(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 947 437**

(51) Int. CI.:

C12N 15/10 (2006.01)
C12N 15/66 (2006.01)
C12N 9/00 (2006.01)
C40B 20/04 (2006.01)
C40B 70/00 (2006.01)
C12Q 1/6806 (2008.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **07.05.2019 PCT/US2019/031161**

(87) Fecha y número de publicación internacional: **14.11.2019 WO19217452**

(96) Fecha de presentación y número de la solicitud europea: **07.05.2019 E 19799951 (9)**

(97) Fecha y número de publicación de la concesión europea: **19.04.2023 EP 3790967**

(54) Título: **Creación de Códigos de barra compartidos en el ADN, en un solo tubo con perlas, para la secuenciación, haplotipado y ensamblaje preciso y rentable**

(30) Prioridad:

08.05.2018 US 201862668757 P
16.05.2018 US 201862672501 P
19.06.2018 US 201862687159 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
09.08.2023

(73) Titular/es:

MGI TECH CO., LTD. (50.0%)
11F-2, Complex Building of Bei Shan, Industrial Zone, No 146 Bei Shan Road, Yantian District Shenzhen, Guangdong 518083, CN y
BGI SHENZHEN (50.0%)

(72) Inventor/es:

DRMANAC, RADOJE T.;
PETERS, BROCK A. y
WANG, OU

(74) Agente/Representante:

SÁEZ MAESO, Ana

DESCRIPCIÓN

Creación de Códigos de barra compartidos en el ADN, en un solo tubo con perlas, para la secuenciación, haplotipado y ensamblaje preciso y rentable

Referencia a solicitudes anteriores

Antecedentes

Hasta la fecha, la gran mayoría de las secuencias individuales del genoma completo carecen de información sobre el orden de las variantes, de una sola o varias bases, transmitidas como bloques contiguos en los cromosomas homólogos. Recientemente se han desarrollado numerosas tecnologías para permitir esto. La mayoría se basan en el proceso de creación de códigos de barras compartidos (13), es decir, la adición del mismo código de barras a los subfragmentos de moléculas únicas de ADN genómico largo. Después de la secuenciación, la información del código de barras puede usarse para determinar qué lecturas se derivan de la molécula de ADN largo original. Este proceso se describió por primera vez por Drmanac (14) y se implementó por Peters y otros, como un ensayo de placas de 384 pocillos. (6). Sin embargo, estos enfoques son difíciles de implementar técnicamente, costosos, tienen una calidad de datos más baja, no proporcionan la creación de código de barras compartidos únicos o alguna combinación de los cuatro. En la práctica, la mayoría de estos enfoques requieren que se genere una secuencia del genoma completo por separado mediante métodos estándar para mejorar la asignación de variantes. Esto ha resultado en el uso limitado de estos métodos, ya que el costo y la facilidad de uso son factores dominantes en las tecnologías que se usan para WGS. El documento WO 2016/061517 describe métodos y composiciones para preparar una biblioteca inmovilizada de fragmentos de ADN con código de barras de un ácido nucleico diana, identificar variantes genómicas, determinar la información de contigüidad, la información de la determinación de haplotipos y el estado de metilación del ácido nucleico diana.

Resumen de la invención

La invención proporciona un método para preparar una biblioteca de secuenciación para secuenciar un ácido nucleico diana sin el uso de nanogotas como se especifica en la reivindicación 1.

Figuras y tablas

Figuras 1(A) a 1(D). Descripción general de stLFR. Figura 1(A) La primera etapa de la stLFR implica insertar una secuencia de hibridación aproximadamente cada 200-1000 pares de bases en las moléculas de ADN genómico largas. Esto se logra mediante el uso de transposones. Luego, el ADN con el transposón integrado se mezcla con perlas que contienen aproximadamente 400 000 copias, cada una, de una secuencia adaptadora que contiene un código de barras único compartido por todos los adaptadores en la perla, un sitio de cebador de PCR común y una secuencia de captura común que es complementaria a la secuencia de los transposones integrados. Después de capturar el ADN genómico en las perlas, los transposones se ligan a los adaptadores con código de barras. Hay algunas etapas adicionales de procesamiento de la biblioteca y luego los subfragmentos con código de barras compartido se secuencian en un secuenciador BGISEQ-500 o equivalente a este. Figura 1(B) El mapeo de datos de lectura mediante código de barras resulta en un agrupamiento de lecturas dentro de regiones de 10 a 350 kb del genoma. La cobertura total y la cobertura de códigos de barras de 4 códigos de barras se muestran para la biblioteca stLFR-1 de 1 ng a través de una pequeña región en el Chr11. La mayoría de los códigos de barras se asocian con un solo agrupamiento de lectura en el genoma. Figura 1(C) El número de fragmentos de ADN largos originales, por código de barras, se grafica para las bibliotecas de 1ng stLFR-1 y stLFR-2 (naranja) y las bibliotecas de 10 ng stLFR-3 y stLFR-4. Más del 80 % de los fragmentos de las bibliotecas stLFR de 1 ng están codificados con un código de barras compartido mediante un solo código de barras único. Figura 1(D) La fracción de lecturas de secuencias sin superposición y los subfragmentos capturados (naranja) que cubren cada fragmento de ADN largo original se grafican para la biblioteca stLFR-1 de 1 ng. Véase también la Figura 14,

Figuras 2(A) a 2(D). Detección de SV. Las deleciones informadas anteriormente en NA12878 también se encontraron mediante el uso de los datos de stLFR. Los mapas de calor de la compartición de códigos de barras para cada deleción pueden encontrarse en la Figura 10. Figura 2(1) Un mapa de calor de la compartición de códigos de barras dentro de ventanas de 2 kb para una región con una deleción heterocigótica de ~ 150 kb en el cromosoma 8, se graficó mediante el uso de un índice de Jaccard, como se describió anteriormente (12). Las regiones de alta superposición se representan en rojo oscuro. Las que no tienen superposición en beige. Las flechas demuestran cómo las regiones que están espacialmente distantes entre sí en el cromosoma 8 tienen una mayor superposición que marca las ubicaciones de la deleción. Figura 2(B) Las lecturas con código de barras compartido se separan por haplotipo y se grafican mediante por el código de barras único en el eje y la posición en el cromosoma 8 en el eje x. La deleción heterocigótica se encuentra en un solo haplotipo. Figura 2(C) También se graficaron los mapas de calor para los códigos de barras superpuestos entre los cromosomas 5 y 12 para una línea celular de un paciente con una translocación conocida (26) y la Figura 2(D) GM20759, una línea celular con una transversión conocida en el cromosoma 2 (27).

Figuras 3(A) y 3(B): Gráficos de distribución de cobertura. La cobertura se graficó para una biblioteca stLFR-2 (A) y una estándar (B) secuenciada en un BGISEQ500. La cobertura se submuestreó a 30X para ambas muestras. La distribución de Poisson para un genoma 30X se graficó en azul.

Figuras 4(A) y 4(B): Superposición de FP entre bibliotecas. (A) Los FP de cada biblioteca stLFR, la biblioteca estándar BGISEQ-500 y una biblioteca sin PCR secuenciada mediante Illumina (biblioteca "HiSeq2500-TruSeq_PCR-Free_DNA_2x251_NA12878" descargada de BaseSpace) se graficaron en un diagrama de Venn. Se comparten 2078 FP entre las cuatro bibliotecas stLFR. (B) La superposición de los FP de la biblioteca stLFR y los FP de la biblioteca Chromium muestra que se comparten 1194 FP entre las dos tecnologías diferentes que usaron ADN aislado de GM12878 en lugar del material de referencia de GIAB para NA12878. Hay 884 FP que son únicos de las bibliotecas stLFR.

Figuras 5(A) a 5(D): métrica de las variantes de stLFR-1. Se analizaron la profundidad de lectura y la profundidad del código de barras para los alelos variantes y de referencia para todas las variantes verdaderas positivas, variantes falsas positivas y variantes falsas positivas compartidas (verde). La profundidad de lectura para los alelos de referencia (A) y alternativos (B) se grafican al igual que los recuentos de códigos de barras para los alelos de referencia (C) y alternativos (D). En general, los falsos positivos compartidos se parecen más a los verdaderos positivos, lo que sugiere que existen algunos criterios de filtrado que pueden diferenciar entre estas variantes y los falsos positivos no compartidos.

Figuras 6(A) a 6(D): métrica de las variantes de stLFR-3. Se analizaron la profundidad de lectura y la profundidad del código de barras para los alelos variantes y de referencia para todas las variantes verdaderas positivas, variantes falsas positivas y variantes falsas positivas compartidas (verde). La profundidad de lectura para los alelos de referencia (A) y alternativos (B) se grafican al igual que los recuentos de códigos de barras para los alelos de referencia (C) y alternativos (D). En general, los falsos positivos compartidos se parecen más a los verdaderos positivos, lo que sugiere que existen algunos criterios de filtrado que pueden diferenciar entre estas variantes y los falsos positivos no compartidos.

Figuras 7(A) y 7(B): Distribución de variantes falsas positivas compartidas. La distancia genómica que separa 2078 variantes FP compartidas se sumó dentro de intervalos consecutivos de 100 pb (azul oscuro), 1000 pb (naranja), 10 000 pb, 100 000 pb y 1 millón de pb. También se graficaron 5 conjuntos de 2078 variantes seleccionadas al azar de la biblioteca stLFR-1. Para cada muestra se grafica el número total de ubicaciones o el número total de variantes. Solo se suman los intervalos o las variantes dentro de los intervalos donde se encuentran 2 o más variantes. (A) Antes del filtrado, hay 219 FP compartidos que parecen estar agrupados estrechamente y probablemente sean el resultado de errores de mapeo. Las 1859 variantes restantes parecen compartir una distribución similar a los conjuntos aleatorios de variantes. (B) Después de filtrar, quedan 1738 FP compartidos, pero solo 72 se agrupan estrechamente.

Figuras 8(A) a 8(T): Detección de la deleción NA12878 mediante el uso de mapas de calor de compartición de códigos de barras. Detección de deleciones en la biblioteca stLFR-1 en chr3:65189000-65213999 mediante el uso de 230 Gb (A) o 100 Gb (B), chr4:116167000-116176999 mediante el uso de 230 Gb (C) o 100 Gb (D), chr4:187094000-187097999 230 Gb (E) o 100 Gb (F), chr7:110182000-110187999 230 Gb (G) o 100 Gb (H), chr16:62545000-62549999 230 Gb (I) o 100 Gb (J), chr1:189704509-189783359 230 Gb (K) o 100 Gb (L), chr3:162512134-162569235 230 Gb (M) o 100 Gb (N), chr5:104432113-104467893 230 Gb (O) o 100 Gb (P), chr6:78967194-79001807, y chr8:39232074-39309652 230 Gb (S) o 100 Gb (T) de los datos de lectura.

Figuras 9(A) a 9(L): Detección de la translocación y la inversión con stLFR. Una línea celular del paciente y la línea celular GM20759 que tienen una translocación entre los cromosomas 5 y 12 y una inversión en el cromosoma 2, respectivamente, se analizaron con stLFR. Para cada biblioteca, se submuestreó la cobertura de secuencia total para investigar la capacidad de detección a coberturas más bajas. La translocación entre los cromosomas 12 y 5 se detectó fácilmente a coberturas de secuencia total de 40 Gb (A), 20 Gb (B), 10 Gb (C) e incluso 5 Gb (D). La inversión en GM20759 también se detectó fácilmente a coberturas de secuencia total de 46 Gb (E), 20 Gb (F), 10 Gb (G) y 5 Gb (H). Además, investigamos estas regiones en la línea celular GM12878, que no se conoce que tenga ninguna de estas SV. La translocación entre los cromosomas 5 y 12 no fue evidente ni en la biblioteca stLFR de 1 ng con 230 Gb de cobertura (I) ni en la biblioteca de 10 ng con 126 Gb de cobertura (J). La transversión tampoco se encontró en la biblioteca stLFR-1 (K) o stLFR-4 (L).

Figura 10(A) a 10(C): Matriz de puntos de alineamiento de los andamios NA12878. Los andamios de SALSA de las bibliotecas stLFR-1 (A) y stLFR-4 (B) se graficaron frente a hg37 del genoma humano de referencia. 734 millones de lecturas HiC de Dixon y otros, (29) también se usaron para generar andamios y también se graficaron frente a hg37 (C). En todos los casos, solo se graficaron los andamios que cubrían el 5 % o más de un cromosoma.

Figura 11: Determinación de haplotipos con LongHap. Puede encontrarse una descripción completa del algoritmo de determinación de haplotipos aplicado con LongHap en la sección Métodos y materiales.

Figura 12: Ensamblaje de secuencias con código de barras. Se usan tres ligazones, necesarias para generar ~3,6 billones de códigos de barras diferentes. La secuencia esperada en cada etapa del ensamblaje del código de barras se muestra como las SEQ. ID NO: 1 a 13.

Figura 13: Diagrama de flujo del ensamblaje de la secuencia con código de barras.

Figura 14: Diagrama de flujo ilustrativo del protocolo de código de barras.

Figura 15: Representación gráfica de la etapa de hibridación.

Figura 16: Representación gráfica de las etapas de ligazón y degradación. Las etapas finales que muestran la desnaturalización y la adición de la cola C son opcionales y no se describen adicionalmente en la presente descripción.

Figura 17: Producción y uso de las DNB de doble hebra.
Figura 18: Amplificación de moléculas largas.
Figura 19: Métodos con nickasa aleatoria.
Figura 20: Métodos de adaptador de horquilla.
Figuras 21(A) y 21(B): La Figura 21(A) es una representación esquemática del ensayo de ligazón en diferentes sustratos de ADN. El donante de ADN de extremo romo es una molécula parcialmente de ADNdh, sintética, con extremos didesoxi 3' (círculos rellenos) para evitar la autoligazón del adaptador. El brazo largo del adaptador está fosforilado en 5'. Los aceptores de ADN se ensamblaron mediante el uso de 2 o 3 oligos (líneas negras, rojas y naranjas) para formar una mella (sin fosfatos), una brecha (1 u 8 nt) o un extremo saliente 5' de 36 nt. Todas las hebras de los sustratos están desfosforiladas y la hebra andamio está protegida con didesoxi en 3'. La Figura 21(B) muestra el análisis del cambio de tamaño de los productos ligados mediante el uso de un gel de poliacrilamida desnaturalizante al 6 %. Los controles negativos sin ligasa (carriles 1, 3, 4, 6, 7, 9, 10, 12 y 13) se cargaron a 1 o 0,5 veces el volumen de sus pruebas experimentales correspondientes. Si ocurre la ligazón, el tamaño del sustrato aumenta en 22 nt. Las puntas de flecha rojas corresponden al sustrato y las puntas de flecha azules corresponden a los sustratos ligados al adaptador. M2 = Patrón de ADN de 25 pb de Thermo Fisher (c) Tabla de los tamaños esperados de los productos de ligazón y estimación de la eficiencia de ligazón mediante el uso de ImageJ. La tasa de eficiencia de la ligazón se estimó al dividir la intensidad de los productos ligados por la intensidad total de los productos ligados y no ligados.
Figuras 22(A) a 22(D): Análisis en gel del cambio de tamaño de los productos ligados mediante el uso de un gel de poliacrilamida al 6 % en TBE. Las puntas de flecha rojas corresponden al sustrato y las puntas de flecha azules corresponden a los sustratos ligados al adaptador: Mella (A, izquierda), saliente 5' (A, derecha), brecha de 1 nt (B), brecha de 2 nt (C) y brecha de 3 nt (D). M2 = patrón de ADN de 25 pb de Thermo Fisher. Se compararon dos secuencias adaptadoras (Ad1 y Ad2) y también se examinaron diferentes bases (A o G) en el extremo 5' de la unión de ligazón de Ad2. ** (e) Tabla de eficiencia de ligazón calculada en base a la intensidad de la banda mediante el uso de ImageJ.
Figuras 23(A) y 23(B): La Figura 23(A) muestra una representación esquemática de la ligazón de rama 3' en un híbrido de ADN/ARN con una región complementaria de 20 pb. Probamos si los adaptadores de extremos romos se ligarían al extremo 3' del ADN en el saliente 5' del ARN y/o al extremo 3' del ARN en el saliente 5' del ADN. (B) Análisis en gel del cambio de tamaño de los productos ligados mediante el uso de un gel de poliacrilamida desnaturalizante al 6 %. La punta de flecha roja corresponde al sustrato de ARN (29 nt) y la punta de flecha verde corresponde al sustrato de ADN (80 nt). La punta de flecha azul corresponde a los sustratos de ARN ligados al adaptador. Si ocurre la ligazón, el tamaño del sustrato aumentaría en 20 nt. Las reacciones 1 y 2 son duplicados. M2 = patrón de ADN de 25 pb de Thermo Fisher.
Figuras 24(A) a 24(C): La Figura 24(A) es una representación esquemática de la inserción del transposón seguida de la ligazón de rama 3'-3' y la amplificación por PCR mediante el uso de Pr-A (flecha azul) y Pr-B (flecha verde). (B) Productos de amplificación después de la inserción del transposón con TnA y/o TnB y/o ligazón de rama 3' de AdB mediante el uso de los cebadores pr-A, pr-B o ambos. Los productos se procesaron en un gel de poliacrilamida al 6 %. M1 = Patrón de ADN de bajo intervalo ThermoFisher MassRuler. (C) Gráfico de la señal de amplificación mediante el uso de pr-A y pr-B después de diversas condiciones de inserción del transposón y ligazón de rama 3'.
Figura 25: Etiquetado de longitud intermedia.
Figuras 26(A) y (B): Ligazón de rama 3' mediante la ligasa de ADN T4 en extremos de ADN no convencionales formados por mellas, brechas y salientes. (A) Representación esquemática del ensayo de ligazón en diferentes tipos de aceptores de ADN. El donante de ADN de extremo romo es una molécula parcialmente de ADNdh, sintética, con extremos didesoxi 3' (círculos rellenos) para evitar la autoligazón del donante de ADN. El brazo largo del donante está fosforilado en 5'. Los aceptores de ADN se ensamblaron mediante el uso de 2 o 3 oligos para formar una mella (sin fosfatos), una brecha (1 u 8 nt) o un extremo recesivo 3' de 36 nt. Todas las hebras de los sustratos están desfosforiladas, y la hebra andamio está protegida con didesoxi en 3'. (B) Análisis del cambio de tamaño de los productos ligados de los sustratos 1, 2, 3 y 4, respectivamente, mediante el uso de un gel de poliacrilamida desnaturalizante al 6 %. Los controles negativos sin ligasa (carriles 1, 3, 4, 6, 7, 9, 10, 12 y 13) se cargaron a 1 o 0,5 veces el volumen de los ensayos experimentales correspondientes. Si ocurre la ligazón, el tamaño del sustrato aumenta en 22 nt. Las puntas de flecha rojas corresponden al sustrato y las puntas de flecha moradas corresponden a los sustratos ligados al donante. Se usó el patrón de ADN de 25 pb de Thermo Fisher. Secuencias de donantes y sustratos en la Tabla S1. La TABLA 8 muestra los tamaños esperados del sustrato y el producto de la ligazón y la eficiencia de ligazón aproximada en cada grupo experimental. La intensidad de cada banda se estimó mediante el uso de ImageJ y se normalizó por su tamaño esperado. La eficiencia de la ligazón se estimó al dividir la intensidad normalizada de los productos ligados por la intensidad total normalizada de los productos ligados y no ligados.
Figuras 27(A) a 27(E): Análisis en gel del cambio de tamaño de los productos ligados mediante el uso de un gel de poliacrilamida al 6 % en TBE. Las puntas de flecha rojas corresponden al sustrato, y las puntas de flecha moradas corresponden a sustratos ligados a donantes: sustrato 5 (mella) (A), sustrato 6 (brecha de 1 nt) (B), sustrato 7 (brecha de 2 nt) (C), sustrato 8 (brecha de 3 nt), (D) y sustrato 9 (extremo recesivo 3') (E). Se usó el patrón de ADN de 25 pb de Thermo Fisher. Se examinaron tres donantes de ADN con diferentes bases en el extremo 5' de la unión de la ligazón (T, A o GA). En la TABLA 9 se muestra la eficiencia de ligazón calculada en base a la intensidad de la banda normalizada mediante el uso de ImageJ.

Figuras 28(A) a 28(D). Ligazón de rama 3' en el extremo 3' del ARN en el híbrido ADN/ARN. Representación esquemática de la ligazón de rama 3' en un híbrido de ADN/ARN con una región complementaria de 20 pb. Probamos si los donantes de ADN de extremo romo se ligarían al extremo recesivo 3' del ADN y/o al extremo recesivo 3' del ARN. El ADN (ON-21) se hibrida con la hebra de ARN (A), mientras que el ADN (ON-23) no puede hibridar con la hebra de ARN (B). Las Figuras 28 (C) y (D) muestran el análisis en gel del cambio de tamaño de los productos ligados mediante el uso del gel de poliacrilamida desnaturalizante al 6 %. Las puntas de flecha rojas corresponden al sustrato de ARN (29 nt) y la punta de flecha verde corresponde al sustrato de ADN (80 nt). La punta de flecha morada corresponde a sustratos de ARN ligados a donantes. Si ocurre la ligazón, el tamaño del sustrato aumentaría en 20 nt. (c) Carril 1 y 2, duplicados experimentales; carriles 7-10, controles sin ligasa; se añadió PEG al 10 % con la ligasa de ADN T4. (d) Carril 1, control sin ligasa; carril 2, 3 y 8, ligasa de ADN T4 con PEG al 10 %; carril 4, 5 y 9, ligasa de ARN T4 1 con DMSO al 20 %; carril 6, 7 y 10, ligasa de ARN T4 2 con DMSO al 20 %. Se usó el patrón de ADN de 25 pb de Thermo Fisher. Puede corresponder a la Figura 23, pero no es exacto en múltiples formas.

Las Figuras 29(A) a 29(C) presentan una representación esquemática de tres métodos de tagmentación por transposones seguidos de amplificación por PCR mediante el uso de Pr-A (flecha azul) y Pr-B (flecha verde). Método de dos transposones (A); tagmentación por transposón Y con relleno de brechas 3' (B); método de un transposón con ligazón del adaptador en la brecha 3'(C). La Figura 29(D) es un gráfico de la señal de amplificación después de la purificación mediante el uso de pr-A o pr-A con pr-B después de las diversas condiciones de ligazones de la brecha y tagmentación. Puede corresponder a la Figura 23, pero no es exacto en múltiples formas.

Figuras 30(A) a 30(C). Sesgo de distribución de bases de la ligazón de brecha Tn5 (A), dos transposones (B) y ligazón TA regular (C). Solo se presentan las primeras 20 bases de cada extremo de la ligazón; adenina, azul; citosina, naranja; guanina, gris; timina, amarillo; se presentan el promedio y la desviación estándar de cinco bibliotecas independientes. No presente en absoluto.

Figuras 31(A) y 31(B). Ligazón de rama 3' del ADN con diferentes condiciones de aditivos. (A) ligazones en el ADN saliente 5' a concentraciones de ATP tituladas. Se realizaron duplicados para ATP 0,01 mM (carriles 4 y 5) y 0,005 mM (carriles 6 y 7). El carril 9 es un control sin donantes. (B) Ligazón de rama 3' del ADN en la mella, la brecha de 1 nt, la brecha de 8 nt, el saliente 5' y el extremo romo, con o sin SSB y ligasa. Las puntas de flecha rojas corresponden al sustrato y las puntas de flecha moradas corresponden a los sustratos ligados al donante. No presente en absoluto.

Tabla 1: Estadísticas de determinación de haplotipos y asignación de variantes. Las lecturas se mapearon a Hg37 con secuencia señuelo y las variantes se asignaron con GATK, con configuraciones predeterminadas, para todas las bibliotecas, excepto donde se describa de cualquier otra manera. Los SNP del VCF de las asignaciones de variantes de alta confianza de GIAB se usaron como entrada para la determinación de haplotipos.

Tabla 2: Estadísticas de andamios.

Tabla 3: El filtrado reduce las asignaciones de falsos positivos. Las asignaciones de FP finales se calcularon al sustraer 1666 de los FP filtrados, excepto para la biblioteca STD que, por definición, no compartió ninguno de estos FP con las bibliotecas stLFR porque se hizo con material de referencia de GIAB.

Tabla 4: Determinación de haplotipos SNP e indel con LongHap.

Tabla 5: Criterios de filtrado. Se usaron diversos criterios de filtrado, explicados en la sección Materiales y métodos, para eliminar los FP.

TABLA 6: Secuencias ilustrativas.

Descripción detallada

1. Proceso de la biblioteca stLFR

1.1 Introducción

Aquí describimos una implementación de la tecnología de lectura de fragmentos largos (stLFR) en un solo tubo (15), un enfoque eficiente para la creación de códigos de barras compartidos en el ADN, con millones de códigos de barras habilitados en un solo tubo. Véase el documento WO 2014/145820 A2 (2014). Esto se logra mediante el uso de la superficie de una microperla como reemplazo de un compartimento (por ejemplo, el pocillo de una placa de 384 pocillos). Cada perla porta muchas copias de una secuencia de código de barras única que se transfiere a los subfragmentos de cada molécula de ADN largo. Estos subfragmentos con códigos de barras compartidos se analizan luego en dispositivos de secuenciación de lecturas cortas comunes, tales como el BGISEQ-500 o equivalente a este. En nuestra implementación de este enfoque, usamos una estrategia de generación de códigos de barras combinatoria basada en la ligazón para crear más de 1,8 billones de códigos de barras diferentes en tres etapas de ligazón. Para una sola muestra usamos ~10-50 millones de estas perlas con código de barras para capturar ~10-100 millones de moléculas de ADN largo en un solo tubo. Es poco frecuente que dos perlas compartan

el mismo código de barras porque tomamos muestras de 10-50 millones de perlas a partir de una biblioteca tan grande de códigos de barras totales. Además, en el caso de usar 50 millones de perlas y 10 millones de fragmentos largos de ADN genómico, la gran mayoría de los subfragmentos de cada fragmento de ADN largo está codificado con código de barras compartido mediante un código de barras único. Esto es análogo a la secuenciación de lecturas largas de una sola molécula y, potencialmente, permite enfoques informáticos poderosos para el ensamblaje *de novo*. Es importante destacar que, la stLFR es simple de realizar y puede implementarse con una inversión relativamente pequeña en oligonucleótidos para generar las perlas con código de barras. Además, la stLFR usa equipos estándar que se encuentran en casi todos los laboratorios de biología molecular y puede analizarse mediante casi cualquier estrategia de secuenciación. Finalmente, la stLFR reemplaza los métodos estándar de preparación de bibliotecas NGS, requiere solo 1 ng de ADN y no aumenta significativamente el costo de los análisis de genoma completo o exoma completo, con un costo total por muestra de menos de 30 dólares.

Como se usa en la presente descripción, "un solo tubo" se refiere al análisis de un gran número de fragmentos de ADN individuales sin necesidad de separar los fragmentos en tubos, recipientes, alícuotas, pocillos o gotas separadas durante las etapas de etiquetado. En cambio, la superficie de una microperla sirve como reemplazo de un compartimento.

La primera etapa en la stLFR es la inserción de una secuencia de hibridación, preferentemente a intervalos regulares, a lo largo de fragmentos de ADN genómico. Los intervalos adecuados pueden variar con la aplicación y el resultado deseado, pero están, típicamente, en el intervalo de 100-1500 pb, a menudo de 200-1000 pb. Esto se logra a través de la incorporación de secuencias de ADN mediante transposición. En una modalidad, la transposasa es la Tn3, Tn5, Tn7 o Mu. A menudo, se usa una transposasa Tn5 (véase Picelli y otros, 2014). El ADN transpuesto, o la secuencia de inserción, comprende una región de simple hebra para la hibridación ("secuencia de hibridación"), así como también una secuencia mosaico de doble hebra que es reconocida por la enzima y permite la reacción de transposición (Figura 1A). Esta etapa de transposición se hace en solución (en lugar de tener la secuencia de inserción enlazada directamente a la perla). Esto permite una incorporación muy eficiente de la secuencia de hibridación a lo largo de las moléculas de ADN genómico. Como se observó anteriormente (10), la enzima transposasa tiene la propiedad de permanecer unida al ADN genómico después del evento de transposición y deja efectivamente intacta la molécula de ADN genómico larga con el transposón integrado.

Después de que el ADN se ha tratado con, por ejemplo, Tn5, se diluye en tampón de hibridación y se combina con perlas con códigos de barras clonales. En un enfoque (Ejemplos, más abajo), se usan 50 millones de perlas de ~ 2,8 µm con códigos de barras clonales, en tampón de hibridación. Cada perla contiene aproximadamente 400 000 adaptadores de captura (también llamados oligos de captura u oligonucleótidos de captura) y cada uno contiene la misma secuencia de código de barras. Una porción del adaptador de captura contiene nucleótidos de uracilo para permitir la destrucción de los adaptadores no usados en una etapa posterior. Por ejemplo, el adaptador de captura puede tener 5-50 % de uracilo, más a menudo 5-50 %, más a menudo 5-20 %. La mezcla se incuba en condiciones óptimas de temperatura y tampón, durante este tiempo el ADN insertado con el transposón se captura en las perlas a través de la secuencia de hibridación.

Se ha sugerido que el ADN genómico en solución forma bolas con ambas colas que sobresalen (16). Esto puede permitir la captura de fragmentos largos de ADN hacia un extremo de la molécula, seguido de un movimiento giratorio que envuelve la molécula de ADN genómico alrededor de la perla. Aproximadamente cada 7,8 nm en la superficie de cada perla hay un oligo de captura. Esto permite una tasa alta y muy uniforme de captura de subfragmentos. Un fragmento genómico de 100 kb envolvería una perla de 2,8 µm aproximadamente 3 veces. En nuestros datos, el tamaño de fragmento más largo capturado es 300 kb, lo que sugiere que pueden ser necesarias perlas más grandes para capturar moléculas de ADN más largas.

En modalidades alternativas, pueden variar parámetros tales como el tamaño de las perlas, la separación de los oligonucleótidos de captura o el número de oligos diferentes por mezcla. Por ejemplo, las perlas usadas pueden tener un diámetro en el intervalo de 1-20 µm, alternativamente de 2-8 µm, 3-6 µm o 1-3 µm. Por ejemplo, la separación de los oligos con código de barras en las perlas puede ser de al menos 1, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6 o al menos 7 nm. En algunas modalidades, la separación es menor que 10 nm (por ejemplo, 5-10 nm), menor que 15 nm, menor que 20 nm, menor que 30 nm, menor que 40 nm o menor que 50 nm. En algunas modalidades, el número de códigos de barras diferentes usados, por mezcla, puede ser >1 M, >10 M, >30 M, >100 M, >300 M o >1 B. Como se discute más abajo, puede producirse un gran número de códigos de barras para su uso en la invención, por ejemplo, mediante el uso de los métodos descritos en la presente descripción. En algunas modalidades, el número de códigos de barras diferentes que se usan por mezcla puede ser >1 M, >10 M, >30 M, >100 M, >300 M o >1 B y las muestras se toman de un conjunto con una diversidad al menos 10 veces superior (por ejemplo, de >10 M, >0,1 B, 0,3 B, >0,5 B, >1 B, >3 B, >10 B códigos de barras diferentes en las perlas.)

