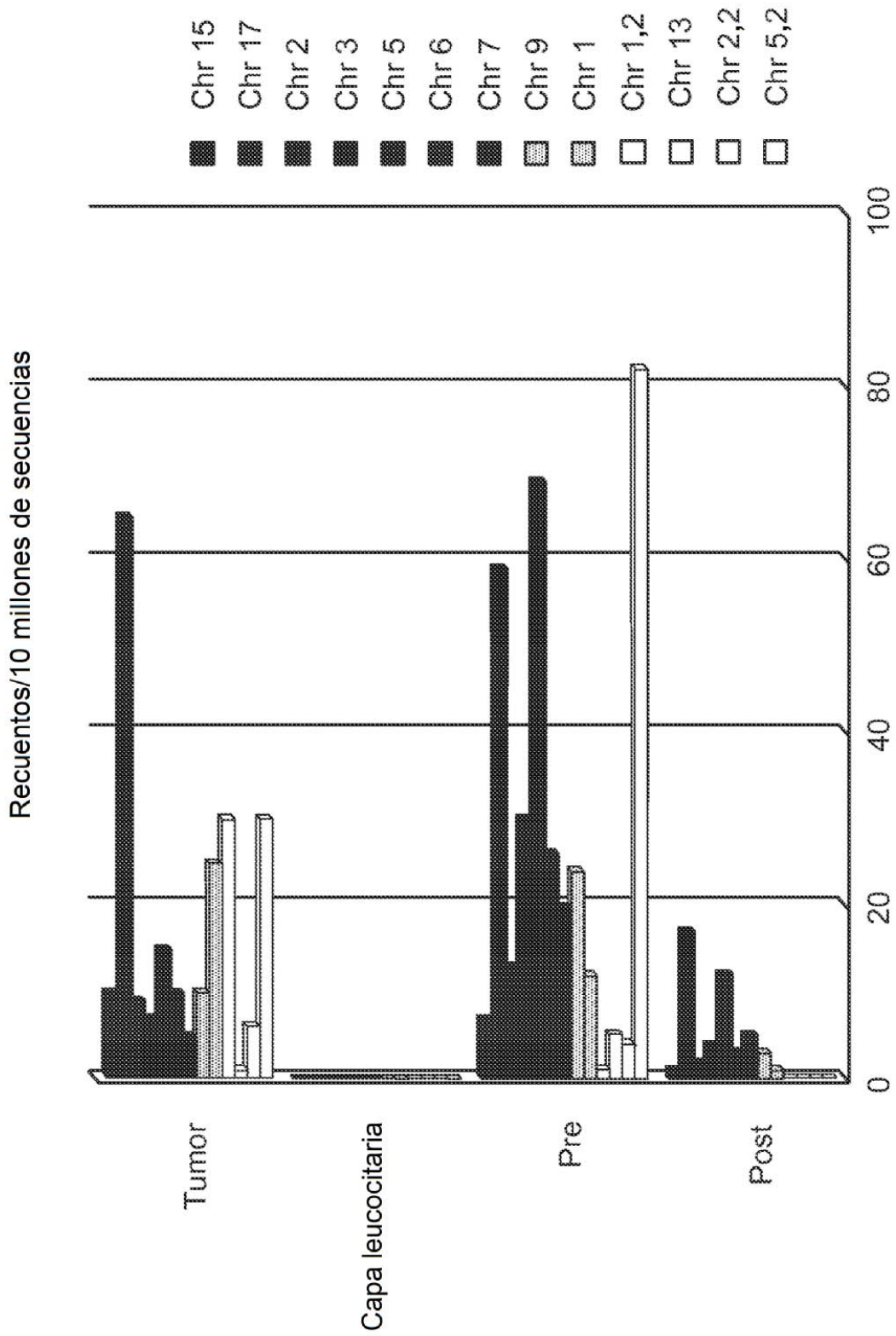


FIG. 5A

