

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 986 436**

51 Int. Cl.:

C12Q 1/6874 (2008.01)

C12Q 1/6806 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **12.04.2012 E 17154750 (8)**

97 Fecha y número de publicación de la concesión europea: **05.06.2024 EP 3246416**

54 Título: **Sistema de secuenciación segura**

30 Prioridad:

15.04.2011 US 201161476150 P

10.05.2011 US 201161484482 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
11.11.2024

73 Titular/es:

**THE JOHNS HOPKINS UNIVERSITY (100.0%)
3400 North Charles Street
Baltimore, MD 21218, US**

72 Inventor/es:

**VOGELSTEIN, BERT;
KINZLER, KENNETH;
PAPADOPOULOS, NICKOLAS y
KINDE, ISAAC**

74 Agente/Representante:

GONZÁLEZ PECES, Gustavo Adolfo

Observaciones:

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

ES 2 986 436 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

Sistema de secuenciación segura

5 **Campo técnico de la invención**

[0001] Esta invención se refiere al área de la secuenciación de ácidos nucleicos. En particular, se refiere a etapas manipulativas y analíticas para analizar y verificar los productos de eventos de baja frecuencia.

10 **Antecedentes de la invención**

[0002] Las mutaciones genéticas subyacen a muchos aspectos de la vida y la muerte, a través de la evolución y la enfermedad, respectivamente. En consecuencia, su medición es fundamental para varios campos de investigación. El clásico análisis de fluctuaciones de Luria y Delbrück es un ejemplo prototípico de la comprensión de los procesos biológicos que se puede obtener simplemente contando el número de mutaciones en experimentos controlados cuidadosamente (1). El recuento de mutaciones *de novo* en seres humanos, no presentes en sus padres, también ha llevado a nuevos conocimientos sobre el ritmo al que nuestra especie puede evolucionar (2, 3). De manera similar, contar los cambios genéticos o epigenéticos en los tumores puede informar sobre cuestiones fundamentales en la biología del cáncer (4). Las mutaciones son de una importancia fundamental en los problemas actuales de abordaje de pacientes con enfermedades víricas tales como el SIDA y la hepatitis a causa de la resistencia a los medicamentos que pueden provocar (5, 6). La detección de tales mutaciones, particularmente en una etapa anterior a que se vuelvan dominantes en la población, probablemente será esencial para optimizar el tratamiento. La detección del ADN del donante en la sangre de pacientes trasplantados de órganos es un indicador importante del rechazo del injerto y la detección del ADN fetal en el plasma materno puede utilizarse para el diagnóstico prenatal de forma no invasiva (7, 8). En las enfermedades neoplásicas, todas ellas provocadas por mutaciones somáticas, hay múltiples aplicaciones de la detección de mutantes raros; se pueden utilizar para ayudar a identificar la enfermedad residual en los márgenes quirúrgicos o en los ganglios linfáticos, para seguir el curso del tratamiento cuando se evalúa en plasma y tal vez para identificar pacientes con enfermedad precoz y curable quirúrgicamente cuando se evalúa en heces, esputo, plasma y otros fluidos corporales (9-11).

[0003] Estos ejemplos resaltan la importancia de identificar mutaciones raras para la investigación tanto básica como clínica. En consecuencia, a lo largo de los años se han ideado formas innovadoras de evaluarlas. Los primeros métodos implicaban ensayos biológicos basados en prototrofia, resistencia a infecciones víricas o fármacos, o ensayos bioquímicos (1, 12-18). La clonación y secuenciación molecular proporcionaron una nueva dimensión al campo, ya que permitieron identificar el tipo de mutación, en lugar de simplemente su presencia (19-24). Algunos de los métodos más novedosos y potentes se basan en la PCR digital, en la que se evalúan moléculas individuales una a una (25). La PCR digital es conceptualmente idéntica al análisis de clones individuales de bacterias, células o virus, pero se realiza completamente *in vitro* con reactivos inanimados definidos. Se han descrito varias implementaciones de la PCR digital, incluido el análisis de moléculas dispuestas en placas de múltiples pocillos, en colonias, en dispositivos de microfluidos y en emulsiones de agua en aceite (25-30). En cada una de estas tecnologías, las plantillas mutantes se identifican mediante su unión a oligonucleótidos específicos de la base potencialmente mutante.

[0004] La secuenciación masiva en paralelo representa una forma de PCR digital particularmente potente, ya que permite analizar cientos de millones de moléculas plantilla una a una. Tiene la ventaja frente a los métodos de PCR digital convencionales de que se pueden consultar múltiples bases de forma secuencial y sencilla de forma automatizada. Sin embargo, la secuenciación masiva en paralelo generalmente no se puede utilizar para detectar variantes raras debido a la alta tasa de error asociada con el proceso de secuenciación. Por ejemplo, con los instrumentos de secuenciación Illumina comúnmente utilizados, esta tasa de error varía de ~1 % (31,32) a ~0,05 % (33, 34), dependiendo de factores como la longitud de lectura (35), el uso de algoritmos de lectura de nucleótidos mejorados (36-38) y el tipo de variantes detectadas (39). Es de suponer que algunos de estos errores se deben a mutaciones introducidas durante la preparación de la plantilla, durante las etapas de preamplificación necesarias para la preparación de la biblioteca y durante la amplificación adicional en fase sólida en el propio instrumento. Otros errores se deben a una mala incorporación de bases durante la secuenciación y a errores en la lectura de nucleótidos. Los avances en las lecturas de nucleótidos pueden mejorar la confianza (p. ej., (36-39)), pero los errores basados en instrumentos siguen siendo limitantes, particularmente en muestras clínicas en donde la prevalencia de la mutación puede ser del 0,01 % o menor (11). En el trabajo que se describe a continuación, mostramos cómo se pueden preparar plantillas y cómo los datos de secuenciación obtenidos a partir de ellas se pueden interpretar de manera más fiable, de modo que se puedan identificar mutaciones relativamente raras con instrumentos disponibles en el mercado. El documento WO-A-2009/152928 describe un método para la determinación cuantitativa de ácidos nucleicos en una muestra, en particular para la determinación cuantitativa de transcripciones de genes tales como, por ejemplo, ARNm, ADNc, microARN, ARN no codificante, así como a la provisión de marcadores para la realización de los métodos analíticos. El documento WO-A-2010/127186 describe construcciones de oligonucleótidos, conjuntos de tales construcciones de oligonucleótidos y métodos para utilizar tales construcciones de oligonucleótidos para proporcionar secuencias validadas o conjuntos de secuencias validadas correspondientes a regiones de interés deseadas ("ROI"). Dichas ROI validadas y construcciones que las

contienen tienen una amplia variedad de usos, incluso en biología sintética, análisis cuantitativo de ácidos nucleicos, detección de polimorfismos y/o mutaciones, y similares.

[0005] Existe una necesidad continua en la técnica de mejorar la sensibilidad y precisión de las determinaciones de secuencias con fines de investigación, clínicos, forenses y genealógicos.

Resumen de la invención

[0006] La invención se expone en las reivindicaciones adjuntas. Las realizaciones de la descripción que no se encuentran dentro del alcance de dichas reivindicaciones se proporcionan únicamente con fines ilustrativos y no forman parte de la presente invención. Según una parte de la divulgación, un método analiza secuencias de ácidos nucleicos. Una secuencia identificadora única (UID) de ácido nucleico se une a un primer extremo de cada uno de una pluralidad de fragmentos de ácido nucleico de analito para formar fragmentos de ácido nucleico de analito identificados de forma única. La secuencia de nucleótidos de un fragmento de ácido nucleico analito identificado de forma única se determina de forma redundante, en donde determinadas secuencias de nucleótidos que comparten una UID forman una familia de miembros. Se identifica que una secuencia de nucleótidos representa con precisión un fragmento de ácido nucleico analito cuando al menos el 1 % de los miembros de la familia contienen la secuencia (proporcionada únicamente con fines ilustrativos y no cubierta por las reivindicaciones).

[0007] Según otra parte de la divulgación, un método analiza secuencias de ácidos nucleicos. Se une una secuencia identificadora única (UID) a un primer extremo de cada uno de una pluralidad de fragmentos de ADN analito usando al menos dos ciclos de amplificación con primer y segundo cebadores para formar fragmentos de ADN analito identificados de forma única. Las UID exceden los fragmentos de ADN analito durante la amplificación. Los primeros cebadores comprenden un primer segmento complementario a un amplicón deseado; un segundo segmento que contiene la UID; y un tercer segmento que contiene un sitio de cebado universal para su posterior amplificación. Los segundos cebadores comprenden un sitio de cebado universal para la amplificación posterior. Cada ciclo de amplificación une un sitio de cebado universal a una cadena. Los fragmentos de ADN analito identificados de forma única se amplifican para formar una familia de fragmentos de ADN analito identificados de forma única a partir de cada fragmento de ADN analito identificado de forma única. Se determinan las secuencias de nucleótidos de varios miembros de la familia.

[0008] La invención proporciona un método para analizar si una secuencia de ADN amplificada representa con exactitud un fragmento de ADN analito que comprende una mutación utilizando secuencias identificadoras únicas (UID) endógenas, según se establece en la reivindicación 1. Se obtiene ADN analito fragmentado que comprende fragmentos de 30 a 2000 bases, ambas inclusive. Cada extremo de un fragmento forma una UID endógena para el fragmento. Los oligonucleótidos adaptadores se unen a los extremos de los fragmentos para formar fragmentos adaptados. Los fragmentos que representan uno o más genes seleccionados se enriquecen opcionalmente mediante la captura de un subconjunto de los fragmentos usando oligonucleótidos de captura complementarios a los genes seleccionados en el ADN analito o amplificando fragmentos complementarios a los genes seleccionados. Los fragmentos adaptados se amplifican usando cebadores complementarios a los oligonucleótidos adaptadores para formar familias de fragmentos adaptados. La secuencia de nucleótidos se determina de una pluralidad de miembros de una familia. Se comparan las secuencias de nucleótidos de varios miembros de la familia. Una secuencia de nucleótidos se identifica como representación exacta de un fragmento de ADN analito que comprende dicha mutación cuando al menos un 95 % de los miembros de la familia comparten dicha mutación.

[0009] Otra parte más de la divulgación (proporcionada únicamente con fines ilustrativos y no cubierta por las reivindicaciones) es una composición que comprende una población de pares de cebadores, en donde cada par comprende un primer y un segundo cebador para amplificar e identificar un gen o una porción de gen. El primer cebador comprende una primera porción de 10 a 100 nucleótidos complementarios al gen o porción del gen y una segunda porción de 10 a 100 nucleótidos que comprende un sitio para la hibridación con un tercer cebador. El segundo cebador comprende una primera porción de 10 a 100 nucleótidos complementarios al gen o porción del gen y una segunda porción de 10 a 100 nucleótidos que comprende un sitio para la hibridación con un cuarto cebador. Interpuesta entre la primera porción y la segunda porción del segundo cebador hay una tercera porción que consiste en 2 a 4000 nucleótidos que forman una identificación única (UID). Las identificaciones únicas en la población tienen al menos 4 secuencias diferentes. El primer y segundo cebador son complementarios de cadenas opuestas del gen o porción del gen. Un kit puede comprender la población de cebadores y el tercer y cuarto cebador complementarios a las segundas porciones de cada uno del primer y segundo cebador.

[0010] La presente divulgación proporciona a la técnica herramientas y métodos para determinar de manera sensible y precisa características o secuencias de ácidos nucleicos.

Breve descripción de los dibujos

[0011]

Figura 1. **Elementos esenciales de Safe-SeqS.** En la primera etapa, se asigna a cada fragmento que se va a analizar una secuencia identificadora única (UID) (barras con rayado metálico o punteado). En la segunda etapa, los fragmentos marcados de forma única se amplifican, produciendo familias de UID, cada miembro de las cuales tiene la misma UID. Un supermutante se define como una familia de UID en la que ≥ 95 % de los miembros de la familia tienen la misma mutación.

Figura 2. **Safe-SeqS con UID endógenas más captura.** Las secuencias de los extremos de cada fragmento producido por cizallamiento aleatorio (barras sombreadas) sirven como identificaciones únicas (UID). Estos fragmentos se unen a adaptadores (barras con rayado tipo tierra y barras con rayado cruzado) para que posteriormente puedan amplificarse mediante PCR. Se produce un fragmento identificable de forma única a partir de cada cadena de la plantilla bicatenaria; solo se muestra una cadena. Los fragmentos de interés se capturan en una fase sólida que contiene oligonucleótidos complementarios a las secuencias de interés. Después de la amplificación por PCR para producir familias de UID con cebadores que contienen secuencias de “injerto” 5' (barras rellenas adhesivas y ligeramente punteadas), se realiza la secuenciación y los supermutantes se definen como en la Figura 1.

Figura 3. **Safe-SeqS con UID exógenas.** El ADN (sometido a cizallamiento o no sometido a cizallamiento) se amplifica con un conjunto de cebadores específicos de genes. Uno de los cebadores tiene una secuencia de ADN aleatoria (p. ej., un conjunto de 14 N) que forma la secuencia identificadora única (UID; barras sombreadas de diversas formas), ubicadas en 5' con respecto a su secuencia específica del gen, y ambas tienen secuencias que permiten la amplificación universal en la siguiente etapa (barras con rayado tipo tierra y barras con rayado cruzado). Dos ciclos de asignación de UID producen dos fragmentos (cada uno con una UID diferente) de cada molécula plantilla bicatenaria, como se muestra. La PCR posterior con cebadores universales, que también contienen secuencias de “injerto” (barras rellenas adhesivas y ligeramente punteadas), produce familias de UID que se secuencian directamente. Los supermutantes se definen como en la leyenda de la Figura 1.

Figuras 4A-4B. **Sustituciones de una sola base identificadas mediante análisis convencional y Safe-SeqS.** La estrategia de UID exógenas representada en la Figura 3 se usó para producir fragmentos de PCR a partir del gen *CTNNA1* de tres individuos normales, no relacionados. Cada posición representa una de las 87 posibles sustituciones de una sola base (3 posibles sustituciones/base x 29 bases analizadas). Estos fragmentos se secuenciaron en un instrumento Illumina GA IIx y se analizaron de manera convencional (Figura 4A) o con Safe-SeqS (Figura 4B). Los resultados de Safe-SeqS se muestran en la misma escala que el análisis convencional para una comparación directa; el recuadro es una vista ampliada. Cabe señalar que es probable que la mayoría de las variantes identificadas mediante el análisis convencional representen errores de secuenciación, como lo indica su alta frecuencia en relación con Safe-SeqS y su consistencia entre muestras no relacionadas.

Figura 5. **Safe-SeqS con UID endógenas más PCR inversa.** La secuencia de los extremos de cada fragmento producido por cizallamiento aleatorio sirve como identificaciones únicas (UID; barras sombreadas de diversas formas). Estos fragmentos se unen a adaptadores (barras con rayado tipo tierra y barras con rayado cruzado) como en una preparación de biblioteca de Illumina estándar. Se produce un fragmento marcado de forma única a partir de cada cadena de la plantilla bicatenaria; solo se muestra una cadena. Después de la circularización con una ligasa, se realiza una PCR inversa con cebadores específicos de genes que también contienen secuencias de “injerto” 5' (barras rellenas adhesivas y ligeramente punteadas). Esta PCR produce familias de UID que se secuencian directamente. Los supermutantes se definen como en la Figura 1.

Figura 6A-6B. **Posición de sustituciones de una sola base frente a la frecuencia de error en oligonucleótidos sintetizados con fosforamiditas y Phusion.** Safe-SeqS analizó una porción representativa del mismo fragmento de ADN de 31 bases sintetizado con fosforamiditas (Figura 6A) o Phusion polimerasa (Figura 6B). Se trazan las medias y las desviaciones estándar de siete experimentos independientes de cada tipo. Hubo un promedio de 1721 ± 383 y 196 ± 143 supermutantes de SBS identificados en los fragmentos sintetizados con fosforamidita y generados por Phusion, respectivamente. El eje y indica la fracción de los errores totales en la posición indicada. Cabe señalar que los errores en el fragmento de ADN sintetizado con fosforamidita fueron coherentes entre las siete réplicas, como sería de esperar si los errores se introdujeran sistemáticamente durante la síntesis misma. Por el contrario, los errores en los fragmentos generados por Phusion parecían ser heterogéneos entre las muestras, como era de esperar de un proceso estocástico (Luria y Delbruck, Genetics 28: 491-511, 1943).

Figura 7. **Distribución de los miembros de la familia de UID.** Se utilizó la estrategia de UID exógenas representada en la Figura 3 para producir fragmentos de PCR de una región de *CTNNA1* de tres individuos normales, no relacionados (Tabla 2B); se muestra un ejemplo representativo de las familias de UID con ≤ 300 miembros (99 % del total de familias de UID) generadas a partir de un individuo. El eje y indica el número de familias de UID diferentes que contenían el número de miembros de la familia que se muestra en el eje x.

Descripción detallada de la invención

[0012] La invención se expone en las reivindicaciones adjuntas. Las realizaciones de la descripción que no se encuentran dentro del alcance de dichas reivindicaciones se proporcionan únicamente con fines ilustrativos y no forman parte de la presente invención. Los inventores han desarrollado un enfoque, llamado “Safe-SeqS” (Safe-Sequencing System, sistema de secuenciación segura). En una realización, implica dos etapas básicas (Figura 1). La primera es la asignación de una identificación única (UID) a cada molécula plantilla de ácido nucleico que se va a analizar. El segundo es la amplificación de cada plantilla marcada de forma única, de modo que se generen muchas moléculas hijas con la secuencia idéntica (definida como una familia de UID). Si una mutación preexistía en la molécula plantilla utilizada para la amplificación, esa mutación debería estar presente en una cierta proporción, o incluso en todas, las moléculas hijas que contengan esa UID (salvo errores posteriores de replicación o secuenciación). Una familia de UID en la que cada miembro de la familia (o una determinada proporción predeterminada) tiene una mutación idéntica se denomina “supermutante”. Las mutaciones que no se produzcan en las plantillas originales, tales como las que ocurren durante las etapas de amplificación o por errores en la lectura de nucleótidos, no deberían dar lugar a supermutantes, es decir, no estarán presentes con la frecuencia predeterminada en una familia de UID.

[0013] El enfoque se puede emplear para cualquier propósito en donde se requiera un nivel muy alto de precisión y sensibilidad a partir de datos de secuencia. Como se muestra a continuación, el enfoque puede utilizarse para evaluar la fidelidad de una polimerasa, la precisión de la síntesis de ácidos nucleicos sintetizados *in vitro* y la prevalencia de mutaciones en ácidos nucleicos nucleares o mitocondriales de células normales. El enfoque puede usarse para detectar y/o cuantificar mosaicismo y mutaciones somáticas.