Las secuencias de códigos de barras individuales se transfieren a intervalos regulares a través de la ligazón del extremo 3' del adaptador de captura al extremo 5' de la secuencia de hibridación con el transposón insertado, mediada por un oligonucleótido puente o férula (términos usados indistintamente) con una primera región complementaria al adaptador de captura y una segunda región complementaria a la secuencia de hibridación (Figura

1A y Figura 15). Las perlas se recolectan y los complejos de ADN/transposasa se rompen, lo que produce subfragmentos de menos de 1 kb de tamaño.

Si se desea, la creación de códigos de barras de muestra se puede lograr en esta etapa. Se usan transposones que portan un código de barras único entre la secuencia mosaico y la secuencia de hibridación. Estos pueden sintetizarse en formato de placa de 96, 384 o 1536 y cada pocillo contiene muchas copias de un transposón que porta el mismo código de barras y cada código de barras es diferente entre los pocillos. Pueden insertarse transposones en diferentes muestras de ADN en formato de placa 96, 384 o 1536 mediante el uso de estos transposones con código de barras. Las muestras etiquetadas con el código de barras de muestra se pueden combinar de cualquier manera.

Debido al gran número de perlas y la alta densidad de oligos de captura por perla, la cantidad de exceso de adaptador es cuatro órdenes de magnitud mayor que la cantidad de producto. Este enorme cantidad de adaptador sin usar puede abrumar las siguientes etapas. Para evitar esto, diseñamos perlas con oligos de captura conectados por el extremo 5'. Esto permitió desarrollar una estrategia de exonucleasa que degradaba específicamente el exceso de oligonucleótidos de captura sin usar. Véanse las Figuras 14 y 16. También puede usarse una uracil-ADN glucosilasa (UDG) para degradar el exceso de adaptadores.

En un aspecto, el método incluye combinar en una sola mezcla (i) los primeros fragmentos del ácido nucleico diana y (ii) una población de perlas, en donde cada perla comprende oligonucleótidos inmovilizados sobre ella, dichos oligonucleótidos comprenden una secuencia que contiene una etiqueta (o adaptadores con código de barras), en donde cada secuencia que contiene la etiqueta comprende una secuencia etiqueta, en donde los oligonucleótidos inmovilizados en la misma perla individual comprenden la misma secuencia que contiene la etiqueta y la mayoría de las perlas tienen secuencias etiqueta diferentes. En algunas modalidades, los fragmentos de ADN son concatémeros de al menos 2, al menos 10, al menos 30 o al menos 100 copias de moléculas de ADN o ADNc. Los monómeros de ácido nucleico pueden tener una longitud de 0,5 kb a 10 kb, o son > 1 kb, o tienen una longitud > 10 kb. En algunos enfoques, la secuencia se determina para >50 % o >70 %, >90 %, 95 %, >99 %, 100 % de bases de las moléculas de ADN o ADNc en una mezcla.

1.1.1 Métodos de dos transposones

En un enfoque de stLFR, se usan dos transposones diferentes en la etapa de inserción inicial, lo que permite realizar la PCR después del tratamiento con exonucleasa. Sin embargo, este enfoque resulta en aproximadamente un 50 % menos de cobertura por molécula de ADN larga, ya que requiere que se inserten dos transposones diferentes uno al lado del otro para generar un producto de PCR apropiado.

1.1.2 Métodos de un solo transposón mediante el uso de la ligazón de rama 3'

Para lograr la cobertura más alta por fragmento de ADN genómico, usamos un solo transposón en la etapa de inserción inicial y añadimos un adaptador adicional a través de la ligazón. Esta ligazón no canónica, denominada ligazón de rama 3', implica la unión covalente del fosfato 5' de un adaptador de extremo romo al hidroxilo 3' empotrado del ADN genómico (Figura 1A). La ligazón de ramas se describe en el Ejemplo 3, más abajo. Véase también la publicación de patente de Estados Unidos US2018/0044668y la solicitud internacional WO 2016/037418. Véase también, la publicación de patente de Estados Unidos. 2018/0044667. Mediante el uso de este método, teóricamente es posible amplificar y secuenciar todos los subfragmentos de una molécula genómica capturada.

Además, esta etapa de ligazón permite colocar un código de barras de muestra adyacente a la secuencia genómica para la combinación de más de una muestra. El beneficio de usar estos adaptadores para la creación de códigos de barras de muestra es que el código de barras se puede colocar adyacente al ADN genómico, de modo que puede usarse el mismo cebador para secuenciar el código de barras y el ADN genómico y no se requiere ningún cebador de secuenciación adicional para leer el código de barras. La creación de códigos de barras de muestra permite que las preparaciones de múltiples muestras se combinen antes de las secuencias y se distingan por el código de barras. Los adaptadores de ligazón de rama 3' pueden sintetizarse en formato de placa de 96, 384 o 1536 y cada pocillo contiene muchas copias del adaptador que portan el mismo código de barras y cada código de barras es diferente entre los pocillos. Después de la captura en las perlas, estos adaptadores pueden usarse para la ligazón en formato de placa 96, 384 o 1536.

Después de esta etapa de ligazón, se realiza la PCR y la biblioteca está lista para ingresar a cualquier flujo de trabajo estándar de secuenciación de próxima generación (NGS). Se apreciará que la PCR (u otra amplificación) puede llevarse a cabo mediante el uso de un primer cebador que se hibrida con un sitio en el oligonucleótido de captura o su complemento (véase la Figura 1A) y un segundo cebador que se hibrida con un sitio en el adaptador de ligazón de rama 3' o su complemento. En el caso del BGISEQ-500, la biblioteca se circulariza como se describió anteriormente (17). A partir de círculos de hebra simple, se fabrican nanobolas de ADN y se cargan en nanomatrices estampadas (17). Estas nanomatrices se someten luego a la secuenciación basada en la síntesis de anclaje de sonda combinatoria (cPAS) en el BGISEQ-500 (18-20). Después de la secuenciación, se extraen las secuencias de códigos de barras. El mapeo de los datos de lectura mediante códigos de barras únicos muestra que la mayoría de

las lecturas con el mismo código de barras se agrupan en una región del genoma correspondiente a la longitud de ADN usada durante la preparación de la biblioteca (Figura 1B). Una descripción detallada de este método, así como también un protocolo para hacer las perlas se describe en los EJEMPLOS 1 y 2.

En algunas modalidades >50 %, >70 %, >80 %, >90 % o >95 % de los fragmentos de ADN con código de barras se codifican con un código de barras único. En algunas modalidades, >50 %, >70 %, >80 % >90 % de los subfragmentos en un fragmento se ligan al oligo de código de barras. En algunas modalidades, >10 % o >20 %, >40 %, >50 %, >60 % de los subfragmentos de fragmentos largos se secuencian, como promedio.

1.2 Cobertura de lectura y asignación de variantes de stLFR

Para demostrar la determinación de haplotipos y la asignación de variantes de stLFR, generamos cuatro bibliotecas con 1 ng (stLFR-1 y stLFR-2) y 10 ng (stLFR-3 y stLFR-4) de ADN de NA12878. Se varió el número de perlas y se usaron 10 millones (stLFR-3), 30 millones (stLFR-4) y 50 millones (stLFR-1 y stLFR-2). Finalmente, se probaron los métodos de ligazón de rama 3' (stLFR-1, stLFR-2 y stLFR-3) y de dos transposones (stLFR-4). Tanto stLFR-1 como stLFR-2 se secuenciaron profundamente a 336 Gb y 660 Gb de cobertura de bases total, respectivamente. También analizamos estos en coberturas submuestreadas. stLFR-3 y stLFR-4 se secuenciaron a niveles más modestos de 117 Gb y 126 Gb, respectivamente. Las lecturas con código de barras compartido se mapearon para construir 37 del genoma de referencia humano mediante el uso de BWA-MEM (21). Debido a que stLFR no requiere ninguna etapa de amplificación previa, la distribución de la cobertura de lectura a través del genoma estuvo cerca de Poisson (Figura 3). La cobertura sin duplicados osciló entre 34-58X y el número de moléculas de ADN largo por código de barras osciló entre 1,2-6,8 (Tabla 1 y Figura 1C). Como se esperaba, las bibliotecas stLFR hechas a partir de 50 millones de perlas y 1 ng de ADN genómico tenían las tasas más altas de creación de códigos de barras compartidos únicos, con más del 80 % (Figura 1C). En estas bibliotecas también se observó la cobertura de lectura sin superposición por molécula de ADN largo promedio más alta de 10,7-12,1 % y la cobertura de bases sin superposición de subfragmentos capturados por molécula de ADN largo promedio más alta de 17,9-18,4 % (Figura 1d). Esta cobertura es~10 veces más alta que lo demostrado anteriormente mediante el uso de 3 ng de ADN y transposones unidos a las perlas (12).

Para cada biblioteca, las variantes se asignaron mediante el uso de GATK (22) y la configuración predeterminada. La comparación de las asignaciones de SNP e indel con Genome in a Bottle (GIAB) (23) permitió determinar las tasas de falsos positivos (FP) y falsos negativos (FN) (TABLA 1). Además, realizamos la asignación de variantes mediante el uso de la misma configuración en GATK en una biblioteca estándar que no es stLFR hecha de~1000 veces más ADN genómico y secuenciada también en un BGISEQ-500 (STD) y una biblioteca Chromium de 10X Genomics (11). También comparamos las tasas de precisión y sensibilidad con las informadas en el estudio de la biblioteca de haplotipado en perlas realizado por Zhang y otros. (12). Nuestro enfoque stLFR y el método descrito por Zhang y otros, demostró tasas más bajas de FP de SNP e Indel que la biblioteca Chromium. stLFR tenía tasas de FP y FN dos veces más altas que la biblioteca STD y, en dependencia de la biblioteca de stLFR en particular y los criterios de filtrado, la tasa de FN era más alta o más baja que la biblioteca Chromium. La tasa de FN más alta en las bibliotecas stLFR en comparación con las bibliotecas estándar se debe principalmente al tamaño de inserción promedio más corto (~200 pb frente a 300 pb en una biblioteca estándar). Dicho esto, stLFR tuvo una tasa de FN mucho más baja que Zhang y otros, para SNP e Indel y una tasa de FN mucho más baja que la biblioteca Chromium para Indel (TABLA 1). En general, la mayoría de las métricas para la asignación de variantes fueron mejores para nuestras bibliotecas stLFR que los resultados publicados por Zhang y otros o las bibliotecas Chromium, especialmente cuando se usaron procesos de mapeo y de asignación de variantes no optimizados (TABLA 1, "Sin filtro").

Un posible problema con el uso de datos de GIAB para medir la tasa de FP es que no pudimos usar el material de referencia de GIAB (NIST RM 8398) debido al tamaño de fragmento bastante pequeño del ADN aislado. Por esta razón, usamos la línea celular GM12878 y aislamos el ADN mediante el uso de un método basado en diálisis capaz de producir ADN de peso molecular muy alto (véanse los métodos). Sin embargo, es posible que nuestro aislado de la línea celular GM12878 pueda tener una serie de mutaciones somáticas únicas en comparación con el material de referencia de GIAB y, por tanto, provoque que el número de FP se infle en nuestras bibliotecas stLFR. Para examinar esto adicionalmente, comparamos la superposición de variantes de FP de un solo nucleótido entre las 4 bibliotecas stLFR y las dos bibliotecas que no son LFR (Figura 4a). En general, se compartieron 544 variantes de FP entre las seis bibliotecas y 2078 FP fueron únicos para las cuatro bibliotecas stLFR. También comparamos los FP de stLFR con la biblioteca Chromium y encontramos que más de la mitad (1194) de estos FP compartidos también estaban presentes en la biblioteca Chromium (Figura 4b). Un examen de la cobertura de lectura y código de barras de estas variantes compartidas mostró que eran más similares a las de las variantes TP (Figura 5-6). También examinamos la distribución a través del genoma de estas variantes de FP compartidas frente a 2078 variantes seleccionadas al azar (Figura 7a). Este análisis mostró 219 variantes que se encuentran en grupos, donde dos o más de estos FP están dentro de los 100 pb uno del otro. Sin embargo, la mayoría (90 %) de las variantes tienen distribuciones que parecen indistinguibles de las variantes seleccionadas al azar. Además, de esos FP compartidos entre las bibliotecas stLFR y Chromium, solo se encontraron 41 agrupados (Figura 7a). Finalmente, 96 de estas variantes se asignan por GIAB pero con una cigosidad diferente a la asignada en las bibliotecas stLFR.

Si aceptamos la evidencia de que estas variantes de FP compartidas son en gran medida reales y no están presentes en el material de referencia de GIAB, la tasa de FP para la stLFR podría ser de hasta 1859 variantes menos de lo que se informa en la TABLA 1 para la detección de SNP. Esto es aún varios miles de variantes de un solo nucleótido más que la biblioteca estándar BGISEQ-500. Para mejorar adicionalmente la tasa de FP en las bibliotecas stLFR, probamos una serie de estrategias de filtrado diferentes para eliminar errores. En última instancia, al aplicar algunos criterios de filtrado basados en relaciones de alelos variantes y de referencia y recuentos de códigos de barras (véanse los ejemplos), pudimos eliminar entre 3647-13 840 variantes de FP en dependencia de la biblioteca y la cantidad de cobertura. Es importante destacar que, esto se logró mientras solo aumentaba la tasa de FN en un 0,10-0,29 % en las bibliotecas stLFR. Después de esta etapa de filtrado, examinamos los FP compartidos entre las cuatro bibliotecas stLFR. El filtrado eliminó solo 340 variantes de FP compartidas, de las cuales 147 se agruparon dentro de 100 pares de bases entre sí y probablemente no eran reales (Figura 7b). Esto sugiere además que la mayoría de estos FP compartidos son variantes reales. Si se tienen en cuenta estas variantes y el número reducido de variantes de FP después del filtrado, resulta en una tasa de FP similar y una tasa de FN 2-3 veces más alta que la biblioteca STD filtrada para la asignación de SNP (Tabla 3). Este aumento en la tasa de FN se debe principalmente a un aumento del mapeo no único de pares de parejas con tamaños de inserción cortos en las bibliotecas stLFR.

1.3 Rendimiento de la determinación de haplotipos en stLFR

Para evaluar el rendimiento de la determinación de variantes de haplotipos, las variantes de alta confianza de GIAB se identificaron mediante el uso del paquete de software HapCut2, disponible públicamente (24). Más del 99 % de todos los SNP heterocigóticos se colocaron en cóntigos con N50 que oscilaban entre 0,6-15,1 Mb, en dependencia del tipo de biblioteca y la cantidad de datos de secuencia (TABLA 1). La biblioteca stLFR-1, con 336 Gb de cobertura de lectura total (cobertura de genoma único 44X), logró el rendimiento de determinación de haplotipos más alto con un N50 de 15,1 Mb. La longitud de N50 parecía afectarse principalmente por la longitud y la cobertura de los fragmentos genómicos largos. Esto puede verse en la disminución de N50 de la stLFR-2, ya que el ADN usado para esta muestra era levemente más antiguo y estaba más fragmentado que el material usado para la stLFR-1 (TABLA 1, longitud promedio de fragmento de 52,5 kb frente a 62,2 kb) y el N50 ~10 veces más corto que las bibliotecas de 10 ng (stLFR-3 y 4). La comparación con los datos de GIAB mostró que las tasas de error de cambio corto y largo fueron bajas y comparables con estudios anteriores (11, 12, 25). El rendimiento de stLFR fue muy similar al de la biblioteca Chromium. Como el método de haplotipado en perlas de Zhang y otros no tenía datos de lectura disponibles, solo pudimos comparar nuestros resultados con los resultados de su algoritmo de determinación de haplotipos, escrito y optimizado específicamente para sus datos. Esto demostró que las bibliotecas stLFR-1 y stLFR-2 tenían un N50 más largo, una tasa de error de cambio corto similar, pero una tasa de error de cambio largo más alta. stLFR-3 y stLFR-4, que usaron más ADN, tenían un N50 similar al de Zhang y otros. Sin embargo, la comparación directa es difícil debido a las diferencias en la entrada de ADN y la cobertura.

Se debe señalar que este resultado de la determinación de haplotipos se logró mediante el uso de un programa que no se escribió para datos de stLFR. Para ver si podía mejorarse este resultado, desarrollamos un programa de determinación de haplotipos, LongHap, y lo optimizamos específicamente para datos de stLFR. Mediante el uso de variantes de GIAB, LongHap pudo identificar el 99 % de los SNP en cóntigos con un N50 de 18,1 Mb (TABLA 1). Es importante destacar que, estos aumentos en las longitudes de cóntigos se lograron mientras disminuían los errores de cambio cortos y largos (TABLA 1). LongHap también puede identificar los indel. La aplicación de LongHap a stLFR-1 mediante el uso de SNP e indel de GIAB resulta en un N50 de 23,4 Mb, pero también resulta en un aumento de las tasas de error de cambio (Tabla 4).

1.4 Detección de variación estructural

Estudios anteriores han demostrado que la información de fragmentos largos puede mejorar la detección de variaciones estructurales (SV) y deleciones grandes descritas (4-155 kb) en NA12878 (11, 12). Para demostrar el poder de stLFR para detectar SV, examinamos los datos de superposición de códigos de barras, como se describió anteriormente (12), para las bibliotecas stLFR-1 y stLFR-4 en estas regiones. En todos los casos se observó la deleción en los datos de stLFR-1, incluso a una cobertura más baja (Figura 2a y Figura 8). Un examen más detenido de las lecturas de secuencias con códigos de barras compartidos, que cubren una deleción de ~ 150 kb en el cromosoma 8, demostró que la deleción era heterocigótica y se encontraba en un solo haplotipo (Figura 2b-c). La biblioteca stLFR-4 de 10 ng también detectó la mayoría de las deleciones, pero las tres más pequeñas fueron difíciles de identificar debido a una cobertura por fragmento más baja (y, por tanto, a la menor superposición de códigos de barras) de esta biblioteca.

Para evaluar el rendimiento de stLFR para detectar otros tipos de SV, hicimos bibliotecas a partir de una línea celular de un paciente con una translocación conocida entre los cromosomas 5 y 12 (26) y GM20759, una línea celular con una inversión conocida en el cromosoma 2 (27). Las bibliotecas stLFR pudieron identificar la inversión y la translocación en las líneas celulares respectivas (Figura 2d-e). El submuestreo de la cantidad de lecturas por biblioteca mostró que se detectó una fuerte señal de las translocaciones incluso con tan solo 5 Gb de datos de lectura (cobertura total ~ 1,7X, Figura 9a-h). Finalmente, el examen de ambas SV en la biblioteca stLFR-1 no resultó

en ningún patrón evidente (Figura 9i-l), lo que sugiere que la tasa de falsos positivos para la detección de estos tipos de SV es baja.

1.5 Andamiaje de cóntigos con stLFR

stLFR es un método poderoso en parte porque usa un número muy grande (por ejemplo, ~1,8 billones) de códigos de barras únicos y permite la creación de códigos de barras compartidos que es específica para cada molécula individual de ADN genómico largo. Este tipo de datos debería ser beneficioso para el ensamblaje de novo del genoma y el andamiaje mejorado. Para demostrar cómo puede usarse stLFR para mejorar los ensamblajes del genoma, usamos lecturas de las bibliotecas stLFR-1 y stLFR-4 y SALSA (28), un programa diseñado para datos de captura de conformación de cromatina (Hi-C), para andamiaje de ensamblajes de lectura de una sola molécula en tiempo real (SMRT) de NA12878 (29). SALSA no fue diseñado para datos stLFR, por lo que es necesario modificar los datos stLFR a una estructura similar a Hi-C. Esto se logró mediante la selección de pares de lecturas que compartían el mismo código de barras y se ubicaban hacia los extremos de la molécula de ADN largo capturada. Estos se marcaron luego como pares de lectura para el programa SALSA. La sustitución de datos de stLFR por datos de Hi-C resultó en un andamiaje excelente. El uso de solo 60 millones de lecturas de stLFR permitió el enlace de 1411 cóntigos en 597 andamios con un N50 de 44,7 Mb. Estos andamios cubrieron 2,84 Gb del genoma. Estas métricas se compararon muy favorablemente con las generadas en el manuscrito de SALSA mediante el uso de los mismos cóntigos y 10 veces más (734 millones) de pares de lectura Hi-C, generados a partir de células madre embrionarias humanas (30) (Tabla 2). La calidad de los andamios de stLFR se analizó además al alinearlos para construir 37 del genoma de referencia humano y al compararlos con el programa dnadiff (31). En general, los andamios stLFR coincidieron estrechamente con el genoma de referencia y el número de puntos de ruptura, translocaciones, reubicaciones e inversiones fue similar al de los andamios generados con las lecturas Hi-C (Tabla 2). Las matrices de puntos de alineamiento demuestran además el alto grado de continuidad entre los andamios de stLFR y el genoma de referencia (Figura 10).

1.6 Discusión

Aquí describimos una tecnología eficiente para la preparación de bibliotecas de secuenciación del genoma completo, stLFR, que permite la creación de códigos de barras compartidos en subfragmentos de moléculas largas de ADN genómico con un solo código de barras clonal único en un proceso de un solo tubo. El uso de microperlas como compartimentos miniaturizados permite usar un número prácticamente ilimitado de códigos de barras clonales por muestra a un costo insignificante. Nuestra captura optimizada basada en la hibridación de ADN insertado con el transposón, en perlas, combinada con la ligazón de rama 3' y la degradación por una exonucleasa del exceso extremo de adaptadores de captura, permite codificar con éxito hasta~20 % de los subfragmentos en moléculas de ADN de hasta 300 kb de longitud. Es importante destacar que, esto se logra sin la amplificación de fragmentos largos de ADN iniciales y el sesgo de representación que esto conlleva. De esta manera, stLFR resuelve el costo y la capacidad limitada de creación de códigos de barras compartidos, de los métodos basados en la emulsión.

La calidad de las asignaciones de variantes mediante el uso de stLFR es muy alta y, posiblemente, con una optimización adicional, se acerque a la de los métodos WGS estándar, pero con el beneficio añadido de que la creación de códigos de barras compartidos permite aplicaciones informáticas avanzadas. Demostramos una determinación de haplotipos del genoma casi completa y de alta calidad en cóntigos largos con tasas de error extremadamente bajas, detección de SV y andamiaje de cóntigos para permitir las aplicaciones del ensamblaje *de novo*. Todo esto se logra a partir de una sola biblioteca que no requiere equipo especial ni aumenta significativamente el costo de preparación de la biblioteca.

Como resultado de la creación de códigos de barras eficiente, usamos con éxito tan solo 1 ng de ADN humano (cobertura de genoma 600 X) para fabricar bibliotecas stLFR y logramos una WGS de alta calidad con la mayoría de los subfragmentos con códigos de barras compartidos únicos. Puede usarse menos ADN, pero stLFR no usa amplificación de ADN durante la creación de códigos de barras compartidos y, por tanto, no crea subfragmentos superpuestos de cada molécula de ADN largo individual. Por esta razón, la cobertura genómica general se afecta a medida que se reduce la cantidad de ADN. Además, se crea un problema de muestreo ya que stLFR actualmente retiene 10-20 % de cada molécula de ADN largo original seguida de amplificación por PCR. Esto resulta en una tasa de duplicación de lecturas relativamente alta y resulta en un costo de secuenciación añadido, pero es posible introducir mejoras. Una solución evidente es eliminar la etapa de PCR. Esto eliminaría el muestreo, pero también podría reducir sustancialmente las tasas de error de falsos positivos y falsos negativos. Además, las mejoras, tales como la optimización de la distancia de inserción entre los transposones y el aumento de la longitud de las lecturas de secuenciación a 200 bases apareadas en los extremos, deberían permitir con facilidad y aumentarán la cobertura y la calidad general. Para algunas aplicaciones, tales como la detección de variaciones estructurales, el uso de menos ADN y menos cobertura puede ser conveniente. Como demostramos en este documento, tan solo 5 Gb de cobertura de secuencia pueden detectar translocaciones intercromosómicas e intracromosómicas y, en estos casos, la tasa de duplicación es insignificante. De hecho, stLFR puede representar un reemplazo simple y rentable para las bibliotecas de pares de parejas largas en un entorno clínico.

Además, creemos que este tipo de datos puede permitir el ensamblaje *de novo* de la identificación diploide completa, a partir de una sola biblioteca stLFR sin necesidad de lecturas físicas largas, tales como las generadas por las tecnologías SMRT o de nanoporos. Una característica interesante de la inserción de transposones es que crea una superposición de secuencias de 9 bases entre subfragmentos adyacentes. Con frecuencia, estos subfragmentos vecinos se capturan y secuencian, lo que permite duplicar sintéticamente la longitud de las lecturas (por ejemplo, para lecturas de 200 bases, dos subfragmentos vecinos capturados crearían dos lecturas de 200 bases con una superposición de 9 bases, o 391 bases). stLFR no requiere equipo especial como los métodos microfluídicos basados en gotas y el costo por muestra es mínimo. En este documento, demostramos el uso de 50 millones de perlas, pero es posible usar más. Esto permitirá muchos tipos de análisis rentables en los que serían útiles cientos de millones de códigos de barras. Prevemos que este tipo de creación de código de barras masivo, barato, puede ser útil para los análisis de ARN, tales como la secuenciación de ARNm de longitud completa a partir de miles de células, en combinación con tecnologías de una sola célula o la secuenciación profunda de poblaciones de ARN 16S en muestras microbianas. El mapeo de la identificación de cromatina mediante el ensayo de cromatina accesible a la transposasa (ATAC-seq) (32) o los estudios de metilación también son posibles con stLFR.

1.7 Ácidos nucleicos diana

Como se usa en la presente descripción, el término "ácido nucleico (o polinucleótido) diana" o "ácido nucleico de interés" se refiere a cualquier ácido nucleico (o polinucleótido) adecuado para procesar y secuenciar mediante los métodos descritos en la presente descripción. El ácido nucleico puede ser de simple hebra o doble hebra y puede incluir ADN, ARN u otros ácidos nucleicos conocidos. Los ácidos nucleicos diana pueden ser los de cualquier organismo, que incluyen, sin limitarse a, virus, bacterias, levaduras, plantas, peces, reptiles, anfibios, aves y mamíferos (que incluyen, sin limitación, ratones, ratas, perros, gatos, cabras, ovejas, vacas, caballos, cerdos, conejos, monos y otros primates no humanos y humanos). Un ácido nucleico diana puede obtenerse de un individuo o de múltiples individuos (es decir, una población). Una muestra de la que se obtiene el ácido nucleico puede contener ácidos nucleicos de una mezcla de células o incluso de organismos, tales como: una muestra de saliva humana que incluye células humanas y bacterianas; un ratón con un xenoinjerto que incluye células de ratón y células de un tumor humano trasplantado; etc. Los ácidos nucleicos diana pueden estar sin amplificar o pueden amplificarse mediante cualquier método de amplificación de ácidos nucleicos adecuado conocido en la técnica. Los ácidos nucleicos diana pueden purificarse de acuerdo con métodos conocidos en la técnica para eliminar contaminantes celulares y subcelulares (lípidos, proteínas, carbohidratos, ácidos nucleicos distintos de los que van a secuenciarse, etc.), o pueden estar sin purificar, es decir, incluyen al menos algunos contaminantes celulares y subcelulares, que incluyen, sin limitación, células intactas que se rompen para liberar sus ácidos nucleicos para su procesamiento y secuenciación. Los ácidos nucleicos diana pueden obtenerse a partir de cualquier muestra adecuada mediante el uso de métodos conocidos en la técnica. Tales muestras incluyen, pero no se limitan a: tejidos, células o cultivos celulares aislados, fluidos corporales (que incluyen, pero no se limitan a, sangre, orina, suero, linfa, saliva, secreciones anales y vaginales, transpiración y semen); muestras de aire, agrícolas, de agua y de suelo, etc. Los ejemplos no limitantes de ácidos nucleicos diana incluyen "ácidos nucleicos circulantes" (CNA), que son ácidos nucleicos que circulan en la sangre humana u otros fluidos corporales, que incluyen, pero no se limitan a, fluido linfático, líquidos, ascitis, leche, orina, heces y lavado bronquial, por ejemplo, y pueden distinguirse como ácidos nucleicos sin células (CF) o asociados a células (revisado enPinzani y otros, Métodos 50: 302-307, 2010).

Los ácidos nucleicos diana pueden ser ADN genómico (por ejemplo, de un solo individuo), ADNc y/o pueden ser ácidos nucleicos complejos, que incluyen ácidos nucleicos de múltiples individuos o genomas. Los ejemplos de ácidos nucleicos complejos incluyen un microbioma, células fetales circulantes en el torrente sanguíneo de una embarazada (véase, por ejemplo, Kavanagh y otros, J. Chromatol. B 878: 1905-1911, 2010), células tumorales circulantes (CTC) del torrente sanguíneo de un paciente con cáncer (véase, por ejemplo, Allard y otros, Clin Cancer Res. 10: 6897-6904, 2004). Otro ejemplo es el ADN genómico de una sola célula o de un pequeño número de células, tales como, por ejemplo, de biopsias (por ejemplo, células fetales de biopsia del trofectodermo de un blastocisto; células cancerosas de una aspiración con aguja de un tumor sólido, etc.). Otro ejemplo son los patógenos, por ejemplo, células bacterianas, virus u otros patógenos, en un tejido, en sangre u otros fluidos corporales, etc. Como se usa en la presente descripción, el término "ácido nucleico complejo" se refiere a grandes poblaciones de ácidos nucleicos o polinucleótidos no idénticos. En determinadas modalidades, el ácido nucleico diana es ADN genómico; ADN del exoma (un subconjunto de ADN genómico completo enriquecido en secuencias transcritas que contiene el conjunto de exones en un genoma); un transcriptoma (es decir, el conjunto de todos los transcritos de ARNm producidos en una célula o población de células, o ADNc producido a partir de tal ARNm); un metiloma (es decir, la población de sitios metilados y el patrón de metilación en un genoma); un exoma (es decir, regiones codificantes de proteínas de un genoma seleccionado mediante un método de enriquecimiento o captura de exones; un microbioma; una mezcla de genomas de diferentes organismos; una mezcla de genomas de diferentes tipos de células de un organismo; y otras mezclas de ácidos nucleicos complejos que comprende un gran número de moléculas de ácido nucleico diferentes (los ejemplos incluyen, sin limitación, un microbioma, un xenoinjerto, una biopsia de tumor sólido que comprende tanto células normales como tumorales, etc.), que incluyen subconjuntos de los tipos de ácidos nucleicos complejos mencionados anteriormente. En una modalidad, tal ácido nucleico complejo tiene una secuencia completa que comprende al menos una gigabase (Gb) (un genoma humano diploide comprende aproximadamente 6 Gb de secuencia).