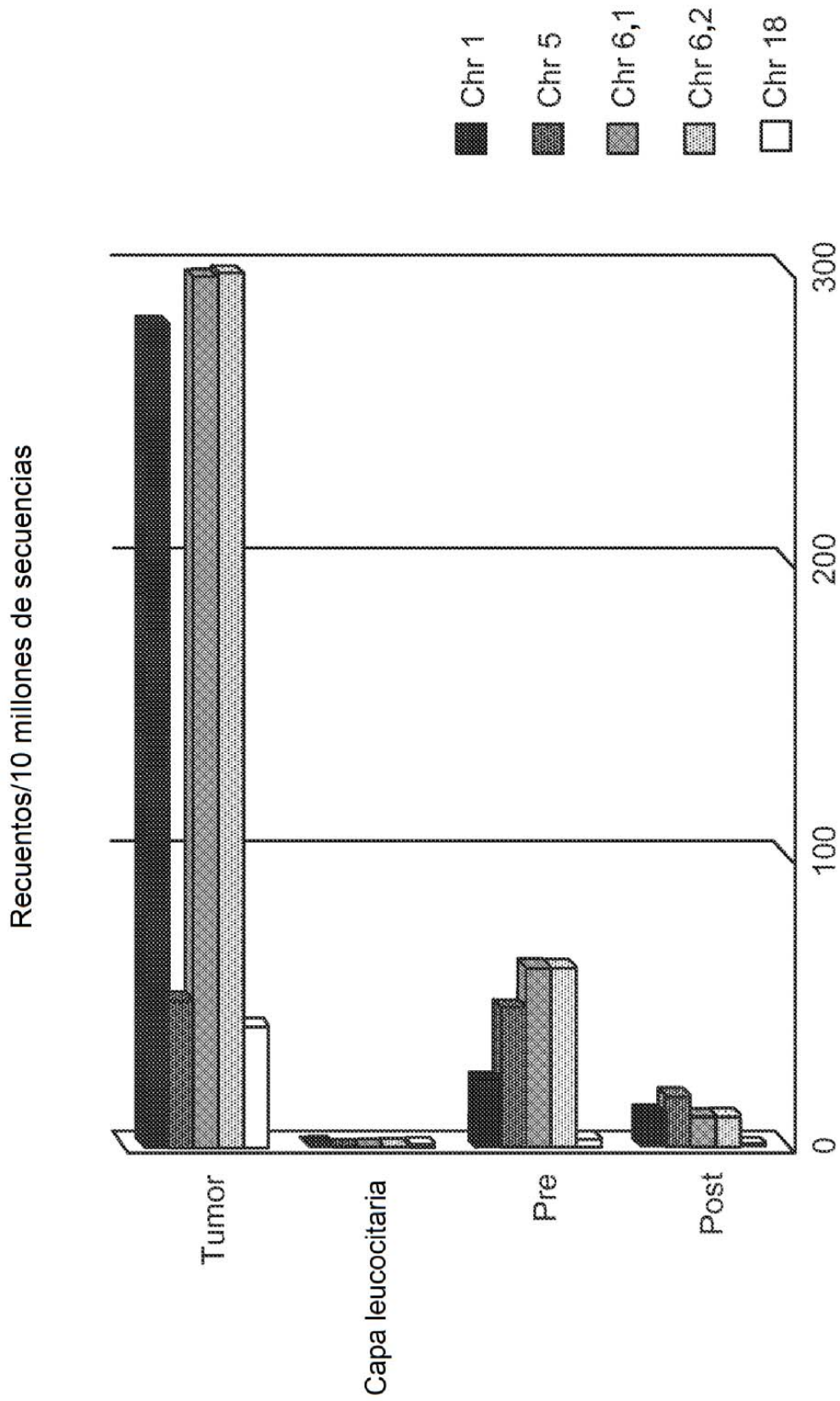


FIG. 5B