[0014] Los fragmentos de ácidos nucleicos se pueden obtener usando una técnica de formación de fragmentos aleatorios tal como cizallamiento mecánico, tratamiento de ultrasonidos o sometimiento de ácidos nucleicos a otras tensiones físicas o químicas. Es posible que los fragmentos no sean estrictamente aleatorios, ya que algunos sitios pueden ser más susceptibles a las tensiones que otros. También se pueden usar endonucleasas que fragmentan aleatoria o específicamente para generar fragmentos. El tamaño de los fragmentos puede variar, pero deseablemente estará en intervalos entre 30 y 5000 pares de bases, entre 100 y 2000, entre 150 y 1000, o dentro de intervalos con diferentes combinaciones de estos valores de acotación. Los ácidos nucleicos pueden ser, por ejemplo, ARN o ADN. También se pueden usar formas modificadas de ARN o ADN.

[0015] La unión de una UID exógena a un fragmento de ácido nucleico del analito que no caracteriza la materia objeto de las reivindicaciones se puede realizar mediante cualquier medio conocido en la técnica, incluidos enzimáticos, químicos o biológicos. Un medio emplea una reacción en cadena de la polimerasa. Otro medio emplea una enzima ligasa. La enzima puede ser, por ejemplo, de mamífero o bacteriana. Los extremos de los fragmentos se pueden reparar antes de unirlos utilizando otras enzimas como el fragmento Klenow de la ADN polimerasa T4. Otras enzimas que pueden usarse para la unión son otras enzimas polimerasas. Se puede agregar una UID a uno o ambos extremos de los fragmentos. Una UID puede estar contenida dentro de una molécula de ácido nucleico que contenga otras regiones para otra funcionalidad prevista. Por ejemplo, se puede agregar un sitio de cebado universal para permitir una amplificación posterior. Otro sitio adicional puede ser una región de complementariedad con una región o gen particular en los ácidos nucleicos del analito. Una UID puede tener una longitud de 2 a 4000, de 100 a 1000, de 4 a 400 bases, por ejemplo.

[0016] Las UID se pueden crear mediante la adición aleatoria de nucleótidos para formar una secuencia corta que se utilizará como identificación. En cada posición de adición, se puede utilizar una selección de uno de cuatro desoxirribonucleótidos. Como alternativa, se puede usar una selección de uno de tres, dos o un desoxirribonucleótidos. Por tanto, la UID puede ser completamente aleatoria, algo aleatoria o no aleatoria en determinadas posiciones. Otra forma de crear UID utiliza nucleótidos predeterminados ensamblados en un chip. En esta manera de fabricación, la complejidad se alcanza de manera planificada. Puede resultar ventajoso adjuntar una UID a cada extremo de un fragmento, lo que aumenta la complejidad de la población de UID en los fragmentos.

[0017] Un ciclo de reacción en cadena de la polimerasa para agregar UID exógenas se refiere a la desnaturalización térmica de una molécula bicatenaria, la hibridación de un primer cebador con una cadena simple resultante, la extensión del cebador para formar una nueva segunda cadena hibridada con la cadena simple original. Un segundo ciclo se refiere a la desnaturalización de la nueva segunda cadena a partir de la cadena única original, la hibridación de un segundo cebador con la nueva segunda cadena y la extensión del segundo cebador para formar una nueva tercera cadena, hibridada con la nueva segunda cadena. Es posible que se requieran múltiples ciclos para aumentar la eficacia, por ejemplo, cuando el analito esté diluido o haya inhibidores presentes.

[0018] En el caso de las UID endógenas, que caracterizan la materia objeto de las reivindicaciones, se pueden añadir adaptadores a los extremos de los fragmentos mediante unión. Se puede reducir la complejidad de los fragmentos de analito mediante una etapa de captura, ya sea en una fase sólida o en una etapa en fase líquida. Normalmente, la etapa de captura empleará hibridación con sondas que representan un gen o conjunto de genes de interés. Si se encuentran en una fase sólida, los fragmentos que no se unen se separan de los fragmentos que se unen. Las fases sólidas adecuadas conocidas en la técnica incluyen filtros, membranas, perlas, columnas, etc. Si

está en una fase líquida, se puede añadir un reactivo de captura que se una a las sondas, por ejemplo, mediante una interacción de tipo biotina-avidina. Después de la captura, los fragmentos deseados se pueden eludir para su posterior procesamiento. El orden de adición de adaptadores y de la captura es irrelevante. Otro medio de reducir la complejidad de los fragmentos de análisis implica la amplificación de uno o más genes o regiones específicas. Una forma de lograr esto es utilizar la PCR inversa. Se pueden utilizar cebadores que sean específicos de genes, logrando así un enriquecimiento mientras se forman bibliotecas. Opcionalmente, los cebadores específicos de genes pueden contener secuencias de injerto para su posterior unión a una plataforma de secuenciación masiva paralela.

[0019] Dado que las UID endógenas proporcionan un número limitado de posibilidades únicas, dependiendo del tamaño del fragmento y la longitud de lectura de la secuencia, se pueden usar combinaciones de UID endógenas y exógenas. La introducción de secuencias adicionales durante la amplificación aumentaría las UID disponibles y, por lo tanto, aumentaría la sensibilidad. Por ejemplo, antes de la amplificación, la plantilla se puede dividir en 96 pocillos y se podrían usar 96 cebadores diferentes durante la amplificación. Esto multiplicaría por 96 las UID disponibles, ya que podrían distinguirse hasta 96 plantillas con la misma UID endógena. Esta técnica también se puede utilizar con UID exógenas, de modo que los cebadores de cada pocillo agreguen una secuencia única y específica del pocillo a los productos de amplificación. Esto puede mejorar la especificidad de la detección de plantillas raras.

[0020] La amplificación de fragmentos que contienen una UID se puede realizar según técnicas conocidas para generar familias de fragmentos. Se puede utilizar la reacción en cadena de la polimerasa. También se pueden utilizar otros métodos de amplificación, según sea conveniente. Se puede utilizar la PCR inversa, al igual que la amplificación por círculo rodante. La amplificación de fragmentos normalmente se realiza utilizando cebadores que son complementarios a los sitios de cebado que están unidos a los fragmentos al mismo tiempo que las UID. Los sitios de cebado están distales a las UID, de modo que la amplificación incluye las UID. La amplificación forma una familia de fragmentos, cada miembro de la familia comparte la misma UID. Debido a que la diversidad de UID excede con creces la diversidad de los fragmentos, cada familia debe derivar de una única molécula de fragmento en el análisis. Los cebadores utilizados para la amplificación pueden modificarse químicamente para hacerlos más resistentes a las exonucleasas. Una de dichas modificaciones es el uso de enlaces fosforotioato entre uno o más nucleótidos 3'. Otra emplea boranofosfatos.

[0021] Los miembros de la familia se secuencian y comparan para identificar cualquier divergencia dentro de una familia. La secuenciación se realiza preferiblemente en una plataforma de secuenciación masiva paralela, muchas de las cuales están disponibles en el mercado. Si la plataforma de secuenciación requiere una secuencia para “injetar”, es decir, unir al dispositivo de secuenciación, dicha secuencia se puede añadir durante la adición de las UID o adaptadores, o por separado. Una secuencia de injerto puede ser parte de un cebador de UID, un cebador universal, un cebador específico de una diana genética, los cebadores de amplificación utilizados para formar una familia o separados. La secuenciación redundante se refiere a la secuenciación de una pluralidad de miembros de una sola familia.

[0022] Se puede establecer un umbral para identificar una mutación en un análisis. Si la “mutación” aparece en todos los miembros de una familia, entonces deriva del análisis. Si aparece en menos de todos los miembros, es posible que se haya introducido durante el análisis. Los umbrales de lectura de una mutación se pueden establecer, por ejemplo, en el 1 %, 5 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 97 %, 98 % o 100 %, en donde, según la presente invención, como se define en las reivindicaciones, el método de la reivindicación 1 se refiere a un umbral de al menos el 95 % como se define en el texto reivindicado “que identifica una secuencia de nucleótidos que representa con precisión dicho fragmento de ADN análisis que comprende dicha mutación, cuando al menos el 95 % de los miembros de la familia comparten dicha mutación”.

[0023] Se pueden usar poblaciones de pares de cebadores (proporcionadas únicamente con fines ilustrativos y no cubiertas por las reivindicaciones) para unir UID exógenas. El primer cebador comprende una primera porción de 10 a 100 nucleótidos complementarios al gen o porción del gen y una segunda porción de 10 a 100 nucleótidos que comprende un sitio para la hibridación con un tercer cebador. El segundo cebador comprende una primera porción de 10 a 100 nucleótidos complementarios al gen o porción del gen y una segunda porción de 10 a 100 nucleótidos que comprende un sitio para la hibridación con un cuarto cebador. Interpuesta entre la primera porción y la segunda porción del segundo cebador hay una tercera porción que consiste en 2 a 4000 nucleótidos que forman una identificación única (UID). Las identificaciones únicas en la población tienen al menos 4, al menos 16, al menos 64, al menos 256, al menos 1024, al menos 4096, al menos 16.384, al menos 65.536, al menos 262.144, al menos 1.048.576, al menos 4.194.304, al menos 16.777.216 o al menos 67.108.864 secuencias diferentes. El primer y segundo cebador son complementarios de cadenas opuestas del gen o porción del gen. Se puede fabricar un kit (proporcionado únicamente con fines ilustrativos y no cubierto por las reivindicaciones) que contenga tanto los cebadores para unir UID exógenas como los cebadores de amplificación, es decir, el tercer y cuarto cebador complementario a las segundas porciones de cada uno del primer y segundo cebador. El tercer y cuarto cebador pueden contener opcionalmente secuencias de injerto o indexación adicionales. La UID puede comprender secuencias seleccionadas al azar, secuencias de nucleótidos predefinidas o tanto secuencias seleccionadas al azar como nucleótidos predefinidos. Si son ambas cosas, estas pueden unirse en bloques o intercalarse.

[0024] Los métodos de análisis se pueden utilizar tanto para cuantificar como para determinar una secuencia. Por ejemplo, se puede comparar la abundancia relativa de dos fragmentos de ADN analito.

[0025] Los resultados que se describen a continuación en los ejemplos demuestran que el enfoque Safe-SeqS puede mejorar sustancialmente la precisión de la secuenciación masiva paralela (Tablas 1 y 2). Puede implementarse a través de UID endógenas (según la invención reivindicada como se define en las reivindicaciones) o introducidas de forma exógena, y puede aplicarse a prácticamente cualquier flujo de trabajo de preparación de muestras o plataforma de secuenciación. Como se demuestra aquí, el enfoque puede utilizarse fácilmente para identificar mutantes raros en una población de plantillas de ADN, para medir las tasas de error de la polimerasa y para juzgar la fiabilidad de las síntesis de oligonucleótidos. Una de las ventajas de la estrategia es que produce el número de plantillas analizadas, así como la fracción de plantillas que contienen variantes de bases. Los métodos *in vitro* descritos anteriormente para la detección de pequeñas cantidades de moléculas plantilla (p. ej., (29, 50)) permiten determinar la fracción de plantillas mutantes, pero no pueden determinar el número de plantillas mutantes y normales en la muestra original.

[0026] Es interesante comparar Safe-SeqS con otros enfoques para reducir errores en la secuenciación de próxima generación. Como se ha mencionado anteriormente, en el contexto de la invención, se han desarrollado algoritmos sofisticados para aumentar la precisión de la lectura de nucleótidos (p. ej., (36-39)). Sin duda, estos pueden reducir las lecturas de falsos positivos, pero su sensibilidad aún está limitada por las mutaciones artefactos que ocurren durante las etapas de la PCR necesarias para la preparación de la biblioteca, así como por (un número reducido de) errores de lectura de nucleótidos. Por ejemplo, el algoritmo empleado en el estudio actual utilizó criterios muy estrictos para las lecturas de nucleótidos y se aplicó a longitudes de lectura cortas, pero aún así no pudo reducir la tasa de error a menos de un promedio de $2,0 \times 10^{-4}$ errores/pb. Esta frecuencia de error es al menos tan baja como la reportada con otros algoritmos. Para mejorar aún más la sensibilidad, estas mejoras de lecturas de nucleótidos se pueden utilizar junto con Safe-SeqS. Travers y col. han descrito otra potente estrategia para reducir los errores (51). Con esta tecnología, ambas cadenas de cada molécula plantilla se secuencian de forma redundante después de una serie de etapas enzimáticas de preparación. Sin embargo, este enfoque solo se puede realizar en un instrumento específico. Además, para muchas aplicaciones clínicas, hay relativamente pocas moléculas plantilla en la muestra inicial y se requiere la evaluación de casi todas ellas para obtener la sensibilidad requerida. El enfoque descrito en el presente documento con UID introducidas exógenamente (Figura 3) cumple este requisito al acoplar el paso de asignación de UID con una amplificación posterior en la que se pierden pocas moléculas. Nuestros enfoques de UID endógenas (Figura 2 y Figura 5) y el descrito por Travers y col. no son idóneos para este fin debido a las inevitables pérdidas de moléculas plantilla durante la unión y otras etapas de preparación.

[0027] ¿Cómo sabemos que las mutaciones identificadas mediante los análisis convencionales en el estudio actual representan artefactos en lugar de verdaderas mutaciones en las plantillas originales? Una prueba contundente que apoya esta afirmación viene dada por la observación de que la prevalencia de mutaciones en todos los experimentos excepto uno fue similar: $2,0 \times 10^{-4}$ a $2,4 \times 10^{-4}$ mutaciones/pb (Tablas 1 y 2). La excepción fue el experimento con oligonucleótidos sintetizados a partir de fosoramiditas, en el que el error del proceso sintético fue aparentemente mayor que la tasa de error del análisis convencional de Illumina cuando se utilizó con criterios estrictos de lectura de nucleótidos. Por el contrario, la prevalencia de mutaciones de Safe-SeqS varió mucho más, de 0,0 a $1,4 \times 10^{-5}$ mutaciones/pb, dependiendo de la plantilla y del experimento. Además, la prevalencia de mutaciones medida por Safe-SeqS en el experimento más controlado, en el que se midió la fidelidad de la polimerasa (Tabla 2A), fue casi idéntica a la predicha en experimentos anteriores en los que la fidelidad de la polimerasa se midió mediante ensayos biológicos. Nuestras mediciones de la prevalencia de mutaciones en el ADN de células normales coinciden con algunos datos experimentales previos. Sin embargo, las estimaciones de estas prevalencias varían ampliamente y pueden depender del tipo de célula y la secuencia analizadas (véase el texto S1). Por lo tanto, no podemos estar seguros de que las pocas mutaciones reveladas por Safe-SeqS representen errores que ocurrieron durante el proceso de secuenciación en lugar de verdaderas mutaciones presentes en las plantillas de ADN originales. Las posibles fuentes de error en el proceso Safe-SeqS se describen en el texto S1.

[0028] Otra posible aplicación de Safe-SeqS es la minimización de la contaminación por PCR, que supone un problema grave para los laboratorios clínicos. Con la asignación de UID endógenas o exógenas, las UID de las plantillas mutantes pueden compararse simplemente con las identificadas en experimentos anteriores; la probabilidad de que la misma mutación de dos muestras independientes tenga la misma UID en diferentes experimentos es insignificante cuando las mutaciones son poco frecuentes. Además, con UID exógenas, un experimento de control con la misma plantilla, pero sin que la UID asigne ciclos de PCR (Figura 3) puede garantizar que no haya contaminación de ADN en esa preparación de plantilla; no se debe amplificar ninguna plantilla en ausencia de ciclos de asignación de UID y, por lo tanto, no se debe observar ningún producto de PCR del tamaño adecuado.

[0029] Como todas las técnicas, Safe-SeqS tiene limitaciones. Por ejemplo, hemos demostrado que la estrategia de UID exógenas se puede utilizar para analizar un único amplicón en profundidad. Es posible que esta tecnología no sea aplicable a situaciones en donde se deben analizar múltiples amplicones a partir de una muestra que contiene un número limitado de plantillas. La multiplexación en los ciclos de asignación de UID (Figura 3) puede proporcionar una

solución a este problema. Una segunda limitación es que la eficacia de la amplificación en los ciclos de asignación de UID es crítica para el éxito del método. Las muestras clínicas pueden contener inhibidores que reducen la eficacia de esta etapa. Presumiblemente, este problema se puede superar realizando más de dos ciclos en la etapa de PCR de asignación de UID (Figura 3), aunque esto complicaría la determinación del número de plantillas analizadas. La especificidad de Safe-SeqS está actualmente limitada por la fidelidad de la polimerasa utilizada en la etapa de PCR de asignación de UID; es decir, $8,8 \times 10^{-6}$ mutaciones/pb en su implementación actual con dos ciclos. Aumentar el número de ciclos en la etapa de PCR de asignación de UID a cinco disminuiría la especificidad general a $\sim 2 \times 10^{-6}$ mutaciones/pb. Sin embargo, esta especificidad puede incrementarse al requerir más de un supermutante para la identificación de la mutación; la probabilidad de introducir la misma mutación artificial dos o tres veces sería extremadamente baja ($[2 \times 10^{-6}]^2$ o $[2 \times 10^{-6}]^3$, respectivamente). En resumen, existen varias formas sencillas de realizar variaciones de Safe-SeqS y variaciones de análisis para satisfacer las necesidades de experimentos específicos.

[0030] Luria y Delbrück, en su artículo clásico de 1943, escribieron que su “predicción no puede verificarse directamente, porque lo que observamos, cuando contamos el número de bacterias resistentes en un cultivo, no es el número de mutaciones que han ocurrido sino el número de bacterias resistentes que han surgido por multiplicación de aquellas que mutaron, la cantidad de multiplicación depende de lo atrás que se produjo la mutación”. El procedimiento Safe-SeqS descrito en el presente documento puede verificar dichas predicciones porque es posible estimar el número y el tiempo de aparición de cada mutación a partir de los datos, como se observa en los experimentos sobre la fidelidad de la polimerasa. Además de las plantillas generadas por polimerasas *in vitro*, se puede aplicar el mismo enfoque al ADN de bacterias, virus y células de mamíferos. Por lo tanto, esperamos que esta estrategia proporcione respuestas definitivas a una variedad de preguntas biomédicas importantes.

[0031] Se puede comprender de manera más completa haciendo referencia a los siguientes ejemplos específicos.