En algunos casos, los ácidos nucleicos diana o los primeros fragmentos son fragmentos genómicos. En algunas modalidades, los fragmentos genómicos tienen más de 10 kb, por ejemplo, 10-100 kb, 10-500 kb, 20-300 kb o más de 100 kb. La cantidad de ADN (por ejemplo, ADN genómico humano) usada en una sola mezcla puede ser <10 ng, <3 ng, <1 ng, <0,3 ng o <0,1 ng de ADN. En algunos casos, los ácidos nucleicos diana o los primeros fragmentos tienen una longitud de 5000 a 100 000 KB

1.8 Enfoques adicionales

Aunque los ejemplos de trabajo descritos en la presente descripción usan la reacción en cadena de la polimerasa, pueden usarse otros métodos de amplificación de ácidos nucleicos. Está dentro de la capacidad de una persona experta en la técnica realizar modificaciones apropiadas a una tecnología de amplificación adecuada.

La Figura 17-B5 ilustra enfoques adicionales. La Figura 17 muestra la producción de DNB de doble hebra, que pueden insertarse con transposones y capturarse por perlas stLFR. Pueden hacerse hasta miles de copias (por ejemplo, 10-10 000 copias, tales como 10-1000 copias o 100-1000 copias) en la misma hebra de ADN. Esto permite una alta cobertura de la molécula original con la secuenciación stLFR. La Figura 18 ilustra que, cuando se dispone de una cantidad limitada de la plantilla de ADN, pueden usarse etapas de amplificación previa limitadas antes de la stLFR. La Figura 19 describe un enfoque en el que se usa una nickasa aleatoria a baja concentración, un fragmento Klenow a concentración media y una ligasa a alta concentración. Las perlas y el ADN están en concentraciones adecuadas para stLFR. A medida se hacen las mellas y se abren las brechas mediante Klenow, la ligazón es inmediata y bloquea los fragmentos largos en perlas. Se permite que prosiga la formación de mellas y se abren más brechas para que se liguen más adaptadores a las brechas. La extensión del cebador resulta en fragmentos de ~500 pares de bases. Un segundo adaptador se liga al extremo romo y la biblioteca puede secuenciarse. La Figura 20 muestra la ligazón de adaptadores de horquilla en el ADN largo y el uso de cebadores en los bucles y Ph29 o una polimerasa similar para crear ADNdh concatenado antes de la creación de códigos de barras. Además de mejorar la cobertura de lectura por molécula, un resultado interesante de este proceso es que al final de 0,5-3 h de reacción de la polimerasa, la "longitud" total (número de bases) para cada concatémero es similar, independientemente de la longitud del fragmento inicial. Esto proporciona una opción para usar las perlas con códigos de barras cuya capacidad de unión corresponde al tamaño de los concatémeros y, por tanto, evita la unión de múltiples concatémeros por perla. Esto reduciría el número de perlas necesarias por reacción, por tanto reduce adicionalmente el costo.

La Figura B5 muestra un enfoque para el etiquetado de longitud intermedia. En un enfoque, 96 o más transposones con códigos de barras diferentes se enlazan en grupos de 10 o menos por un resto enlazador (por ejemplo, ADN, moléculas inertes largas tales como dextrina o polietilenglicol (PEG), o proteínas largas tales como queratina o colágeno). La hibridación y la ligazón pueden usarse para unir transposones al ADN enlazador. Los otros métodos pueden ser la unión a través de enlaces químicos o mediante la unión de una avidina a estas moléculas y la unión de una biotina al transposón. Esto logra dos cosas, controla la distancia de inserción entre los transposones y brinda información de proximidad de lectura intermedia (10 kb o menos). Esto es útil para el análisis de secuencias repetidas (repeticiones en tándem, mapeo de trinucleótidos, etc.). El ADN que comprende secuencias de inserción puede capturarse en perlas como en otros enfoques de stLFR descritos en la presente descripción y en otros lugares. Véase Joseph C. Mellor, y otros, "Phased NGS Library Generation via Tethered Synaptic Complexes," seqWell (2017), disponible en la red mundial (http://) en seqwell.com/wp-content/uploads/2017/02/seqWell_LongBow_poster_AGBT2017.pdf (último acceso el 16 de mayo de 2018).

1.9 Referencias para la Sección 1

1. K. Zhang y otros, Long-range polony haplotyping of individual human chromosome molecules. Nat Genet 38, 382-387 (2006).
2. L. Ma y otros, Direct determination of molecular haplotypes by chromosome microdissection. Not Methods 7, 299-301 (2010).
3. J. O. Kitzman y otros, Haplotype-resolved genome sequencing of a Gujarati Indian individual. Not Biotechnol 29, 59-63 (2011).
4. E. K. Suk y otros, A comprehensively molecular haplotype-resolved genome of a European individual. Genome Res 21, 1672-1685 (2011).
5. H. C. Fan, J. Wang, A. Potanina, S. R. Quake, Whole-genome molecular haplotyping of single cells. Not Biotechnol 29, 51-57 (2011).
6. B. A. Peters y otros, Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. Nature 487, 190-195 (2012).
7. J. Duitama y otros, Fosmid-based whole genome haplotyping of a HapMap trio child: evaluation of Single Individual Haplotyping techniques. Nucleic Acids Res 40, 2041-2053 (2012).
8. S. Selvaraj, R. D. J, V. Bansal, B. Ren, Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. Not Biotechnol 31, 1111-1118 (2013).
9. V. Kuleshov y otros, Whole-genome haplotyping using long reads and statistical methods. Nat Biotechnol 32, 261-266 (2014).

10. S. Amini y otros, Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. Nat Genet 46, 1343-1349 (2014).
11. G. X. Zheng y otros, Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nat Biotechnol, (2016).
12. F. Zhang y otros, Haplotype phasing of whole human genomes using bead-based barcode partitioning in a single tube. Not Biotechnol 35, 852-857 (2017).
13. B. A. Peters, J. Liu, R. Drmanac, Co-barcoded sequence reads from long DNA fragments: a cost-effective solution for "perfect genome" sequencing. Frontiers in genetics 5, 466 (2014).
14. R. Drmanac. Nucleic Acid Analysis by Random Mixtures of Non-Overlapping Fragments. documento WO 2006/138284 A2 (2006).
15. R. Drmanac, Peters, B.A., Alexeev, A. Multiple tagging of long DNA fragments. documento WO 2014/145820 A2 (2014).
16. K. Jo, Y. L. Chen, J. J. de Pablo, D. C. Schwartz, Elongation and migration of single DNA molecules in microchannels using oscillatory shear flows. Lab Chip 9, 2348-2355 (2009).
17. R. Drmanac y otros, Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 327, 78-81 (2010).
18. T. Fehlmann y otros, cPAS-based sequencing on the BGISEQ-500 to explore small non-coding RNAs. Clin Epigenetics 8, 123 (2016).
19. J. Huang y otros, A reference human genome dataset of the BGISEQ-500 sequencer. Gigascience 6, 1-9 (2017).
20. S. S. T. Mak y otros, Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. Gigascience 6, 1-13 (2017).
21. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760 (2009).
22. A. McKenna y otros, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-1303 (2010).
23. J. M. Zook y otros, Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. Not Biotechnol 32, 246-251 (2014).
24. P. Edge, V. Bafna, V. Bansal, HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. Genome Res 27, 801-812 (2017).
25. Q. Mao y otros, The whole genome sequences and experimentally phased haplotypes of over 100 personal genomes. Gigascience 5, 1-9 (2016).
26. Z. Dong y otros, Low-pass whole-genome sequencing in clinical cytogenetics: a validated approach. Genet Med 18, 940-948 (2016).
27. Z. Dong y otros, Identification of balanced chromosomal rearrangements previously unknown among participants in the 1000 Genomes Project: implications for interpretation of structural variation in genomes and the future of clinical cytogenetics. Genet Med, (2017).
28. J. Ghurye, M. Pop, S. Koren, D. Bickhart, C. S. Chin, Scaffolding of long read assemblies using long range contact information. BMC Genomics 18, 527 (2017).
29. M. Pendleton y otros, Assembly and diploid architecture of an individual human genome via single-molecule technologies. Not Methods 12, 780-786 (2015).
30. J. R. Dixon y otros, Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376-380 (2012).
31. A. M. Phillippy, M. C. Schatz, M. Pop, Genome assembly forensics: finding the elusive mis-assembly. Genome biology 9, R55 (2008).
32. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Not Methods 10, 1213-1218 (2013).

Ejemplos

2. Ejemplo 1: Materiales y métodos

2.1 Aislamiento de ADN de alto peso molecular

Se aisló ADN genómico largo a partir de líneas celulares de acuerdo con una versión modificada del protocolo del kit de aislamiento de ADN RecoverEase™ (Agilent Technologies, La Jolla, CA) (1).

Brevemente, se sedimentaron aproximadamente 1 millón de células y se lisaron con 500 µl de tampón de lisis. Después de una incubación de 10 minutos a 4 °C, se añadieron 20 µl de cóctel de ribonucleasa RNase-IT en 4 ml de tampón de digestión directamente a las células lisadas y se incubaron en un bloque térmico a 50 °C. Después de 5 minutos, se añadieron 4,5 ml de solución de proteinasa K (~ 1,1 mg/ml de proteinasa K, SDS al 0,56 % y TE 0,89X) y la mezcla se incubó a 50 °C durante 2 horas adicionales. Luego, el ADN genómico se transfirió a un tubo de diálisis con un valor de corte para el peso molecular de 1000 kD (Spectrum Laboratories, Inc., Rancho Dominguez, CA) y se dializó durante la noche a temperatura ambiente en tampón TE 0,5X.

2.2 Construcción de perlas con códigos de barras

Las perlas con códigos de barras se construyen a través de una estrategia de ligazón de división y combinación, mediante el uso de tres conjuntos de moléculas de ADN de doble hebra con código de barras. Véanse las Figuras 12 y 13. Una secuencia adaptadora común que comprende un sitio de hibridación del cebador de PCR, se unió a perlas magnéticas con estreptavidina M-280 Dynabeads™ (ThermoFisher, Waltham, MA), con un enlazador de doble biotina en 5'. Integrated DNA Technologies (Coralville, IA) construyó tres conjuntos de 1536 oligos con código de barras que contenían regiones de secuencia superpuesta. Las ligazones se realizaron en placas de 384 pocillos en una reacción de 15 µl que contenía Tris-HCl 50 mM (pH 7,5), $MgCl_2$ 10 mM, ATP 1 mM, PEG-8000 al 2,5 %, 571 unidades de ligasa T4, 580 pmol de oligo con código de barras y 65 millones de perlas M-280. Las reacciones de ligazón se incubaron durante 1 hora a temperatura ambiente en un rotador. Entre ligazones, las perlas se combinaron en un solo recipiente a través de la centrifugación, se recolectaron en el costado del recipiente con un imán y se lavaron una vez con tampón de lavado con alto contenido de sales (Tris-HCl 50 mM (pH 7,5), NaCl 500 mM, EDTA 0,1 mM y Tween 20 al 0,05 %) y dos veces con tampón de lavado con bajo contenido de sales (Tris-HCl 50 mM (pH 7,5), NaCl 150 mM y Tween 20 al 0,05 %). Las perlas se resuspendieron en tampón de ligazón 1X y se distribuyeron en placas de 384 pocillos y las etapas de ligazón se repitieron.

Determinados "códigos de barras", a los que se hace referencia en la presente descripción, son "códigos de barras tripartitos". Tripartito se refiere a su estructura y/o a su síntesis. Como se muestra en la Figura 12, los códigos de barras tripartitos pueden sintetizarse mediante ligazones sucesivas de secuencias más cortas (por ejemplo, de 4-20 nucleótidos). En una modalidad, los códigos de barras más cortos tienen una longitud de 10 bases. Como se muestra en la figura, una estructura ilustrativa comprendía CS1-BC1-CS2-BC2-CS3-BC3-CS4, en donde CS es una secuencia constante presente en todos los adaptadores de captura y las secuencias BC son códigos de barras diversos de 10 bases, como se discute en la presente descripción. El código de barras tripartito puede construirse mediante el uso de oligonucleótidos parcialmente de doble hebra con la estructura CSa-BC-CSb, hibridados con un oligonucleótido más corto que es el complemento de BC (es decir, BC'), como se muestra en las figuras.

En un aspecto, la invención proporciona una composición que comprende perlas con oligonucleótidos de captura que comprenden códigos de barras clonales unidos, donde la composición comprende más de 3 billones de códigos de barras diferentes y donde los códigos de barras son códigos de barras tripartitos con la estructura 5'-CS1-BC1-CS2-BC2-CS3-BC3-CS4. En algunas modalidades, CS1 y CS4 son más largas que CS2 y CS3. En algunas modalidades CS2 y CS3 tienen 4-20 bases, CS1 y CS4 tienen 5 o 10 a 40 bases, por ejemplo, 20-30, y las secuencias BC tienen 4-20 bases (por ejemplo, 10 bases) de longitud. En algunas modalidades, CS4 es complementario a un oligonucleótido férula. En algunas modalidades, la composición comprende oligonucleótidos puente. En algunas modalidades, la composición comprende oligonucleótidos puente, perlas que comprenden un código de barras tripartito como se discutió anteriormente, y ADN genómico que comprende secuencias de hibridación con una región complementaria a los oligonucleótidos puente.

2.3 stLFR con dos transposones

Se insertaron 2 pmol de transposones acoplados a Tn5 en 40 ng de ADN genómico en una reacción de 60 µl de TAPS-NaOH 10 mM (pH 8,5), $MgCl_2$ 5 mM y DMF al 10 %, a 55 °C durante 10 minutos. Se transfirieron 1,5 µl de ADN insertado con el transposón a 248,5 µl de tampón de hibridación que consistía en Tris-HCl 50 mM (pH 7,5), $MgCl_2$ 100 mM, y TWEEN®20 al 0,05 %. Se resuspendieron de 10-50 millones de perlas con código de barras en el mismo tampón de hibridación. El ADN diluido se añadió a las perlas con código de barras y la mezcla se calentó a 60 °C durante 10 minutos con una ligera mezcla ocasional. La mezcla de perlas y ADN se transfirió a un rotador de tubos en un horno de laboratorio y se incubó a 45 °C durante 50 minutos. Se añadieron, directamente a la mezcla de perlas y ADN, 500 µl de mezcla de ligazón que contenía Tris-HCl 50 mM (pH 7,8), DTT 10 mM, ATP 1 mM, PEG-8000 al 2,5 % y 4000 unidades de ligasa T4. La reacción de ligazón se incubó a temperatura ambiente en un rotador durante 1 hora. Se añadieron 110 µl de SDS al 1 % y la mezcla se incubó a temperatura ambiente durante 10 minutos para eliminar la enzima Tn5. Las perlas se recolectaron en el costado del tubo a través de un imán y se lavaron una vez con tampón de lavado con bajo contenido de sales y una vez con tampón NEB2 (New England Biolabs, Ipswich, MA). El exceso de oligos de código de barras se eliminó mediante el uso de 10 unidades de UDG (New England Biolabs, Ipswich, MA), 30 unidades de APE1 (New England Biolabs, Ipswich, MA) y 40 unidades de exonucleasa 1 (New England Biolabs, Ipswich, MA) en 100 µl de tampón NEB2 1X. Esta reacción se incubó a 37 °C durante 30 minutos. Las perlas se recolectaron en el costado del tubo y se lavaron una vez con tampón de lavado con bajo contenido de sales y una vez con tampón de PCR 1X (tampón PfuCx 1X (Agilent Technologies, La Jolla, CA), DMSO al 5 %, betaína 1 M, $MgSO_4$ 6 mM y dNTP 600 µM). La mezcla de PCR que contenía tampón de PCR 1X, 400 pmol de cada cebador y 6 µl de enzima PfuCx (Agilent Technologies, La Jolla, CA) se calentó a 95 °C durante 3 minutos y luego se enfrió hasta temperatura ambiente. Esta mezcla se usó para resuspender las perlas y la mezcla combinada se incubó a 72 °C durante 10 minutos, seguido por 12 ciclos de 95 °C durante 10 segundos, 58 °C durante 30 segundos y 72 °C durante 2 minutos.

2.4 stLFR con adaptador de ligazón de rama 3'

Este método comienza con las mismas condiciones de inserción de hibridación pero usa solo un transposón en lugar de dos transposones. Después de las etapas de captura y ligazón del código de barras, como se describió anteriormente, las perlas se recolectaron en el costado del tubo y se lavaron con tampón de lavado bajo en sales. Una mezcla de digestión de adaptadores de 90 unidades de exonucleasa I (New England Biolabs, Ipswich, MA) y 100 unidades de exonucleasa III (New England Biolabs, Ipswich, MA) en 100 µl de tampón TA 1X (Teknova, Hollister, CA) se añadió a las perlas y se incubó a 37 °C durante 10 minutos. La reacción se detiene y se elimina la enzima Tn5 mediante la adición de 11 µl de SDS al 1 %. Las perlas se recolectaron en el costado del tubo y se lavaron una vez con tampón de lavado con bajo contenido de sales y una vez con tampón NEB2 1X (New England Biolabs, Ipswich, MA). El exceso de oligonucleótido de captura se eliminó mediante la adición de 10 unidades de UDG (New England Biolabs, Ipswich, MA) y 30 unidades de APE1 (New England Biolabs, Ipswich, MA) en 100 µl de tampón NEB2 1X (New England Biolabs, Ipswich, MA) y la incubación a 37 °C durante 30 minutos. Las perlas se recolectaron en el costado del tubo y se lavaron una vez con tampón de lavado con alto contenido de sales y una vez con tampón de lavado con bajo contenido de sales. Se ligaron 300 pmol del segundo adaptador a los subfragmentos unidos a las perlas, con 4000 unidades de ligasa T4 en 100 µl de tampón de ligasa que contenía Tris-HCl 50 mM (pH 7,8), $MgCl_2$ 10 mM, DTT 0,5 mM, ATP 1 mM y PEG-8000 al 10 %, en un rotador durante 2 horas a temperatura ambiente. Las perlas se recolectaron en el costado del tubo y se lavaron una vez en tampón de lavado con alto contenido de sales y una vez en tampón de PCR 1X. La mezcla y las condiciones de la PCR fueron las mismas que las del proceso de dos transposones descrito anteriormente.

Un adaptador de ligazón de rama 3' ilustrativo comprende los oligonucleótidos del adaptador de ligazón de rama 3' F (/5Fos/CTGATGGCGCGAGGGAGGC) y el adaptador de ligazón de rama 3' R (TCGCGCCATCA/3'dd/G) que se muestran en la Tabla 6. En este ejemplo, la secuencia del adaptador F comprende una secuencia de hibridación del cebador de PCR. Opcionalmente, puede incluirse un código de barras (por ejemplo, un código de barras de muestra) entre el fosfato 5' y la secuencia que se muestra. En este ejemplo, la secuencia del adaptador R es más corta que la secuencia de hibridación del cebador, de manera que se separará en las condiciones en que se hibrida el cebador de la PCR.

2.5 Mapeo de secuencias y asignación de variantes

Los datos de lectura sin procesar primero se individualizaron mediante la secuencia de código de barras asociada, mediante el uso de la herramienta de división de código de barras (disponible en GitHub https://github.com/stLFR/stLFR_read_demux). El código de barras asignado y las lecturas recortadas se mapearon al genoma de referencia hs37d5 con BWA-MEM (2). El archivo BAM resultante se ordenó luego por coordenadas cromosómicas con SAMtools (3) y los duplicados se marcaron con la función picard MarkDuplicate (http://broadinstitute.github.io/picard). La asignación de variantes cortas (SNP e indel) se realizó mediante el uso de HaplotypeCaller dentro de GATK4.0.3.0 (4). El archivo vcf generado a partir de la etapa anterior se comparó luego con la lista de variantes de alta confianza Genome in a Bottle (GIAB) (ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/release/NA12878_HG001/latest /GRCh37/) (5) mediante el uso de la función rtgtools vcfeval (6). Después de la evaluación comparativa, las bibliotecas stLFR se analizaron mediante el uso de GATK VariantRecalibrator, y se usó la configuración verdadera de GIAB para entrenar el modelo de mezcla gaussiana. Luego, los VCF se filtraron mediante el uso de GATK ApplyVQSR. En casi todos los casos, el tramo 99,9 se aplicó a los vcf sin procesar, con la excepción de la biblioteca stLFR-1 de 100 Gb y la biblioteca STD, donde se aplicó el tramo 100. Luego, establecimos y aplicamos criterios de filtrado estrictos adicionales basados en la puntuación GQ, la relación de profundidad de referencia a alternativa y el soporte del código de barras, como se indica en la Tabla 5:

2.20 Determinación de variantes de haplotipos con Hapcut2

Los SNP se identificaron con Hapcut2 (https://github.com/vibansal/HapCUT2) (7) mediante el uso de su estructura de datos 10X Genomics. El archivo BAM se convirtió primero a un formato que porta la información de código de barras en un formato similar al de un BAM con código de barras de 10X Genomics. Específicamente, se añadió un campo 'BX' a cada línea que refleja la información del código de barras de esa lectura. Las variantes de GIAB o las variantes asignadas por GATK para cada biblioteca se usaron como entrada para la determinación de haplotipos, y el resultado de la determinación de haplotipos se resumió y se comparó con el archivo vcf de identificación de GIAB (5) mediante el uso de la herramienta calculate_haplotype_statistics.py de Hapcut2.

2.3 LongHap

La estrategia de extensión de semillas se usa en el proceso de determinación de haplotipos de LongHap. Inicialmente comienza a partir de un par de semillas, compuestas por la variante heterocigótica que se encuentra más hacia la dirección aguas arriba en el cromosoma. Las semillas se extienden mediante el enlace con las otras variantes candidatas que se encuentran aguas abajo, hasta que no pueden añadirse más variantes a las semillas que se extienden (Figura 11). En este proceso de extensión, las variantes candidatas en diferentes loci no se tratarán por igual (es decir, la variante aguas arriba tiene una prioridad más alta en comparación con las variantes aguas abajo a través del cromosoma). Cada uno de dos loci heterocigóticos tienen dos combinaciones posibles a lo largo de los dos alelos diferentes. Si se toma la variante T_2/G_2 y G_3/C_3, por ejemplo (Figura 11), un patrón de combinación es T_2-G_3 y G_2-C_3, mientras que el otro es T_2-C_3 y G_2-G_3. La puntuación de cada combinación se calcula

mediante el número de fragmentos de ADN largo que abarcan los dos loci, lo que equivale al número de códigos de barras únicos con lecturas mapeadas a estos dos loci. Como se muestra en la Figura 11, la puntuación final de la primera combinación es 3, que es tres veces más que la segunda. La variante T_2/G_2se añade a las semillas que se extienden y el proceso se repite. En particular, si algún código de barras soporta ambos alelos en un locus específico, se ignorará al calcular la puntuación de enlace. Esto ayuda a disminuir la tasa de error de cambio. Cuando ocurre un conflicto al enlazar las variantes candidatas aguas abajo, como la variante A_4/C_4en la Figura 11, se tomará una decisión simple al comparar el número de loci enlazados para permitir una extensión adicional de las variantes candidatas. En este caso, hay dos loci enlazados en el escenario de la izquierda, mientras que solo hay uno en el escenario de la derecha. LongHap elegirá el patrón de combinación izquierdo como resultado final de la determinación de haplotipos.

2.8 Detección de SV

Las variantes estructurales se detectaron mediante el cálculo de códigos de barras compartidos entre regiones del genoma como se describió anteriormente (8). Primero se eliminaron las lecturas duplicadas. Las lecturas mapeadas con código de barras compartidos se exploraron mediante el uso de una ventana deslizante (el valor predeterminado es 2 kb) a lo largo del genoma, cada ventana registró cuántos códigos de barras se encontraron dentro de esta ventana de 2 kb y se calculó un índice de Jaccard para la relación de códigos de barras compartidos entre los pares de ventanas. Los eventos de variantes estructurales se identificaron mediante el índice de Jaccard, al compartir la métrica entre los pares de ventanas.

Para cada par de ventanas (X, Y) a través del genoma, el índice de Jaccard se calcula de la siguiente manera:

$$\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, \ldots \mathrm{x}_n);\ \mathrm{Y} = (y_1, y_2, \ldots \mathrm{y}_n)$$

$$\text{índice de Jaccard}_{ij} = \begin{cases} \dfrac{x_i \cap y_j}{x_i \cup y_j} & (\text{si } x_i > 0 \ \text{ o } \ y_j > 0) \\ 0 & (\text{si } x_i = y_j = 0) \end{cases}$$

2.9 Andamiaje del cóntigo con SALSA

Las lecturas de secuenciación de las bibliotecas stLFR se usaron para el andamiaje de un ensamblaje de NA12878 que contenía 18 903 cóntigos con NG50 de 26,83 Mb (9) (cóntigos descargados del sitio web de genomas del NCBI mediante el uso del programa de andamiaje SALSA (10). Para imitar la estructura de secuencia HiC adecuada para SALSA, las lecturas de secuencia de stLFR se seleccionaron a partir de fragmentos de tamaño >= 5 kb. A partir de cada fragmento con una longitud > = 5 kb, se seleccionó la 'primera' y la 'última' lectura para formar un par de lectura. Subsecuentemente, tales pares de lectura artificial se seleccionaron al moverse hacia adentro en estos fragmentos a intervalos de 2 kb. Luego, estos pares de lectura se mapearon en los cóntigos de NA12878 y el andamiaje se realizó con SALSA. Luego, los andamios resultantes se alinearon y compararon con el genoma de referencia hg19 mediante el uso de nucmer y dnadiff del programa MUMmer 4 (11).

2.1 Referencias para el Ejemplo 1

a. I. Agent Technologies, RecoverEase DNA Isolation Kit. Revision C.0, (2015).
b. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760 (2009).
c. H. Li y otros, The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079 (2009).
d. A. McKenna y otros, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-1303 (2010).
e. J. M. Zook y otros, Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. Not Biotechnol 32, 246-251 (2014).
f. J. G. Cleary y otros, Comparing Variant Call Files for Performance Benchmarking of Next-Generation Sequencing Variant Calling Pipelines. bioRxiv, (2015).
g. P. Edge, V. Bafna, V. Bansal, HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. Genome Res 27, 801-812 (2017).
h. F. Zhang y otros, Haplotype phasing of whole human genomes using bead-based barcode partitioning in a single tube. Not Biotechnol 35, 852-857 (2017).
i. M. Pendleton y otros, Assembly and diploid architecture of an individual human genome via single-molecule technologies. Not Methods 12, 780-786 (2015).
j. J. Ghurye, M. Pop, S. Koren, D. Bickhart, C. S. Chin, Scaffolding of long read assemblies using long range contact information. BMC Genomics 18, 527 (2017).
k. S. Kurtz y otros, Versatile and open software for comparing large genomes. Genome biology 5, R12 (2004).

Ejemplo 2: Protocolo detallado

3.1 Materiales

Patrón de ADN de 1 Kb Plus (ThermoFisher, núm. de cat. 10787018)
Tubo de diálisis CE MWCO 100 Kd Biotech (Spectrum Labs, núm. de cat. 131486)
Placa de PCR de 384 pocillos Armadillo (ThermoFisher, núm. de cat. AB2384)
Perlas AMPure XP Agencourt® (Beckman Coulter, núm. de cat. A63882)
APE 1 (10 000 unidades/ml) (New England Biolabs, núm. de cat. M0282L)
ATP (100 mM) (Teknova, núm. de cat. A1210)
Oligonucleótidos para la construcción de perlas con código de barras (IDT) (véase la nota)
Betaína (5 M) (Sigma-Aldrich, núm. de cat. B0300-5VL)
BSA (20 mg/ml) (New England Biolabs, núm. de cat. B9000S)
Oligonucleótidos adaptadores comunes (IDT)
DMF (~100 %) (Sigma-Aldrich, núm. de cat. D4551-250ML)
DMSO (100 %) (Sigma-Aldrich, núm. de cat. D9170-5VL)
dNTP (25 mM) (ThermoFisher, núm. de cat. R1121)
Tubo de diálisis (1000 kD MWCO) (Spectrum Laboratories, Inc., núm. de cat. 131486)
DTT (Sigma-Aldrich, núm. de cat. 11583786001)
Estreptavidina M-280 Dynabeads™ (ThermoFisher, núm. de cat. 60210)
EDTA (0,5 M, pH 8,0) (Sigma-Aldrich, núm. de cat. 03690-100ML)
Exonucleasa I (20 000 unidades/ml) (New England Biolabs, núm. de cat. M0293L)
Exonucleasa III (100 000 unidades/ml) (New England Biolabs, núm. de cat. M0206L)
Formamida (100 %, 250 ml) (Sigma-Aldrich, núm. de cat. 47671-250ML-F)
Glicerol (100 %) (Sigma-Aldrich, núm. de cat. G5516-100ML)
KCl (Sigma-Aldrich, núm. de catálogo P9333-1KG)
KH2PO4 (Sigma-Aldrich, núm. de cat. 795488-1KG)
KOH (Sigma-Aldrich, núm. de cat. P5958-1KG)
MgCl2 (1 M) (Sigma-Aldrich, núm. de cat. 63069-500ML)
MgSO4 (1 M) (Sigma-Aldrich, núm. de cat. M3409-100ML)
Película adhesiva transparente MicroAmp (ThermoFisher, núm. de cat. 4306311)
NaCl (5M) (ThermoFisher, núm. de cat. AM9760G)
Na2HPO4 (Sigma-Aldrich, núm. de cat. S7907-1KG)
NaOH (10M) (Sigma-Aldrich, núm. de cat. 72068-100ML)
Tampón NEB2 (10X) (New England Biolabs, núm. de cat. B7002S)
PEG-8000 (50 %) (Rigaku, núm. de cat. 1008063)
ADN polimerasa Pfu Turbo Cx Hotstart (Agilent, núm. de cat. 600414)
Proteinasa K, recombinante, solución de grado PCR (14-22 mg/ml) (Roche, núm. de cat. 03115844001)
Patrón de ARN de bajo intervalo RiboRuler (Thermofisher, núm. de cat. SM1831)
Cóctel de ribonucleasa RNase-IT (Agilent, núm. de cat. 400720)
SDS (10 %) (ThermoFisher, núm. de cat. 15553027)
Sacarosa (Sigma-Aldrich, núm. de cat. S7903-1KG)
Ligasa de ADN T4 (2x106 unidades/ml) (New England Biolabs, núm. de cat. M0202M)
Tampón TA (10X) (Teknova, núm. de cat. T0379)
TAPS-NaOH (1 M, pH 8,5) (Boston BioProducts, núm. de cat. BB-2375)
TBE (10X) (ThermoFisher, núm. de cat. 15581028)
Tampón TE (10X) (Fisher Scientific, núm. de cat. BP24771)
Enzima Tn5
Oligonucleótidos de transposones (IDT)
Tris-HCl (1 M, pH 7,5) (ThermoFisher, núm. de cat. 15567027)
Tris-HCl (2 M, pH 7,8) (Amresco, núm. de cat. J837-500ML)
Triton™X-100 (10 %) (Sigma-Aldrich, núm. de cat. 93443-100ML)
TWEEN®20 (10 %) (Roche, núm. de cat. 11332465001)
UDG (5000 unidades/ml) (New England Biolabs, núm. de cat. M0280L)

3.2 Equipos

Contenedor de poliestireno alto de 2,4 l (Click Clack, núm. de cat. 659030) o equivalente
Imán DynaMag™-2 (ThermoFisher, núm. de cat. 12321D)
Imán Easy 50 EasySep™ (Stem Cell Technologies, núm. de cat. 18002) o equivalente
Horno de laboratorio capaz de albergar un rotador de tubos.
Agitador magnético de placas
Barra agitadora magnética de tamaño mediano
Agitador de laboratorio estándar tipo vórtex
Termociclador Tetrad PCR (Bio-Rad, núm. de cat. PTC0240) o equivalente con capacidad para volúmenes de reacción de 100 µl por pocillo

Rotador de tubos (Thermo Fisher, núm. de cat. 88881001) o equivalente

3.3 Configuración de reactivos

Tampón de hibridación (3X)
3 ml de Tris-HCl 1 M, pH 7,5
6 ml de NaCl 5 M
91 ml de dH2O estéril
Almacenar a temperatura ambiente durante 1 año.