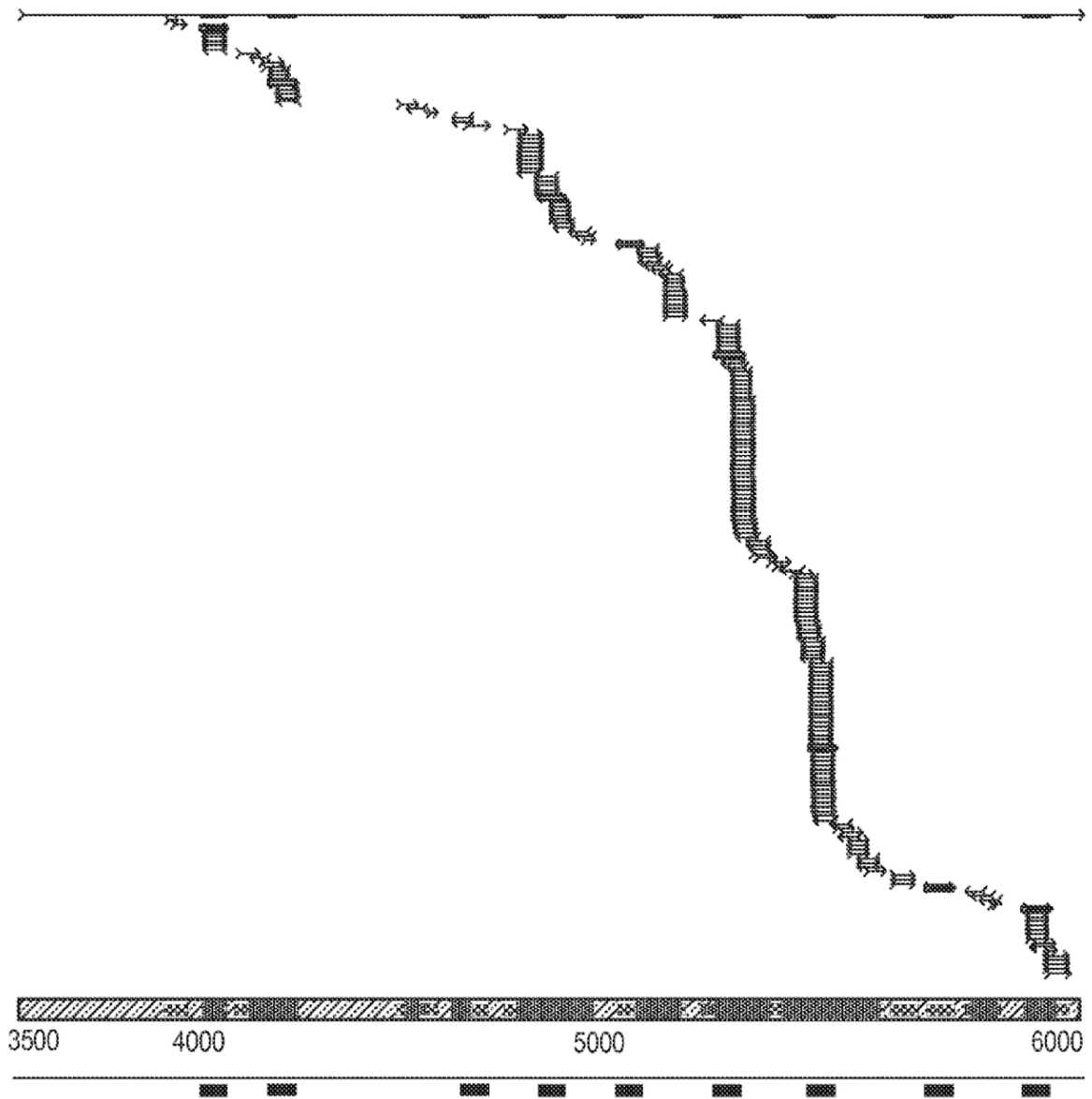


FIG. 6