Ejemplo 1: UID endógenas

[0032] Las UID, a veces denominadas códigos de barras o índices, se pueden asignar a fragmentos de ácido nucleico de muchas maneras. Estos incluyen la introducción de secuencias exógenas mediante PCR (40, 41) o unión (42, 43). Aún más simple, el ADN genómico cortado aleatoriamente contiene inherentemente UID que consisten en las secuencias de los dos extremos de cada fragmento sometido a cizallamiento (Figura 2 y Figura 5). La secuenciación de extremos emparejados de estos fragmentos produce familias de UID que pueden analizarse como se ha descrito anteriormente. Para emplear dichas UID endógenas en Safe-SeqS, utilizamos dos enfoques independientes: uno diseñado para evaluar muchos genes simultáneamente y el otro diseñado para evaluar un solo fragmento de gen en profundidad (Figura 2 y Figura 5, respectivamente).

[0033] Para la evaluación de múltiples genes, unimos adaptadores de secuenciación estándar de Illumina a los extremos de fragmentos de ADN sometido a cizallamiento para producir una biblioteca de secuenciación estándar y luego capturamos genes de interés en una fase sólida (44). En este experimento, se utilizó una biblioteca creada a partir del ADN de aproximadamente 15.000 células normales y se seleccionaron 2594 pb de seis genes para su captura. Tras excluir los polimorfismos de nucleótido único conocidos, también se identificaron 25.563 mutaciones aparentes, correspondientes a $2,4 \times 10^{-4} \pm$ mutaciones/pb (Tabla 1). Según análisis previos de frecuencias de mutación en células humanas, era probable que al menos el 90 % de estas mutaciones aparentes representarían mutaciones introducidas durante la preparación de plantillas y bibliotecas o errores de lectura de nucleótidos. Cabe señalar que la tasa de error determinada en el presente documento ($2,4 \times 10^{-4}$ mutaciones/pb) es considerablemente menor de lo que normalmente se informa en experimentos que utilizan el instrumento Illumina porque utilizamos criterios muy estrictos para la lectura de nucleótidos.

Tabla 1. Safe-SeqS con UID endógenas

Análisis convencional	Captura	PCR inversa
pb de alta calidad	106.958.863	1.041.346.645
Profundidad de lectura media de pb de alta calidad	38.620x	2.085.600x
Mutaciones identificadas	2,5.563	234.352
Mutaciones/pb	2,4E-04	2,3E-04
Análisis Safe-SeqS		
pb de alta calidad	106.958.863	1.041.346.645
Profundidad de lectura media de pb de alta calidad	38.620x	2.085.600x
Familias de UID	69.505	1057
N.º promedio de miembros/familia de UID	40	21.688
Mediana del n.º de miembros/familia de UID	19	4
Supermutantes identificados	8	0

Análisis convencional	Captura	PCR inversa
Supermutantes/pb	3,5E-06	0,0

[0034] Con el análisis Safe-SeqS de los mismos datos, determinamos que en este experimento se evaluaron 69.505 moléculas plantilla originales (es decir, se identificaron 69.505 familias de UID, con una media de 40 miembros por familia, Tabla 1). Todas las variantes polimórficas identificadas mediante análisis convencional también fueron identificadas por Safe-SeqS. Sin embargo, solo se observaron 8 supermutantes entre estas familias, correspondientes a $3,5 \times 10^{-6}$ mutaciones/pb. Por lo tanto, Safe-SeqS redujo los presuntos errores de secuenciación al menos 70 veces.

[0035] El análisis Safe-SeqS también puede determinar qué cadena de una plantilla está mutada, por lo que un criterio adicional para la lectura de mutaciones podría requerir que la mutación aparezca solo en una o en ambas cadenas de la plantilla originalmente bicatenaria. Los secuenciadores masivamente paralelos pueden obtener información de secuencia de ambos extremos de una plantilla en dos lecturas secuenciales. (Este tipo de experimento de secuenciación se denomina ejecución de “extremo emparejado” en la plataforma Illumina, pero se pueden realizar experimentos similares en otras plataformas de secuenciación donde puede que tengan otro nombre). Las dos cadenas de una plantilla bicatenaria se pueden diferenciar por la orientación observada de las secuencias y el orden en que aparecen cuando se obtiene información de la secuencia de ambos extremos. Por ejemplo, un par de cadenas de UID podría consistir en los dos grupos de secuencias siguientes cuando cada extremo de una plantilla se secuencia en lecturas secuenciales: 1) Una secuencia en orientación sentido que comienza en la posición 100 del cromosoma 2 en la primera lectura seguida de una secuencia en orientación antisentido que comienza en la posición 400 del cromosoma 2 en la segunda lectura; y 2) una secuencia en orientación antisentido que comienza en la posición 400 del cromosoma 2 en la primera lectura seguida de una secuencia en orientación sentido que comienza en la posición 100 del cromosoma 2 en la segunda lectura. En el experimento de captura descrito anteriormente, 42.222 de 69.505 UID (que representan 21.111 moléculas bicatenarias originales) en la región de interés representaban pares de cadenas de UID. Estas 42.222 UID abarcaban 1.417.838 bases en la región de interés. Cuando se permitió que una mutación solo ocurriera dentro de los pares de cadenas de UID (ya sea en una o ambas cadenas), se observaron dos supermutantes, lo que produjo una frecuencia de mutación de $1,4 \times 10^{-6}$ supermutantes/pb. Cuando se requirió que ocurriera una mutación en solo una cadena de un par de cadenas de UID, solo se observó un supermutante, lo que produjo una frecuencia de mutación de $7,1 \times 10^{-7}$ supermutantes/pb. Cuando se requirió que ocurriera una mutación en ambas cadenas de un par de cadenas de UID, solo se observó un supermutante, lo que produjo una frecuencia de mutación de $7,1 \times 10^{-7}$ supermutantes/pb. Por lo tanto, exigir que se produzcan mutaciones solo en una o en ambas cadenas de plantillas puede aumentar aún más la especificidad de Safe-SeqS.

[0036] También se utilizó una estrategia que emplea UID endógenas para reducir las mutaciones falsas positivas tras la secuenciación profunda de una única región de interés. En este caso, se utilizó una biblioteca preparada como se ha descrito anteriormente a partir de -1750 células normales como plantilla para la PCR inversa empleando cebadores complementarios a un gen de interés, por lo que los productos de la PCR pudieron utilizarse directamente para la secuenciación (Figura 5). Con el análisis convencional, se observó un promedio de $2,3 \times 10^{-4}$ mutaciones/pb, similar a la observada en el experimento de captura (Tabla 1). Dado que en este experimento solo se evaluaron 1057 moléculas independientes de células normales, según lo determinado mediante el análisis Safe-SeqS, todas las mutaciones observadas con el análisis convencional probablemente representaron falsos positivos (Tabla 1). Con el análisis Safe-SeqS de los mismos datos, no se identificaron supermutantes en ninguna posición.

Ejemplo 2: UID exógenas

[0037] Aunque los resultados descritos anteriormente muestran que Safe-SeqS puede aumentar la fiabilidad de la secuenciación masiva paralela, la cantidad de moléculas diferentes que pueden examinarse utilizando UID endógenas es limitada. Para fragmentos sometidos a cizallamiento a un tamaño promedio de 150 pb (intervalo 125-175), la secuenciación de extremos emparejados de 36 bases puede evaluar un máximo de ~7200 moléculas diferentes que contienen una mutación específica ($2 \text{ lecturas} \times 2 \text{ orientaciones} \times 36 \text{ bases/lectura} \times 50 \text{ bases de variación en cada extremo del fragmento}$). En la práctica, el número real de UID es menor porque el proceso de cizallamiento no es completamente aleatorio.

[0038] Para hacer un uso más eficiente de las plantillas originales, desarrollamos una estrategia Safe-SeqS que empleaba una cantidad mínima de etapas enzimáticas. Esta estrategia también permitió el uso de ADN degradado o dañado, como el que se encuentra en muestras clínicas o después del tratamiento con bisulfito, para el examen de la metilación de la citosina (45). Como se muestra en la Figura 3, esta estrategia emplea dos conjuntos de cebadores de PCR. El primer conjunto se sintetiza con precursores de fosforamida estándar y contiene secuencias complementarias al gen de interés en el extremo 3' y colas diferentes en los extremos 5' de los cebadores directo e inverso. Las diferentes colas permitieron la amplificación universal en la siguiente etapa. Finalmente, había un tramo de 12 a 14 nucleótidos aleatorios entre la cola y los nucleótidos específicos de secuencia en el cebador directo (40). Los nucleótidos aleatorios forman las UID. Una forma equivalente de asignar UID a fragmentos, que no se utiliza en este estudio, emplearía 10.000 cebadores directos y 10.000 cebadores inversos sintetizados en una micromatriz. Cada uno de estos 20.000 cebadores tendría cebadores específicos de genes en sus extremos 3' y una de las 10.000 secuencias UID específicas,

predeterminadas y no superpuestas en sus extremos 5', lo que permitiría 10^8 (es decir, $[10^4]^2$) posibles combinaciones de UID. En cualquier caso, se realizan dos ciclos de PCR con los cebadores y una polimerasa de alta fidelidad, produciendo un fragmento de ADN bicatenario con marcador único de cada una de las dos cadenas de cada molécula plantilla original (Figura 3). Los cebadores de asignación de UID residuales y no utilizados se eliminan mediante digestión con una exonucleasa específica monocatenaria, sin purificación adicional, y se añaden dos nuevos cebadores. Como alternativa o además de dicha digestión, se puede usar una columna de sílice que retenga selectivamente fragmentos de mayor tamaño o se pueden usar perlas de inmovilización reversible en fase sólida (SPRI) en condiciones que retengan selectivamente fragmentos de mayor tamaño para eliminar artefactos de amplificación más pequeños e inespecíficos. Esta purificación puede ayudar a reducir la acumulación de dímero de cebador en las etapas posteriores. Los nuevos cebadores, complementarios a las colas introducidas en los ciclos de asignación de UID, contienen secuencias de injerto en sus extremos 5', lo que permite la amplificación en fase sólida en el instrumento Illumina, y residuos de fosforotioato en sus extremos 3' para hacerlos resistentes a cualquier exonucleasa restante. Después de 25 ciclos adicionales de PCR, los productos se cargan en el instrumento Illumina. Como se muestra a continuación, esta estrategia nos permitió evaluar la mayoría de los fragmentos de entrada y se utilizó para varios experimentos ilustrativos.

Ejemplo 3: análisis de la fidelidad de la ADN polimerasa.

[0039] La medición de las tasas de error de las ADN polimerasas es esencial para su caracterización y dicta las situaciones en las que se pueden utilizar estas enzimas. Se eligió medir la tasa de error de la polimerasa Phusion, ya que esta polimerasa tiene una de las frecuencias de error más bajas de cualquier enzima disponible en el mercado y, por lo tanto, plantea un desafío particular para un enfoque *in vitro*. Primero amplificamos una única molécula plantilla de ADN humano, que comprende un segmento de un gen humano elegido arbitrariamente, a través de 19 series de PCR. Los productos de PCR de estas amplificaciones, en su totalidad, se utilizaron como plantillas para Safe-SeqS como se describe en la Figura 3. En siete experimentos independientes de este tipo, el número de familias de UID identificadas mediante secuenciación fue de 624.678 ± 421.274 , lo que coincide con una eficiencia de amplificación del $92 \pm 9,6 \%$ por serie de PCR.

[0040] El fabricante informa que la tasa de error de la polimerasa Phusion, estimada mediante la clonación de productos de PCR que codifican la β -galactosidasa en vectores plasmídicos y la transformación en bacterias, es $4,4 \times 10^{-7}$ errores/pb/ciclo de PCR. Incluso con lecturas de nucleótidos de muy alta rigurosidad, el análisis convencional de los datos de secuenciación de Illumina reveló una tasa de error aparente de $9,1 \times 10^{-6}$ errores/pb/ciclo de PCR, más de un orden de magnitud mayor que la tasa de error de la polimerasa Phusion informada (Tabla 2A). Por el contrario, el Safe-SeqS de los mismos datos reveló una tasa de error de $4,5 \times 10^{-7}$ errores/pb/ciclo de PCR, casi idéntica a la medida para la polimerasa Phusion en ensayos biológicos (Tabla 2A). La gran mayoría (>99 %) de estos errores fueron sustituciones de una sola base (Tabla 3A), lo que coincide con datos previos sobre los espectros de mutación creados por otras ADN polimerasas procariotas (15, 46, 47).

Tabla 2A-2C. Safe-SeqS con UID exógenas

2A. Fidelidad de la polimerasa		
Análisis convencional de 7 réplicas		
pb de alta calidad	996.855.791	64.030.757
Mutaciones totales identificadas	198.638	22.515
Mutaciones/pb	2,0E-04	1,7E-05
Tasa de error de Phusion calculada (errores/pb/ciclo)	9,1E-06	7,7E-07
Análisis Safe-SeqS de 7 réplicas		
pb de alta calidad	996.855.791	64.030.757
Familias de UID	624.678	421.274
Miembros/familia de UID	107	122
Total de supermutantes identificados	197	143
Supermutantes/pb	9,9E-06	2,3E-06
Tasa de error de Phusion calculada (errores/pb/ciclo)	4,5E-07	1,0E-07
2B. Mutaciones de CTNNB1 en el ADN de células humanas normales		
Análisis convencional de 3 individuos		
pb de alta calidad	559.334.774	66.600.749
Mutaciones totales identificadas	118.488	11.357
Mutaciones/pb	2,1E-04	1,6E-05
Análisis Safe-SeqS de 3 individuos		

2A. Fidelidad de la polimerasa		Media	Desviación estándar
Análisis convencional de 7 réplicas			
pb de alta calidad		559.334.774	66.600.749
Familias de UID		374.553	263.105
Miembros/familia de UID		68	38
Total de supermutantes identificados		99	78
Supermutantes/pb		9,0E-06	3,1E-06
2C. Mutaciones mitocondriales en el ADN células humanas normales			
Análisis convencional de 7 individuos			
pb de alta calidad		147.673.456	54.308.546
Mutaciones totales identificadas		30.599	12.970
Mutaciones/pb		2,1E-04	9,4E-05
Análisis de 7 individuos con Safe-SeqS			
pb de alta calidad		147.673.456	54.308.546
Familias de UID		515.600	89.985
Miembros/familia de UID		15	6
Total de supermutantes identificados		135	61
Supermutantes/pb		1,4E-05	6,8E-06

Tabla 3A-C. Fracción de sustituciones, inserciones y deleciones de una sola base con UID exógenas

3A. Fidelidad de la polimerasa		Media	Desviación estándar
Análisis convencional de 7 réplicas			
Mutaciones totales identificadas		198.638	22.515
Fracción de mutaciones representadas por sustituciones de una sola base		99 %	0 %
Fracción de mutaciones representadas por deleciones		1 %	0 %
Fracción de mutaciones representadas por inserciones		0 %	0 %
Análisis Safe-SeqS de 7 réplicas			
Total de supermutantes identificados		197	143
Fracción de supermutantes representada por sustituciones de una sola base		99 %	2 %
Fracción de supermutantes representada por deleciones		1 %	2 %
Fracción de supermutantes representada por inserciones		0 %	0 %
3B. Mutaciones de CTNNB1 en el ADN de células humanas normales			
Análisis convencional de 3 individuos			
Mutaciones totales identificadas		118.488	11.357
Fracción de mutaciones representadas por sustituciones de una sola base		97 %	0 %
Fracción de mutaciones representadas por deleciones		3 %	0 %
Fracción de mutaciones representadas por inserciones		0 %	0 %
Análisis Safe-SeqS de 3 individuos			
Total de supermutantes identificados		99	78
Fracción de supermutantes representada por sustituciones de una sola base		100 %	1 %
Fracción de supermutantes representada por deleciones		0 %	1 %
Fracción de supermutantes representada por inserciones		0 %	0 %
3C. Mutaciones mitocondriales en el ADN de células humanas normales			
Análisis convencional de 7 individuos			
Mutaciones totales identificadas		30.599	12.970
Fracción de mutaciones representadas por sustituciones de una sola base		98 %	1 %
Fracción de mutaciones representadas por deleciones		2 %	1 %
Fracción de mutaciones representadas por inserciones		0 %	0 %
Análisis de 7 individuos con Safe-SeqS			
Total de supermutantes identificados		135	61
Fracción de supermutantes representada por sustituciones de una sola base		99 %	1 %
Fracción de supermutantes representada por deleciones		1 %	1 %
Fracción de supermutantes representada por inserciones		0 %	0 %

[0041] Safe-SeqS también permitió determinar el número total de eventos mutacionales distintos y una estimación del ciclo de PCR en el que ocurrió la mutación. En estos experimentos se realizaron 19 ciclos de PCR en pocillos que contenían una única molécula plantilla. Cuando ocurrió un error de polimerasa en el ciclo 19, solo se produjo un supermutante (a partir de la cadena que contenía la mutación). Si el error ocurrió en el ciclo 18, debería haber dos supermutantes (derivados de las cadenas mutantes producidas en el ciclo 19), etc. En consecuencia, el ciclo en el que ocurrió el error está relacionado con la cantidad de supermutantes que contienen ese error. Los datos de siete experimentos independientes demuestran un número relativamente uniforme de errores totales de polimerasa observados ($2,2 \pm 1,1 \times 10^{-6}$ mutaciones distintas/pb), en buena concordancia con el número esperado de observaciones de las simulaciones ($1,5 \pm 0,21 \times 10^{-6}$ mutaciones distintas/pb). Los datos también muestran un momento muy variable de aparición de errores de polimerasa entre experimentos (Tabla 4), como se predice a partir del análisis de fluctuación clásico (1). Este tipo de información es difícil de obtener mediante el análisis convencional de los mismos datos de secuenciación de próxima generación, en parte debido a la frecuencia de mutación aparente prohibitivamente alta mencionada anteriormente.