3.4 Tampón D (10X)

224 mg de KOH
50 µl de EDTA 0,5 M
2,45 ml de dH2O estéril
Hacer alícuotas y almacenar a -20 °C durante 1 mes.

3.5 Tampón de acoplamiento (1X)

5 ml de TE 1X
5 ml de glicerol al 100 %
Almacenar a -20 °C durante 1 año.

3.6 Tampón de digestión (1X, pH 8,0)

1,75 g de Na2HPO4
0,2 g de KCl
0,2 g de KH2PO4
27,4 ml de NaCl 5 M
20 ml de EDTA 0,5 M (pH 8,0)
800 ml de dH2O estéril
Ajuste el pH a 8,0 con NaOH 1 M.
Añadir dH2O estéril hasta un volumen final de 1 litro.
Esterilizar por filtración.
Almacenar a temperatura ambiente durante 1 año.

3.7 Tampón de ligazón de rama 3' (3X)

6 ml de PEG-8000 al 50 %
0,75 ml de Tris-HCl 2 M (pH 7,8)
0,3 ml de MgCl2 1 M
0,3 ml ATP 0,1 M
15 µl de DTT 1 M
75 µl de BSA a 20 mg/ml
2,560 ml de dH2O estéril
Almacenar a -20 °C durante 1 año.

3.8 Tampón de unión de perlas con alto contenido de sales (1X)

5 ml de Tris-HCl 1 M (pH 7,5)
6 ml de NaCl 5 M
20 µl de EDTA 0,5 M
88,98 ml de dH2O estéril
Almacenar a temperatura ambiente durante 1 año.

3.9 Tampón de lavado con alto contenido de sales (1X)

5 ml de Tris-HCl 1 M, pH 7,5
10 ml de NaCl 5 M
20 µl de EDTA 0,5 M
0,5 ml de TWEEN® 20 al 10 %
84,48 ml de dH2O estéril
Almacenar a temperatura ambiente durante 1 año.

3.10 Tampón de hibridación (1X)

50 ml de Tris-HCl 1 M, pH 7,5
100 ml de MgCl2 1 M
5 ml de TWEEN® 20 al 10 %
845 ml de agua
Almacenar a temperatura ambiente durante 1 año

3.11 Tampón de ligazón (10X)

25 ml de PEG-8000 al 50 %
12,5 ml de Tris-HCl 2 M (pH 7,8)
5 ml de ATP 100 mM
5 ml de MgCl2 1 M
2,5 ml de dH2O estéril
Almacenar a -20 °C durante 1 año.

3.12 Tampón de ligazón, sin MgCl2 (10X)

25 ml de PEG-8000 al 50 %
12,5 ml de Tris-HCl 2 M (pH 7,8)
5 ml de ATP 100 mM
5 ml de DTT 1M
2,5 ml de dH2O estéril
Almacenar a -20 °C durante 1 año.

3.13 Tampón de lavado con bajo contenido de sales (1X)

5 ml de Tris-HCl 1 M, pH 7,5
3 ml de NaCl 5 M
0,5 ml de TWEEN® 20 al 10 %
91,5 ml de dH2O estéril
Almacenar a temperatura ambiente durante 1 año.

3.14 Tampón de lisis (1X, pH 8,3)

0,22 g de KCl
120 g de sacarosa
13 ml de Tris-HCl 1 M (pH 7,5)
2 ml de EDTA 0,5 M (pH 8,0)
28 ml de NaCl 5 M
10 ml de Triton® X-100
800 ml de dH2O estéril
Ajuste el pH a 8,3
Añadir dH2O estéril hasta un volumen final de 1 litro
Esterilizar por filtración
Almacenar a 4 ° C durante 1 año.

3.15 Tampón de transposasa (5X)

0,5 ml de TAPS-NaOH 1 M (pH 8,5)
0,25 ml de MgCl2 1 M
5 ml de DMF al 100 %
4,25 ml de dH2O estéril
Almacenar a -20 °C durante 1 año.

3.16 Mezcla PfuCx (2X)

2 ml de tampón PfuCx 10X (incluido con la enzima)
0,5 ml de DMSO al 100 %
2 ml de betaína 5 M
60 µl de MgSO4 1 M
240 µl de dNTP 25 mM
5,2 ml de dH2O estéril

3.16 Oligos de construcción de perlas con códigos de barras

Todos los oligonucleótidos con código de barras se sintetizaron a una escala de 100 nmol en un formato de 384 pocillos con desalinización estándar y se suministraron a una concentración de 200 µM en TE 1X (pH 8,0) por Integrated DNA Technologies (Coralville, IA). Hay un total de 1536 oligos con códigos de barras únicos para cada conjunto de códigos de barras y hay 3 conjuntos de códigos de barras. Esto permite hasta ~3,6 billones de combinaciones de códigos de barras diferentes. Esto puede ser más de lo necesario para algunas aplicaciones y pueden lograrse menos combinaciones de códigos de barras al pedir menos placas de oligonucleótidos. Este diseño en particular requiere que se use al menos un oligonucleótido con códigos de barras de cada conjunto para crear la secuencia final apropiada; sin embargo, pueden hacerse ligeras modificaciones de las secuencias de superposición de 6 bases entre los conjuntos de códigos de barras para eliminar un conjunto de códigos de barras completo.

3.2 Procedimiento

Aislamiento de ADN de alto peso molecular a partir de células

Este método se basa en el protocolo del kit de aislamiento de ADN RecoverEase™ (26), pero se realiza con volúmenes mucho mayores para reducir la viscosidad de la solución resultante.

1. Sedimente hasta 1 × 107 células nucleadas dispersas en un tubo cónico de 15 o 50 ml (500xg durante 5 min). Elimine el sobrenadante. Añada 500 µl de tampón de lisis al sedimento celular y agite la muestra mediante vórtex brevemente durante 3-5 segundos a velocidad media y coloque el tubo cónico en el refrigerador durante ~10 minutos, revuelva ocasionalmente.
2. Prepare la solución de proteinasa K mediante la combinación de 250 µl de SDS al 10 %, 250 µl de proteinasa K y 4 ml de TE 1X. Colóquela en un bloque térmico a 50 °C y caliente brevemente (~5 minutos).
3. Prepare la solución de digestión mediante la combinación de 20 µl de cóctel de ribonucleasa RNase-It con 4 ml de tampón de digestión.
4. Añada ~4 ml de la solución de digestión preparada a las células lisadas y el tampón de la etapa uno y sacuda suavemente el tubo cónico.
5. Coloque el tubo cónico en un bloque térmico a 50 °C, después de 5 minutos añada 4,5 ml de la solución de proteinasa K calentada al sedimento que flota libremente. Revuelva el tubo cónico suavemente para mezclar.
6. Vuelva a tapar el tubo e incube en un bloque térmico a 50 °C durante 2 horas, revuelva suavemente el tubo cada 30 minutos.
7. Corte aproximadamente 13 cm del tubo de diálisis (tiene una capacidad de aproximadamente 1 ml/cm). Permita que se equilibre en TE 0,5X durante 30 minutos. Selle un extremo con un clip de diálisis.
8. Vierta al menos 1 l de tampón TE 0,5X en un depósito de diálisis.
9. Vierta con cuidado el ADN genómico viscoso del tubo cónico en el extremo abierto del tubo de diálisis. Selle el extremo abierto del tubo de diálisis con el clip de diálisis. Fije el flotador a un clip. Coloque el tubo de diálisis con flotador en el depósito de diálisis.
10. Dialice el ADN genómico a temperatura ambiente durante 24 a 48 horas mientras agita suavemente el tampón con una barra agitadora magnética. Reemplace el tampón TE una vez durante el período de diálisis para maximizar la pureza del ADN recuperado.
11. Al finalizar la diálisis, retire el tubo de diálisis del tampón TE, retire el flotador y el clip de la parte superior del tubo de diálisis y vierta suavemente en un tubo cónico de 15 ml. El ADN puede usarse inmediatamente sin fragmentarse.

3.3 Perlas con código de barras

Las perlas con código de barras se construyen mediante una estrategia de división y combinación con 3 conjuntos de moléculas de ADN de doble hebra con código de barras. Los adaptadores de longitud completa se construyen a través de ligazones sucesivas (Figura 12 y Figura 13). Los oligonucleótidos con código de barras se suministran en placas de 384 pocillos (véanse las notas sobre reactivos). Los oligonucleótidos adaptadores comunes se suministran en tubos. En dependencia de la tecnología de secuenciación que se use, puede ser necesario alterar la secuencia del cebador de PCR dentro del oligo adaptador común.

12. Mezcle 10 µl de oligonucleótido complementario de cada pocillo de las placas de origen de 384 pocillos en placas de PCR de 384 pocillos con 10 µl de tampón de hibridación 3X. Mezcle 30 µl de oligonucleótidos adaptadores comunes en un pocillo de una tira de tubos de PCR de 8 pocillos.
13. Incube a 70 °C durante 3 minutos seguido de una rampa lenta de 0,1 °C/s a 20 °C en un termociclador de PCR. Los oligonucleótidos con código de barras hibridados tienen una concentración final de 66 µM.
14. Mezcle 4,725 ml (157,5 µmol) de enlazador de perlas hibridado que contiene una doble biotina en 5' con 3,225 ml de tampón de ligazón (10X), 460,8 µl (921 600 unidades) de ligasa de ADN T4 y 9,67 ml de dH_2O hasta un volumen total de 18,081 ml.
15. Dispense 11,2 µl de la mezcla de ligazón en cada pocillo de cuatro placas de PCR de 384 pocillos nuevas. Luego, añada 8,8 µl (580 pmol) de cada pocillo de las placas con el primer código de barras hibridado a cada pocillo que contenga la mezcla de ligazón del enlazador de perlas. Selle con una película adhesiva transparente MicroAmp, agite mediante vórtex, centrifugue e incube a temperatura ambiente durante 1 hora.

16. Recolecte 100 billones (143 ml) de perlas magnéticas recubiertas de estreptavidina M-280 mediante la transferencia de 50 ml de perlas a un tubo de centrífuga vacío de 50 ml. Coloque el tubo de 50 ml con perlas en el imán Easy 50 EasySep™ durante 5 minutos para recolectar las perlas en el costado del tubo. Elimine con cuidado el sobrenadante con una pipeta. Transfiera los segundos 50 ml de perlas al tubo en el imán. Deje reposar durante 5 minutos en el imán y elimine con cuidado el sobrenadante. Transfiera los últimos 43 ml de perlas al tubo de 50 ml. Deje reposar durante 5 minutos en el imán y elimine con cuidado el sobrenadante. Lave las perlas dos veces con tampón de lavado con bajo contenido de sales y luego resuspenda bien en 8 ml de tampón de unión de perlas con alto contenido de sales.

17. Dispense 5 µl de perlas en tampón de unión de perlas con alto contenido de sales en cada pocillo de las placas que contienen el producto de la ligazón. Agite mediante vórtex, ocasionalmente, el tubo de origen con las perlas mientras dispensa para mantener las perlas bien suspendidas.

18. Selle las placas con una película adhesiva transparente MicroAmp, agite mediante vórtex y colóquelas en el rotador de tubos para la incubación a temperatura ambiente durante 1 hora en el modo de "oscilación".

19. Centrifugue las placas a 300 x g durante 5 segundos para retirar las perlas del sello, pero no permita que se forme un sedimento. Retire el sello y añada 2,8 µl de SDS al 0,1 % a cada pocillo. Vuelva a sellar las placas con una película adhesiva transparente MicroAmp, agite brevemente mediante vórtex e incube a temperatura ambiente durante 10 minutos.

20. Agite mediante vórtex y luego centrifugue las placas a 300 x g durante 5 segundos para retirar las perlas del sello de la placa. Retire el sello de cada placa e invierta las placas en una bandeja de recolección. Centrifugue a 500 x g durante 2 minutos. Mediante el uso de una pipeta serológica de 10 ml, recolecte las perlas en un tubo nuevo de 50 ml.

21. Recolecte las perlas al costado del tubo en el imán Easy 50 EasySep™ durante 5 minutos. Deseche el sobrenadante. Lave una vez con 10 ml de tampón de lavado con alto contenido de sales y luego dos veces con tampón de lavado con bajo contenido de sales. Resuspenda las perlas con 8 ml de tampón de ligazón 1X.

22. Dispense 5 µl de perlas en cada pocillo de cuatro placas de PCR de 384 pocillos nuevas. Agite mediante vórtex, ocasionalmente, el tubo de origen con las perlas mientras dispensa para mantener las perlas bien suspendidas.

23. Para ligar el segundo conjunto de códigos de barras, haga una mezcla que contenga 3,225 ml de tampón de ligazón (10X), 460,8 µl (921 600 unidades) de ligasa de ADN T4 y 6,33 ml de dH2O hasta un volumen total de 10,02 ml. Dispense 6,2 µl de la segunda mezcla de ligazón en cada pocillo de las cuatro placas de PCR de 384 pocillos que contienen las perlas. A continuación, añada 8,8 µl (580 pmol) de cada pocillo de las placas con el segundo código de barras hibridado a los pocillos correspondientes de las placas de PCR de 384 pocillos que contienen las perlas y la mezcla ligazón.

24. Repita las etapas 18-22.

25. Para ligar el tercer conjunto de códigos de barras, haga una mezcla de ligazón que contenga 3,225 ml de tampón de ligazón (10X), 460,8 µl (921 600 unidades) de ligasa de ADN T4 y 6,33 ml de dH2O hasta un volumen total de 10,02 ml. Dispense 6,2 µl de la tercera mezcla de ligazón en cada pocillo de las cuatro placas de PCR de 384 pocillos que contienen las perlas. A continuación, añada 8,8 µl (580 pmol) de cada pocillo de las placas con el tercer código de barras hibridado a los pocillos correspondientes de las placas de PCR de 384 pocillos que contienen las perlas y la mezcla ligazón.

26. Repita las etapas 18-22. Las perlas ahora pueden almacenarse a 4 ° C hasta por un año. En la forma actual, las perlas son casi completamente de doble hebra y aún no están en la forma correcta para su uso en la stLFR.

27. Cuente las perlas con un hemocitómetro y extraiga 5 millones de perlas para la etapa de QC. Coloque el tubo con perlas en el imán DynaMag™-2 durante 5 minutos. Deseche el sobrenadante. Añada 5 µl de formamida al 100 %, 4 µl de dH2O y 1 µl de tampón de carga 10X. Incube a 95 °C durante 3 minutos en un termociclador de PCR. Coloque inmediatamente en hielo durante 2 minutos. Coloque el tubo con perlas en el imán DynaMag™-2 durante 5 minutos. Recolecte el sobrenadante, cárguelo en un gel de TBU al 15 % y procéselo a 200 V durante 40 minutos para comprobar la longitud y la cantidad de oligonucleótidos. Alternativamente, las perlas pueden examinarse mediante el uso de un citómetro de flujo mediante la hibridación de un oligonucleótido marcado con fluorescencia al extremo 3' de la secuencia adaptadora de las perlas. Típicamente, vemos que alrededor del 25 % del total de sitios unidos a estreptavidina tienen una secuencia adaptadora construida de longitud completa.

3.20 Preparación de perlas para stLFR

Para preparar perlas para stLFR, primero deben desnaturalizarse a ADN de simpe hebra y luego volver a hibridarse con el oligo puente.

28. Añada con una pipeta 500 millones de perlas con código de barras construidas de la etapa 26 de la sección anterior en un tubo de microcentrífuga estándar de 1,5 ml.

29. Colóquelo en el imán DynaMag™-2 durante 2 minutos para recolectar las perlas en el costado del tubo. Elimine el sobrenadante.

30. Añada 1 ml de una dilución 1X de tampón D. Agite mediante vórtex brevemente e incube durante 2 minutos a temperatura ambiente.

31. Colóquelo en el imán DynaMag™-2 durante 2 minutos para recolectar las perlas en el costado del tubo. Elimine el sobrenadante.

32. Repita las etapas 30 y 31 una vez más.

33. Lave una vez en tampón de hibridación 1X. Colóquelo en el imán DynaMag™-2 durante 2 minutos para recolectar las perlas en el costado del tubo. Elimine el sobrenadante.
34. Mezcle 36 µl del oligo puente 100 µM, 333,33 µl de tampón de hibridación (3X) y 630,67 µl de dH2O para obtener un volumen final de 1 ml. Añada la mezcla a las perlas. Agite mediante vórtex brevemente.
35. Incube a 60 °C durante 5 minutos y a temperatura ambiente durante 50 minutos.
36. Colóquelo en el imán DynaMag™-2 durante 2 minutos para recolectar las perlas en el costado del tubo. Elimine el sobrenadante y resuspenda en 500 µl de tampón de lavado con bajo contenido de sales. Estas perlas ahora están listas para stLFR y pueden almacenarse durante 3 meses a 4 °C.

3.21 Protocolo stLFR de dos transposones

El protocolo utiliza dos transposones para crear secuencias de hibridación y sitios de cebadores de PCR a lo largo de las moléculas de ADN genómico. Este es el método stLFR más simplificado y rápido, pero tiene, posiblemente, un 50 % menos de cobertura por fragmento largo de ADN que el protocolo de ligazón de rama 3'. Puede ser necesario alterar parte de la secuencia del transposón después de la región mosaico para que sea compatible con las tecnologías de secuenciación distintas de BGISEQ-500. Compruebe los cebadores de secuenciación que se usan antes de pedir estos oligonucleótidos. La información sobre todas las secuencias de oligonucleótidos está disponible en los materiales complementarios.
37. Hibride los oligos de transposones de captura mediante la combinación de 10 µl de Transposón1T (100 µM), 10 µl de TransposónB (100 µM), 10 µl de tampón de hibridación (3X) en el primer pocillo de una tira de tubos de PCR de 8 pocillos y los oligos de transposones que no son de captura mediante la combinación de 10 µl de Transposón1T (100 µM), 10 µl de TransposónB (100 µM), 10 µl de tampón de hibridación (3X) en el segundo pocillo de la misma tira de tubos de PCR.
38. Incube a 70 °C durante 3 minutos seguido de una rampa lenta de 0,1 °C/s a 20 °C en un termociclador de PCR. Combine los dos transposones en el tercer pocillo de la tira de tubos de PCR.
39. Acople la enzima Tn5 a la mezcla de transposones mediante la combinación de 9,6 µl de transposones mezclados con 23,53 µl de Tn5 (13,6 pmol/µl) y 46,87 µl de tampón de acoplamiento (1X).
40. Incube a 30 ° C durante 1 hora. Use inmediatamente o almacene a -20 °C hasta por 1 mes. Para obtener un rendimiento y una consistencia óptimos entre los experimentos, sugerimos hacer alícuotas antes del almacenamiento.
41. Incorpore los transposones en el ADN genómico largo mediante la combinación de 12 µl de tampón de transposasa (5X), 0,5 µl del transposón acoplado de la etapa 40 y 40 ng de ADN en un volumen total de 60 µl en un pocillo de una tira de tubos de 8 pocillos. Nota: esta cantidad de ADN y la cantidad de transposón acoplado pueden ajustarse en esta etapa. Será necesario titular la cantidad de enzima Tn5 usada ya que puede haber variabilidad entre lotes. Además, es posible comenzar con menos ADN, pero a los fines de la titulación es útil usar 40 ng, de modo que parte del material se pueda procesar en un gel de agarosa para determinar la eficiencia de la incorporación del transposón (véanse las etapas posteriores).
42. Incube a 55 °C durante 10 minutos.
43. Transfiera 40 µl del material con el transposón incorporado a un pocillo de una nueva tira de tubos de 8 pocillos. Añada 4 µl de SDS al 1 % e incube a temperatura ambiente durante 10 minutos.
44. Cargue el material de la etapa 43 en un gel de agarosa al 1 % en TBE 0,5X y procéselo a 150 V durante 40 minutos. El ADN transpuesto debe tener entre 200 y 1500 pb en el gel. Típicamente, queremos ver la parte más brillante de la mancha de ADN alrededor de 600 pb, esto podría ser diferente en base a la tecnología de secuenciación que se elija. Típicamente, cargamos los controles que se someten a las mismas etapas pero que carecen del transposón, la enzima Tn5 o el ADN genómico. Si el tamaño de los productos integrados con transposones parece correcto, vaya a la etapa 45. De lo contrario, repita las etapas anteriores pero ajuste la concentración del producto de acoplamiento hasta que la mancha tenga el tamaño deseado.
45. Diluya 1,5 µl del producto restante de la etapa 42, con 248,5 µl de tampón de hibridación 1x.
46. Transfiera 50 µl de perlas (50 millones) de la etapa 36 a un tubo de microcentrífuga de 1,5 ml. Colóquelo en el imán DynaMag™-2 durante 2 minutos para recolectar las perlas en el costado del tubo. Elimine el sobrenadante y resuspenda en 250 µl de tampón de hibridación (1X).
47. Caliente el ADN y las perlas por separado a 60 °C durante 30 segundos.
48. Añada 250 µl de ADN diluido a los 250 µl de perlas, mezcle suavemente mediante golpes con un dedo a la parte inferior del tubo y continúe la incubación a 60 °C durante 10 minutos. Mezcle ligeramente el tubo cada pocos minutos con el dedo.
49. Colóquelo en el rotador de tubos para la incubación en el horno a 45 °C durante 50 minutos en modo de "oscilación".
50. Haga una mezcla de ligazón mediante la combinación de 100 µl de tampón de ligazón, sin MgCl2 (10X), 2 µl de ligasa de ADN T4 (2x106 unidades/ml) y 398 µl de dH2O. Retire el tubo del rotador y añada la mezcla de ligazón para obtener un volumen total de 1 ml.
51. Incube en el rotador de tubos durante 1 hora en modo de "oscilación" a temperatura ambiente.
52. Añada 110 µl de SDS al 1 % al tubo e incube durante 10 minutos a temperatura ambiente.
53. Colóquelo en el imán DynaMag™-2 durante 2 minutos para recolectar las perlas en el costado del tubo. Elimine el sobrenadante y lave una vez con 500 µl de tampón de lavado con bajo contenido de sales y una vez con 500 µl de tampón NEB2 (1X).

54. Prepare una mezcla de digestión de oligonucleótidos de captura mediante la combinación de 10 µl de tampón NEB2 (10X), 2 µl de UDG (5000 U/ml), 3 µl de APE1 (10 000 U/ml), 2 µl de exonucleasa 1 (20 000 unidades/ml) y 83 µl de dH2O. Elimine el tampón de lavado y añada la mezcla de digestión a las perlas.
55. Agite mediante vórtex ligeramente para resuspender las perlas e incube a 37 °C durante 30 minutos.
56. Colóquelo en el imán DynaMag™-2 durante 2 minutos para recolectar las perlas en el costado del tubo. Elimine el sobrenadante y lave una vez con 500 µl de tampón de lavado con bajo contenido de sales y una vez con 500 µl de tampón PfuCx (1X).
57. Prepare la mezcla maestra de PCR mediante la adición de 150 µl de mezcla de PCR (2X), 4 µl de cebador 1 de PCR (100 µM), 4 µl de cebador 2 de PCR (100 µM), 6 µl de enzima PfuCx y 136 µl de dH2O. Caliente previamente la mezcla maestra de PCR a 95 °C durante 3 minutos. Colóquelo en el imán DynaMag™-2 durante 2 minutos para recolectar las perlas en el costado del tubo. Elimine el tampón de lavado y añada la mezcla maestra de PCR a las perlas.
58. Agite mediante vórtex ligeramente para resuspender las perlas y configure la reacción de PCR con las siguientes condiciones:

Etapa 1	72 °C	10 minutos
Etapa 2	95 °C	10 segundos
Etapa 3	58 °C	30 segundos
Etapa 4	72 °C	2 minutos
Etapa 6	Repita las etapas 2-5 de 10-12 veces	

59. El PCR debe resultar en ˜500 ng de ADN, procese 20 ng de producto en un gel de agarosa al 1 % en TBE 0,5X durante 40 minutos a 150 V. El material debe ser una mancha con un pico alrededor de los 500 pb.
60. Purifique el producto de PCR con 300 µl de perlas Agencourt XP de acuerdo con el protocolo del fabricante. Este producto purificado ya está listo para entrar en el proceso de secuenciación.

Protocolo stLFR de ligazón de rama 3' con un solo transposón

Este protocolo se basa en la inserción de un solo transposón y nuevos métodos de ligazón de adaptadores en una brecha de ADN y puede permitir una cobertura más alta por fragmento, lo que puede ser importante para algunas estrategias de secuenciación, tales como el ensamblaje de novo. Esta estrategia es levemente más costosa debido a los reactivos adicionales. También tarda 2,5 horas más.
61. Hibride los oligos de transposones de captura mediante la combinación de 10 µl de Transposón1T (100 µM), 10 µl de TransposónB (100 µM), 10 µl de tampón de hibridación (3X) en el primer pocillo de una tira de tubos de PCR de 8 pocillos y el adaptador de ligazón de brechas mediante la combinación de 10 µl de RamaT (100 µM), 10 µl de RamaB (100 µM), 10 µl de tampón de hibridación (3X) en el segundo pocillo de la misma tira de tubos de PCR.
62. Incube a 70 °C durante 3 minutos seguido de una rampa lenta de 0,1 °C/s a 20 °C en un termociclador de PCR.
63. Acople la enzima Tn5 al transposón mediante la combinación de 9,6 µl de transposón de captura hibridado en la etapa 61 con 23,53 µl de Tn5 (13,6 pmol/µl) y 46,87 µl de tampón de acoplamiento (1X).
64. Incube a 30 ° C durante 1 hora. Use inmediatamente o almacene a -20 °C hasta por 1 mes.
65. Siga las etapas 41-51.
66. Colóquelo en el imán DynaMag™-2 durante 2 minutos para recolectar las perlas en el costado del tubo. Elimine el sobrenadante y lave una vez con 500 µl de tampón de lavado con bajo contenido de sales.
67. Haga una mezcla de digestión de oligonucleótidos adaptadores mediante la combinación de 10 µl de tampón TA (10X), 4,5 µl de exonucleasa I (20 000 U/ml), 1 µl de exonucleasa III (100 000 U/ml) y 74,5 µl de dH2O. Elimine el tampón de lavado y añada la mezcla de digestión a las perlas.
68. Agite mediante vórtex ligeramente para resuspender las perlas e incube en el rotador de tubos durante 10 minutos a 37 °C en modo de "oscilación".
69. Añada 11 µl de SDS al 1 % e incube durante 10 minutos a temperatura ambiente.
70. Colóquelo en el imán DynaMag™-2 durante 2 minutos para recolectar las perlas en el costado del tubo. Elimine el sobrenadante y lave una vez con 500 µl de tampón de lavado con bajo contenido de sales y una vez con 500 µl de tampón NEB2 (1X).
71. Haga una mezcla de digestión de oligonucleótidos de captura mediante la combinación de 10 µl de tampón NEB2 (10X), 2 µl de UDG (5000 U/ml), 3 µl de APE1 (10 000 U/ml) y 85 µl de dH2O. Elimine el tampón de lavado y añada la mezcla de digestión a las perlas.
72. Agite mediante vórtex ligeramente para resuspender las perlas e incube a 37 °C durante 30 minutos.
73. Colóquelo en el imán DynaMag™-2 durante 2 minutos para recolectar las perlas en el costado del tubo. Elimine el sobrenadante y lave una vez con 500 µl de tampón de lavado con alto contenido de sales y una vez con 500 µl de tampón de lavado con bajo contenido de sales (1X).
74. Prepare la mezcla de ligazón de rama 3' mediante la combinación de 33,4 µl de tampón de ligazón de rama 3' (3X), 18 µl del adaptador de ligazón de rama 3' (16,7 µM) preparado en la etapa 61, 2 µl de ligasa de ADN T4 (2x106 unidades/ml), y 46,6 µl de dH2O. Elimine el tampón de lavado y añada la mezcla de ligazón a las perlas.

75. Agite mediante vórtex ligeramente para resuspender las perlas e incube en el rotador de tubos durante 2 horas a 25 ° C en modo de "oscilación".
76. Colóquelo en el imán DynaMag™-2 durante 2 minutos para recolectar las perlas en el costado del tubo. Elimine el sobrenadante y lave una vez con 500 µl de tampón de lavado con alto contenido de sales y una vez con 500 µl de tampón de PCR (1X).
77. Prepare la mezcla maestra de PCR mediante la adición de 150 µl de tampón de PCR 2X, 4 µl de cebador 1 de PCR (100 µM), 4 µl de cebador 2 de PCR (100 µM), 6 µl de enzima de PCR y 136 µl de dH2O. Elimine el tampón de lavado y añada la mezcla maestra de PCR a las perlas.
78. Agite mediante vórtex ligeramente para resuspender las perlas y configure la reacción de PCR con las siguientes condiciones:

Etapa 1	95 oC	3 minutos
Etapa 2	95 oC	10 segundos
Etapa 3	58 oC	30 segundos
Etapa 4	72 oC	2 minutos
Etapa 5	Repita las etapas 2-4 de 10 a 12 veces	

79. Siga las etapas 59-60 anteriores.

3.4 Análisis de datos de stLFR

El punto de partida de este proceso es un archivo FASTQ. Este es un formato estándar para leer datos generado por la mayoría de las tecnologías de secuenciación. El software que usamos para la deconvolución de la información del código de barras toma el archivo FASTQ y espera que se adjunten 42 bases del código de barras y la secuencia adaptadora común al final de la primera lectura. Hace coincidir los datos de lectura del código de barras con las 1536 secuencias esperadas en cada posición del código de barras. La estrategia de creación de códigos de barras usada por stLFR permite la corrección de errores de códigos de barras que tienen una sola base que no coincide. El resultado final de nuestro software es un archivo FASTQ con la información del código de barras adjunta al final de la ID de lectura con el formato #Barcode1ID_Barcode2ID_Barcode3ID, donde BarcodeID es un número de 0-1536. Cero para una ID de código de barras significa que no coincide con ninguna de las secuencias de códigos de barras esperadas. Recomendamos usar BWA-mem27 para el mapeo, GATK28 para la asignación de variantes y HapCUT229 para la determinación de haplotipos. También recomendamos mapear a Hg19 con secuencias señuelo.