Tabla 4A-4G. Número observado y esperado de errores generados por la polimerasa Phusion

5	4A. Experimento 1	Observado	Esperado (media $\pm$ DE)*
	Mutaciones representadas por 1 supermutante	10	19 $\pm$ 3,7
	Mutaciones representadas por 2 supermutantes	8	5,8 $\pm$ 2,3
	Mutaciones representadas por 3 supermutantes	4	1,3 $\pm$ 1,1
10	Mutaciones representadas por 4 supermutantes	4	1,8 $\pm$ 1,3
	Mutaciones representadas por 5 supermutantes	2	0,61 $\pm$ 0,75
	Mutaciones representadas por 6 supermutantes	2	0,22 $\pm$ 0,44
	Mutaciones representadas por 7 supermutantes	0	0,01 $\pm$ 0,10
15	Mutaciones representadas por 8 supermutantes	0	0,87 $\pm$ 0,86
	Mutaciones representadas por 9 supermutantes	2	0,28 $\pm$ 0,51
	Mutaciones representadas por 10 supermutantes	0	0,14 $\pm$ 0,38
	Mutaciones representadas por >10 supermutantes	3	1,5 $\pm$ 2,7
20			
	Mutaciones distintas	35	32 $\pm$ 4,2
25	4B. Experimento 2		
	Mutaciones representadas por 1 supermutante	19	23 $\pm$ 4,1
	Mutaciones representadas por 2 supermutantes	5	9,5 $\pm$ 2,8
	Mutaciones representadas por 3 supermutantes	4	2,7 $\pm$ 1,6
30	Mutaciones representadas por 4 supermutantes	7	2,7 $\pm$ 1,7
	Mutaciones representadas por 5 supermutantes	2	0,88 $\pm$ 0,94
	Mutaciones representadas por 6 supermutantes	1	0,40 $\pm$ 0,60
	Mutaciones representadas por 7 supermutantes	3	0,16 $\pm$ 0,42
35	Mutaciones representadas por 8 supermutantes	1	0,99 $\pm$ 1,0
	Mutaciones representadas por 9 supermutantes	1	0,39 $\pm$ 0,68
	Mutaciones representadas por 10 supermutantes	0	0,17 $\pm$ 0,43
	Mutaciones representadas por >10 supermutantes	9	1,8 $\pm$ 3,4
40			
	Mutaciones distintas	52	43 $\pm$ 5,1
45	4C. Experimento 3		
	Mutaciones representadas por 1 supermutante	7	17 $\pm$ 3,4
	Mutaciones representadas por 2 supermutantes	9	5,4 $\pm$ 2,0
	Mutaciones representadas por 3 supermutantes	4	1,2 $\pm$ 1,1
50	Mutaciones representadas por 4 supermutantes	4	1,7 $\pm$ 1,4
	Mutaciones representadas por 5 supermutantes	2	0,50 $\pm$ 0,70
	Mutaciones representadas por 6 supermutantes	0	0,17 $\pm$ 0,45
	Mutaciones representadas por 7 supermutantes	1	0,03 $\pm$ 0,17
55	Mutaciones representadas por 8 supermutantes	0	0,59 $\pm$ 0,74
	Mutaciones representadas por 9 supermutantes	0	0,24 $\pm$ 0,50
	Mutaciones representadas por 10 supermutantes	1	0,07 $\pm$ 0,29
	Mutaciones representadas por >10 supermutantes	5	1,5 $\pm$ 2,6
60			
	Mutaciones distintas	33	28 $\pm$ 3,7
65	4D. Experimento 4		
	Mutaciones representadas por 1 supermutante	7	15 $\pm$ 3,7
	Mutaciones representadas por 2 supermutantes	8	4,1 $\pm$ 1,7
	Mutaciones representadas por 3 supermutantes	2	0,70 $\pm$ 0,74
65	Mutaciones representadas por 4 supermutantes	1	1,5 $\pm$ 1,3
	Mutaciones representadas por 5 supermutantes	3	0,21 $\pm$ 0,52

	4A. Experimento 1	Observado	Esperado (media $\pm$ DE)*
	Mutaciones representadas por 6 supermutantes	2	0,08 $\pm$ 0,27
	Mutaciones representadas por 7 supermutantes	1	0,0 $\pm$ 0,0
5	Mutaciones representadas por 8 supermutantes	2	0,65 $\pm$ 0,77
	Mutaciones representadas por 9 supermutantes	2	0,17 $\pm$ 0,43
	Mutaciones representadas por 10 supermutantes	0	0,05 $\pm$ 0,22
	Mutaciones representadas por >10 supermutantes	1	0,92 $\pm$ 2,1
10			
	Mutaciones distintas	29	23 $\pm$ 3,2
	4E. Experimento 5		
	Mutaciones representadas por 1 supermutante	9	23 $\pm$ 4,1
15	Mutaciones representadas por 2 supermutantes	6	9,5 $\pm$ 2,8
	Mutaciones representadas por 3 supermutantes	5	2,7 $\pm$ 1,6
	Mutaciones representadas por 4 supermutantes	3	2,7 $\pm$ 1,7
	Mutaciones representadas por 5 supermutantes	6	0,88 $\pm$ 0,94
20	Mutaciones representadas por 6 supermutantes	2	0,40 $\pm$ 0,60
	Mutaciones representadas por 7 supermutantes	1	0,16 $\pm$ 0,42
	Mutaciones representadas por 8 supermutantes	2	0,99 $\pm$ 1,0
	Mutaciones representadas por 9 supermutantes	2	0,39 $\pm$ 0,68
25	Mutaciones representadas por 10 supermutantes	3	0,17 $\pm$ 0,43
	Mutaciones representadas por >10 supermutantes	7	1,8 $\pm$ 3,4
	Mutaciones distintas	46	43 $\pm$ 5,1
30			
	4F. Experimento 6		
	Mutaciones representadas por 1 supermutante	4	6,7 $\pm$ 2,8
	Mutaciones representadas por 2 supermutantes	7	1,5 $\pm$ 1,2
35	Mutaciones representadas por 3 supermutantes	1	0,10 $\pm$ 0,33
	Mutaciones representadas por 4 supermutantes	2	0,60 $\pm$ 0,82
	Mutaciones representadas por 5 supermutantes	0	0,07 $\pm$ 0,26
	Mutaciones representadas por 6 supermutantes	0	0,01 $\pm$ 0,10
40	Mutaciones representadas por 7 supermutantes	1	0,0 $\pm$ 0,0
	Mutaciones representadas por 8 supermutantes	1	0,39 $\pm$ 0,60
	Mutaciones representadas por 9 supermutantes	0	0,01 $\pm$ 0,10
	Mutaciones representadas por 10 supermutantes	0	0,0 $\pm$ 0,0
45	Mutaciones representadas por >10 supermutantes	2	0,50 $\pm$ 1,1
	Mutaciones distintas	18	9,9 $\pm$ 1,4
50	4G. Experimento 7		
	Mutaciones representadas por 1 supermutante	8	2,9 $\pm$ 1,6
	Mutaciones representadas por 2 supermutantes	2	0,61 $\pm$ 0,79
	Mutaciones representadas por 3 supermutantes	0	0,04 $\pm$ 0,24
55	Mutaciones representadas por 4 supermutantes	0	0,41 $\pm$ 0,59
	Mutaciones representadas por 5 supermutantes	1	0,01 $\pm$ 0,10
	Mutaciones representadas por 6 supermutantes	0	0,0 $\pm$ 0,0
	Mutaciones representadas por 7 supermutantes	0	0,0 $\pm$ 0,0
	Mutaciones representadas por 8 supermutantes	0	0,14 $\pm$ 0,35
60	Mutaciones representadas por 9 supermutantes	0	0,01 $\pm$ 0,10
	Mutaciones representadas por 10 supermutantes	0	0,0 $\pm$ 0,0
	Mutaciones representadas por >10 supermutantes	0	0,32 $\pm$ 0,93
65	Mutaciones distintas	11	4,5 $\pm$ 0,62

4A. Experimento 1	Observado	Esperado (media $\pm$ DE)*
*Véase el texto SI para obtener detalles de las simulaciones		

Ejemplo 4: Análisis de la composición de oligonucleótidos.

[0042] Para la mayoría de las aplicaciones, tales como la PCR de rutina o la clonación, se tolera un pequeño número de errores durante la síntesis de oligonucleótidos a partir de precursores de fosforamidita. Sin embargo, para la biología sintética, en donde se deben unir muchos oligonucleótidos, estos errores representan un obstáculo importante para el éxito. Se han ideado estrategias inteligentes para hacer más eficiente el proceso de construcción de genes (48, 49), pero todas esas estrategias se beneficiarían de una síntesis más precisa de los propios oligonucleótidos. Determinar el número de errores en los oligonucleótidos sintetizados es difícil porque la fracción de oligonucleótidos que contienen errores puede ser menor que la sensibilidad de los análisis de secuenciación convencionales de próxima generación.

[0043] Para determinar si Safe-SeqS podría usarse para esta determinación, utilizamos química de fosforamidita estándar para sintetizar un oligonucleótido que contiene 31 bases que fueron diseñadas para ser idénticas a las analizadas en el experimento de fidelidad de la polimerasa descrito anteriormente. En el oligonucleótido sintético, las 31 bases estaban rodeadas por secuencias complementarias a los cebadores que podrían usarse para las etapas de asignación de UID de Safe-SeqS (Figura 3). Al realizar Safe-SeqS en ~ 300.000 oligonucleótidos, encontramos que había $8,9 \pm 0,28 \times 10^{-4}$ supermutantes/pb y que estos errores ocurrieron en todos los oligonucleótidos (Figura 6A). Los oligonucleótidos contenían una gran cantidad de errores de inserción y delección, representando el $8,2 \pm 0,63\%$ y $25 \pm 1,5\%$ del total de supermutantes, respectivamente. Es importante destacar que tanto la posición como la naturaleza de los errores fueron altamente reproducibles entre siete réplicas independientes de este experimento realizadas en el mismo lote de oligonucleótidos (Figura 6A). Esta naturaleza y distribución de errores tenían poco en común con la de los errores producidos por la polimerasa Phusion (Figura 6B y Tabla 5), que se distribuyeron en el patrón estocástico esperado entre experimentos replicados. El número de errores en los oligonucleótidos sintetizados con fosforamiditas fue aproximadamente 60 veces mayor que en los productos equivalentes sintetizados por la polimerasa Phusion. Estos datos, en conjunto, indican que la gran mayoría de los errores en las primeras se generaron durante su síntesis y no durante el procedimiento Safe-SeqS.

Tabla 5. ADN sintetizado con fosforamidita o Phusion: Comparación de transiciones y transversiones

Fosforamiditas	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7	Promedio	Desviación estándar
Supermutantes de transición:	496	509	471	396	323	273	470	420	92
Supermutantes de transversión:	1494	1499	1521	1154	944	907	1626	1306	298
valor de p	3,4E-05								
Phusion									
Supermutantes de transición:	63	275	127	5	87	182	103	120	87
Supermutantes de transversión:	14	124	77	12	57	191	63	77	63
valor de p	0,08								
*Los valores de p se calcularon utilizando una prueba de la t para datos emparejados bilateral									

[0044] ¿Safe-SeqS preserva la proporción de secuencias normales mutantes en las plantillas originales? Para abordar esta pregunta, sintetizamos dos oligonucleótidos de 31 bases de secuencia idéntica a excepción del nt 15 (C/G 50:50 en lugar de T) y los mezclamos en fracciones nominales mutantes/normales del 3,3 % y 0,33 %. Mediante el análisis Safe-SeqS de las mezclas de oligonucleótidos, encontramos que las proporciones eran del 2,8 % y el 0,27 %, respectivamente. Concluimos que los procedimientos de amplificación y asignación de UID utilizados en Safe-SeqS no alteran en gran medida la proporción de variantes de secuencias y, por lo tanto, proporcionan una estimación fiable de esa proporción cuando se desconoce. Esta conclusión también está avalada por la reproducibilidad de variantes de fracciones cuando se analizan en experimentos independientes de Safe-SeqS (Figura 6A).

Ejemplo 5 - Análisis de secuencias de ADN de células humanas normales.

[0045] A continuación, se utilizó la estrategia de UID exógenas (Figura 3) para determinar la prevalencia de mutaciones raras en una pequeña región del gen *CTNNB1* de ~ 100.000 células humanas normales de tres individuos no relacionados. Comparando con el número de familias de UID obtenidas en los experimentos Safe-SeqS (Tabla 2B), calculamos que la mayoría ($78 \pm 9,8\%$) de los fragmentos de entrada se convirtieron en familias

de UID. Había un promedio de 68 miembros/familia de UID, lo que cumplía fácilmente la redundancia requerida para Safe-SeqS (Figura 7). El análisis convencional de los datos de secuenciación de Illumina reveló un promedio de 118.488 ± 11.357 mutaciones entre los ~560 Mb de secuencia analizados por muestra, lo que corresponde a una prevalencia de mutación aparente de $2,1 \pm 0,16 \times 10^{-4}$ mutaciones/pb (Tabla 2B). Solo se observó un promedio de 99 ± 78 supermutantes en el análisis Safe-SeqS. La gran mayoría (>99 %) de los supermutantes fueron sustituciones de una sola base y la frecuencia de mutación calculada fue de $9,0 \pm 3,1 \times 10^{-6}$ mutaciones/pb (Tabla 3B). De este modo, Safe-SeqS redujo la frecuencia aparente de mutaciones en el ADN genómico al menos 24 veces (Figura 4).

[0046] Una posible estrategia para aumentar la especificidad de Safe-SeqS es realizar la amplificación de la biblioteca (y posiblemente los ciclos de asignación de UID) en varios pocillos. Esto se puede lograr en tan solo 2 o hasta 384 pocillos utilizando placas de PCR estándar, o ampliarse a muchos más pocillos cuando se utiliza un dispositivo de microfluidos (de miles a millones). Cuando se realiza de esta manera, se pueden introducir secuencias de indexación en las plantillas que son exclusivas de los pocillos en los que se amplifica la plantilla. Por lo tanto, las mutaciones raras deberían dar lugar a dos supermutantes (es decir, uno de cada cadena), ambos con la misma secuencia índice de pocillos. Al realizar Safe-SeqS con UID exógenas en las plantillas de *CTNNB1* descritas anteriormente y diluidas en 10 pocillos (cada pocillo produce plantillas amplificadas con una secuencia índice diferente), la frecuencia de mutación se redujo aún más de $9,0 \pm 3,1 \times 10^{-6}$ a $3,7 \pm 1,2 \times 10^{-6}$ supermutantes/pb. Por lo tanto, analizar plantillas en múltiples compartimentos (de una manera que produzca plantillas codificadas diferencialmente conforme al compartimento en el que se amplificaron las plantillas) puede ser una estrategia adicional para aumentar la especificidad de Safe-SeqS.

Ejemplo 6: Análisis de secuencias de ADN a partir de ADN mitocondrial

[0047] Aplicamos la misma estrategia a un segmento corto de ADN mitocondrial en ~1000 células de cada uno de siete individuos no relacionados. El análisis convencional de las bibliotecas de secuenciación de Illumina producidas con el procedimiento Safe-SeqS (Figura 3) reveló un promedio de 30.599 ± 12.970 mutaciones entre los ~150 Mb de secuencia analizados por muestra, correspondiente a una prevalencia de mutación aparente de $2,1 \pm 0,94 \times 10^{-4}$ mutaciones/pb (Tabla 2C). Solo se observaron 135 ± 61 supermutantes en el análisis Safe-SeqS. Al igual que con el gen *CTNNB1*, la gran mayoría de las mutaciones fueron sustituciones de una sola base, aunque también se observaron delecciones ocasionales de una sola base (Tabla 3C). La frecuencia de mutación calculada en el segmento analizado de ADNmt fue $1,4 \pm 0,68 \times 10^{-5}$ mutaciones/pb (Tabla 2C). Por lo tanto, Safe-SeqS redujo la frecuencia aparente de mutaciones en el ADN genómico al menos 15 veces.

Ejemplo 7: Materiales y métodos

[0048] UID endógenas. Se preparó ADN genómico de páncreas humano o células linfoblastoides cultivadas utilizando kits de Qiagen. El ADN del páncreas se utilizó para el experimento de captura y las células linfoblastoides se utilizaron para el experimento de PCR inversa. El ADN se cuantificó mediante absorbancia óptica y con qPCR. El ADN se fragmentó hasta un tamaño promedio de ~200 pb mediante cizallamiento acústico (Covaris) y, a continuación, se reparó el extremo, se le aplicó una cola A y se unió a adaptadores en forma de Y según los protocolos estándar de Illumina. Los extremos de cada molécula plantilla proporcionan UID endógenas correspondientes a sus posiciones cromosómicas. Después de la amplificación mediada por PCR de las bibliotecas con secuencias de cebadores dentro de los adaptadores, se capturó el ADN (1) con un filtro que contenía 2594 nt correspondientes a seis genes cancerosos. Después de la captura, se realizaron 18 ciclos de PCR para garantizar cantidades suficientes de plantilla para la secuenciación en un instrumento Illumina GA IIx.

[0049] Para los experimentos de PCR inversa (Figura 5), unimos adaptadores personalizados (IDT, Tabla 6) en lugar de adaptadores Illumina en forma de Y estándar al ADN celular sometido a cizallamiento. Estos adaptadores conservaban la región complementaria al cebador de secuenciación universal, pero carecían de las secuencias de injerto necesarias para la hibridación con la celda de flujo Illumina GA IIx. El ADN unido se diluyó en 96 pocillos y el ADN de cada columna de 8 pocillos se amplificó con un cebador directo único que contenía una de las 12 secuencias índice en su extremo 5' más un cebador inverso estándar (Tabla 6). Las amplificaciones se realizaron utilizando Phusion HotStart I (NEB) en reacciones de 50 µl que contenían tampón Phusion HF 1X, dNTP 0,5 mM, 0,5 µM de cada cebador directo e inverso (ambos 5'-fosforilados) y 1 U de polimerasa Phusion. Se utilizaron las siguientes condiciones de ciclado: un ciclo de 98 °C durante 30 s; y 16 ciclos de 98 °C durante 10 s, 65 °C durante 30 s y 72 °C durante 30 s. Las 96 reacciones se combinaron y luego se purificaron utilizando un kit de purificación por PCR Qiagen MinElute (n.º de catálogo 28004) y un kit de extracción en gel QIAquick (n.º de catálogo 28704). Para preparar las plantillas circulares necesarias para la PCR inversa, se diluyó el ADN a ~1 ng/µl y se unió con ADN Ligasa T4 (Enzymatics) durante 30 min a temperatura ambiente en una reacción de 600 µl que contenía 1X tampón de unión de ADN T4 y 18.000 U de ADN Ligasa T4. La reacción de unión se purificó usando un kit Qiagen MinElute. La PCR inversa se realizó utilizando Phusion Hot Start I en 90 ng de plantilla circular distribuida en doce reacciones de 50 µl, cada una de las cuales contenía tampón Phusion HF 1X, dNTP 0,25 mM, 0,5 µM de cada uno de los cebadores *KRAS* directo e inverso (Tabla 6) y 1 U de polimerasa Phusion. Ambos cebadores específicos de *KRAS* contenían secuencias de injerto para la hibridación con la celda de flujo Illumina GA IIx (Tabla 6). Se utilizaron las siguientes condiciones de ciclado: un ciclo de 98 °C durante 2 min; y 37 ciclos de 98 °C durante 10 s, 61 °C durante 15 s y 72 °C durante 10 s. La purificación final se realizó con un kit NucleoSpin Extract II (Macherey-Nagel) y se

eluyó en 20 ul de tampón NE. Los fragmentos de ADN resultantes contenían UID compuestas por tres secuencias: dos endógenas, representadas por los dos extremos de los fragmentos sometidos a cizallamiento originales más la secuencia exógena introducida durante la amplificación de indexación. Como se utilizaron 12 secuencias exógenas, esto aumentó 12 veces el número de UID distintas con respecto al obtenido sin UID exógenas. Este número podría aumentarse fácilmente utilizando un mayor número de cebadores distintos.