3.5 Referencias para el Ejemplo 2

1. Zhang, K. y otros. Long-range polony haplotyping of individual human chromosome molecules. Nat Genet 38, 382-387 (2006).
2. Ma, L. y otros. Direct determination of molecular haplotypes by chromosome microdissection. Nat Methods 7, 299-301 (2010).
3. Kitzman, J. O. y otros. Haplotype-resolved genome sequencing of a Gujarati Indian individual. Nat Biotechnol 29, 59-63 (2011).
4. Suk, E. K. y otros. A comprehensively molecular haplotype-resolved genome of a European individual. Genome Res 21, 1672-1685 (2011).
5. Fan, H. C., Wang, J., Potanina, A. y Quake, S. R. Whole-genome molecular haplotyping of single cells. Nat Biotechnol 29, 51-57 (2011).
6. Peters, B. A. y otros. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. Nature 487, 190-195 (2012).
7. Duitama, J. y otros. Fosmid-based whole genome haplotyping of a HapMap trio child: evaluation of Single Individual Haplotyping techniques. Nucleic Acids Res 40, 2041-2053 (2012).
8. Selvaraj, S., J, R. D., Bansal, V. y Ren, B. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. Nat Biotechnol 31, 1111-1118 (2013).
9. Kuleshov, V. y otros. Whole-genome haplotyping using long reads and statistical methods. Nat Biotechnol 32, 261-266 (2014).
10. Amini, S. y otros. Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. Nat Genet 46, 1343-1349 (2014).
11. Zheng, G. X. y otros. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nat Biotechnol (2016).
12. Zhang, F. y otros. Haplotype phasing of whole human genomes using bead-based barcode partitioning in a single tube. Nat Biotechnol 35, 852-857 (2017).
13. Peters, B. A., Liu, J. y Drmanac, R. Co-barcoded sequence reads from long DNA fragments: a cost-effective solution for "perfect genome" sequencing. Frontiers in genetics 5, 466 (2014).
14. Drmanac, R. Nucleic Acid Analysis by Random Mixtures of Non-Overlapping Fragments. documento WO 2006/138284 A2 (2006).

15. McElwain, M. A., Zhang, R. Y., Drmanac, R. y Peters, B. A. Long Fragment Read (LFR) Technology: Cost-Effective, High-Quality Genome-Wide Molecular Haplotyping. Methods Mol Biol 1551, 191-205 (2017).
16. Schaaf, C. P. y otros. Truncating mutations of MAGEL2 cause Prader-Willi phenotypes and autism. Nat Genet 45, 1405-1408 (2013).
17. Peters, B. A. y otros. Detection and phasing of single base de novo mutations in biopsies from human in vitro fertilized embryos by advanced whole-genome sequencing. Genome Res 25, 426-434 (2015).
18. Ciotlos, S. y otros. Whole genome sequence analysis of BT-474 using complete Genomics' standard and long fragment read technologies. Gigascience 5, 8 (2016).
19. Hellner, K. y otros. Premalignant SOX2 overexpression in the fallopian tubes of ovarian cancer patients: Discovery and validation studies. EBioMedicine 10, 137-149 (2016).
20. Mao, Q. y otros. The whole genome sequences and experimentally phased haplotypes of over 100 personal genomes. Gigascience 5, 1-9 (2016).
21. Gulbahce, N. y otros. Quantitative Whole Genome Sequencing of Circulating Tumor Cells Enables Personalized Combination Therapy of Metastatic Cancer. Cancer Res 77, 4530-4541 (2017).
22. Walker, R. F. y otros. Clinical and genetic analysis of a rare syndrome associated with neoteny. Genetics In Medicine (2017).
23. Mao, Q. y otros. Advanced Whole-Genome Sequencing and Analysis of Fetal Genomes from Amniotic Fluid. Clinical chemistry (2018).
24. Drmanac, R., Peters, B.A., Alexeev, A. Multiple tagging of individual long DNA fragments. documento WO 2014/145820 A2 (2013).
25. Picelli, S. y otros. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res 24, 2033-2040 (2014).
26. Agent Technologies, I. RecoverEase DNA Isolation Kit. Revision C.0 (2015).
27. Li, H. y Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760 (2009).
28. McKenna, A. y otros. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-1303 (2010).
29. Edge, P., Bafna, V. y Bansal, V. HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. Genome Res 27, 801-812 (2017).

El archivo BAM "NA12878_WGS_v2_phased_possorted_bam.bam" de un conjunto de datos de Chromium reciente se descargó del sitio web de 10X Genomics y se procesó de la misma manera que las bibliotecas stLFR. Para obtener resultados filtrados, usamos el archivo VCF "NA12878_WGS_v2_phased_variants.vcf.gz" de la misma biblioteca Chromium. Este VCF contiene datos que se procesaron a través de la estructura optimizada de 10X Genomics. El tamaño del fragmento para la biblioteca Chromium se copió del sitio web de 10X Genomics. 10 Genomics usa una media ponderada de longitud para calcular el tamaño de los fragmentos, lo que puede resultar en un tamaño mayor que el tamaño promedio de los fragmentos.[2]Los datos de lectura no estaban disponibles, esto es lo que se informa en Zhang y otros w (12).[3]Datos de una biblioteca estándar procesada en un BGISEQ-500.

La Tabla 6 muestra secuencias ilustrativas que pueden usarse en los métodos de stLFR descritos en la presente descripción.

Ejemplo 3: Ligazón de rama 3', un nuevo método para ligar ADN a los extremos 3'OH de ADN o ARN, y sus aplicaciones

4.1 Introducción

Este ejemplo describe la ligazón de rama 3' en general. La ligazón de rama 3' se usa para añadir un adaptador adicional (adaptador de ligazón de rama 3') en la modalidad de stLFR descrita en la presente descripción. Véase, por ejemplo, §1.1.2.

Las ligasas unen rupturas en los ácidos nucleicos, lo cual es esencial para la viabilidad y vitalidad celular. Las ligasas de ADN catalizan la formación de un enlace fosfodiéster entre los extremos del ADN y desempeñan funciones cruciales en la reparación, recombinación y replicación del ADN *in vivo.* Las ligasas de ARN unen los extremos de ARN 5'-fosforilo (5'PO4) y 3'-hidroxilo (3'OH) a través de enlaces fosfodiéster y participan en la reparación, corte y empalme y edición del ARN. Pueden utilizarse *in vitro* ligasas de los tres reinos de organismos (bacterias, arqueobacterias y eucariotas), como herramientas moleculares importantes para aplicaciones como clonación, amplificación o detección basada en ligasas, biología sintética, etc.

Una de las ligasas más ampliamente usadas *in vitro* es la ligasa de ADN del bacteriófago T4, que es un solo polipéptido de 55 kDa y requiere ATP como fuente de energía. La ligasa de ADN T4 típicamente une a los extremos 5'PO4 y 3'OH adyacentes del ADN bicatenario. Además de sellar mellas o ligar extremos cohesivos, la ligasa de ADN T4 también puede catalizar eficientemente la unión de extremos romos, lo que no se ha observado en todas las otras ligasas de ADN. Algunas propiedades catalíticas inusuales de esta ligasa se informaron anteriormente, tales como el sellado de brechas de simple hebra en el ADN bicatenario, el sellado de mellas adyacentes a sitios abásicos en el ADN de doble hebra (ADNdh), la promoción de la formación de bucles intramoleculares de ADN parcialmente

de doble hebra y la unión de hebras de ADN que contienen extensiones de rama 3'. (Nilsson y Magnusson, Nucleic Acids Res 10:1425-1437, 1982; Goffin y otros, Nucleic Acids Res 15:8755-8771, 1987; Mendel-Hartvig y otros, Nucleic Acids Res. 32:e2, 2004; Western y Rose, Nucleic Acids Res., 19:809-813, 1991). Los investigadores también observaron ligazones independientes de la plantilla mediadas por la ligasa T4, tales como el sellado de mellas mal apareadas en el ADNdh (Alexander, 2003, Nucleic Acids Res. 15 de junio de 2003;31(12):3208-16) o incluso la ligazón de ADN de simple hebra (ADNsh), aunque con muy baja eficiencia (H. Kuhn, 2005, FEBS J. diciembre de 2005;272(23):5991-6000). Estos resultados sugieren que el apareamiento perfecto de bases complementarias en la unión de la ligazón, o adyacente a esta, no es críticamente necesario para alguna actividad no convencional de la ligasa de ADN de T4. Las ligasas de ARN T4 1 y 2 son los productos de los genes 63 y 24, respectivamente, del fago T4. Ambas requieren un extremo 5'PO4 y 3'OH adyacente para una ligazón exitosa con hidrólisis de ATP a AMP y PPi. Los sustratos para la ligasa de ARN de T4 1 incluyen ARN y ADN de simple hebra, mientras que la ligasa de ARN T4 2 sella preferentemente las mellas en el ARNdh en lugar de ligar los extremos del ARNsh.

Aquí demostramos un evento de unión de extremos no convencional mediado por la ligasa de ADN T4, que llamamos ligazón de rama 3' (3'BL). Puede unir fragmentos de ADN o ADN/ARN en las mellas, brechas de simple hebra o regiones salientes 5' para formar una estructura de rama. Este informe estudió exhaustivamente una amplia variedad de cofactores y activadores de la ligazón y optimizó las condiciones de ligazón para este tipo de ligazón nueva. Con nuestro protocolo 3'BL, no se requirió el apareamiento de bases y la ligazón puede completarse en más del 90 % incluso para una brecha de 1 nt. Una de sus aplicaciones es unir adaptadores al ADN o ARN en la preparación de bibliotecas NGS. Varias estructuras genómicas que anteriormente se consideraba que no podían ligarse, ahora pueden convertirse en un sustrato para 3'BL, lo que resulta en una alta tasa de conversión del ADN de entrada en moléculas ligadas al adaptador, mientras se evitan las quimeras. Demostramos que 3'BL podría acoplarse con la inserción de transposones. La estrategia de inserción de transposones direccional que proponemos puede producir, teóricamente, plantillas, el 100 % de las cuales pueden utilizarse para la secuenciación. Aplicaciones de microARN. Nuestro estudio demostró el valor de esta nueva técnica para la preparación de bibliotecas NGS, así como también el potencial para avanzar en muchas otras aplicaciones moleculares, tales como el marcaje radiactivo de los extremos 3' del ARN.

4.2 Ligazón de rama 3', una nueva forma de ligar los extremos del ADN

Convencionalmente, la ligazón de ADN implica la unión de los extremos de ADN 5'PO4 y 3'OH de fragmentos de extremos cohesivos o romos. La ligazón de extremos cohesivos es generalmente más rápida y menos dependiente de la concentración de enzimas que la unión de extremos romos. Ambos procesos pueden ser catalizados por la ligasa de ADN del bacteriófago T4, que usa ATP como cofactor productor de energía y requiere Mg2+. También se informó que la ligasa de ADN T4 liga oligos de simple hebra específicos o degenerados a sustratos parcialmente de simple hebra a través de la hibridación. Aquí demostramos una ligazón sin precedentes, mediada por la ligasa de ADN T4, que no requiere apareamiento de bases complementarias y puede ligar un donante de ADN de extremos romos al extremo 3'OH de un aceptor de ADN bicatenario en las mellas, brechas o salientes 5', para formar una estructura de rama (Figura 21a). Por lo tanto, usamos el término ligazón de rama 3' (3'BL) para describir estas ligazones. El ADN donante sintético que usamos contenía un extremo bicatenario romo y un extremo de ADNsh. Los sustratos aceptores contenían una de las siguientes estructuras: una mella desfosforilada, una brecha de 1 u 8 nucleótidos (nt) o un saliente 5' de 36 nt. La ligasa T4 ayuda a unir el 5'PO4 de la hebra adaptadora al único 3'OH que se puede ligar de la hebra sustrato para formar un producto de ligazón en forma de horquilla.

Para optimizar la eficiencia de la ligazón, probamos exhaustivamente una serie de factores que afectan la eficiencia general de la ligazón, que incluyen la relación adaptador::sustrato de ADN, cantidad de ligasa T4, concentración final de ATP, concentración de Mg^{2+}, pH, tiempo de incubación y presencia de diferentes aditivos tales como polietilenglicol-8000 (PEG-8000) y proteína de unión a simple hebra (SSB) (Figuras 1 y 2 complementarias). Descubrimos que la adición de PEG-8000 a una concentración final del 10 % podría aumentar sustancialmente la eficiencia de ligazón desde menos del 10 % hasta más del 90 % (Figura 21). Un amplio intervalo de concentración de ATP (de 1 µM a 1 mM) y concentración de Mg^{2+} (3 mM a 10 mM) también funcionó con la ligazón de rama 3'. La cantidad de ligasa necesaria para la 3'BL fue comparable a la de la ligazón de extremos romos. En nuestras condiciones optimizadas, usamos relaciones molares de adaptador::sustrato de ADN de 10 a 100, y realizamos las reacciones a 37 °C durante una hora a pH 7,8 con ATP 1 mM, $MgCl_2$ 10 mM y PEG-8000 al 10 %. Se usaron como controles la ligazón de los mismos adaptadores a sustratos de extremos romos así como también las reacciones sin ligasa. Para analizar los rendimientos del producto de ligazón, las reacciones se procesaron en geles de poliacrilamida desnaturalizantes (Figura 21b) o geles en TBE (Figura 22a-d). La relación de intensidad de producto a sustrato se usó para cuantificar la eficiencia de ligazón a través de ImageJ (Figura 21b-c). La ligazón del saliente 5' (carril 11 en la Figura 1b) pareció estar completa en más del 90 %, incluso más que el control de ligazón de extremos romos (carril 14, 76,9 %), lo que sugiere una eficiencia de ligazón notablemente alta de los salientes 5' de ADN. Los sustratos con brechas de 1 u 8 nt (carriles 5 y 8) mostraron una buena eficiencia de ligazón de alrededor del 60 %. Sin embargo, la eficiencia de ligazón de mellas (carril 2) fue la más baja, alrededor del 20 %. Pero el rendimiento de ligazón podría mejorarse si incubamos la ligazón de mellas durante 12 horas, lo que sugiere una cinética más lenta para la reacción de ligazón de mellas (Figura 22).

También extendimos nuestro estudio a diferentes secuencias de adaptadores y sustratos (Figura 22). Los extremos 5' PO_4 de tres adaptadores diferentes (Ad-T, Ad-A o Ad-GA) contenían una sola T o A o un dinucleótido GA antes de una secuencia consenso CTGCTGA como ligazón. Se ligaron individualmente a los extremos 3'OH de las plantillas aceptoras con una T en la unión de ligazón. En general, se observó una eficiencia de ligazón más alta (70-90 %) en todos los casos, excepto en las ligazones de mellas, con Ad-T y Ad-A que con Ad-GA (Figura 22), lo que indica algunas preferencias de nucleótidos de la ligasa de ADN T4 en las uniones de ligazón. A pesar de las secuencias del adaptador y del sustrato, las ligazones de rama 3' o saliente 5' siempre mostraron mejores eficiencias (60-90 %), mientras que la ligazón de la mella fue bastante ineficiente con 1 hora de incubación. Presumimos que estas discrepancias en la eficiencia de ligazón se deben a la flexión del ADN donde comienza la mella/la brecha/el saliente y expone el grupo 3'OH para la ligazón. Probablemente la región de ADNsh más larga haga que los extremos 3' sean más accesibles en la ligazón y resulte en una eficiencia de ligazón más alta. También probamos si podría ocurrir un evento de unión de extremos similar, como la ligazón de rama 5'. Por el contrario, no se observó ligazón evidente de un adaptador de extremo romo al extremo 5'PO_4 en la brecha o el saliente 3', lo que indica que la ligasa de ADN T4 posiblemente tenga una estructura terciaria más estricta en los extremos 5' del donante que en los extremos 3'.

4.3 Ligazón de rama 3', una nueva forma de ligar ADN a ARN

Investigamos adicionalmente la 3'BL en híbridos de ADN/ARN (ON21/22) que forman un saliente 5' de ADN y uno de ARN (Figura 23a). Los controles negativos de ligazón incluyeron híbridos de ADN/ARN, oligos de ADNsh o ARNsh, individualmente o incubados con adaptadores (carriles 3, 4 y 5 en la Figura 23b). Curiosamente, cuando los híbridos de ADN/ARN se incubaron con adaptadores, vimos un cambio de tamaño del oligonucleótido de ARN desde 29 nt originales a 49 nt con una eficiencia >90 %, lo que sugiere que la ligasa de ADN T4 puede ligar eficientemente el adaptador al ARN. Sin embargo, el sustrato de ADN permaneció sin cambios (carriles 1 y 2 en la Figura 23b). Esto sugiere que los adaptadores de ADN de extremos romos se ligaron al extremo 3' del ARN en el saliente 5' del ADN, pero no al extremo 3' del ADN en el saliente 5' del ARN. Para confirmar que se necesitaba la estructura saliente 5' para la 3'BL, llevamos a cabo la misma reacción de ligazón mediante el reemplazo del oligonucleótido de ADN original (ON21) con otra plantilla de ADN largo (ON23) que no es complementaria con el ARN ON22. Como era de esperar, no se observó ligazón con el uso de la plantilla de ADN ON23, lo que sugiere que la 3'BL solo puede ocurrir con salientes 5'. Nuestro hallazgo indica que la ligasa de ADN T4 tiene determinadas preferencias de sustrato, posiblemente causadas por diferencias en las afinidades de unión entre proteína y sustrato.

Un estudio anterior sugirió que la ligasa de ADN T4 y la ligasa de ARN T4 2, pero no la ligasa de ARN T4 1, pueden unir un extremo 5'PO_4 de ADN a un extremo 3'OH de ADN o ARN yuxtapuesto en un híbrido ARN/ADN bicatenario, pero no a un 3'OH de ARN (Bullard 2006, Biochem J 398: 135-144). Realizamos la misma prueba de ligazón mediante el uso de las ligasas de ARN T4 1 y 2 (Figura 24C). Parecía que las ligasas de ARN T4 1 y 2 podían ligar el adaptador de extremo romo al ARN, pero la eficiencia de la ligazón era muy baja (<10 %).

4.4 Construcción de la biblioteca de inserción de transposones direccional

Dado que se demostró que la ligazón de rama 3' es útil para ligar adaptadores a varias estructuras genómicas con alta eficiencia, exploramos su aplicación en los flujos de trabajo de NGS. El método de construcción de bibliotecas basado en transposones es eficiente en cuanto a tiempo y consume menos ADN de entrada que la preparación de bibliotecas NGS convencionales. Sin embargo, mediante el uso de sistemas comerciales de preparación de bibliotecas basados en transposones, solo la mitad de las moléculas etiquetadas están flanqueadas por dos secuencias adaptadoras diferentes, y el ADN etiquetado está flanqueado por regiones autocomplementarias que pueden formar estructuras de horquilla estables que pueden comprometer la calidad de la secuenciación (Gorbacheva, 2015, Biotechniques abril; 58(4): 200-202). Además, la incorporación de secuencias adaptadoras mediada por PCR, no se ha adaptado para la secuenciación con bisulfito del genoma completo ni para la construcción de bibliotecas de NGS sin PCR.

Para superar estas limitaciones, hemos desarrollado un nuevo protocolo para la construcción de bibliotecas NGS basadas en transposones que incorpora 3'BL. Los transposones Tn5 y MuA funcionan a través de un mecanismo de "cortar y pegar", donde la secuencia adaptadora del transposón se une al extremo 5' del ADN diana, lo que crea una brecha de 9 pb o 5 pb, respectivamente, en el extremo 3' del ADN genómico (Figura 24). Luego, se usó 3'BL para añadir otra secuencia adaptadora al extremo 3' del ADN genómico en la brecha para completar la ligazón del adaptador direccional. Comparamos la eficiencia del enfoque 3'BL con la de un enfoque de inserción de doble transposón, que usa dos adaptadores basados en Tn5 diferentes, TnA y TnB. El ADN genómico humano se incubó con el complejo de transposoma TnA solo o con cantidades equimolares de los complejos de transposoma TnA y TnB. El producto de la fragmentación del transposoma TnA se usó además para la 3'BL con el adaptador de extremo romo AdB, que comparte una secuencia adaptadora común con TnB. La amplificación por PCR mediante el uso de dos cebadores, Pr-A y Pr-B, diseñados para adaptadores TnA y AdB/TnB, respectivamente, mostró un rendimiento de PCR similar (Figura 4b, carriles 9 y 10, y Figura 4c) lo que sugiere que estos dos enfoques tenían la misma eficiencia. No se observó una amplificación significativa cuando solo se usó un cebador específico para los adaptadores TnA o AdB/TnB (Figura 4b y Figura 4c). Como se esperaba debido a la supresión de la PCR, tanto el enfoque de ligazón 3' como el enfoque de inserción de doble transposón mostraron una eficiencia de PCR

significativamente más alta que la reacción de inserción del transposón con solo el complejo de transposoma TnA o TnB solos (Figura 24b, carril 3 y carril 8, y Figura 24c).

4.1 Materiales y métodos

Ligazón de rama 3' para ADN bicatenario

Los sustratos para 3'BL estaban compuestos por 2 pmol de ON1 u ON9 mezclados con 4 pmol cada uno de uno o dos oligos adicionales en tampón Tris-EDTA (TE) de pH 8 (Life Technologies). Sustrato 1 y 5 (mella): ON1/2/3 y ON9/10/11; sustrato 2 y 6 (brecha de 1 pb): ON1/2/4 y ON9/10/12; sustrato 3 (brecha de 8 pb): ON1/4/5; sustrato 4 y 9 (saliente 5'): ON1/2 y ON9/10; sustrato 7 (brecha de 2 pb): ON9/10/13; sustrato 8 (brecha de 3 pb): ON9/10/14; control de extremo romo: ON1/6 (Figura 1, Tabla complementaria 1). La plantilla se ligó a 180 pmol de adaptador (Ad-C: ON7/8, Ad-T: ON15/16, Ad-A: ON17/8 o Ad-GA: ON19/20) mediante el uso de 2400 unidades de ligasa T4 (Enzymatics Inc) en tampón 3'BL [0,05 mg/ml de BSA (New England Biolabs), Tris-Cl 50 mM, pH 7,8 (Amresco), MgCl2 10 mM (EMD Millipore), DTT 0,5 mM (VWR Scientific), PEG -8000 al 10 % (Sigma Aldrich) y ATP 1 mM (Sigma Aldrich)]. Las pruebas de optimización se realizaron mediante la alteración de la concentración de ATP desde 1 µM a 1 mM, la concentración de Mg^{2+}, el valor de pH, la temperatura desde 12 a 42 °C y aditivos como PEG-8000 desde 2,5 % a 10 % y SSB desde 2,5 a 20 ng/µl. La mezcla de ligazón se preparó en hielo y se incubó a 37 °C durante 1 a 12 horas, seguido de inactivación por calor a 65 °C durante 15 min. Las muestras se purificaron mediante el uso de perlas Axygen (Corning) y se eluyeron en 40 µl de tampón TE. Todas las reacciones de ligazón se procesaron en geles de poliacrilamida desnaturalizantes o en TBE al 6 % (Life Technologies) y se visualizaron en un Alpha Imager (Alpha Innotech). Se cargó un control de entrada en una cantidad igual o la mitad de la cantidad de plantillas usadas para la ligazón. Se estimó una tasa de eficiencia de ligazón al dividir la intensidad de los productos ligados por la intensidad total de los productos ligados y no ligados, mediante el uso del software ImageJ (NIH).

Ligazón de rama 3' para el híbrido de ADN/ARN

Los sustratos para 3'BL estaban compuestos por 10 pmol de oligonucleótido de ARN ON22 mezclados con 2 pmol de oligo de ADN ON21 u ON23. Para la 3'BL mediada por la ligasa de ADN T4, el sustrato se incubó con Ad-T (ON15/16) en tampón 3'BL como se describe anteriormente y se incubó a 37 ºC durante 1 hora. Se realizó la 3'BL mediante el uso de la ligasa de ARN T4 1 o 2 en su propio tampón de ligasa de ARN 1x (NEB) junto con DMSO al 20 %. Todos los productos de ligazón se ensayaron en geles de poliacrilamida desnaturalizantes al 6 %.

Construcción de la biblioteca de inserción de transposones direccional

Los oligonucleótidos de transposones usados en este experimento se sintetizaron por Sangon Biotech. Para los 2 experimentos de transposones que usan TnA y TnB, los oligos TnA, TnB y MErev se hibridaron en una relación de 1:1:2. Para el experimento de un solo transposón con tn1, tn1 y MErev se hibridaron en una relación de 1:1.

El ensamblaje del transposoma se realizó mediante la mezcla de 15 pmol de adaptadores hibridados previamente, 7 µl de transposasa Tn5 (Vazemy) y 5,5 µl de glicerol para obtener una reacción de 20 µl que se incubó a 30 °C durante 1 hora. La inserción de transposones en el ADN genómico (Coriell 19240) se llevó a cabo en reacciones de 20 µl que contenían 100 ng de ADNg, tampón TAG (Vazyme) y 2 µl del transposoma ensamblado. La reacción se incubó a 55 °C durante 10 min, seguido de la adición de 100 µl de tampón PB (Qiagen) para retirar el complejo de transposoma del ADN tagmentado y la purificación mediante el uso de perlas Agencourt AMPure XP (Beckman Coulter). La ligazón de rama 3' de AdB (ONB1, ONB2) al ADN tagmentado se realizó en reacciones que contenían 100 pmol del adaptador, 600 U de la ligasa de ADN T4 (Enzymatics Inc.) y tampón 3'BL incubado a 25 °C durante 1 hora. Las reacciones se purificaron mediante el uso de perlas AMPure XP. La amplificación por PCR del ADN tagmentado y con brechas ligadas se hizo en reacciones de 50 µl que contenían 2 µl del ADN tagmentado o con brechas ligadas, tampón TAB, 1 µl de enzima de amplificación TruePrep (Vazyme), dNTP 200 mM (Enzymatics Inc.) y 400 mM de cada cebador Pr-A y Pr-B. Las reacciones tagmentadas se procesaron a 72 °C durante 3 min; 98 °C durante 30 seg; 98 °C durante 10 seg, 58 °C durante 30 seg, 72 °C durante 2 min, por 8 ciclos; y 72 °C para una extensión de 10 minutos. Las reacciones de ligazón de brechas se procesaron mediante el uso del mismo programa sin la extensión inicial de 3 min a 72 °C. Las reacciones de PCR se purificaron mediante el uso de perlas AMPure XP, ya sea en una selección de tamaño de una sola etapa o a través de fraccionamiento doble. Los productos purificados se cuantificaron mediante el uso del kit de ADN de alta sensibilidad Qubit (Invitrogen).

Ejemplo 4: Ligazón de rama 3': Un nuevo método para ligar ADN no complementario a extremos 3'OH internos o empotrados en el ADN o ARN

Las ligasas de ácido nucleico son enzimas cruciales que reparan rupturas en el ADN o ARN durante la síntesis, reparación y recombinación. Se han desarrollado diversas herramientas moleculares mediante el uso de las diversas actividades de las ligasas de ADN/ARN. Sin embargo, quedan por descubrir actividades de la ligasa adicionales. En la presente descripción, demostramos la capacidad no convencional de la ligasa de ADN T4 para unir ADN de doble hebra de extremo romo fosforilado en 5' a rupturas de ADN en extremos recesivos 3', brechas o mellas para formar una estructura de rama 3'. Por lo tanto, esta ligazón independiente del apareamiento de bases se denomina ligazón

de rama 3' (3'BL). En un estudio exhaustivo de las condiciones óptimas de ligazón, similar a la ligazón de extremos romos, la presencia de PEG-8000 al 10 % en el tampón de ligazón, aumentó significativamente la eficiencia de la ligazón. Se observó cierta preferencia de nucleótidos en los sitios de unión mediante el uso de diferentes ADN sintéticos, lo que indica un nivel de sesgo de ligazón para la 3'BL. Además, descubrimos que la ligasa de ADN T4 ligaba eficientemente el ADN al extremo 3' del ARN en un híbrido ADN/ARN, mientras que las ligasas de ARN son menos eficientes en esta reacción. Estas nuevas propiedades de la ligasa de ADN T4 pueden utilizarse como una técnica molecular amplia en muchas aplicaciones importantes. Realizamos un estudio de prueba de concepto de un nuevo protocolo de tagmentación direccional para la construcción de bibliotecas de secuenciación de próxima generación (NGS) que elimina los adaptadores invertidos y permite la inserción de códigos de barras de muestra adyacentes al ADN genómico. La 3'BL, después de la tagmentación de un solo transposón, puede lograr, teóricamente, una plantilla que se puede usar al 100 %, y nuestros datos empíricos demuestran que el nuevo enfoque produjo un mayor rendimiento en comparación con la tagmentación tradicional de doble transposón o transposón Y. Exploramos además, el posible uso de 3'BL para preparar bibliotecas NGS de ARN dirigidas con una mitigación del sesgo basado en la estructura y de los problemas con dímeros de adaptadores.

5.1 Introducción

Las ligasas reparan rupturas en los ácidos nucleicos, y esta actividad es esencial para la viabilidad y vitalidad celular. Las ligasas de ADN catalizan la formación de un enlace fosfodiéster entre los extremos del ADN y desempeñan funciones cruciales en la reparación, recombinación y replicación del ADN in vivo (1-3). Las ligasas de ARN unen los extremos de ARN 5'-fosforilo (5'PO4) y 3'-hidroxilo (3'OH) a través de enlaces fosfodiéster y participan en la reparación, corte y empalme y edición del ARN (4). Las ligasas de los tres reinos de organismos (bacterias, arqueobacterias y eucariotas) pueden utilizarse in vitro como herramientas moleculares importantes para aplicaciones tales como la clonación, amplificación o detección basada en ligasas y biología sintética (5-7).

Una de las ligasas más ampliamente usadas in vitro es la ligasa de ADN del bacteriófago T4, que es un solo polipéptido de 55 kDa que requiere ATP como fuente de energía (8). La ligasa de ADN T4 típicamente une a los extremos 5'PO4 y 3'OH adyacentes del ADN bicatenario. Además de sellar las mellas y ligar extremos cohesivos, la ligasa de ADN T4 también puede catalizar eficientemente la unión de extremos romos, lo que no se ha observado en ninguna otra ligasa de ADN (9,10). Algunas propiedades catalíticas inusuales de esta ligasa se informaron anteriormente, tales como el sellado de brechas de simple hebra en el ADN bicatenario, el sellado de mellas adyacentes a sitios abásicos en el ADN de doble hebra (ADNdh), la promoción de la formación de bucles intramoleculares con ADN parcialmente de doble hebra y la unión de hebras de ADN que contienen extensiones de rama 3' (11-13). Los investigadores también observaron la ligazón independiente de la plantilla mediada por la ligasa T4, tal como el sellado de mellas mal apareadas en el ADNdh (14) o incluso la ligazón de ADN de simple hebra (ADNsh), aunque con una eficiencia muy baja (15). Estos resultados sugieren que el apareamiento perfecto de bases complementarias en la unión de ligazón o adyacente a esta, no es críticamente necesario para algunas actividades no convencionales de la ligasa de ADN T4. Las ligasas de ARN T4 1 y 2 son los productos de los genes 63 y 24, respectivamente, del fago T4. Ambos requieren un extremo 5'PO4 y 3'OH adyacente para una ligazón exitosa con la hidrólisis concurrente de ATP a AMP y PPi. Los sustratos para la ligasa de ARN T4 1 incluyen ARN y ADN de simple hebra, mientras que la ligasa de ARN T4 2 sella preferentemente las mellas en el ARNdh en lugar de ligar los extremos del ARNsh (16,17).