5

10

15

20

25

30

35

40

45

50

55

60

65

Tabla 6. Oligonucleótidos utilizados

5	Leyenda tipográfica:	Leyenda de símbolos:
	REGIÓN COMPLEMENTARIA A LAS PLANTILLAS SECUENCIA UID ESPECÍFICA DE LA PLANTILLA	/5Phos/ = Fosfato 5' * = Enlace fosforotioato
10	SECUENCIA UNIVERSAL SECUENCIA ÍNDICE ESPECÍFICA DEL EXPERIMENTO CEBADORES DE INJERTO ILLUMINA (PARA HIBRIDACIÓN A LA CELDA DE FLUJO)	
	UID endógenas	
15	Captura	Secuencia (SEQ ID NO: 1-81, respectivamente)
	Adaptador: cadena 1	/5Phos/GATCGGAAGAGCGGTTACAGCAGGAATGCCGAG
20	Adaptador: cadena 2	ACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
	Amplificación del genoma completo: directo	AATGATACGGCGACCACCGAGATCTACACA CACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
25	Amplificación del genoma completo: inverso	CAAGCAGAAGACGGCATACGAGAT CTCGGCATTCTGCTGAACCGCTCTTCCGAT*C*T
	Amplificación tras la captura: directo	AATGATACGGCGACCACCGAGATCTACACA CACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
30	Amplificación tras la captura: inverso	CAAGCAGAAGACGGCATACGAGAT CTCGGCATTCTGCTGAACCGCTCTTCCGAT*C*T
	Cebador de secuenciación, Lectura 1 (Illumina; San Diego, CA)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
35	Cebador de secuenciación, Lectura 2 (Illumina; San Diego, CA)	CTCGGCATTCTGCTGAACCGCTCTTCCGATCT
	PCR inversa	
40	Adaptador: cadena 1	/5Phos/GATCGGAAGAGCGGTTACAGCAGGAATGCCGAG
	Adaptador: cadena 2	ACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
45	Amplificación del genoma completo: directo-1	/5Phos/ CGTGAT ACACTCTTTCCCTACAACGACGCTCTTCCGAT*C*T
	Amplificación del genoma completo: directo-2	/5Phos/ ACATCG ACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
50	Amplificación del genoma completo: directo-3	/5Phos/ GCCTAA CACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
	Amplificación del genoma completo: directo-4	/5Phos/ TGGTCA CACTCTTTCCCTACAACGACGCTCTTCCGAT*C*T
55	Amplificación del genoma completo: directo-5	/5Phos/ CACGTG ACACTCTTTCCCTACAACGACGCTCTTCCGAT*C*T
	Amplificación del genoma completo: directo-6	/5Phos/ ATTGGC ACACTCTTTCCCTACAACGACGCTCTTCCGAT*C*T
60	Amplificación del genoma completo: directo-7	/5Phos/ GATCTG ACACTCTTTCCCTACAACGACGCTCTTCCGAT*C*T
	Amplificación del genoma completo: directo-8	/5Phos/ TCAAGT ACACTCTTTCCCTACAACGACGCTCTTCCGAT*C*T
65	Amplificación del genoma completo: directo-9	/5Phos/ CTGATC ACACTCTTTCCCTACAACGACGCTCTTCCGAT*C*T
	Amplificación del genoma completo: directo-10	/5Phos/ AAGCTA CACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
70	Amplificación del genoma completo: directo-11	/5Phos/ GTAGCC ACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
	Amplificación del genoma completo: directo-12	/5Phos/ TACAAG ACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
75	Amplificación del genoma completo: inverso	/5Phos/CTCGGCATTCTGCTGAACCGCTCTTCCGAT*C*T
	PCR inversa: antisentido	AATGATACGGCGACCACCGAGATCTACACC AGCAGGCGTTATAATAAAAAATATGA
80	PCR inversa: directo	CAAGCAGAAGACGGCATACGAGAT TGACTGAATATAAACCTTGTTGGTAGTTG
	Cebador de secuenciación 1 (para leer las secuencias internas)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
85	Cebador de secuenciación 2 (para leer las secuencias internas)	CTCGGCATTCTGCTGAACCGCTCTTCCGATCT
	Cebador índice 1 (para leer los índices experimentales)	CGGAAGAGCGTCGTGTAGGGAAGAGTGT
90	Cebador índice 2 (para leer los índices experimentales)	CGGAAGAGCGGTTACAGCAGGAATGCCGAG
	UID exógenas	
95	Fidelidad de la polimerasa	
	Amplificación mediante PCR digital: directo	GGTTACAGGCTCATGATGTAACC
100	Amplificación mediante PCR digital: inverso	GATACCAGCTTGGTAATGGCA
	Amplificación de asignación de UID: directo	CGACGTAACGACGCGCCAGTNNNNNNNNNNNGGTTACAGGCTCATGATGTAACC
105	Amplificación de asignación de UID: inverso	CACACAGGAAACAGCTATGACCATGGATACCGCTTGGTAATGGCA
	Amplificación de biblioteca: directo-1	AATGATACGGCGACCACCGAGATCTACACCGTGAT CGACGTAACGACGCGCAACGACGGCCA*G*T
110	Amplificación de biblioteca: directo-2	AATGATACGGCGACCACCGAGATCTACACACATCG CGACGTAACGACGCGCAACGACGGCCA*G*T
	Amplificación de biblioteca: directo-3	AATGATACGGCGACCACCGAGATCTACACGGCTAAC CGACGTAACGACGCGCAACGACGGCCA*G*T
115	Amplificación de biblioteca: directo-4	AATGATACGGCGACCACCGAGATCTACACTGGTCA CGACGTAACGACGCGCAACGACGGCCA*G*T
	Amplificación de biblioteca: directo-5	AATGATACGGCGACCACCGAGATCTACACCACTGT CGACGTAACGACGCGCAACGACGGCCA*G*T
120	Amplificación de biblioteca: directo-6	AATGATACGGCGACCACCGAGATCTACACATTGGC CGACGTAACGACGCGCAACGACGGCCA*G*T
	Amplificación de biblioteca: directo-7	AATGATACGGCGACCACCGAGATCTACACGATCTG CGACGTAACGACGCGCAACGACGGCCA*G*T
125	Amplificación de biblioteca: directo-8	AATGATACGGCGACCACCGAGATCTACACTCAAGTC GACGTAACGACGCGCAACGACGGCCA*G*T
	Amplificación de biblioteca: directo-9	AATGATACGGCGACCACCGAGATCTACACCTGAT CCGACGTAACGACGCGCAACGACGGCCA*G*T

	Amplificación de biblioteca: directo-10	AATGATACGGCGACCACCGAGATCTACACAAGCTAC CGACGTAAACGACGGCCA*G*T
	Amplificación de biblioteca: inverso	CAAGCAGAAGACGGCATACGAGAT CACACAGGAAACAGCTATGACCA*T*G
	Cebador de secuenciación (para leer las secuencias UID e internas)	CGACGTAAACGACGGCCAGT
5	Cebador índice (para leer los índices experimentales)	ACTGGCGTCGTTTTACGTCG
	Mutaciones de <i>CTNNB1</i> en ADN de células humanas normales	
	Amplificación de asignación de UID: directo	CGACGTAAACGACGGCCAGT NNNNNNNNNNNNNN GCAGCAACAGTCTTACCTGCACT
	Amplificación de asignación de UID: inverso	CACACAGGAAACAGCTATGAOCATGTCACATCCTCTTCTCAGGATT
10	Amplificación de biblioteca: directo	AATGATACGGCGACCACCGAGATCTACAC CGACGTAAACGACGGCCA*G*T
	Amplificación de biblioteca: inverso-1	CAAGCAGAAGACGGCATACGAGATATCACG CACACAGGAAACAGCTATGACCA*T*G
	Amplificación de biblioteca: inverso-2	CAAGCAGAAGACGGCATACGAGATCGATGT CACACAGGAAACAGCTATGACCA*T*G
	Amplificación de biblioteca: inverso-3	CAAGCAGAAGACGGCATACGAGATTGACCA CACACAGGAAACAGCTATGACCA*T*G
	Amplificación de biblioteca: inverso-4	CAAGCAGAAGACGGCATACGAGATGCCAAT CACACAGGAAACAGCTATGACCA*T*G
15	Amplificación de biblioteca: inverso-5	CAAGCAGAAGACGGCATACGAGATCAGATCC CACACAGGAAACAGCTATGACCA*T*G
	Amplificación de biblioteca: inverso-6	CAAGCAGAAGACGGCATACGAGATACTTGAC CACACAGGAAACAGCTATGACCA*T*G
	Amplificación de biblioteca: inverso-7	CAAGCAGAAGACGGCATACGAGATGATCAGC CACACAGGAAACAGCTATGACCA*T*G
	Amplificación de biblioteca: inverso-8	CAAGCAGAAGACGGCATACGAGATTAGCTT CACACAGGAAACAGCTATGACCA*T*G
20	Amplificación de biblioteca: inverso-9	CAAGCAGAAGACGGCATACGAGATGGCTACC CACACAGGAAACAGCTATGACCA*T*G
	Amplificación de biblioteca: inverso-10	CAAGCAGAAGACGGCATACGAGATCTTGTA CACACAGGAAACAGCTATGACCA*T*G
	Cebador de secuenciación (para leer las secuencias UID e internas)	CGACGTAAACGACGGCCAGT
	Cebador índice (para leer los índices experimentales)	CATGGTCATAGCTTTCTGTGTG
25	Mutaciones mitocondriales en el ADN de células humanas normales	
	Amplificación de asignación de UID: directo	CGACGTAAACGACGGCCAGT NNNNNNNNNNNNNN TTACCGAGAAAGCTCACAAGAA
	Amplificación de asignación de UID: inverso	CACACAGGAAACAGCTATGACCATGATGCTAAGGCGAGGATGAAA
	Amplificación de biblioteca: directo-1	AATGATACGGCGACCACCGAGATCTACACACA TCGCGACGTAAACGACGGCCA*G*T
30	Amplificación de biblioteca: directo-2	AATGATACGGCGACCACCGAGATCTACACGCCTAA CGACGTAAACGACGGCCA*G*T
	Amplificación de biblioteca: directo-3	AATGATACGGCGACCACCGAGATCTACACTGGTCA CGACGTAAACGACGGCCA*G*T
	Amplificación de biblioteca: directo-4	AATGATACGGCGACCACCGAGATCTACACATTGGC CGACGTAAACGACGGCCA*G*T
	Amplificación de biblioteca: directo-5	AATGATACGGCGACCACCGAGATCTACACGATCTG CGACGTAAACGACGGCCA*G*T
35	Amplificación de biblioteca: directo-6	AATGATACGGCGACCACCGAGATCTACACTCAAGT CGACGTAAACGACGGCCA*G*T
	Amplificación de biblioteca: directo-7	AATGATACGGCGACCACCGAGATCTACACCTGATC CGACGTAAACGACGGCCA*G*T
	Amplificación de biblioteca: inverso	CAAGCAGAAGACGGCATACGAGAT CACACAGGAAACAGCTATGACCA*T*G
	Cebador de secuenciación 1 (para leer las UID)	CGACGTAAACGACGGCCAGT
40	Cebador de secuenciación 2 (para leer las secuencias internas)	CCTAATTCCCCCATCCTTAC
	Cebador índice (para leer los índices experimentales)	ACTGGCGTCGTTTTACGTCG
	Análisis de la composición del oligonucleótido de fosforamidita	
	Plantilla sintetizada, TN	GGTTACAGGCTCATGATGTAACCTCTGTGCTTGGTG IAACTTTAAACATATTTTGCCATTAC
45	Plantilla sintetizada, mut (S = mezcla 50/50 de C y G)	GGTTACAGGCTCATGATGTAACCTCTGTGCTTGGTG SAACTTTAAACATATTTTGCCATTAC
	Amplificación de asignación de UID: directo	CAAGCTGGTATC
	Amplificación de asignación de UID: inverso	ACACTCTTTCCCTACACGACGCT NNNNNNNNNNNN GGTGAGTCTGTGCAGGCAT
	Amplificación de biblioteca: directo	CTCGAGCACTGTCTGACTGAGACGATACACGCTTGGTAATGGCA
	Amplificación de biblioteca: inverso	AATGATACGGCGACCACCGAGATCTACAC CGTGATACACTCTTTCCCTACACGACGC*T*C
50	Cebador de secuenciación (para leer las secuencias UID e internas)	CAAGCAGAAGACGGCATACGAGAT CTCGAGCACTGTCTGACTGAG*A*C
		ACACTCTTTCCCTACACGACGCTC

[0050] UID exógenas. Se preparó ADN genómico de mucosas colónicas humanas normales o linfocitos sanguíneos utilizando kits de Qiagen. El ADN de la mucosa del colon se utilizó para los experimentos con *CTNNB1* y ADN mitocondrial, mientras que el ADN de linfocitos se utilizó para los experimentos en *CTNNB1* y sobre la fidelidad de la polimerasa. El ADN se cuantificó con PCR digital (2) utilizando cebadores que amplificaron genes de copia única de células humanas (análisis de fidelidad de la polimerasa y *CTNNB1*), qPCR (ADN mitocondrial) o por absorbancia óptica (oligonucleótidos). Cada cadena de cada molécula plantilla se codificó con una UID de 12 o 14 bases utilizando dos ciclos de PCR específica de amplicón, como se describe en el texto y en la Figura 3. Ambos cebadores específicos de amplicón contenían secuencias marcadoras universales en sus extremos 5' para una etapa de amplificación posterior. Las UID constituyeron 12 o 14 secuencias de nucleótidos aleatorias unidas al extremo 5' de los cebadores directos específicos del amplicón (Tabla 6). Estos cebadores pueden generar 16,8 y 268 millones de UID distintos, respectivamente. Es importante que el número de UID distintos supere con creces el número de moléculas plantilla originales para minimizar la probabilidad de que dos plantillas originales diferentes adquieran la misma UID. Los ciclos de PCR de asignación de UID incluyeron Phusion Hot Start II (NEB) en una reacción de 45 ul que contenía tampón IX Phusion HF, dNTP 0,25 mM, 0,5 uM de cada uno de cebadores directo (que contiene 12-14 N) e inverso, y 2 U de

polimerasa Phusion. Para mantener las concentraciones finales de la plantilla $<1,5$ ng/ul, se utilizaron varios pocillos para crear algunas bibliotecas. Se emplearon las siguientes condiciones de ciclado: una incubación a 98°C durante 30 s (para activar Phusion Hot Start II); y dos ciclos de 98°C durante 10 s, 61°C durante 120 s y 72°C durante 10 s. Para garantizar la eliminación completa de los cebadores de la primera serie, cada pocillo se digirió con 60 U de una nucleasa específica de ADN monocatenario (Exonucleasa-I; Enzymatics) a 37°C durante 1 h. Después de una inactivación por calor de 5 min a 98°C , se agregaron cebadores complementarios a los marcadores universales introducidos (Tabla 6) hasta una concentración final de 0,5 uM cada uno. Estos cebadores contenían dos fosforotioatos terminales para hacerlos resistentes a cualquier actividad residual de la exonucleasa-I. También contenían secuencias de injerto 5' necesarias para la hibridación con la celda de flujo Illumina GA IIx. Finalmente, contenían una secuencia índice entre la secuencia de injerto y la secuencia marcadora universal. Esta secuencia índice permite analizar simultáneamente los productos de PCR de múltiples individuos diferentes en el mismo compartimento de celda de flujo del secuenciador. Se utilizaron las siguientes condiciones de ciclado para los 25 ciclos posteriores de PCR: 98°C durante 10 s y 72°C durante 15 s. No se realizaron etapas de purificación intermedias en un esfuerzo por reducir las pérdidas de moléculas plantilla.

[0051] Tras la segunda serie de amplificación, los pocillos se consolidaron y purificaron utilizando un kit de purificación por PCR Qiagen QIAquick (n.º de catálogo 28104) y eluyeron en 50 ul de tampón EB (Qiagen). Los fragmentos del tamaño esperado se purificaron después de la electroforesis en gel de agarosa (bibliotecas de ADNmt) o poliácridamida (todas las demás bibliotecas). Para la purificación en gel de agarosa, las ocho alícuotas de 6 ul se cargaron en pocillos de un gel Size Select Gel al 2 % (Invitrogen) y se recogieron bandas del tamaño esperado en tampón EB según lo especificado por el fabricante. Para la purificación del gel de poliácridamida, se cargaron diez alícuotas de 5 ul en pocillos de un gel de poliácridamida TBE al 10 % (Invitrogen). Se cortaron, trituraron y eluyeron rodajas de gel que contenían los fragmentos de interés esencialmente como se describe (3).

[0052] Análisis de la fidelidad de la polimerasa Phusion. La amplificación de un fragmento de ADN genómico humano dentro del gen BMX (acceso RefSeq NM_203281.2) se realizó primero utilizando las condiciones de PCR descritas anteriormente. Se diluyó la plantilla de modo que estuviera presente un promedio de una molécula plantilla en cada 10 pocillos de una placa de PCR de 96 pocillos. A continuación, se realizaron reacciones de PCR de cincuenta ul en tampón Phusion HF 1X, dNTP 0,25 mM, cebadores directo e inverso 0,5 uM (Tabla 6) y 2 U de polimerasa Phusion. Las condiciones de ciclado fueron un ciclo de 98°C durante 30 s; y 19 ciclos de 98°C durante 10 s, 61°C durante 120 s y 72°C durante 10 s. Los cebadores se eliminaron mediante digestión con 60 U de exonucleasa I a 37°C durante 1 h, seguido de una inactivación por calor de 5 min a 98°C . No se realizó ninguna purificación del producto de la PCR, ni antes ni después de la digestión con exonucleasa-I. Después se utilizó todo el contenido de cada pocillo como plantilla para la estrategia de UID exógenas descrita anteriormente.

[0053] Secuenciación. La secuenciación de todas las bibliotecas descritas anteriormente se realizó utilizando un instrumento Illumina GA IIx según lo especificado por el fabricante. La longitud total de las lecturas utilizadas para cada experimento varió de 36 a 73 bases. La lectura de nucleótidos y la alineación de secuencias se realizaron con la conducción Eland (Illumina). Para el análisis posterior solo se utilizaron lecturas de alta calidad que cumplieran con los siguientes criterios: (i) las primeras 25 bases pasaron el filtro de castidad estándar de Illumina; (ii) cada base de la lectura tuvo una puntuación de calidad ≥ 20 ; y (iii) ≤ 3 discrepancias con las secuencias esperadas. Para las bibliotecas de UID exógenas, además es necesario que las UID tengan una puntuación de calidad ≥ 30 . Apreciamos una frecuencia relativamente alta de errores en los extremos de las lecturas en las bibliotecas de UID endógenas preparadas con el protocolo estándar de Illumina, presumiblemente introducidos durante el cizallamiento o la reparación final, por lo que la primera y las últimas tres bases de estos marcadores se excluyeron del análisis.

[0054] Análisis Safe-SeqS. Las lecturas de alta calidad se agruparon en familias de UID en función de sus UID endógenas (según la invención tal como se define en las reivindicaciones) o exógenas. Solo se consideraron las familias de UID con dos o más miembros. Estas familias de UID incluían la gran mayoría (≥ 99 %) de las lecturas de secuenciación. Para garantizar que se utilizaran los mismos datos para el análisis convencional y Safe-SeqS, también excluimos del análisis convencional las familias de UID que solo contenían un miembro. Además, solo identificamos una base como “mutante” en el análisis de secuenciación convencional si se identificó la misma variante en al menos dos miembros de al menos una familia de UID (es decir, dos mutaciones) al comparar el análisis convencional con el de Safe-SeqS con UID exógenas. Para comparar con Safe-SeqS de UID endógenas, era necesario al menos dos miembros de cada una de dos familias de UID (es decir, cuatro mutaciones) para identificar una posición como “mutante” en el análisis convencional. Ya sea con UID endógenas o exógenas, un supermutante se definió como una familia UID en la que ≥ 95 % de los miembros compartían la mutación idéntica. Por lo tanto, las familias de UID con <20 miembros tenían que ser 100 % idénticas en la posición mutante, mientras que en las familias de UID con más miembros se permitía una tasa combinada de error de replicación y secuenciación del 5 %. Para determinar la fidelidad de la polimerasa utilizando Safe-SeqS y comparar los resultados con análisis anteriores de fidelidad de la polimerasa Phusion, fue necesario darse cuenta de que los análisis anteriores solo detectarían mutaciones presentes en ambas cadenas de los productos de PCR (4). Esto equivaldría a analizar los productos de PCR generados con un ciclo menos con Safe-SeqS, y la corrección adecuada se realizó en la Tabla 2A. A menos que se especifique lo contrario, todos los valores enumerados en el texto y las tablas representan medias y desviaciones estándar.

Ejemplo 8: Procesos generadores de errores

[0055] Las mutaciones aparentes, definidas como cualquier lectura de nucleótidos que varía de la base esperada en una posición definida, pueden proceder de una variedad de procesos:

1. Las mutaciones presentes en el ADN plantilla. Para las plantillas derivadas de células humanas normales, estas incluyen mutaciones que estuvieron presentes en el cigoto, ocurrieron más tarde durante el desarrollo embrionario y adulto, o estuvieron presentes en un contaminante introducido inadvertidamente en la muestra. Se espera que estas mutaciones estén presentes en ambas cadenas de las plantillas en cuestión. Si la mutación se produjo solo en el último ciclo celular de una célula cuyo ADN se utilizó como plantilla, la mutación estaría presente solo en una cadena de la plantilla.
2. Las bases modificadas químicamente presentes en las plantillas. Se ha estimado que hay miles de bases oxidadas presentes en cada célula humana (5). Cuando dicho ADN es amplificado por la polimerasa Phusion, puede producirse una mutación aparente en una cadena.
3. Los errores introducidos durante el proceso de cizallamiento hicieron necesario generar pequeños fragmentos para la secuenciación. El cizallamiento acústico genera altas temperaturas de corta duración que pueden dañar el ADN.
4. Los errores introducidos durante la reparación final de los fragmentos sometidos a cizallamiento. La fuente de estos errores puede ser la infidelidad de la polimerasa o la incorporación de bases modificadas químicamente en los dNTP utilizados para la polimerización.
5. Los errores introducidos por otras etapas enzimáticas, particularmente si las enzimas son impuras y están contaminadas con nucleasas, polimerasas o ligasas.
6. Los errores introducidos durante la amplificación por PCR para preparar las bibliotecas para la captura o para la PCR inversa.
7. Los errores producidos durante la PCR después de la captura o durante la amplificación por PCR inversa.
8. Los errores introducidos en los ciclos de asignación de UID de Safe-SeqS (Figura 3).
9. Los errores introducidos en los ciclos de amplificación de la biblioteca de Safe-SeqS realizados con UID exógenas. Cabe señalar que si los cebadores de asignación de UID del proceso n.º 8 no se eliminan por completo, y podrían amplificar fragmentos de ADN que contienen errores introducidos durante estos ciclos, creando un nuevo supermutante.
10. Los errores introducidos en el primer ciclo de PCR puente en la celda de flujo de Illumina. Si la amplificación es ineficaz, un error introducido en el segundo ciclo de PCR puente también podría dar lugar a un grupo que contenga una mutación en la mayoría de las moléculas que lo componen.
11. Los errores producidos en la lectura de nucleótidos.

Ejemplo 9: Lograr precisión con Safe-SeqS

[0056] Con los enfoques convencionales de secuenciación por síntesis, todos los procesos productores de errores descritos anteriormente son relevantes, lo que da como resultado un número relativamente alto de lecturas de mutaciones falsas positivas (Tablas 1 y 2). Safe-SeqS reduce al mínimo la cantidad de lecturas de mutaciones falsas positivas de varias maneras. Safe-SeqS con UID exógenas produce la menor cantidad de lecturas de mutaciones falsas positivas porque requiere la menor cantidad de etapas enzimáticas. Con las UID exógenas, los procesos generadores de errores n.º 3 a n.º 7 se eliminan por completo al no realizarse estas etapas. Safe-SeqS con UID exógenas también reduce drásticamente los errores resultantes de los procesos generadores de errores n.º 10 y n.º 11 debido a la forma en que se analizan los datos.

[0057] Después de Safe-SeqS con UID exógenas, los únicos errores falsos positivos restantes deben ser los introducidos durante los ciclos de PCR de asignación de UID (proceso de generación de errores n.º 8) o los cebadores que contienen UID residuales durante los ciclos de amplificación de la biblioteca (proceso de generación de errores n.º 9). En teoría, los errores del proceso generador de errores n.º 8 pueden eliminarse requiriendo al menos dos supermutantes para identificar una posición como “mutante”. Este requisito es razonable porque cada mutación preexistente en una plantilla de ADN bicatenaria debería dar lugar a dos supermutantes, uno de cada cadena. Además, este requisito eliminaría el proceso generador de errores n.º 2 (bases dañadas en las plantillas originales) porque dichas bases, cuando se copian, deberían dar lugar a un solo supermutante. Finalmente, los errores generados durante los ciclos de amplificación de la biblioteca (proceso n.º 9) no serán amplificados por los cebadores que contienen UID residuales si esos cebadores se eliminan por completo, como se realiza aquí con exceso de Exonucleasa-I.

[0058] Con las UID endógenas, los errores introducidos por los procesos n.º 10 y n.º 11 se reducen drásticamente debido a la forma en que se analizan los datos (como ocurre con las UID exógenas). Los errores introducidos en los procesos n.º 2 a n.º 7 se pueden reducir al mínimo exigiendo que se observe una mutación en al menos dos familias de UID, por los motivos indicados en el párrafo anterior. Con este requisito, en teoría, deberían identificarse pocas mutaciones falsas positivas.

[0059] En la práctica, la situación se complica por el hecho de que las distintas amplificaciones no son perfectas, por lo que cada cadena de cada molécula plantilla original no se recupera como una familia de UID. Esta eficiencia puede variar de una muestra a otra, dependiendo en parte de la concentración de inhibidores presentes en las muestras clínicas. Además, con las UID exógenas, un error de la polimerasa durante la etapa de amplificación de la biblioteca puede crear una nueva familia de UID que no estaba representada en la etapa de asignación de UID. Si este error ocurriera en una plantilla mutante, se crearía un supermutante artificial adicional.

[0060] Estos factores pueden gestionarse incorporando varios criterios adicionales en los análisis. Por ejemplo, se podría exigir que las familias de UID contengan más de dos, cinco o diez miembros. Otro requisito podría ser que las UID exógenas de los supermutantes no estén relacionadas con ninguna otra UID de la biblioteca por una diferencia de una base. Esto eliminaría los supermutantes artificiales generados durante las etapas de amplificación de la biblioteca (indicadas en el párrafo anterior). Establecimos este requisito de forma rutinaria en nuestros análisis de Safe-SeqS, pero apenas se apreció la diferencia (<1 %) en la cantidad de supermutantes identificados. La especificidad de las mutaciones se puede aumentar aún más al requerir que más de un supermutante identifique una posición como “mutante”, como se ha descrito anteriormente para las UID endógenas. Cuando se requieren múltiples supermutantes, la especificidad se puede aumentar aún más al requerir que cada cadena de la plantilla bicatenaria original contenga la mutación o, cuando las bibliotecas se amplifican usando múltiples pocillos, que las mutaciones raras compartan una secuencia introducida que identifique el pocillo en que las mutaciones (es decir, una de cada cadena) fueron amplificadas. Estas decisiones implican el equilibrio habitual entre especificidad y sensibilidad. En nuestros experimentos con UID exógenas (Tabla 2), solo requerimos un supermutante para identificar una posición como “mutante” e incluimos todas las familias de UID con más de un miembro. Como las UID endógenas se asociaron con más procesos generadores de errores que las UID exógenas, requerimos dos supermutantes para identificar una posición como mutante en los experimentos informados en la Tabla 1 y también incluimos a todas las familias de UID con más de un miembro.

Ejemplo 10: Prevalencias de mutaciones en tejidos humanos normales

[0061] Los experimentos presentados en las Tablas 1 y 2, en los que se evaluaron >10.000 plantillas, muestran que hay mutaciones presentes en el ADN nuclear de células humanas normales con una frecuencia de $3,5 \times 10^{-6}$ a $9,0 \times 10^{-6}$ mutantes/pb dependiendo de la región analizada. Es imposible determinar si este bajo nivel representa mutaciones genuinas presentes en las plantillas originales o la suma de mutaciones genuinas más mutaciones artefactos de los procesos generadores de errores descritos anteriormente. La prevalencia de mutaciones en células humanas no se ha investigado ampliamente, en parte porque son muy poco frecuentes. Sin embargo, se han ideado varias técnicas inteligentes para identificar mutantes raros y, en principio, pueden utilizarse con fines comparativos. Desafortunadamente, las estimaciones de la prevalencia de mutaciones humanas varían ampliamente, desde hasta 10^{-5} mutantes/pb hasta 10^{-8} mutantes/pb (6-12). En varios de estos estudios, las estimaciones se complican por la falta de datos sobre la naturaleza de las mutaciones reales: en algunos casos podrían ser causadas por pérdidas de cromosomas completos, en otros por mutaciones de aminoácidos y en otros principalmente por mutaciones terminadoras o pequeñas inserciones o deleciones. Además, estos estudios utilizaron varias fuentes de células normales y examinaron diferentes genes, lo que dificulta las comparaciones directas. Las estimaciones de las prevalencias y frecuencias de mutaciones del ADN mitocondrial varían de manera similar (13-19). Será interesante analizar en futuros trabajos las mismas plantillas de ADN y genes con diversas tecnologías para determinar la base de estas diferentes estimaciones.

[0062] Sin embargo, supongamos que todas las mutaciones identificadas con Safe-SeqS representan mutaciones genuinas presentes en las plantillas de ADN originales de células normales. ¿Qué nos dice esto acerca del número de generaciones por las que han pasado estas células desde que se concibió el organismo? Existe una relación simple entre la frecuencia de mutación y la prevalencia de la mutación: la prevalencia de la mutación es igual al producto de la frecuencia de mutación por el número de generaciones por las que ha pasado la célula desde su concepción. En estudios previos se ha determinado que la frecuencia de mutación somática es $\sim 10^{-9}$ mutantes/pb/generación, aunque esta estimación también varía de un estudio a otro por razones relacionadas con las mencionadas anteriormente con respecto a la prevalencia de mutaciones. La combinación de esta estimación de la frecuencia de mutación derivada de la bibliografía con nuestras estimaciones de prevalencia de mutación indica que las células normales analizadas (linfocitos, estirpes celulares linfoblastoides o mucosas del colon) habían pasado por entre 3500 y 8900 generaciones, lo que representa células que se dividen cada 3 a 7 días para los individuos examinados en este estudio (edad promedio de 65 años).

Ejemplo 11: Simulación asistida por ordenador de errores introducidos por la polimerasa

[0063] El momento de las mutaciones introducidas por las polimerasas altera en gran medida el número final de mutaciones observadas (20). Por ejemplo, dos mutaciones diferirían en cuanto a la prevalencia aproximadamente ~64 veces si se introdujeran con 6 ciclos de diferencia (2^6). Debido a que las polimerasas introducen mutaciones de manera estocástica, se empleó un método de Monte Carlo simple para las simulaciones. En estas simulaciones, utilizamos la estimación del fabricante de la tasa de error de la polimerasa Phusion con un ajuste apropiado para la capacidad de Safe-SeqS para detectar mutaciones en una sola cadena (4). Cabe señalar que los errores introducidos en el ciclo 19, así como en los dos ciclos de asignación de UID, darían lugar a cambios en solo una cadena del dúplex, es decir, darían lugar a un supermutante en lugar de dos. En cada experimento, supusimos que había una eficiencia de amplificación constante dada por el número total de plantillas obtenidas al final del experimento (es decir, si el número de familias de UID era N, entonces supusimos que el número de plantillas aumentó por un factor de $N/2^{19}$ en cada ciclo). Se realizaron mil simulaciones para cada uno de los siete experimentos y los resultados se presentan en la Tabla 4.

Referencias (solo para los Ejemplos 8-11)

[0064]

1. Herman D. S. y col. (2009) "Filter-based hybridization capture of subgenomes enables resequencing and copy-number detection". *Nat Methods* 6:507-510.
2. Vogelstein B. y Kinzler K. W. (1999) "Digital PCR". *Proc Natl Acad Sci EE. UU.* 96:9236-9241.
3. Chory J. y Pollard J. D., Jr. (2001) "Separation of small DNA fragments by conventional gel electrophoresis". *Curr Protoc Mol Biol*, capítulo 2: unidad 2.7.
4. Barnes W. M. (1992) "The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion". *Gene* 112:29-35.
5. Collins A. R. (1999) "Oxidative DNA damage, antioxidants, and cancer". *Bioessays* 21:238-246.
6. Morley A. A., Cox S., y Holliday R. (1982) "Human lymphocytes resistant to 6-thioguanine increase with age". *Mech Ageing Dev* 19:21-26.
7. Trainor K. J. y col. (1984) "Mutation frequency in human lymphocytes increases with age". *Mech Ageing Dev* 27:83-86.
8. Grist S. A., McCarron M., Kutlaca A., Turner D. R. y Morley A. A. (1992) "In vivo human somatic mutation: frequency and spectrum with age". *Mutat Res* 266:189-196.
9. Williams G. T., Geraghty J. M., Campbell F., Appleton M.A. y Williams E. D. (1995) "Normal colonic mucosa in hereditary non-polyposis colorectal cancer shows no generalised increase in somatic mutation". *Br J Cancer* 71: 1077-1080.
10. Campbell F., Appleton M. A., Shields C. J. y Williams G. T. (1998) "No difference in stem cell somatic mutation between the background mucosa of right- and leftsided sporadic colorectal carcinomas". *J Pathol* 186:31-35.
11. Araten D. J., Nafa K., Pakdeesuwan K. y Luzzatto L. (1999) "Clonal populations of hematopoietic cells with paroxysmal nocturnal hemoglobinuria genotype and phenotype are present in normal individuals". *Proc Natl Acad Sci EE. UU.* 96:5209-5214.
12. Araten D. J. y col. (2005) "A quantitative measurement of the human somatic mutation rate". *Cancer Res* 65:8111-8117.
13. Monnat R. J., Jr. y Loeb L. A. (1985) "Nucleotide sequence preservation of human mitochondrial DNA". *Proc Natl Acad Sci EE. UU.* 82:2895-2899.
14. Bodenteich A., Mitchell L. G. y Merrill C. R. (1991) "A lifetime of retinal light exposure does not appear to increase mitochondrial mutations". *Gene* 108:305-309.
15. Howell N., Kubacka I. y Mackey D. A. (1996) "How rapidly does the human mitochondrial genome evolve?" *Am J Hum Genet* 59:501-509.
16. Khrapko K y col. (1997) "Mitochondrial mutational spectra in human cells and tissues". *Proc Natl Acad Sci EE. UU.* 94:13798-13803.
17. Heyer E. y col. (2001) "Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees". *Am J Hum Genet* 69:1113-1126.

18. Howell N. y col. (2003) "The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates". *Am J Hum Genet* 72:659-670.

19. Taylor R. W. y col. (2003) "Mitochondrial DNA mutations in human colonic crypt stem cells". *J Clin Invest* 112:1351-1360.

20. Luria S. E. y Delbruck M. (1943) "Mutations of Bacteria from Virus Sensitivity to Virus Resistance". *Genetics* 28:491-511.

Referencias (Para todo a excepción de los Ejemplos 8-11)

[0065]

1. Luria S. E. y Delbruck M. (1943) "Mutations of Bacteria from Virus Sensitivity to Virus Resistance". *Genetics* 28:491-511.

2. Roach J. C. y col. (2010) "Analysis of genetic inheritance in a family quartet by whole-genome sequencing". *Science* 328:636-639.

3. Durbin R. M. y col. (2010) "A map of human genome variation from populationscale sequencing". *Nature* 467:1061-1073.

4. Shibata D. (2011) "Mutation and epigenetic molecular clocks in cancer". *Carcinogenesis* 32:123-128.

5. McMahon M. A. y col. (2007) "The HBV drug entecavir - effects on HIV-1 replication and resistance". *N Engl J Med* 356:2614-2621.