Aquí, demostramos un evento de unión de extremos no convencional mediado por la ligasa de ADN T4 que llamamos ligazón de rama 3' (3'BL). Este método puede unir fragmentos de ADN o ADN/ARN en las mellas, brechas de simple hebra o extremos recesivos 3' para formar una estructura de rama. Este informe incluye un estudio exhaustivo de una amplia variedad de cofactores y activadores de la ligazón y la optimización de las condiciones de la ligazón para este tipo de ligazón nueva. Con nuestro protocolo 3'BL, no se requirió apareamiento de bases, y la ligazón puede completarse en un 70-90 % en la mayoría de los casos, incluso para una brecha de 1 nt. Una aplicación de este método es la unión de adaptadores a ADN o ARN durante la preparación de la biblioteca NGS. Varias estructuras genómicas, que anteriormente se consideraba que no se podían ligar, ahora pueden usarse como sustratos para 3'BL, lo que resulta en una alta tasa de conversión de ADN de entrada en moléculas ligadas al adaptador, mientras se evitan las quimeras. Demostramos que 3'BL puede acoplarse con la tagmentación de transposones para aumentar el rendimiento de la biblioteca. La estrategia de tagmentación direccional que proponemos producirá, teóricamente, plantillas, el 100 % de las cuales puede utilizarse para la secuenciación. Nuestro estudio demostró el valor de esta nueva técnica para la preparación de bibliotecas NGS y el potencial para avanzar en muchas otras aplicaciones moleculares.

5.2 Resultados: Ligazón de rama 3', un nuevo método para ligar extremos de ADN

Convencionalmente, la ligazón de ADN implica la unión de los extremos de ADN 5'PO4 y 3'OH de fragmentos de extremos cohesivos o romos. La ligazón de extremos cohesivos es generalmente más rápida y menos dependiente de la concentración de enzimas en comparación con la unión de extremos romos. Ambos procesos pueden catalizarse por la ligasa de ADN del bacteriófago T4, que usa ATP como cofactor productor de energía y requiere Mg2+ (8). También se informó que la ligasa de ADN T4 liga oligos de simple hebra específicos o degenerados a sustratos parcialmente de simple hebra a través de la hibridación (18,19). Aquí, demostramos una ligazón mediada

por la ligasa de ADN T4 no convencional que no requiere apareamiento de bases complementarias y puede ligar un donante de ADN de extremo romo al extremo 3'OH de un aceptor de ADN bicatenario en hebras empotradas 3', brechas o mellas (Figura 26a). Por lo tanto, usamos el término ligazón de rama 3' (3'BL) para describir estas ligazones. El ADN donante sintético que usamos contenía un extremo bicatenario romo en 5' y un extremo de ADNsh en 3'. Los sustratos aceptores contenían una de las siguientes estructuras: una mella desfosforilada, una brecha de 1 u 8 nucleótidos (nt) o un extremo empotrado 3' de 36 nt (Tabla complementaria 1). La ligasa T4 ayuda a unir el 5'PO_4de la hebra donadora al único 3'OH que se puede ligar de la hebra aceptora para formar un producto de ligazón en forma de rama.

Para optimizar la eficiencia de la ligazón, probamos exhaustivamente una serie de factores que afectan la eficiencia general de la ligazón, que incluyen la relación adaptador:sustrato de ADN, la cantidad de ligasa T4, la concentración final de ATP, la concentración de Mg2+, el pH, el tiempo de incubación y diferentes aditivos, tales como el polietilenglicol-8000 (PEG-8000) y la proteína de unión a simple hebra (SSB). La adición de PEG-8000 hasta una concentración final del 10 % aumentó sustancialmente la eficiencia de la ligazón desde menos del 10 % hasta más del 80 % (Figuras 26 y 27). Una amplia gama de concentraciones de ATP (desde 1□M a 1 mM) y concentraciones de Mg2+ (3 mM a 10 mM) fueron compatibles con la 3'BL. La cantidad de ligasa necesaria para la 3'BL fue comparable a la de la ligazón de extremos romos. En nuestras condiciones optimizadas, usamos relaciones molares de ADN donante:sustrato de 30 a 100, y realizamos las reacciones a 37 °C durante una hora a pH 7,8 con ATP 1 mM, MgCl2 10 mM y PEG-8000 al 10 %. La ligazón de los mismos donantes a sustratos de extremos romos y las reacciones sin ligasa se usaron como controles positivo y negativo, respectivamente.

El donante de ligazón (Ad-G) es de doble hebra en un extremo (fosforilado en 5' y protegido con didesoxi en 3') y de simple hebra (protegido con didesoxi en 3') en el otro extremo (Figura 26). Los sustratos de ligazón están compuestos por la misma hebra inferior (ON1) con diferentes hebras superiores para componer estructuras de mellas, brechas y salientes. Para cuantificar los rendimientos de los productos de la ligazón, los productos de reacción se separaron en geles de poliacrilamida desnaturalizantes al 6 % (Figura 26b). La eficiencia de la ligazón se calculó como la relación entre la intensidad del producto y el sustrato mediante el uso de ImageJ (Figura 26b-c). La ligazón recesiva 3' (carril 11 en la Figura 26b) pareció completarse en aproximadamente un 90 %, lo que es incluso más alto que el control de ligazón de extremos romos (carril 14, 72,74 %) y sugiere una eficiencia de ligazón notablemente alta con extremos recesivos 3' de ADN. Los sustratos con brecha de 1 u 8 nt (carriles 5 y 8) mostraron una buena eficiencia de ligazón de aproximadamente el 45 %. La eficiencia de la ligazón de mellas (carril 2) fue la más baja con aproximadamente un 13 %. Sin embargo, este rendimiento de ligazón mejoró cuando la reacción de ligazón de mellas se incubó durante más tiempo, lo que sugiere una cinética más lenta para la reacción de ligazón de mellas.

También extendimos nuestro estudio a diferentes secuencias de adaptadores y sustratos (Figura 27). Los extremos 5'PO4 de tres adaptadores diferentes (Ad-T, Ad-A o Ad-GA en la Tabla complementaria 1) contenían una sola T o A o el dinucleótido GA en la unión de ligazón antes de una secuencia CTGCTGA consenso. Estos extremos 5'PO4 se ligaron individualmente a los extremos 3'OH de las plantillas aceptoras con una T en la unión de ligazón. En general, se observó una alta eficiencia de ligazón (70-90 %) en la mayoría de los casos excepto para ligazones de mellas o 3'BL con el uso de Ad-GA (Figura 27f), lo que indica cierta preferencia de nucleótidos de ligasa de ADN T4 en las uniones de ligazón. Independientemente de las secuencias del adaptador y del sustrato, las ligazones del extremo recesivo 3' o de las brechas siempre mostraron mejores eficiencias (60-90 %), mientras que la ligazón de mellas fue bastante ineficiente, en una incubación de una hora. Presumimos que estas discrepancias en la eficiencia de la ligazón se deben a la flexión del ADN donde comienza la mella/la brecha/el saliente y expone el grupo 3'OH para la ligazón. Probablemente la región de ADNsh más larga haga que los extremos 3' sean más accesibles en la ligazón y resulte en una eficiencia de ligazón más alta. También probamos si podría ocurrir un evento de unión de extremos similar, como la ligazón de rama 5'. Por el contrario de 3'BL, no se observó ligazón evidente de un adaptador de extremo romo al extremo 5'PO4 en la brecha o el extremo recesivo 5'. Este resultado sugiere un mayor impedimento estérico de la ligasa de ADN T4 en los extremos 5' del donante en comparación con los extremos 3'.

5.3 Ligazón de rama 3' para ligar ADN a ARN

Investigamos adicionalmente la 3'BL en híbridos de ADN/ARN (ON-21/ON-23 en la Tabla 3) que forman un saliente 5' de ADN y uno de ARN (Figura 28a). La ligazón en híbridos de ADN/ADN sirvió como control positivo, mientras que los controles negativos de ligazón incluyeron híbridos de ADN/ARN, ADNsh u oligos de ARNsh incubados individualmente o con adaptadores (carriles 3, 4 y 5 en la Figura 28c y). Curiosamente, cuando los híbridos de ADN/ARN se incubaron con un donante de ADNdh de extremo romo, observamos un cambio de tamaño del oligo de ARN desde los 29 nt originales a 49 nt tras la ligazón. Sin embargo, el sustrato de ADN permaneció sin cambios (carriles 1 y 2 en la Figura 28c). Este resultado sugiere que el donante de ADNdh de extremo romo se ligó al extremo 3' del ARN en el extremo recesivo 3' del ADN pero no al extremo 3' del ADN en los extremos recesivos 3' del ARN. Como control positivo, los híbridos de ADN/ADN con extremos recesivos 3' en cada lado mostraron cambios de banda a especies más grandes en ambas hebras con una eficiencia de casi el 100 %. Para confirmar que se necesitaba la estructura recesiva 3' para 3'BL, realizamos la misma reacción de ligazón mientras se reemplazaba el oligo de ADN original (ON-21) con otra plantilla de ADN largo (ON-23) que no es complementaria al ARN ON-22 (Figura 28b). Como era de esperar, no se observó ligazón con el uso de la plantilla de ADN ON-23

(carril 10-13 en la Figura 28c). Nuestro hallazgo indica que la ligasa de ADN T4 puede promover la 3'BL en híbridos de ADN/ARN y que esta actividad tiene determinadas preferencias de sustrato estéricas que pueden verse afectadas por las diferencias en las afinidades de unión del sustrato y la ligasa de ADN T4.

Un estudio anterior informó que para sellar mellas en híbridos de ADN/ARN, la ligasa de ADN T4 y la ligasa de ARN T4 2, pero no la ligasa de ARN T4 1, pueden unir efectivamente un extremo de ADN 5'PO4 a un extremo 3'OH de ADN o ARN yuxtapuesto cuando la hebra complementaria es ARN pero no ADN (17). Por lo tanto, realizamos la misma prueba de ligazón mediante el uso de las ligasas de ARN T4 1 y 2 en DMSO al 20 % (Figura 28d) o en PEG al 10 %. En ambas pruebas, la ligasa de ARN T4 1 y la ligasa de ARN T4 2 ligaron levemente los adaptadores de extremo romo al extremo 3' del ARN en un híbrido ADN/ARN. En particular, en los controles solo de ARN, la ligasa de ARN T4 2 podía unir a un adaptador de ADNdh de extremo romo a ARNsh. En conclusión, la ligasa de ADN T4, pero no la ligasa de ARN T4, es competente para ligar eficientemente el ADNdh de extremo romo al extremo 3' del ARN a través de 3'BL.

5.4 Construcción de bibliotecas de tagmentación direccional

Debido a que 3'BL es útil para ligar adaptadores a varias estructuras genómicas con alta eficiencia, exploramos su aplicación en los flujos de trabajo NGS. La construcción de bibliotecas basadas en transposones es rápida y consume menos ADN de entrada en comparación con la preparación de bibliotecas NGS convencionales. Sin embargo, al usar sistemas comerciales de preparación de bibliotecas basados en transposones, solo la mitad de las moléculas etiquetadas están flanqueadas por dos secuencias adaptadoras diferentes (Figura 29a), y el ADN etiquetado está flanqueado por regiones autocomplementarias que podrían formar estructuras de horquilla estables y comprometer la calidad de la secuenciación (20). Además, la incorporación de secuencias adaptadoras mediada por PCR no se ha adaptado para la secuenciación con bisulfito del genoma completo o la construcción de bibliotecas NGS sin PCR.

Para superar estas limitaciones, desarrollamos un nuevo protocolo para la construcción de bibliotecas NGS basadas en transposones mediante la incorporación de 3'BL. Los transposones Tn5 y MuA funcionan a través de un mecanismo de "cortar y pegar", en el que una secuencia adaptadora de transposón se une al extremo 5' del ADN diana para crear una brecha de 9 pb o 5 pb, respectivamente, en el extremo 3' del ADN genómico (Figura 29a). Subsecuentemente, puede usarse 3'BL para añadir otra secuencia adaptadora al extremo 3' del ADN genómico en la brecha para completar la ligazón direccional del adaptador (Figura 29c). Usamos transposones Tn5 en este manuscrito para comparar la eficiencia del enfoque de tagmentación simple + 3'BL (Figura 29c) con la de un enfoque de tagmentación doble, que usa los dos adaptadores TnA y TnB basados en Tn5 diferentes (Figura 29a), y a la de otra estrategia de tagmentación simple direccional que usa adaptadores Y que contienen dos secuencias adaptadoras diferentes (Figura 29b). El ADN genómico humano se incubó con cantidades equimolares de complejos de transposomas TnA y TnB, con el complejo de transposoma TnA solo o con el complejo de transposoma TnY (TnA/B).

El producto de la fragmentación solo del transposoma TnA se usó además como una plantilla para la 3'BL con el adaptador de extremo romo AdB, que comparte una secuencia adaptadora común con TnB. La amplificación por PCR se realizó mediante el uso de dos cebadores, Pr-A y Pr-B, diseñados para reconocer los adaptadores TnA y AdB/TnB, respectivamente. Los datos de cuantificación sugirieron que TnA&AdB tenían la eficiencia más alta en comparación con TnA&TnB y TnY (TnA/B) (Figura 29d). No se observó una amplificación significativa cuando solo se usó un cebador específico para el adaptador TnA (Figura 29d). Como se esperaba debido a la supresión de la PCR, el enfoque TnA-3'BL, el enfoque de tagmentación doble y el enfoque TnY mostraron una eficiencia de PCR significativamente más alta que la reacción de tagmentación con solo el complejo de transposoma TnA o TnB solos (Figura 29d).

También secuenciamos estas bibliotecas mediante el uso de BGISEQ-500 y comparamos el sesgo posicional de base entre el extremo interferido por el transposón, el extremo 3'BL y el extremo de ligazón TA regular (Figura 30). Es evidente que el sesgo posicional en el extremo 3'BL es menor que en el extremo Tn5 (Figura 30a-b), lo que ocurre porque el extremo 3'BL está influenciado tanto por la interrupción del transposón como por la 3'BL. Debido a que solo los primeros 6 nt (posición 1-6) del extremo 3'BL mostraron un sesgo de base y el sesgo fue similar, pero no exactamente igual, al de su extremo Tn5 hibridado (posición 30-35, después del saliente de 9-nt), concluimos que el sesgo posicional que observamos en el extremo 3'BL es causado principalmente por el transposón Tn5. Por lo tanto, 3'BL causa un sesgo mínimo y es similar a la ligazón TA regular (Figura 30c).

5.5 Discusión

Una propiedad importante de la ligasa de ADN T4 es su unión eficiente de ADNdh de extremos romos (21,22), que no se ha observado con otras ligasas de ADN. También se informó que esta ligasa media en algunos eventos catalíticos inusuales, tales como la ligazón de brechas de simple hebra o bases no coincidentes en el ADN bicatenario (11,12), la formación de una molécula de tallo y bucle a partir de ADN parcialmente de doble hebra (13), o la ligazón ineficiente de ADNsh de manera independiente de la plantilla (20).

Aquí, demostramos que la ligasa de ADN T4 catalizó la unión de ADNdh de extremo romo al extremo 3'OH del ADNdh con una mella y la unión de ADN bicatenario parcialmente de simple hebra con una brecha o saliente 5'. Por el contrario, no se observó ligazón al extremo 5'PO4 en los extremos 5' empotrados o en las brechas, lo que indica que después de unirse al extremo 5'PO4 del adaptador de ADNdh, la ligasa de ADN T4 puede acceder al extremo 3' empotrado cuando el ADN se flexiona. Con nuestro método 3'BL, no se requirió apareamiento de bases, e incluso para una brecha de 1 nt, se logró completar más del 70 % con condiciones optimizadas. Sin embargo, se observaron diferentes eficiencias de ligazón para ligar T, A o GA 5' a T 3' (Figura 2), lo que indica cierta preferencia de secuencia en la unión de la ligazón. A pesar del sesgo de ligazón reconocido (23), la ligasa de ADN T4 se usa comúnmente en la etapa de adición del adaptador durante la preparación de la biblioteca NGS. Con su capacidad para realizar 3'BL, la ligasa T4 puede ligar adaptadores a varias estructuras genómicas que anteriormente se consideraba que no se podían ligar, lo que resulta una mayor tasa de uso de la plantilla. 3'BL también se puede acoplar con la tagmentación de transposones. En la estrategia tradicional de doble transposón solo el 50 % de las moléculas tagmentadas son susceptibles para la siguiente etapa de amplificación. Sin embargo, cuando la tagmentación de ADN se realiza mediante el uso de un transposón con 3'BL subsecuente, puede adquirirse un aumento del rendimiento de moléculas con diferentes adaptadores en cada extremo del inserto (Figura 4). Además, los productos 3'BL tagmentados pueden cargarse directamente en la celda de flujo de Illumina como bibliotecas WGS sin PCR, lo que era difícil de lograr mediante el uso de la estrategia de doble transposón.

Se han propuesto otros protocolos de transposones direccionales mediante el uso de un transposón Y compuesto por dos secuencias adaptadoras diferentes o mediante el reemplazo de la hebra no enlazada de un solo transposón con un segundo oligo adaptador seguido de relleno de brechas y ligazón (24). Sin embargo, estos enfoques conservan las secuencias adaptadoras invertidas y no pueden insertar códigos de barras de muestra adyacentes al ADN genómico como lo hace el protocolo de 3'BL tagmentado. En base a los datos de NGS, los extremos genómicos ligados por 3'BL también demostraron menos posiciones con sesgo de composición de base posicional, y el primer sesgo de 6 nt fue leve y causado principalmente por la interrupción del transposón, lo que sugiere que 3'BL tiene un sesgo posicional mínimo. Mediante el uso de este nuevo método de construcción de bibliotecas, Wang y otros, lograron exitosamente, una asignación de variantes completa y altamente precisa en WGS y una determinación de haplotipos casi perfecta de las variantes en cóntigos largos con tamaño N50 de hasta 23,4 Mb para la lectura de fragmentos largos (BioRxiv, https://doi.org/10.1101/324392).

En este estudio, también investigamos 3'BL mediante el uso de plantillas de un ADN/ARN bicatenario quimérico que forma un saliente 5' de ADN y un saliente 5' de ARN (Figura 3). Inesperadamente, el ADNdh de extremos romos se ligó eficientemente a los extremos 3' del ARN, pero no al ADN, lo que sugiere que la ligasa T4 tiene preferencia por la formación de complejos ternarios. La eficiencia de ligazón se redujo en gran medida si se usaba ligasa de ARN T4 I o II para unir los extremos. Otra aplicación preliminar, pero importante, de 3'BL con la ligasa de ADN T4 es el enriquecimiento de ARNm o la construcción de bibliotecas de ARN dirigidas, especialmente para miARN, los pequeños ARN reguladores cuya expresión descontrolada conduce a una serie de enfermedades (25,26). Por tanto, nuestra técnica 3'BL puede aplicarse fácilmente a la detección de cáncer y enfermedad de Alzheimer mediante el uso de miARN. La hibridación con sondas de ADN dirigidas a la cola de poli(A) o secuencias específicas de miARN puede usarse para crear híbridos de ADN-ARN con un saliente 5' de ADN, seguido de la ligazón a secuencias adaptadoras con códigos de barras de muestra y/o UID a través de 3'BL. Estas secuencias comunes luego pueden someterse a transcripción inversa para producir el ADNc de las secuencias de ARN diana. En comparación con las tecnologías actuales de captura de miARN, el uso de 3'BL mediado por la ligasa de ADN T4 podría proporcionar posiblemente varias ventajas para la construcción de bibliotecas de ARN NGS. En primer lugar, la hibridación con una hebra de ADN evitaría la formación de estructuras secundarias por parte de la hebra de ARN y, por lo tanto, mitigaría el sesgo introducido por otros protocolos. En segundo lugar, la ligasa de ADN T4 permite la adición de adaptadores con alta eficiencia a través de 3'BL, lo que evita las interacciones intramoleculares de ARN que pueden promover las ligasas de ARN. En tercer lugar, los dímeros de adaptadores pueden eliminarse de manera efectiva, lo que posiblemente haga innecesaria la purificación inconveniente en gel. Este nuevo método podría conducir a mejores perfiles de expresión de microARN no sesgados, con flujos de trabajo simples y escalables y, por tanto, los estudios de investigación a gran escala serían más asequibles.

Los hallazgos de este estudio se suman a la creciente comprensión de las actividades de la ligasa de ADN T4. Prevemos que la ligazón de rama 3' se convierta en una herramienta general en biología molecular que impulsará el desarrollo de nuevos métodos de ingeniería de ADN más allá de las aplicaciones NGS descritas.

5.6 Materiales y métodos

Ligazón de rama 3' para ADN bicatenario

Los sustratos para 3'BL estaban compuestos por 2 pmol de ON1 u ON9 mezclados con 4 pmol cada uno de uno o dos oligos adicionales en tampón Tris-EDTA (TE) de pH 8 (Life Technologies) de la siguiente manera: sustrato 1 y 5 (mella), ON-1/2/3 y ON-9/10/11; sustrato 2 y 6 (brecha de 1 nt), ON1/2/4 y ON9/10/12; sustrato 3 (brecha de 8 nt), ON1/4/5; sustrato 4 y 9 (saliente 5'), ON1/2 y ON9/10; sustrato 7 (brecha de 2 nt), ON9/10/13; sustrato 8 (brecha de 3 nt), ON9/10/14; control de extremo romo, ON1 y ON6 (Figura 26, Tabla complementaria 1). La plantilla se ligó a 180 pmol de adaptador (Ad-G: ON7/8, Ad-T: ON15/16, Ad-A: ON17/8 o Ad-GA: ON19/20) mediante el uso de 2400

unidades de ligasa T4 (Enzymatics Inc.) en tampón 3'BL [0,05 mg/ml de BSA (New England Biolabs), Tris-Cl 50 mM pH 7,8 (Amresco), MgCl2 10 mM (EMD Millipore), DTT 0,5 mM (VWR Scientific), PEG-8000 al 10 % (Sigma Aldrich) y ATP 1 mM (Sigma Aldrich)]. Las pruebas de optimización se realizaron mediante la alteración de la concentración de ATP desde 1 µM a 1 mM, la concentración de Mg2+ desde 3 a 10 mM, el valor de pH desde 3 a 9, la temperatura desde 12 a 42 °C y el ajuste de aditivos tales como PEG-8000 desde 2,5 % a 10 % y SSB desde 2,5 a 20 ng/µl. La mezcla de ligazón se preparó en hielo y se incubó a 37 °C durante 1 a 12 horas antes de la inactivación por calor a 65 °C durante 15 min. Las muestras se purificaron mediante el uso de perlas Axygen (Corning) y se eluyeron en 40 µl de tampón TE. Todas las reacciones de ligazón se procesaron en geles de poliacrilamida desnaturalizantes o en TBE al 6 % (Life Technologies) y se visualizaron en un Alpha Imager (Alpha Innotech). Se cargó un control de entrada en una cantidad igual o la mitad de la cantidad de plantilla usada para la ligazón. Se estimó una tasa de eficiencia de ligazón al dividir la intensidad de los productos ligados por la intensidad total de los productos ligados y no ligados, mediante el uso del software ImageJ (NIH).

Ligazón de rama 3' para el híbrido de ADN/ARN

Los sustratos para 3'BL estaban compuestos por 10 pmol de oligo de ARN ON-21 mezclado con 2 pmol de oligo de ADN ON-21 u ON-23. Para la 3'BL mediada por la ligasa de ADN T4, el sustrato se incubó con Ad-T (ON15/16) en tampón 3'BL como se describe anteriormente y se incubó a 37 °C durante 1 hora. Se realizó la 3'BL mediante el uso de la ligasa de ARN T4 1 o 2 en tampón de ligasa de ARN 1X (NEB) con DMSO al 20 % o PEG al 25 %. Todos los productos de ligazón se ensayaron en geles de poliacrilamida desnaturalizantes al 6 %.

Construcción de la biblioteca de tagmentación direccional

Los oligonucleótidos de transposones usados en este experimento se sintetizaron por Sangon Biotech. Para los 2 experimentos de transposones que usan TnA/TnB, los oligos para TnA (ON24), TnB (ON25) y MErev (ON26) se hibridaron en una relación de 1:1:2. Para el experimento de un solo transposón con TnA, ON24 y ON26 se hibridaron en una relación de 1:1. Para el experimento del transposón Y (TnA&TnB), ON24 y ON27 se hibridaron en una relación de 1:1.

El ensamblaje del transposón se realizó mediante la mezcla de 100 pmol de adaptadores hibridados previamente, 7 µl de transposasa Tn5 y suficiente glicerol para obtener un total de 20 µl de reacción, que se incubó a 30 °C durante 1 hora. La tagmentación del ADN genómico (Coriell 12878) se realizó en reacciones de 20 µl que contenían 100 ng de ADNg, tampón TAG (casero) y 1 µl del transposón ensamblado. La reacción se incubó a 55 °C durante 10 min; luego se añadieron 40 µl de clorhidrato de guanidina 6 M (Sigma) para retirar el complejo de transposón del ADN tagmentado y el ADN se purificó mediante el uso de perlas Agencourt AMPure XP (Beckman Coulter). La ligazón de la brecha de AdB (ON28 y ON29) al ADN tagmentado se realizó a 25 °C durante 1 hora en reacciones que contenían 100 pmol del adaptador, 600 U de ligasa de ADN T4 (Enzymatics Inc.) y tampón 3'BL. Las reacciones se purificaron mediante el uso de perlas AM Pure XP. La amplificación por PCR del ADN tagmentado y con brechas ligadas se realizó en reacciones de 50 µl que contenían 2 µl del ADN tagmentado o con brechas ligadas, tampón TAB, 1 µl de enzima de amplificación TruePrep (Vazyme), dNTP 200 mM (Enzymatics Inc.) y 400 mM de cada uno de los cebadores Pr-A y Pr-B. Las reacciones tagmentadas se incubaron de la siguiente manera: 72 °C durante 3 minutos; 98 °C durante 30 seg; 8 ciclos de 98 °C durante 10 seg, 58 °C durante 30 seg y 72 °C durante 2 min; y 72 °C para una extensión de 10 minutos. Las reacciones de ligazón de brechas se procesaron mediante el uso del mismo programa sin la extensión inicial de 3 min a 72 °C. Las reacciones de PCR que usaban prA (ON30) o tanto prA como prB (ON31) se purificaron mediante el uso de perlas AM Pure XP. Los productos purificados se cuantificaron mediante el uso del kit de ADN de alta sensibilidad Qubit (Invitrogen).

5.6 Referencias para el ejemplo 4

1. Lehnman, I. R. DNA ligase: structure, mechanism, and function. Science (80-.). 186, 790-797 (1974).
2. Tomkinson, A. E. y Mackey, Z. B. Structure and function of mammalian DNA ligases. Mutat. Res. Repair 407, 1-9 (1998).
3. Timson, D. J., Singleton, M. R. y Wigley, D. B. DNA ligases in the repair and replication of DNA. Mutat. Res. Repair 460, 301-318 (2000).
4. Ho, C. K., Wang, L. K., Lima, C. D. y Shuman, S. Structure and mechanism of RNA ligase. Structure 12, 327-339 (2004).
5. Tomkinson, A. E., Vijayakumar, S., Pascal, J. M. y Ellenberger, T. DNA ligases: structure, reaction mechanism, and function. Chem. Rev. 106, 687-699 (2006).
6. Pascal, J. M. DNA and RNA ligases: structural variations and shared mechanisms. Curr. Opin. Struct. Biol. 18, 96-105 (2008).
7. Shuman, S. DNA ligases: progress and prospects. J. Biol. Chem. 284, 17365-17369 (2009).
8. Dickson, K. S., Burns, C. M. y Richardson, J. P. Determination of the free-energy change for repair of a DNA phosphodiester bond. J. Biol. Chem. 275, 15828-15831 (2000).
9. Cai, L., Hu, C., Shen, S., Wang, W. y Huang, W. Characterization of bacteriophage T3 DNA ligase. J. Biochem. 135, 397-403 (2004).

10. Ampligase® Thermostable DNA Ligase. Disponible en: http://www.epibio.com/enzymes/ligases-kinases-phosphatases/dna-ligases/ampligase-thermostable-dna-ligase?details.
11. Nilsson, S. V y Magnusson, G. Sealing of gaps in duplex DNA by T4 DNA ligase. Nucleic Acids Res. 10, 1425-1437 (1982).
12. Goffin, C., Bailly, V. y Verly, W. G. Nicks 3' or 5' to AP sites or to mispaired bases, and one-nucleotide gaps can be sealed by T4 DNA ligase. Nucleic Acids Res. 15, 8755-8771 (1987).
13. Mendel-Hartvig, M., Kumar, A. y Landegren, U. Ligase-mediated construction of branched DNA strands: a novel DNA joining activity catalyzed by T4 DNA ligase. Nucleic Acids Res. 32, e2-e2 (2004).
14. Alexander, R. C., Johnson, A. K., Thorpe, J. A., Gevedon, T. y Testa, S. M. Canonical nucleosides can be utilized by T4 DNA ligase as universal template bases at ligation junctions. Nucleic Acids Res. 31, 3208-3216 (2003).
15. Kuhn, H. y Frank-Kamenetskii, M. D. Template-independent ligation of single-stranded DNA by T4 DNA ligase. FEBS J. 272, 5991-6000 (2005).
16. Ho, C. K. y Shuman, S. Bacteriophage T4 RNA ligase 2 (gp24. 1) exemplifies a family of RNA ligases found in all phylogenetic domains. Proc. Natl. Acad. Sci. 99, 12709-12714 (2002).
17. Bullard, D. R. y Bowater, R. P. Direct comparison of nick-joining activity of the nucleic acid ligases from bacteriophage T4. Biochem. J. 398, 135-144 (2006).
18. Broude, N. E., Sano, T., Smith, C. L. y Cantor, C. R. Enhanced DNA sequencing by hybridization. Proc. Natl. Acad. Sci. 91, 3072-3076 (1994).
19. Gunderson, K. L. y otros. Mutation detection by ligation to complete n-mer DNA arrays. Genome Res. 8, 1142-1153 (1998).
20. Gorbacheva, T., Quispe-Tintaya, W., Popov, V. N., Vijg, J. y Maslov, A. Y. Improved transposon-based library preparation for the Ion Torrent platform. Biotechniques 58, 200 (2015).
21. Sgaramella, V. y Khorana, H. G. CXII. Total synthesis of the structural gene for an alanine transfer RNA from yeast. Enzymic joining of the chemically synthesized polydeoxynucleotides to form the DNA duplex representing nucleotide sequence 1 to 20. J. Mol. Biol. 72, 427-444 (1972).
22. SGARAMELLA, V. y EHRLICH, S. D. Use of the T4 Polynucleotide Ligase in The Joining of Flush-Ended DNA Segments Generated by Restriction Endonucleases. FEBS J. 86, 531-537 (1978).
23. Seguin-Orlando, A. y otros. Ligation bias in illumina next-generation DNA libraries: implications for sequencing ancient genomes. PLoS One 8, e78575 (2013).
24. Goryshin, I., Baas, B., Vaidyanathan, R. y Maffitt, M. Oligonucleotide replacement for di-tagged and directional libraries. (2016).
25. Bushati, N. y Cohen, S. M. microRNA functions. Annu. Rev. Cell Dev. Biol. 23, 175-205 (2007).
26. Mallory, A. C. y Vaucheret, H. Functions of microRNAs and related small RNAs in plants. Nat. Genet. 38, S31 (2006).