6. Eastman P. S. y col. (1998) "Maternal viral genotypic zidovudine resistance and infrequent failure of zidovudine therapy to prevent perinatal transmission of human immunodeficiency virus type 1 in pediatric AIDS Clinical Trials Group Protocol 076". *J Infect Dis* 177:557-564.

7. Chiu R. W. y col. (2008) "Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma". *Proc Natl Acad Sci EE. UU.* 105:20458-20463.

8. Fan H. C., Blumenfeld Y. J., Chitkara U., Hudgins L., y Quake S. R. (2008) *Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood*. *Proc Natl Acad Sci EE. UU.* 105:16266-16271.

9. Hoque M. O. y col. (2003) "High-throughput molecular analysis of urine sediment for the detection of bladder cancer by high-density single-nucleotide polymorphism array". *Cancer Res* 63:5723-5726.

10. Thunnissen F. B. (2003) "Sputum examination for early detection of lung cancer". *J Clin Pathol* 56:805-810.

11. Diehl F. y col. (2008) "Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients". *Gastroenterology* 135:489-498.

12. Barnes W. M. (1992) "The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion". *Gene* 112:29-35.

13. Araten D. J. y col. (2005) "A quantitative measurement of the human somatic mutation rate". *Cancer Res* 65:8111-8117.

14. Campbell F., Appleton M. A., Shields C. J. y Williams G. T. (1998) "No difference in stem cell somatic mutation between the background mucosa of right- and leftsided sporadic colorectal carcinomas". *J Pathol* 186:31-35.

15. Tindall K. R. y Kunkel T. A. (1988) "Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase". *Biochemistry* 27:6008-6013.

16. Kunkel T. A. (1985) "The mutational specificity of DNA polymerase-beta during in vitro DNA synthesis. Production of frameshift, base substitution, and deletion mutations". *J Biol Chem* 260:5787-5796.

17. van Dongen J. J. y Wolvers-Tettero I. L. (1991) "Analysis of immunoglobulin and T cell receptor genes. Part II: Possibilities and limitations in the diagnosis and management of lymphoproliferative diseases and related disorders". *Clin Chim Acta* 198:93-174.

18. Grist S. A., McCarron M., Kutlaca A., Turner D. R. y Morley A. A. (1992) "In vivo human somatic mutation: frequency and spectrum with age". *Mutat Res* 266:189-196.

19. Liu Q. y Sommer S. S. (2004) "Detection of extremely rare alleles by bidirectional pyrophosphorolysis-activated polymerization allele-specific amplification (Bi-PAP-A): measurement of mutation load in mammalian tissues". *Biotechniques* 36:156-166.
- 5 20. Monnat R. J., Jr. y Loeb L. A. (1985) "Nucleotide sequence preservation of human mitochondrial DNA". *Proc Natl Acad Sci EE. UU.* 82:2895-2899.
21. Shi C. y col. (2004) "LigAmp for sensitive detection of single-nucleotide differences". *Nat Methods* 1:141-147.
- 10 22. Keohavong P. y Thilly W. G. (1989) "Fidelity of DNA polymerases in DNA amplification". *Proc Natl Acad Sci EE. UU.* 86:9253-9257.
23. Sidransky D. y col. (1991) "Identification of p53 gene mutations in bladder cancers and urine samples". *Science* 252:706-709.
- 15 24. Bielas J. H. y Loeb L. A. (2005) "Quantification of random genomic mutations". *Nat Methods* 2:285-290.
- 25 25. Vogelstein B. y Kinzler K. W. (1999) "Digital PCR". *Proc Natl Acad Sci EE. UU.* 96:9236-9241.
- 20 26. Mitra R. D. y col. (2003) "Digital genotyping and haplotyping with polymerase colonies". *Proc Natl Acad Sci EE. UU.* 100:5926-5931.
27. Chetverina H. V., Samatov T. R., Ugarov V. I. y Chetverin A. B. (2002) "Molecular colony diagnostics: detection and quantitation of viral nucleic acids by in-gel PCR". *Biotechniques* 33:150-152, 154, 156.
- 25 28. Zimmermann B. G. y col. (2008) "Digital PCR: a powerful new tool for noninvasive prenatal diagnosis?" *Prenat Diagn* 28:1087-1093.
- 30 29. Dressman D., Yan H., Traverso G., Kinzler K. W. y Vogelstein B. (2003) "Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations". *Proc Natl Acad Sci EE. UU.* 100:8817-8822.
- 30 30. Ottesen E. A., Hong J. W., Quake S. R. y Leadbetter J. R. (2006) "Microfluidic digital PCR enables multigene analysis of individual environmental bacteria". *Science* 314:1464-1467.
- 35 31. Quail M. A. y col. (2008) "A large genome center's improvements to the Illumina sequencing system". *Nat Methods* 5:1005-1010.
- 40 32. Nazarian R. y col. (2010) "Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation". *Nature* 468:973-977.
33. He Y. y col. (2010) "Heteroplasmic mitochondrial DNA mutations in normal and tumour cells". *Nature* 464:610-614.
- 45 34. Gore A. y col. (2011) "Somatic coding mutations in human induced pluripotent stem cells". *Nature* 471:63-67.
- 35 35. Dohm J. C., Lottaz C., Borodina T. y Himmelbauer H. (2008) "Substantial biases in ultrashort read data sets from high-throughput DNA sequencing". *Nucleic Acids Res* 36:e105.
- 50 36. Erlich Y., Mitra P. P., delaBastide M., McCombie W. R. y Hannon G. J. (2008) "Alta-Cyclic: a self-optimizing base caller for next-generation sequencing". *Nat Methods* 5:679-682.
37. Rougemont J. y col. (2008) "Probabilistic base calling of Solexa sequencing data". *BMC Bioinformatics* 9:431.
- 55 38. Druley T. E. y col. (2009) "Quantification of rare allelic variants from pooled genomic DNA". *Nat Methods* 6:263-265.
39. Vallania F. L. y col. (2010) "High-throughput discovery of rare insertions and deletions in large cohorts". *Genome Res* 20:1711-1718.
- 60 40. McCloskey M. L., Stoger R., Hansen R. S. y Laird C. D. (2007) "Encoding PCR products with batch-stamps and barcodes". *Biochem Genet* 45:761-767.
41. Parameswaran P. y col. (2007) "A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing". *Nucleic Acids Res* 35:e130.
- 65

42. Craig D. W. y col. (2008) "Identification of genetic variants using bar-coded multiplexed sequencing". *Nat Methods* 5:887-893.
43. Miner B. E., Stoger R. J., Burden A. F., Laird C. D. y Hansen RS (2004) "Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR". *Nucleic Acids Res* 32:e135.
44. Herman D. S. y col. (2009) "Filter-based hybridization capture of subgenomes enables resequencing and copy-number detection". *Nat Methods* 6:507-510.
45. Jones P. A. y Baylin S. B. (2007) "The epigenomics of cancer". *Cell* 128:683-692.
46. de Boer J. G. y Ripley L. S. (1988) "An in vitro assay for frameshift mutations: hotspots for deletions of 1 bp by Klenow-fragment polymerase share a consensus DNA sequence". *Genetics* 118:181-191.
47. Eckert K. A. y Kunkel T. A. (1990) "High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase". *Nucleic Acids Res* 18:3739-3744.
48. Kosuri S. y col. (2010) "Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips". *Nat Biotechnol* 28:1295-1299.
49. Matzas M. y col. (2010) "High-fidelity gene synthesis by retrieval of sequenceverified DNA identified using high-throughput pyrosequencing". *Nat Biotechnol* 28:1291-1294.
50. Li J. y col. (2008) "Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing". *Nat Med* 14:579-584.
51. Eid J. y col. (2009) "Real-time DNA sequencing from single polymerase molecules". *Science* 323:133-138.

REIVINDICACIONES

1. Un método para analizar si una secuencia de ADN amplificada representa con exactitud un fragmento de ADN analito que comprende una mutación utilizando secuencias identificadoras únicas (UID) endógenas, que comprende:
 - unir oligonucleótidos adaptadores a extremos de fragmentos de ADN analito de entre 30 y 2000 bases, ambos valores inclusive, para formar fragmentos adaptados, en donde cada extremo de un fragmento antes de dicha unión es una UID endógena para el fragmento;
 - amplificar los fragmentos adaptados usando cebadores complementarios a los oligonucleótidos adaptadores para formar familias de fragmentos adaptados;
 - determinar la secuencia de nucleótidos de una pluralidad de miembros de una familia;
 - comparar las secuencias de nucleótidos de la pluralidad de miembros de la familia; e
 - identificar una secuencia de nucleótidos como representación exacta de dicho fragmento de ADN analito que comprende dicha mutación cuando al menos el 95 % de los miembros de la familia comparten dicha mutación.
2. El método de la reivindicación 1, en donde el método comprende además el enriquecimiento en fragmentos que representan uno o más genes seleccionados mediante la captura de un subconjunto de los fragmentos usando oligonucleótidos de captura complementarios a genes seleccionados en el ADN analito.
3. El método de la reivindicación 2, en donde la etapa de unión es previa a la etapa de enriquecimiento.
4. El método de la reivindicación 1, en donde el método comprende además el enriquecimiento en fragmentos que representan uno o más genes seleccionados mediante la amplificación de fragmentos complementarios a genes seleccionados.
5. El método de la reivindicación 4, en donde la etapa de unión es previa a la etapa de enriquecimiento.
6. El método de la reivindicación 1, en donde los fragmentos se forman mediante cizallamiento.
7. El método de la reivindicación 1, en donde antes de la amplificación, el ADN analito se trata con bisulfito para convertir las bases de citosina no metiladas en uracilo.
8. El método de la reivindicación 1, que comprende además la etapa de comparar el número de familias que representan un primer fragmento de ADN analito con el número de familias que representan un segundo fragmento de ADN analito para determinar una concentración relativa de un primer fragmento de ADN analito con respecto a un segundo fragmento de ADN analito en la pluralidad de fragmentos de ADN analito.
9. El método de la reivindicación 1, en donde la identificación de una secuencia de nucleótidos que representa con exactitud un fragmento de ADN analito requiere la observación de una mutación en al menos dos familias.
10. El método de cualquier reivindicación anterior, en donde la etapa de determinación de la secuencia de nucleótidos de una pluralidad de miembros de una familia se lleva a cabo mediante secuenciación masiva en paralelo.