TABLA 1A: Estadísticas de la determinación de haplotipos y asignación de variantes

		stLFR-1			stLFR-2			stLFR-3	stLFR-4
Estadísticas de la biblioteca	Total de bases secuenciadas (Gb)	336	230	100	660	200	100	117	126
	ADN genómico de entrada (ng)	1	1	1	1	1	1	10	10
	Tamaño de fragmento genómico promedio (kb)	66,2	66,3	66,4	52,5	52,7	52,6	30,2	46,8
	Cobertura de genoma único	44X	38X	24X	58X	37X	23X	37X	34X
	Tasa de duplicados	59,4%	49,6%	29,4%	70,88%	41,05%	25,37%	5,4%	15,0%
	Longitud de lectura	PE100	PE100	PE100	PE100	PE100	PE100	PE100	PE100
	Compartimentos únicos	10 186 086	10 007 746	9 427 999	11 823 872	10 932 966	10 297 180	30 544 841	10 577 590
	Promedio de fragmentos por compartimento	1,18	1,18	1,17	1,25	1,23	1,22	2,87	6,84
	Promedio de lecturas con código de barras compartido por fragmento	80,7	71,5	47,4	88,3	60,2	40.7	7,5	8,9
Sin filtro	Precisión de SNP	0,997	0,997	0,995	0,997	0.997	0,995	0,997	0,993
	Sensibilidad de SNP	0,996	0,995	0,988	0,997	0,994	0,986	0,996	0,991
	Precisión de indel	0,934	0,935	0,924	0,938	0,938	0,924	0,960	0,948
	Sensibilidad de indel	0,956	0,951	0,914	0,965	0,950	0,912	0,961	0,925
Filtrado	Precisión de SNP	0,999	0,998	0,997	0,999	0,998	0,996	0,999	0,997
	Sensibilidad de SNP	0,995	0,994	0,985	0,995	0,993	0,985	0,995	0,989
	Precisión de indel	0,971	0,965	0,943	0,974	0,964	0,942	0,978	0,964
	Sensibilidad de indel	0,943	0,940	0,902	0,958	0,940	0,902	0,952	0,917
HapCut2	% de SNP heterocigóticos identificados	99,9%	99,9%	99,8%	99,9%	99,7%	99,7%	98,9%	98,7%
	Tamaño de cóntigo N50 (Mb)	15,1	12,9	8,6	6,4	4,2	2,6	0,6	1,2
	Tasa de error de cambio corto	0,00273	0,00272	0,00272	0,00261	0,00272	0,00271	0,00272	0,00571
	Tasa de error de cambio largo	0,00571	0,00571	0,00570	0,00553	0,00570	0,00570	0,00574	0,00276
LongHap	% de SNP heterocigóticos identificados	0,999	0,9988	0,9966	0,9991	0,9984	0,9952	0,9895	0,9879
	Tamaño de cóntigo N50 (Mb)	18,1	16,6	10,7	8	5,2	3,3	1,1	1,9
	Tasa de error de cambio corto	0,0025748	0,0025949	0,0026139	0,0025228	0,0025307	0,0025773	0,0027524	0,0030534
	Tasa de error de cambio largo	0,0017183	0,0017073	0,0017638	0,0017197	0,0017038	0,0017101	0,0019273	0,0020666

Tabla 1B

10X Genomics[1]	Haplotipado con perlas de Illumina[2]	BGISEQ500 STD[3]
128	99	132
1,25	3	1,000
85,7	-	N/A
33X	19X	43X
6,0 %	21,0 %	3,7 %
PE150	PE76	PE100
1,538,345	147,456	N/A
8,32	~100	N/A
49,8	5	N/A
0,952	0,997	0,998
0,996	0,952	0,998
0,639	0,932	0,960
0,864	0,832	0,972
0,994	-	0,999
0,997	-	0,997
0,916	-	0,991
0,871	-	0,962
99,9 %	98,0 %	N/A
12,8	1,14	N/A
0,00273	0,0013	N/A
0,00572	0,000085	N/A
N/A	N/A	N/A
N/A	N/A	N/A
N/A	N/A	N/A
N/A	N/A	N/A

Tabla 2. Estadísticas de andamiaje

	stLFR-1	stLFR-4	HiC[1]	HiC[2]
Pares de lectura (M)	60	134	734	734
Longitud total del andamio (Gb)	2,84	2,72	2,92	2,92
Andamio N50 (Mb)	44,7	42,8	68,3	60,02
% bases alineadas	98,61 %	98,56 %	98,22 %	94,52 %
Recuento de andamios	597	699	1,411	1,555
Cóntigos en andamios	1,411	1,586	3,096	18,903
Puntos de ruptura	31,386	30,501	35,132	33,079
Reubicaciones	296	327	430	136

(continuación)

Translocaciones	179	189	406	96
Inversiones	624	656	898	408

[1]Los pares de lectura HiC de células madre embrionarias humanas (hESC) (30) se descargaron y usaron para el andamiaje de las lecturas SMRT mediante el uso de SALSA (28) y el mismo proceso que se usó para las bibliotecas stLFR. [2]Los resultados informados por Ghurye y otros, (28) mediante el uso de los mismos pares de lectura HiC para el andamiaje de las lecturas SMRT mediante el uso de SALSA.

Tabla 3

	stLFR-1			stLFR-2			stLFR-3	stLFR-4	BGISEQ-500 STD
Bases totales secuenciadas (Gb)	336	230	100	660	200	100	117	126	132
Asignaciones de PF	10,579	10,498	14,602	11,068	11,012	15,022	8,422	22,404	5,438
Asignaciones de FN	13,023	15,106	40,088	11,218	18,511	46,182	14,205	27,792	7,816
Asignaciones de FP filtradas	4,491	5,443	9,503	4,606	6,326	11,326	4,775	8,564	3,111
Asignaciones de FN filtradas	16,988	19,014	49,330	15,302	22,152	49,443	17,436	34,482	8,984
Cambio en las asignaciones de FP	-6,088	-5,055	-5,099	-6,462	-4,686	-3,696	-3,647	-13 840	-2,327
Cambio en las asignaciones de FN	3,965	3,908	9,242	4,084	3,641	3,261	3,231	6,690	1,168
Asignaciones de FP finales con FP compartidos eliminados	2,825	3,777	7,837	2,940	4,660	9,660	3,109	6,898	3,111

Tabla 4

	stLFR-1			stLFR-2			stLFR-3	stLFR-4
Bases totales secuenciadas (Gb)	336	230	100	660	200	100	117	126
% de SNP heterocigóticos identificados	99,9 %	99,9 %	99,7 %	99,9 %	99,9 %	99,6 %	99,1 %	99,0 %
% de indel heterocigóticos identificados	96,8 %	96,6 %	94,9 %	97,1 %	96,2 %	94,1 %	93,9 %	90,9 %
Tamaño de cóntigo N50 (Mb)	23,4	19,7	13	10,5	7,3	4,1	1,2	2,1
Tasa de error de cambio corto	0,00939	0,00938	0,00988	0,00943	0,00935	0,01002	0,01171	0,01212
Tasa de error de cambio largo	0,00332	0,00337	0,00340	0,00313	0,00337	0,00321	0,00390	0,00426

Tabla 5

	Criterios	stLFR-1 336 Gb	stLFR-1 230 Gb	stLFR-1 100 Gb	stLFR-2 660 Gb	stLFR-2 200 Gb	stLFR-2 100 Gb	stLFR-3	stLFR-4	STD
SNP	GQ mín	23	18	18	41	12	0	13	3	41
SNP	Ref/Alt máx	0,125	0,125	0,15	0,2	0,1	0,07	0,105	0,11	0,2
SNP	Ref/Alt	6,68	6,68	5	6,7	6,68	6,67	6,5	4,8	5,3
SNP	Código de barras es	ref < 1	ref < 1	ref < 1	ref < 2	ref < 2	ref < 2	ref < 1	alt < 1	NA
indel	GQ mín	70	60	45	80	65	40	60	50	95
indel	Ref/Alt máx	0,3	0,27	0,2	0,27	0,28	0,2	0,3	0,22	0,4
indel	Ref/Alt	3,2	3,5	5	3,2	4,2	5	3,5	5	3

Tabla 6

BeadCommonT	SEQ. ID NO:14	/52-Bio/AAAAAAAAAATGTGAGCCAAGGAGTTG
BeadCommonB	SEQ. ID NO:15	CCAGAGCAACTCCTTGGCTCACA
Puente	SEQ. ID NO:16	GCACUGACGACAUGAUCACCAAGGAUCGCCAUAGUCCAUGCUA
Para BGISEQ-500		
Transposón1T	SEQ. ID NO:17	/5Fos/CGATCCTTGGTGATCATGTCGTCAGTGCTTGTCTTCCTAAGATGTGTATAAGAGACAG
Transposón2T	SEQ. ID NO:18	GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG
TransposónB	SEQ. ID NO:19	/5Fos/CTGUCTCUTATACACAUCT
PCR1	SEQ. ID NO:20	TGTGAGCCAAGGAGTTG
PCR2	SEQ. ID NO:21	GCCTCCCTCGCGCCATCAG

Cebadores de secuenciación		
Cebador de secuenciación R1 BGI	SEQ. ID NO:22	GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG
Cebador de secuenciación con código de barras de stLFR BGI	SEQ. ID NO:23	CGAGAACGTCTTGTGAGCCAAGGAGTTGCTCTGG
Cebador de secuenciación R2 BGI	SEQ. ID NO:24	CGTCAGTGCTTGTCTTCCTAAGATGTGTATAAGAGACAG
Cebador 1 MDA BGI	SEQ. ID NO:25	TGATCACCAAGGATCGCCATAGTCCATGCTA
Cebador 2 MDA BGI	SEQ. ID NO:26	CTGTCTCTTATACACATCTTAGGAAGACAAGCACTGACGA

Para ligazón de rama 3'		
Adaptador de ligazón de rama 3' F	SEQ. ID NO:27	/5Fos/CTGATGGCGCGAGGGAGGC
Adaptador de ligazón de rama 3' R	SEQ. ID NO:28	TCGCGCCATCA/3'dd/G

Cebador de secuenciación		
Brecha del cebador de secuenciación R1	SEQ. ID NO:29	CAACTCCTTGGCTCACACGGAGGGAGCGCGGTAGTC

Tabla 7: Eficiencia de ligazón

Eficiencia de ligazón	Sustrato				
	1	2	3	4	5
Adaptador	Mella	saliente	brecha de 1 nt	brecha de 2 nt	brecha de 3 nt
Ad-T	15,2 %	79,5 %	89,6 %	88,9 %	83,9 %
Ad-A	12,0 %	88,6 %)	77,5 %	68,3 %	83,9 %
Ad-GA	7,7 %	58,9 %	80,5 %	56,4 %	59,2 %

carril#	Tipo de sustrato	Sustrato			Producto de ligazón			Eficiencia de ligazón
		Tamaño (nt)	Intensidad (pixel)	Intensidad normalizada (pixel/nt)	Tamaño (nt)	Intensidad (pixel)	Intensidad normalizada (pixel/nt)	
2	Mella	27	19 044 ,75	705,36	49	5062,00	103,31	12,77%
5	Brecha de 1 nt	27	13 120 ,29	485,94	49	22 807 ,69	465,46	48,92%
8	Brecha de 8 nt	25	14 042 ,49	561,70	47	19 060 ,60	405,54	41,93%
11	Extremo recesivo 3´	27	1376,23	50,97	49	17 684 ,29	360,90	87,62%
14	Control de extremo romo	40	5311,44	132,79	62	21 973 ,00	354,40	72,74%

Tabla 8

Tipo de sustrato	Tipo de donante	Sustrato			Producto de ligazón			Eficiencia de ligazón
		Tamaño (nt)	Intensidad (píxel)	Intensidad normalizada (píxel/nt)	Tamaño (nt)	Intensidad (píxel)	Intensidad normalizada (píxel/nt)	
Mella	Ad-T	124	11801,78	95,18	156	1744,01	11,18	10,51 %
	Ad-A	124	13130,49	105,89	182	1091,70	6,00	5,36 %
	Ad-GA	124	12810,37	103,31	184	603,87	3,28	3,08 %
brecha de 1 nt	Ad-T	123	2561,08	20,82	155	23719,00	153,03	88,02 %
	Ad-A	123	2058,55	16,74	181	7034,96	38,87	69,90 %
	Ad-GA	123	1709,67	13,90	183	8340,98	45,58	76,63 %
brecha de 2 nt	Ad-T	122	1164,89	9,55	154	6909,36	44,87	82,45 %
	Ad-A	122	3882,74	31,83	180	7688,03	42,71	57,30 %
	Ad-GA	122	6573,57	53,88	182	8495,74	46,68	46,42 %
brecha de 3 nt	Ad-T	121	2344,08	19,37	153	11764,83	76,89	79,88 %
	Ad-A	121	1974,72	16,32	179	9738,98	54,41	76,93 %
	Ad-GA	121	8896,47	73,52	181	10145,81	56,05	43,26 %
extremo recesivo 3'	Ad-T	108	1934,79	17,91	140	8791,10	62,79	77,80 %
	Ad-A	108	1070,23	9,91	166	7834,38	47,20	82,65 %
	Ad-GA	108	5675,05	52,55	168	7206,26	42,89	44,94 %

REIVINDICACIONES

1. Un método para preparar una biblioteca de secuenciación para secuenciar un ácido nucleico diana sin el uso de nanogotas que comprende:

(a) transponer una secuencia de inserción en los primeros fragmentos del ácido nucleico diana, en donde la secuencia de inserción comprende una secuencia de hibridación, y en donde la transposición produce mellas en los primeros fragmentos e hidroxilo 3' en dichas mellas;
(b) combinar en una sola mezcla (i) los primeros fragmentos del ácido nucleico diana de (a), (ii) un oligonucleótido férula, y (iii) una población de perlas, en donde cada perla comprende oligonucleótidos de captura inmovilizados sobre ella, dichos oligonucleótidos de captura comprenden

1) una secuencia que contiene un código de barras, en donde los oligonucleótidos inmovilizados en la misma perla individual comprenden la misma secuencia que contiene un código de barras y la mayoría de las perlas tienen diferentes secuencias que contienen un código de barras,
2) una secuencia complementaria común a al menos una porción del oligonucleótido férula, donde una segunda porción del oligonucleótido férula es complementaria a al menos una porción de la secuencia de hibridación;
3) un primer sitio de hibridación del cebador de PCR;

(c) ligar oligonucleótidos de captura de perlas individuales a secuencias de hibridación insertadas de los primeros fragmentos individuales para producir productos ligados;
(d) combinar un adaptador de ligazón de rama 3' (3'BL) con los productos ligados de la etapa (c), en donde el adaptador 3'BL comprende un primer oligonucleótido y un segundo oligonucleótido, hibridados para formar un extremo bicatenario romo y un extremo de ADN de simple hebra, en donde el primer oligonucleótido comprende un fosfato 5' en el extremo bicatenario romo;
ligar el primer oligonucleótido a los primeros fragmentos en las mellas de los primeros fragmentos, en donde la ligazón es una ligazón de rama 3' que une covalentemente el fosfato 5' del primer oligonucleótido al hidroxilo 3' en las mellas de los primeros fragmentos,
y en donde el primer oligonucleótido comprende un segundo sitio de hibridación del cebador de PCR.

2. El método de la reivindicación 1, en donde el primer sitio de hibridación del cebador de PCR y el segundo sitio de hibridación del cebador de PCR tienen secuencias diferentes.

3. El método de las reivindicaciones 1 o 2, en donde el primer oligonucleótido adaptador comprende una secuencia de código de barras de muestra.

4. El método de las reivindicaciones 1-3 que comprende la producción de ampliaciones que comprende hibridar los primeros cebadores de PCR a los sitios de hibridación de los primeros cebadores de PCR, hibridar los segundos cebadores de PCR a los sitios de hibridación de los segundos cebadores de PCR y llevar a cabo la amplificación por PCR.

5. El método de las reivindicaciones 1 a 3, en donde en la etapa de transponer la secuencia de inserción en los primeros fragmentos, una enzima transposasa permanece unida a los primeros fragmentos y el método comprende además retirar la transposasa de los primeros fragmentos, de esta manera se producen subfragmentos.

6. El método de las reivindicaciones 1-5, en donde se usa Tn5 para transponer una secuencia de inserción en los primeros fragmentos del ácido nucleico diana.

7. El método de cualquiera de las reivindicaciones 1 a 6, en donde después de la etapa (c) los oligonucleótidos de captura se eliminan enzimáticamente mediante el uso de la exonucleasa, opcionalmente la exonucleasa I o III o ambas.

8. El método de cualquiera de las reivindicaciones 1 a 6, en donde los oligonucleótidos de captura y/o los oligonucleótidos férula comprenden uracilo, de manera que el tratamiento con Uracil-ADN glucosilasa (UDG) degrada los oligonucleótidos que contienen uracilo.

9. El método de la reivindicación 6 que comprende amplificar los subfragmentos para producir amplicones.

10. El método de la reivindicación 4 o la reivindicación 9 que comprende la secuenciación de los amplicones para producir lecturas de secuencias en donde las lecturas de secuencias con la misma secuencia de código de barras son del mismo primer fragmento.

11. El método de cualquiera de las reivindicaciones 1 a 10, en donde el ácido nucleico diana es ADN genómico, opcionalmente ADN genómico humano.

12. El método de una cualquiera de las reivindicaciones 1 a 11, en donde las perlas comprenden al menos 100 000 copias del oligonucleótido de captura.

13. El método de una cualquiera de las reivindicaciones 1 a 12, que comprende además:

(e) asignar la mayoría de las lecturas de secuencia a los primeros fragmentos correspondientes; y
(f) ensamblar las lecturas de secuencia para producir una secuencia ensamblada de la diana.

14. El método de una cualquiera de las reivindicaciones 1 a 13, en donde los primeros fragmentos son de una sola célula.

15. El método de una cualquiera de las reivindicaciones 1 a 14, en donde la mezcla sola contiene de 5-100 equivalentes de genoma de ADN humano.

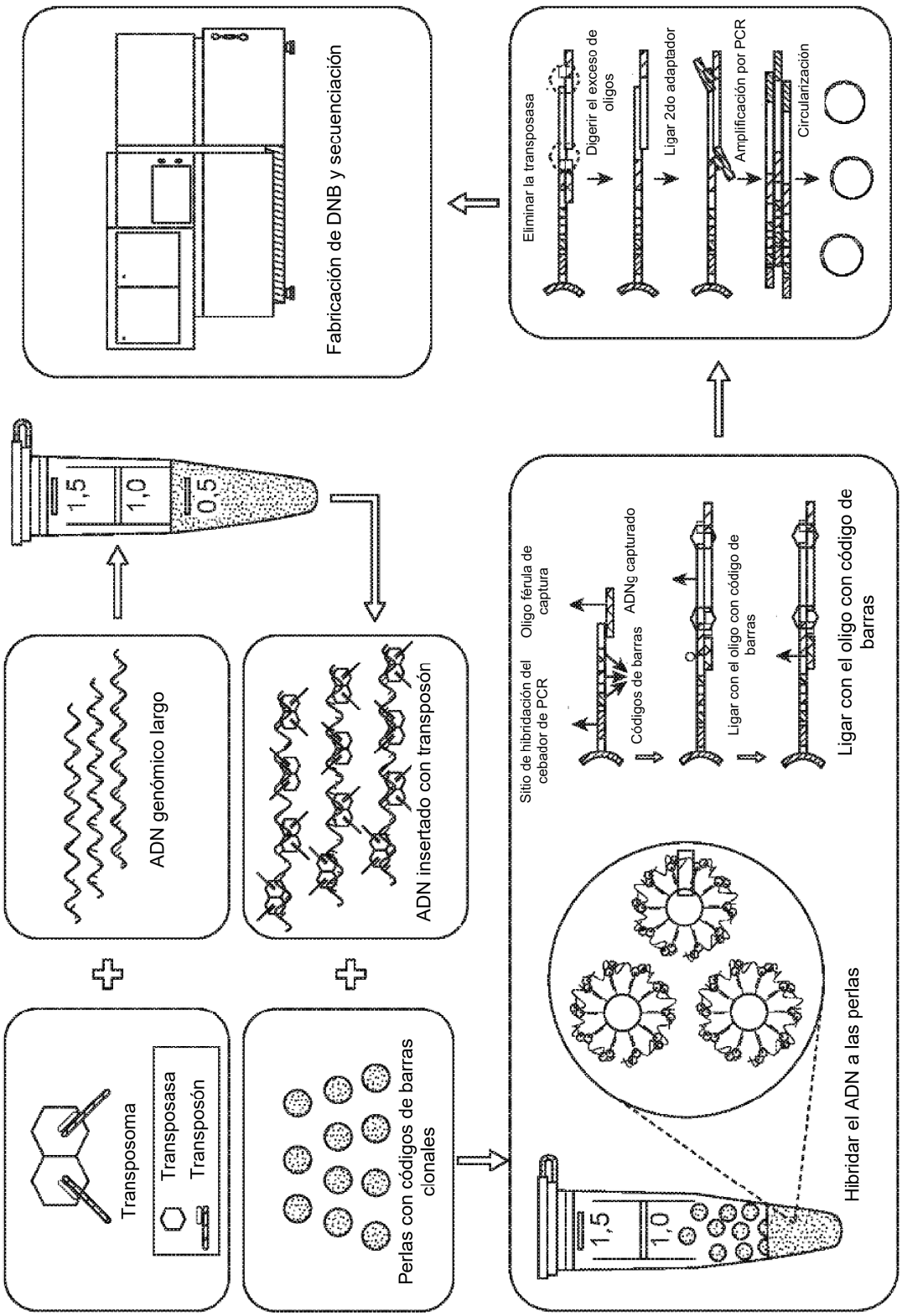

Transposoma
Transposasa
Transposón
Perlas con códigos de barras clonales
ADN genómico largo
ADN insertado con transposón
1,5
1,0
0,5
Hibridar el ADN a las perlas
Sitio de hibridación del cebador de PCR
Oligo férula de captura
Códigos de barras
ADNg capturado
Ligar con el oligo con código de barras
Ligar con el oligo con código de barras
Eliminar la transposasa
Digerir el exceso de oligos
Ligar 2do adaptador
Amplificación por PCR
Circularización
Fabricación de DNB y secuenciación

Figura 1A

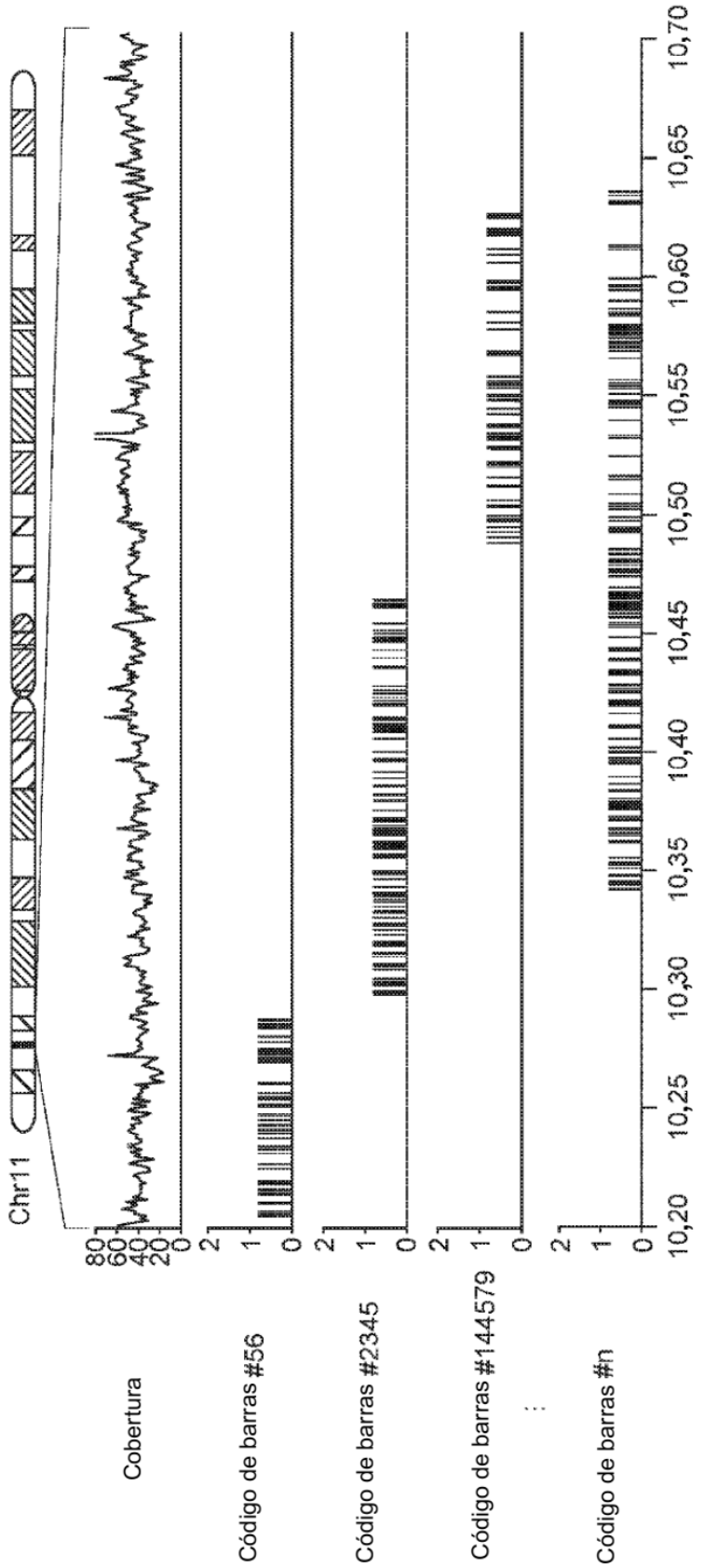

Chr11
Cobertura
80
60
40
20
0
Código de barras #56
Código de barras #2345
Código de barras #144579
Código de barras #n
2
1
0
10,20
10,25
10,30
10,35
10,40
10,45
10,50
10,55
10,60
10,65
10,70
Posición (Mb)