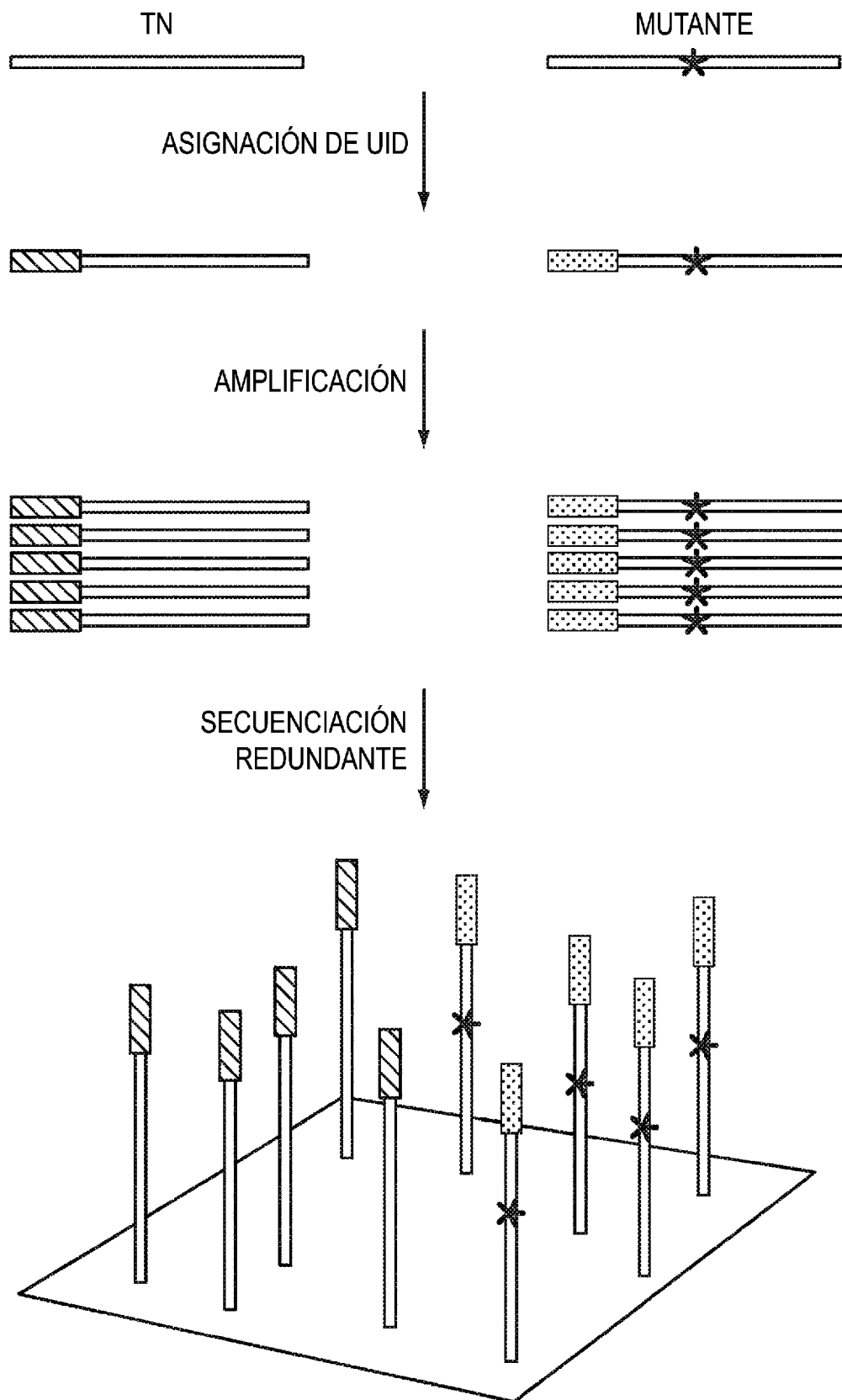


Figura 1

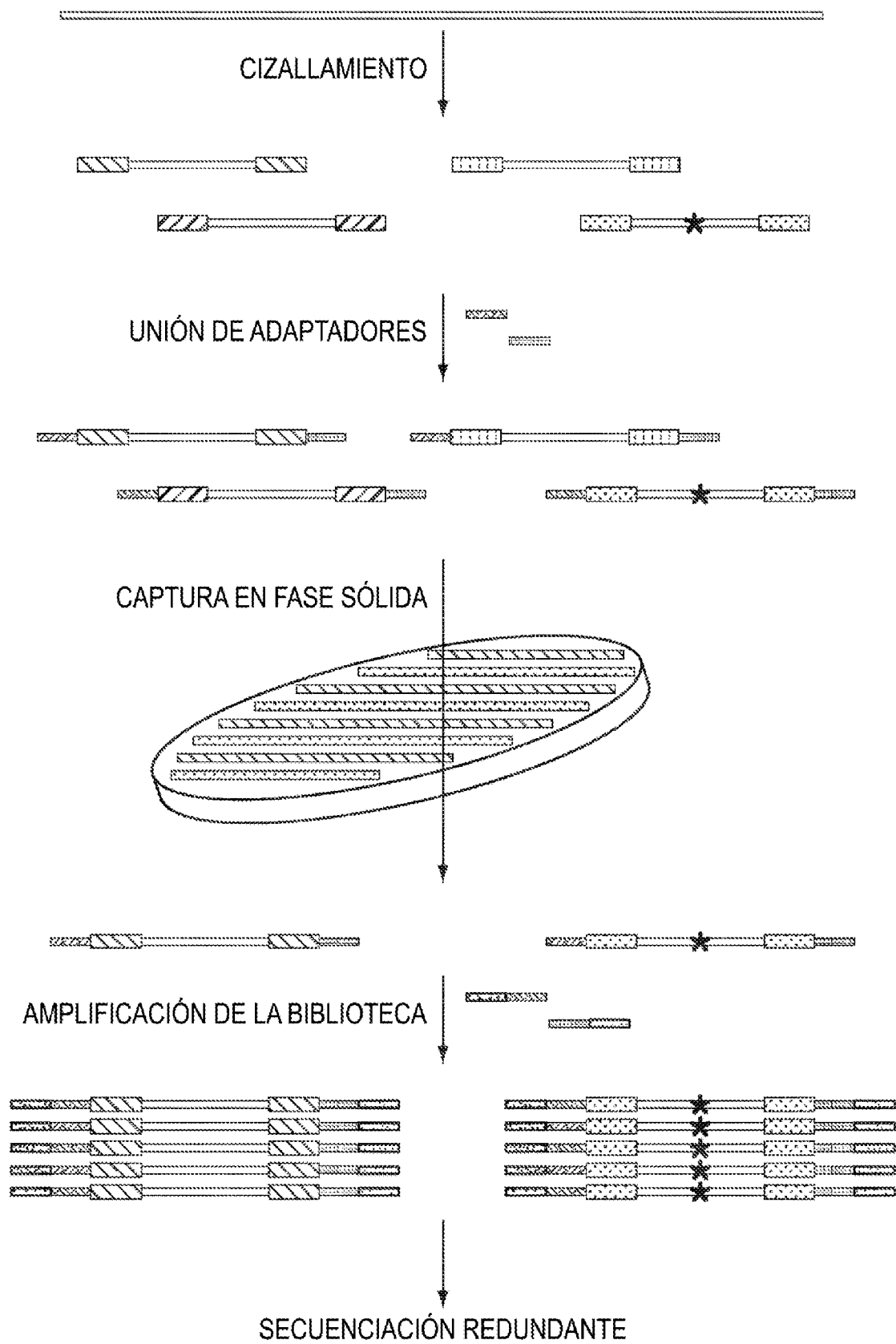


Figura 2

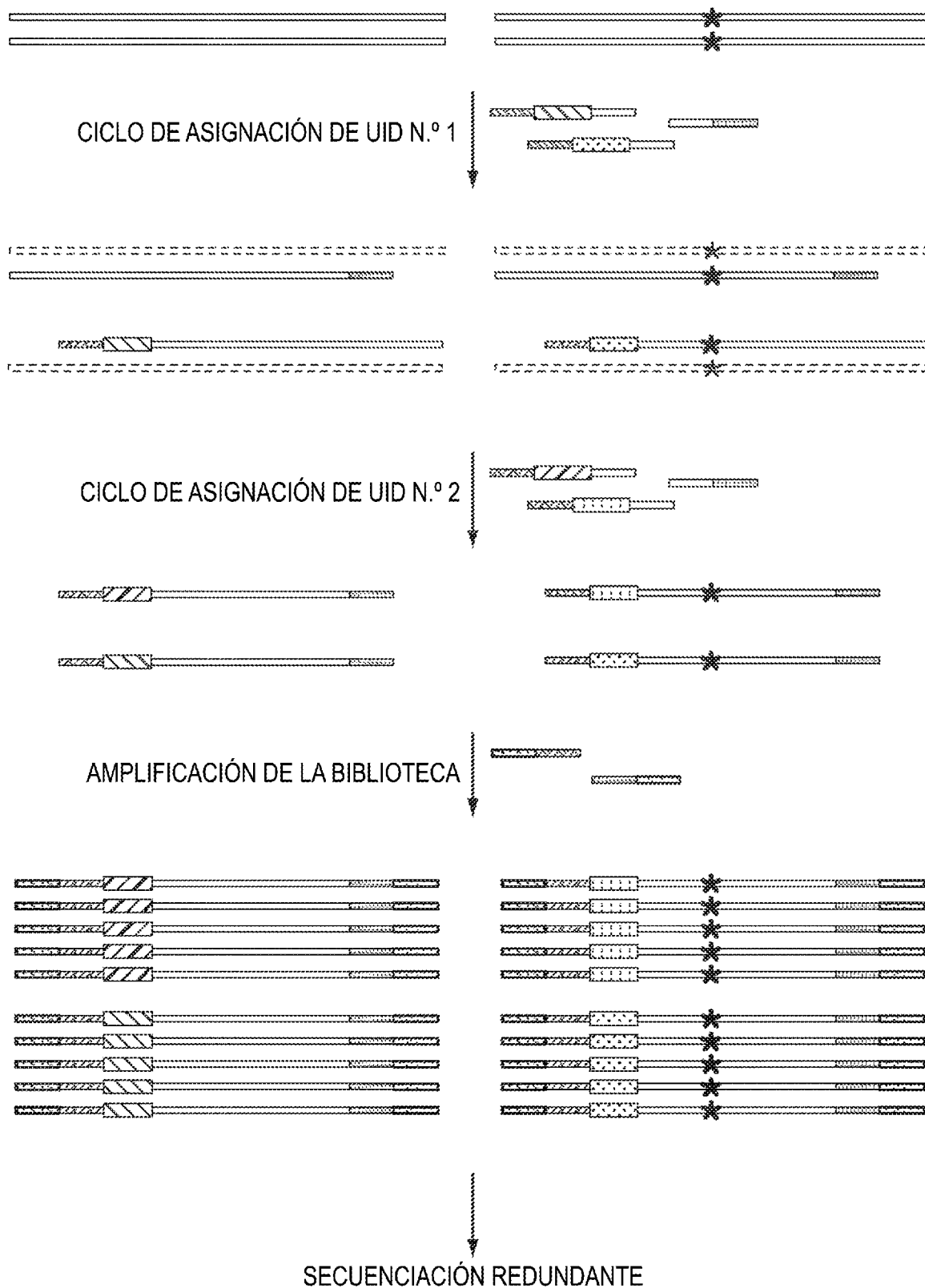


Figura 3

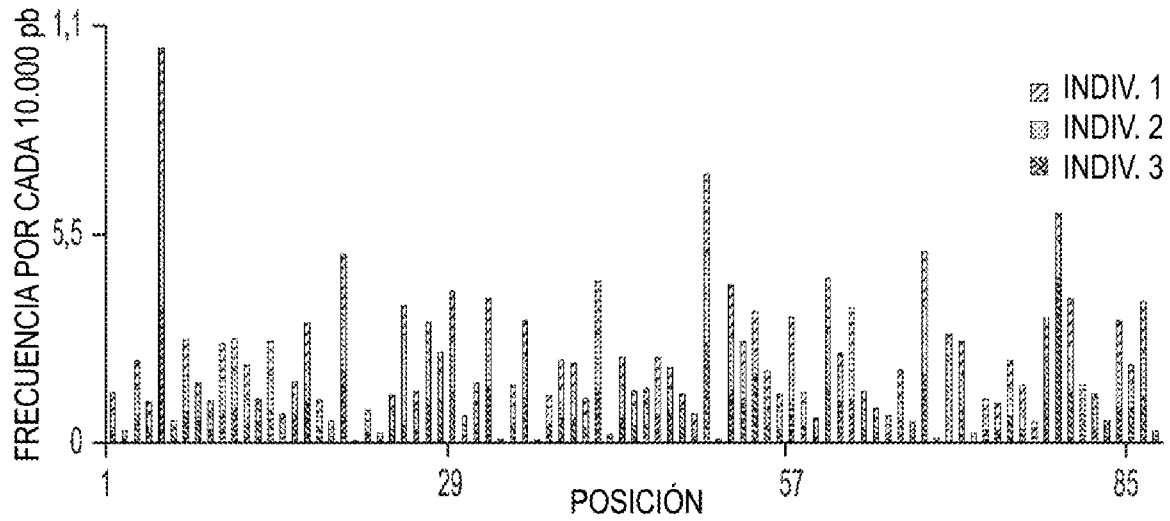


Figura 4A

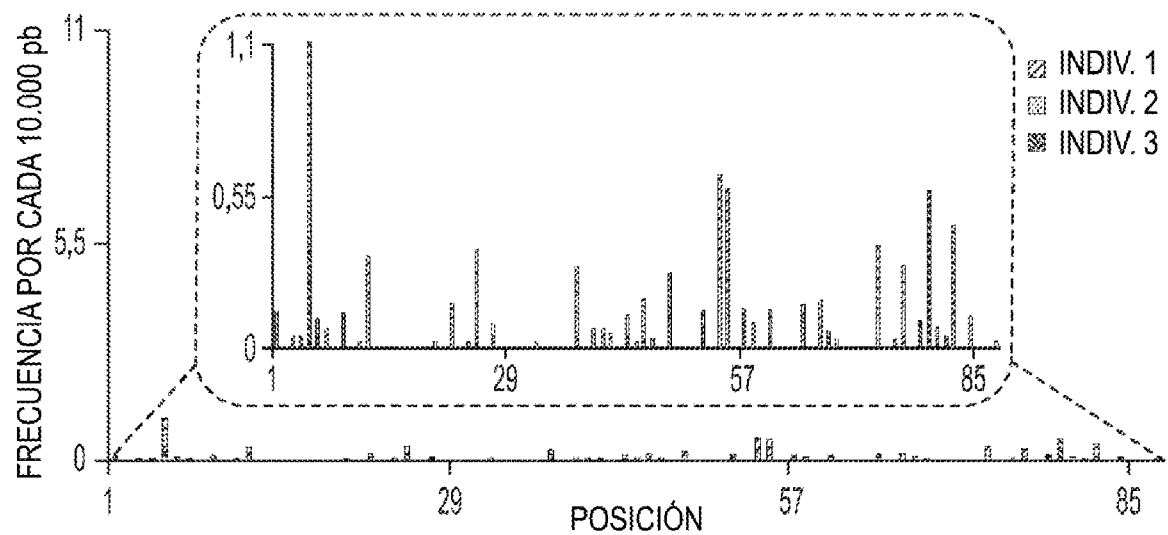


Figura 4B

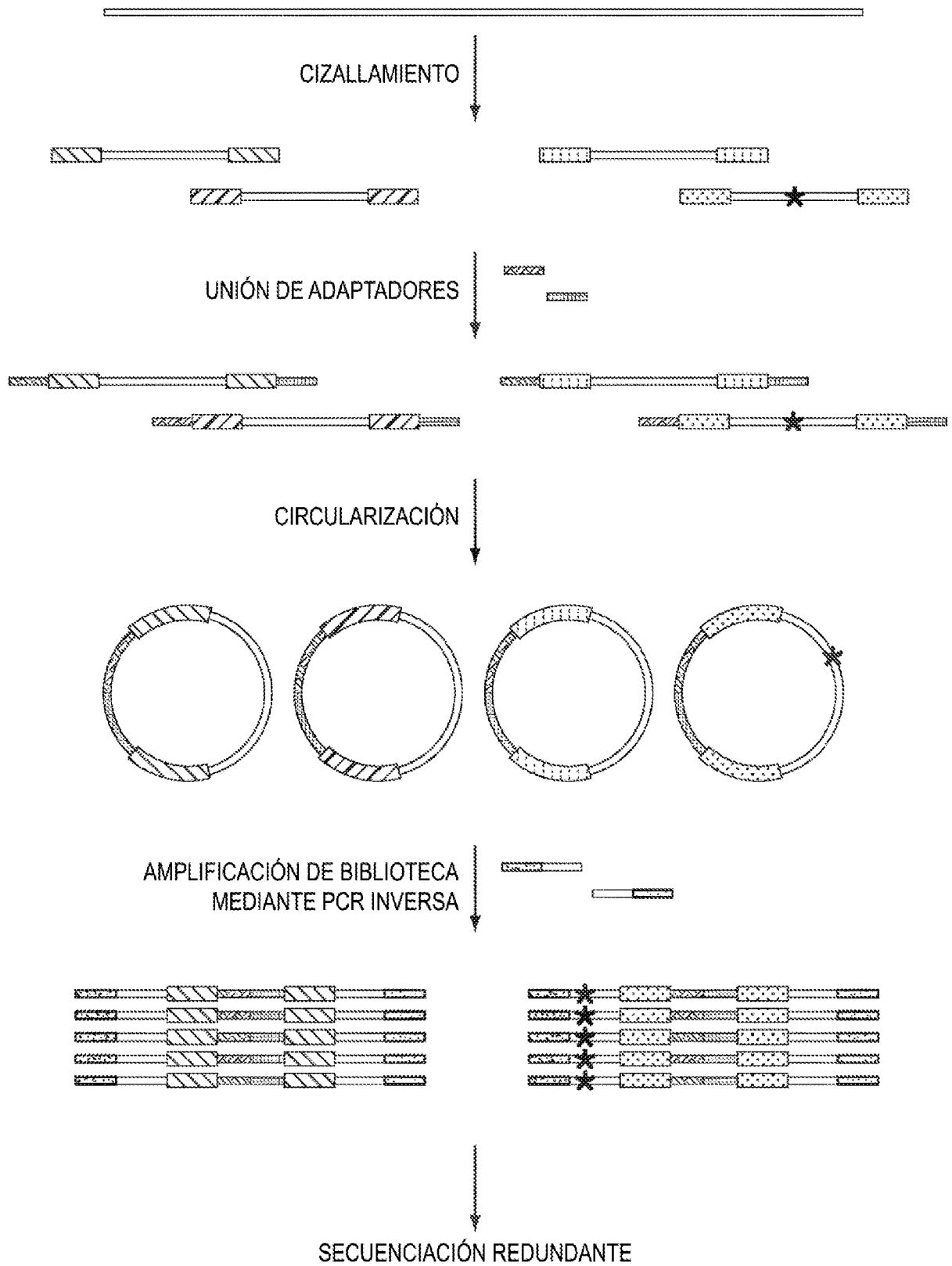


Figura 5

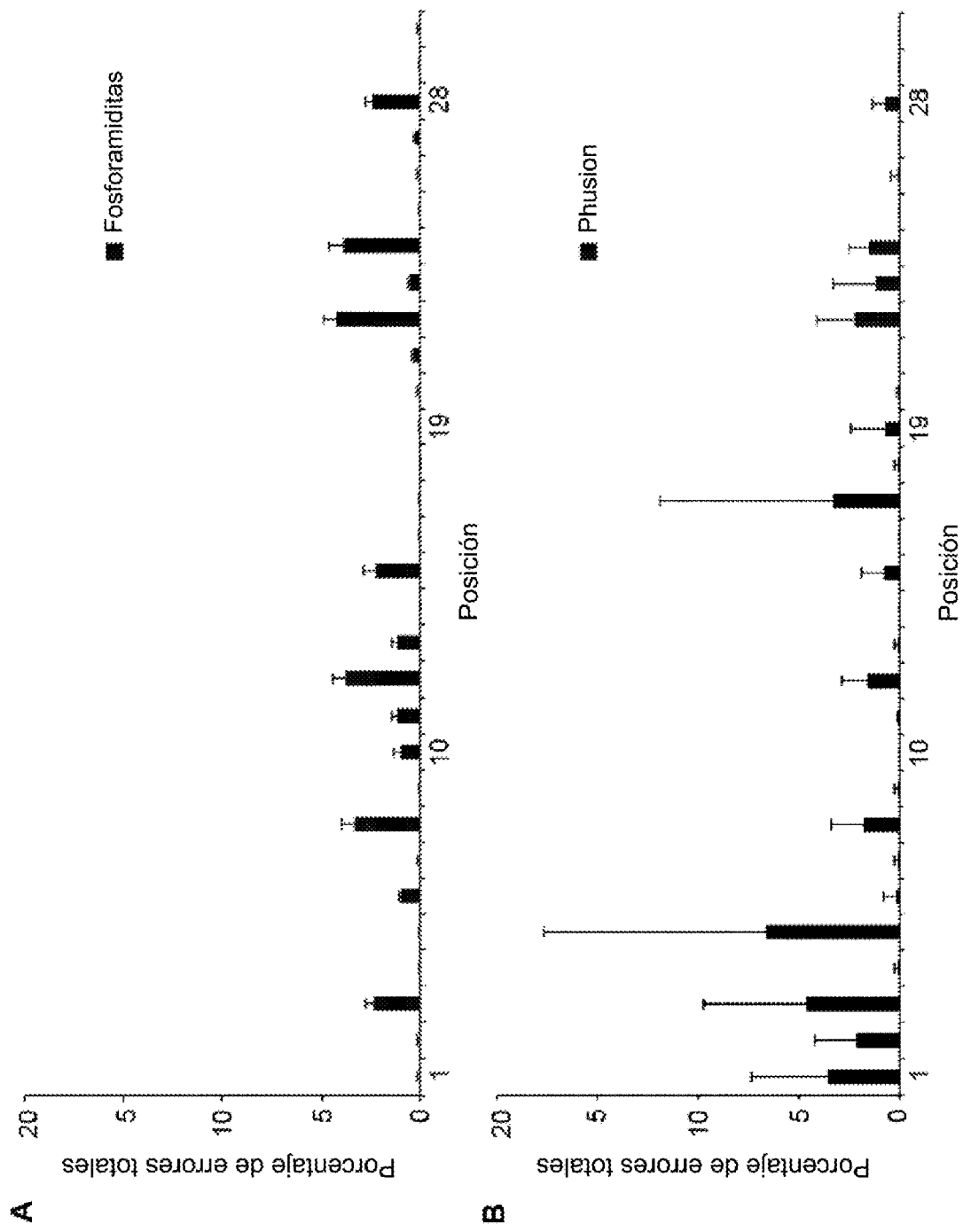


Figura 6A-6B

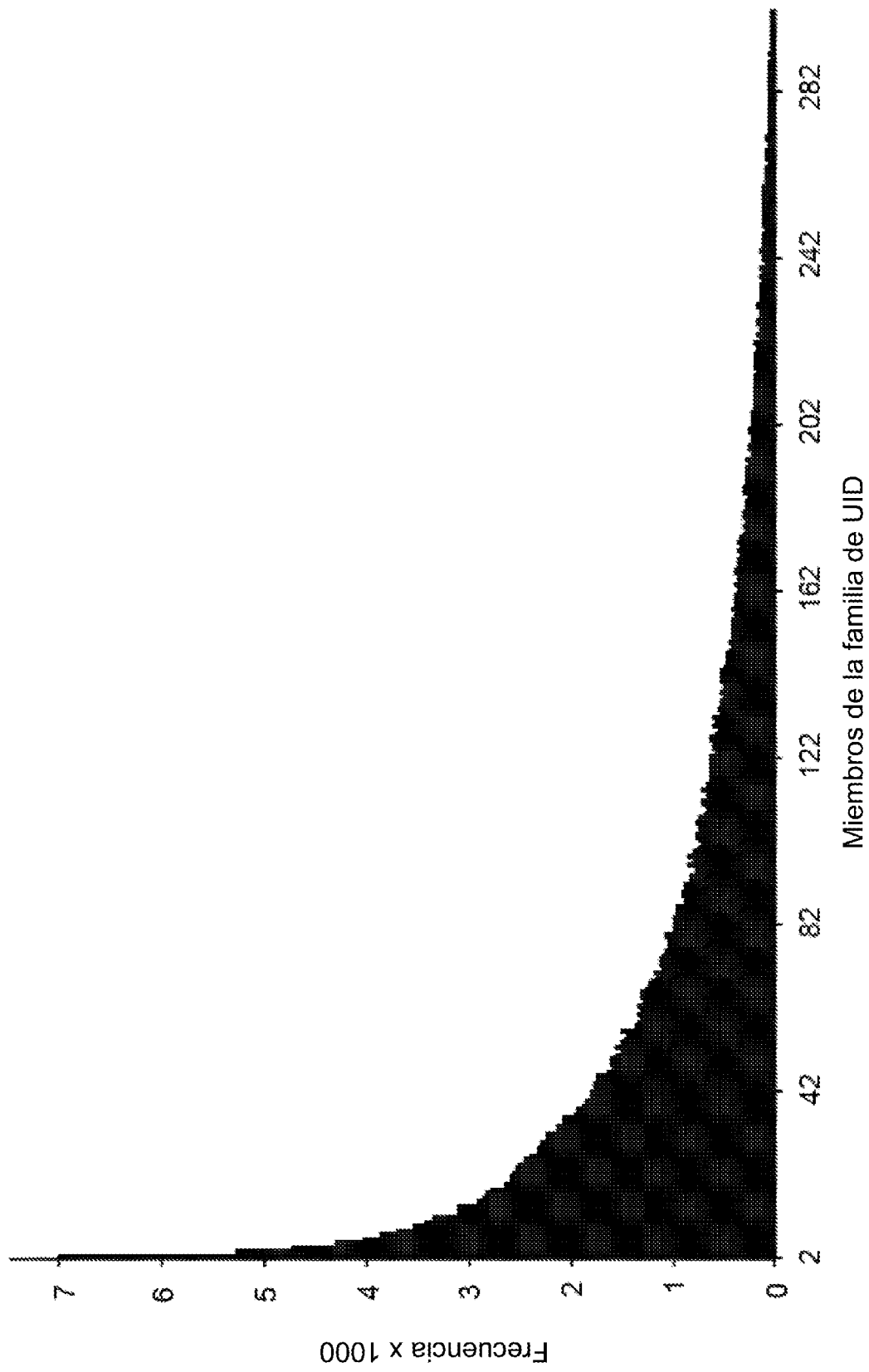


Figura 7