Figura 1B

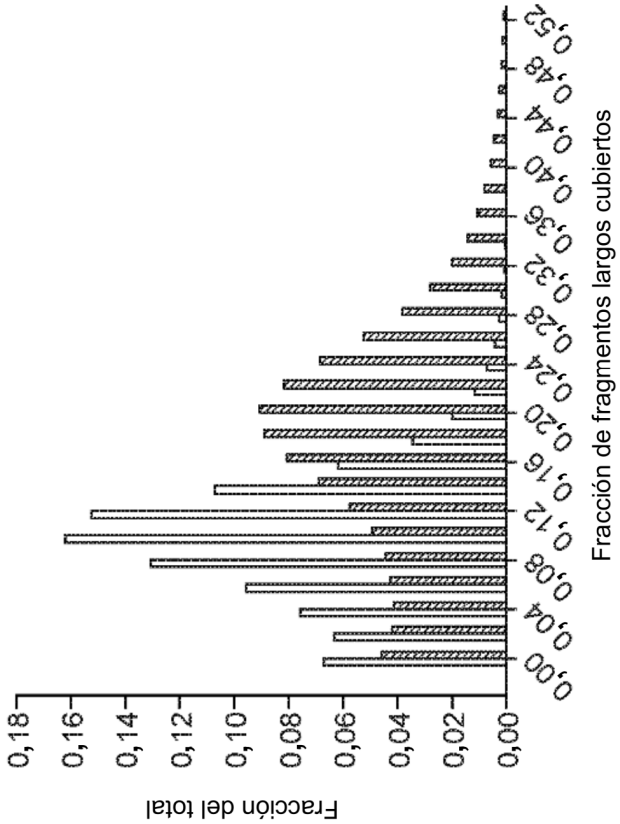


Figura 1D

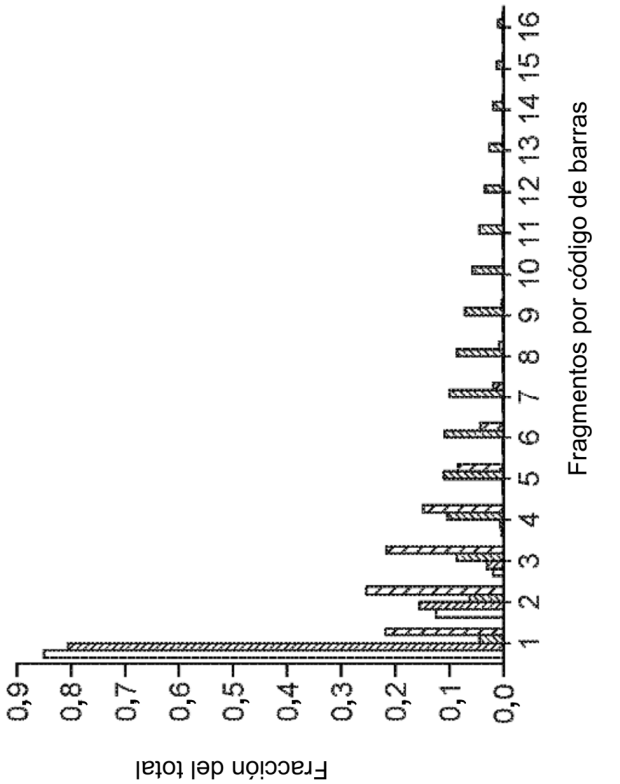


Figura 1C

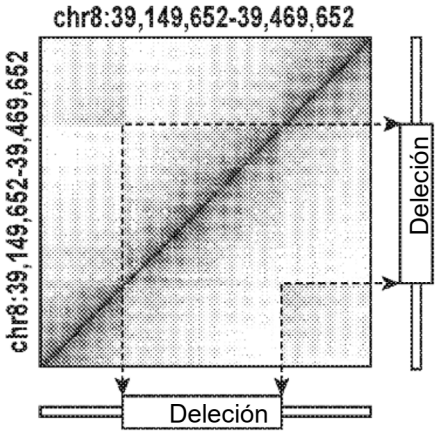


Figura 2A

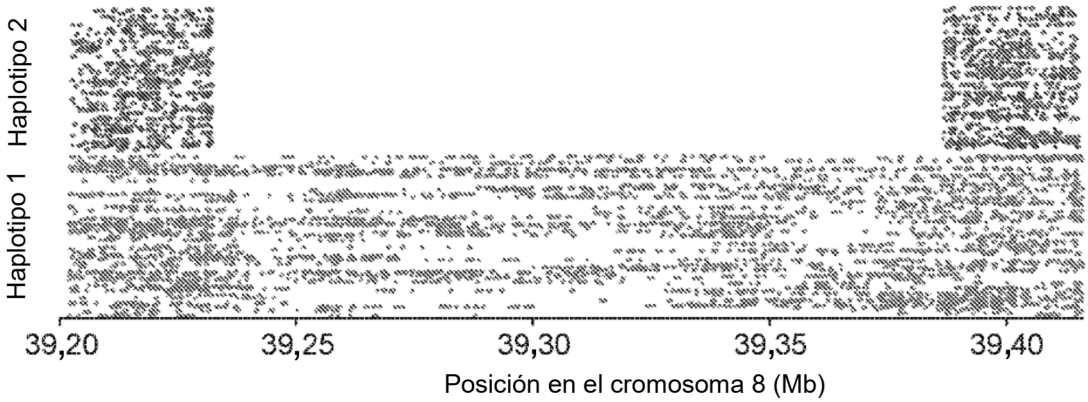


Figura 2B

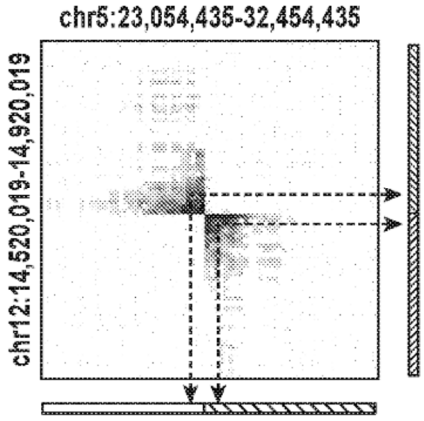


Figura 2C

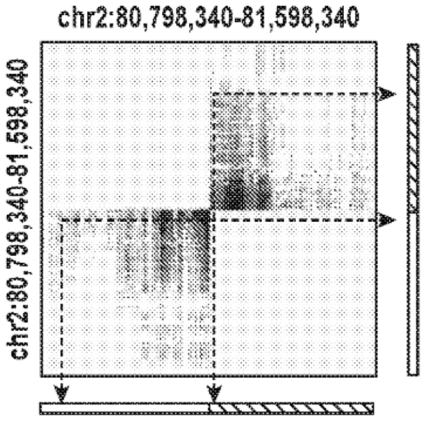


Figura 2D

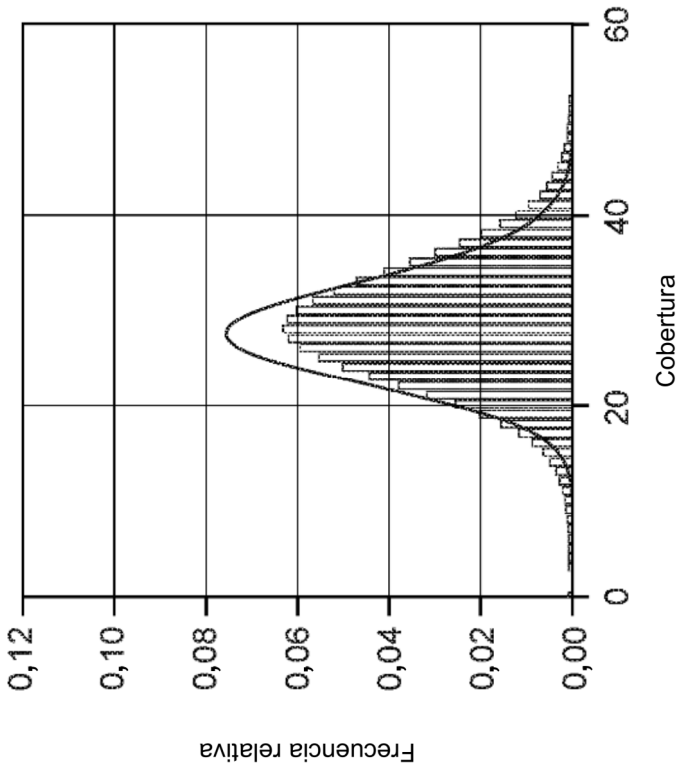


Figura 3B

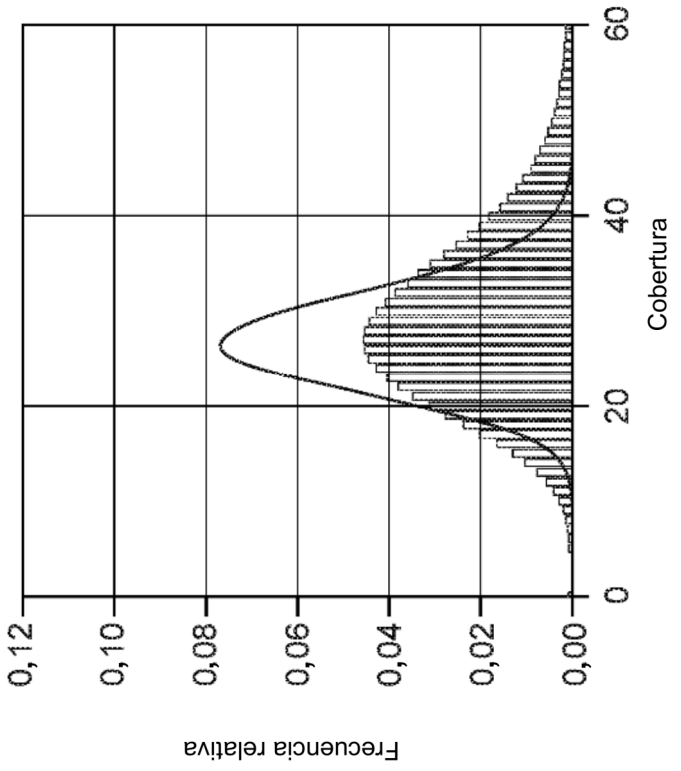


Figura 3A

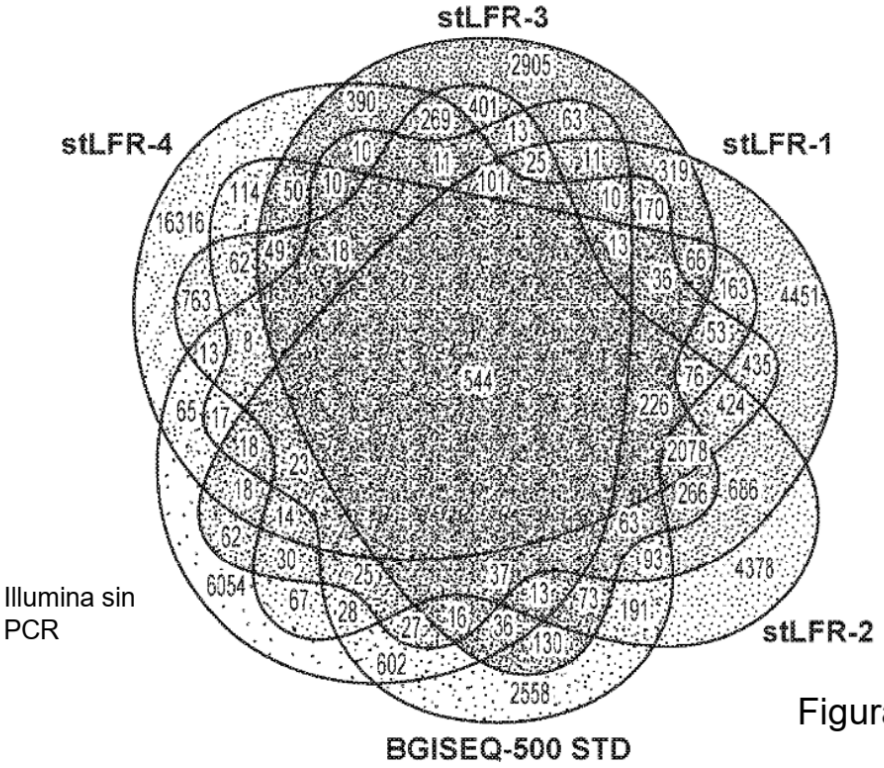


Figura 4A

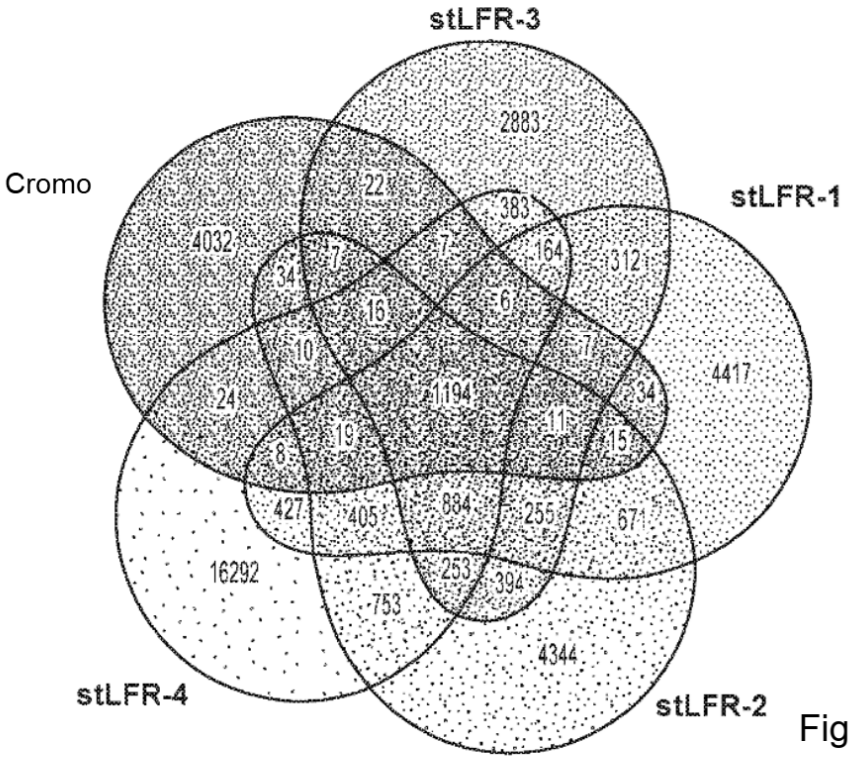


Figura 4B

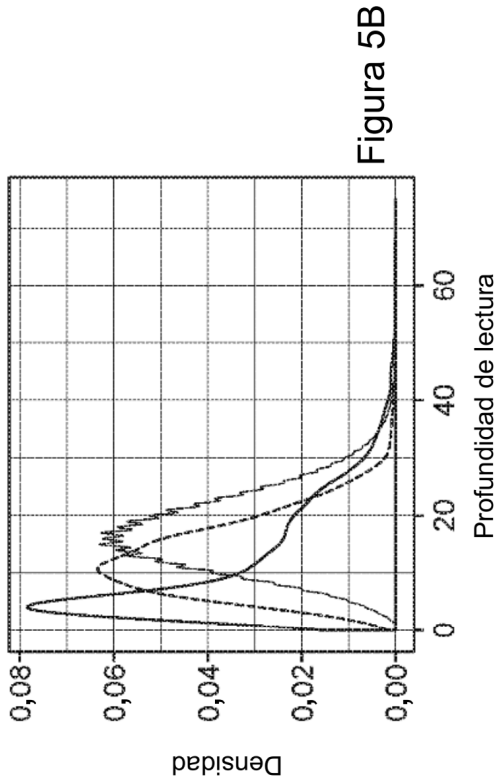


Figura 5B

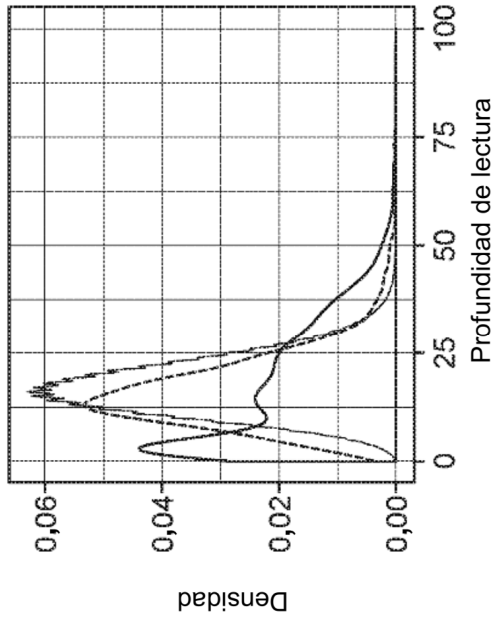


Figura 5A

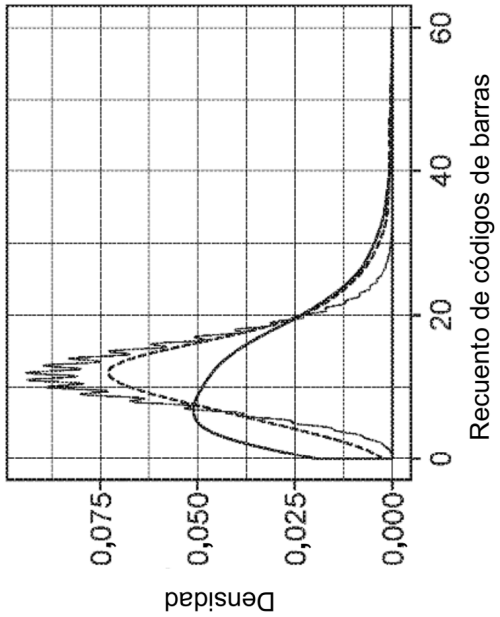


Figura 5D

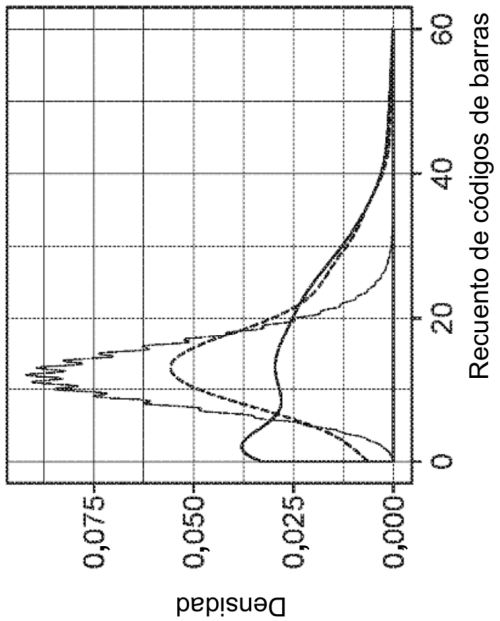


Figura 5C

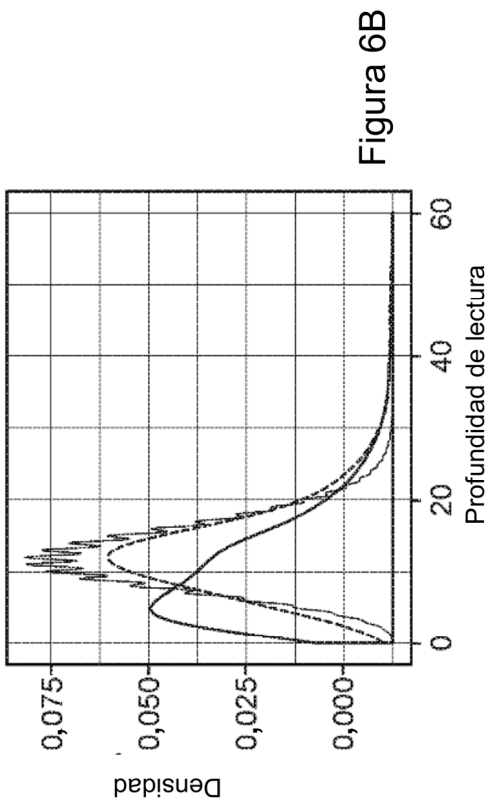


Figura 6B

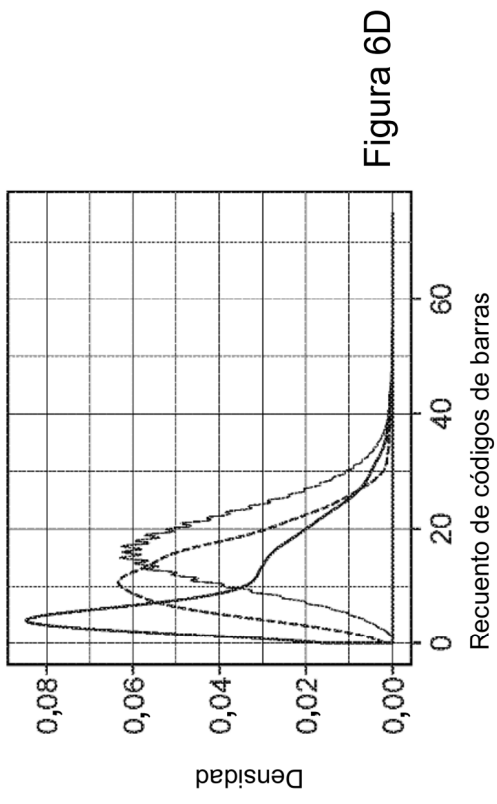


Figura 6D

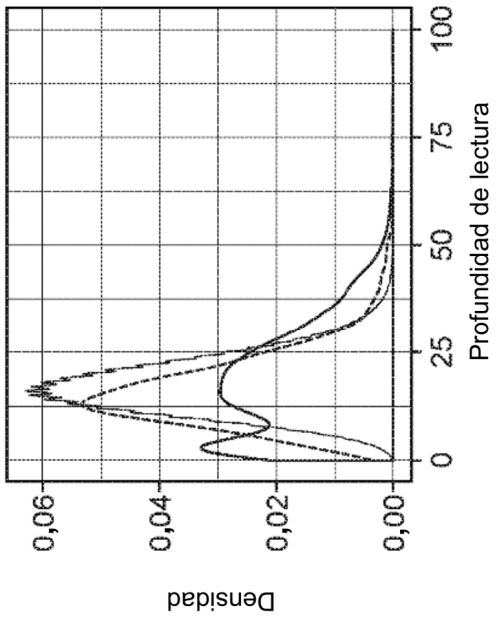


Figura 6A

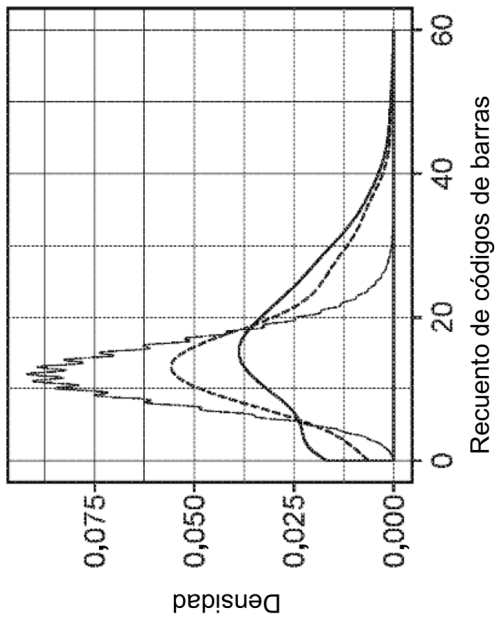


Figura 6C

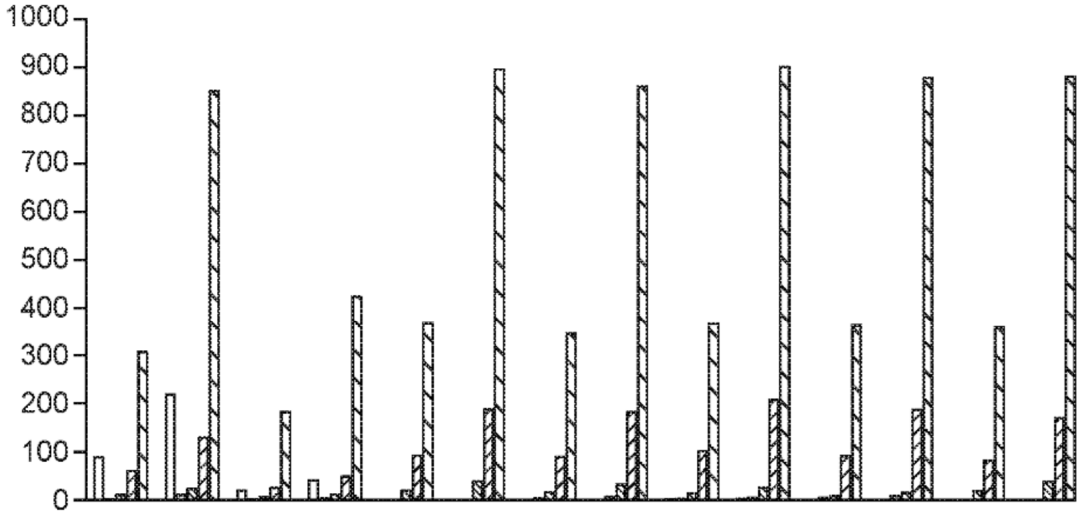


Figura 7A

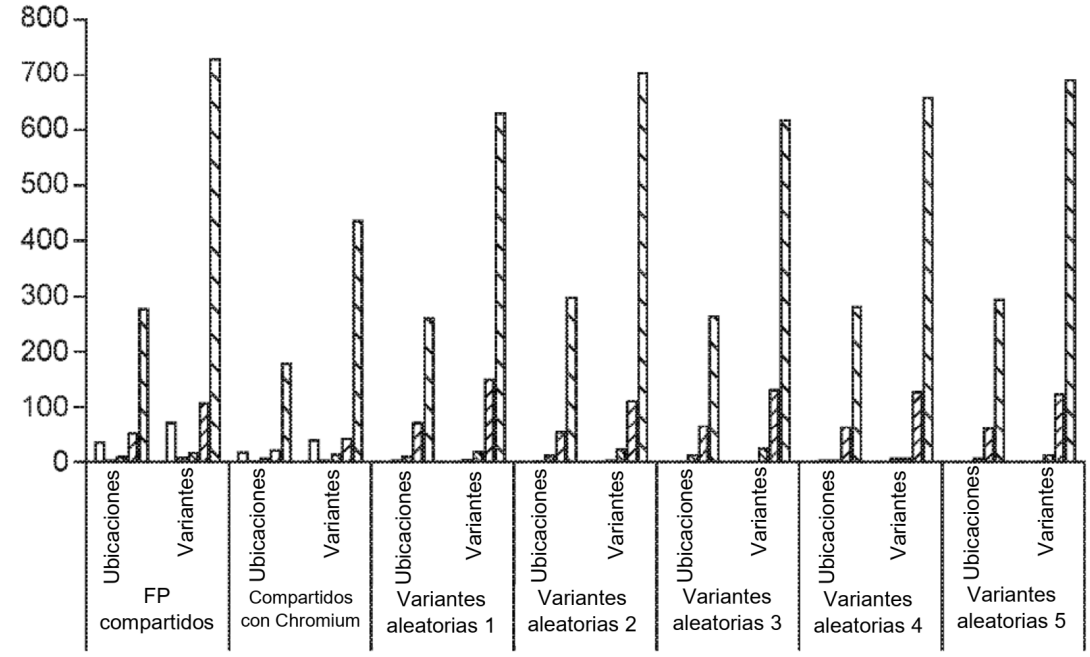


Figura 7B

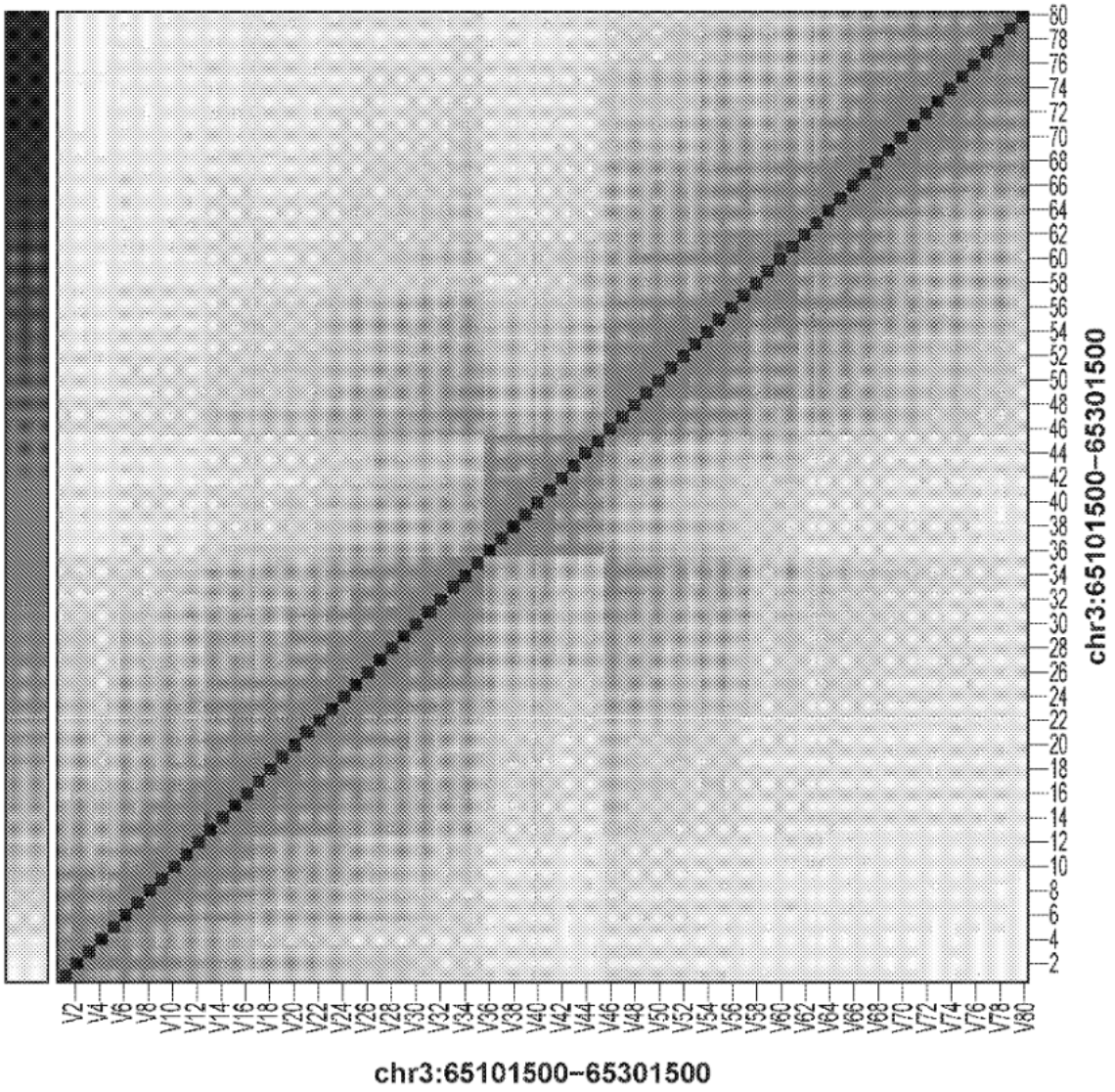


Figura 8A

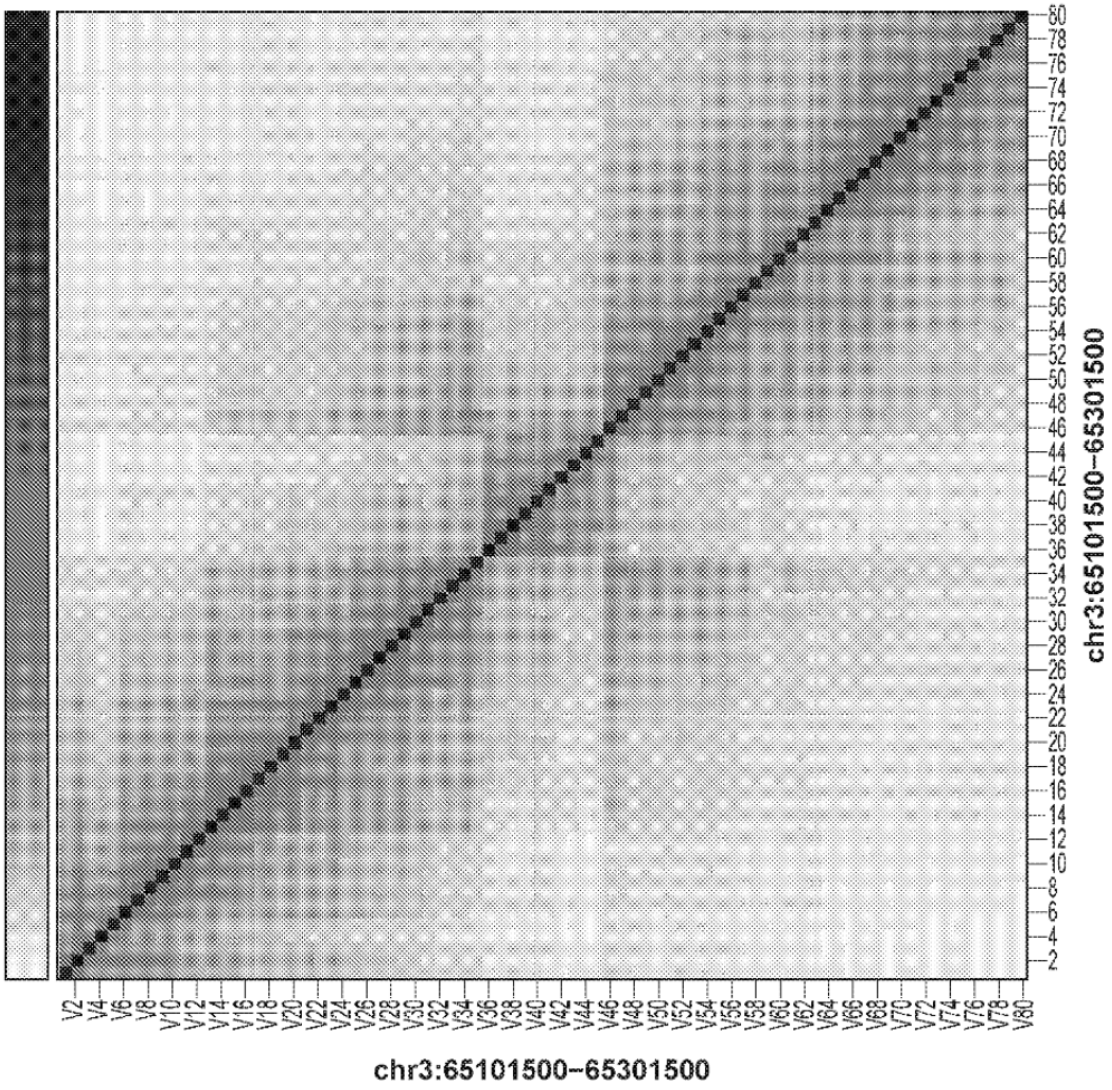


Figura 8B

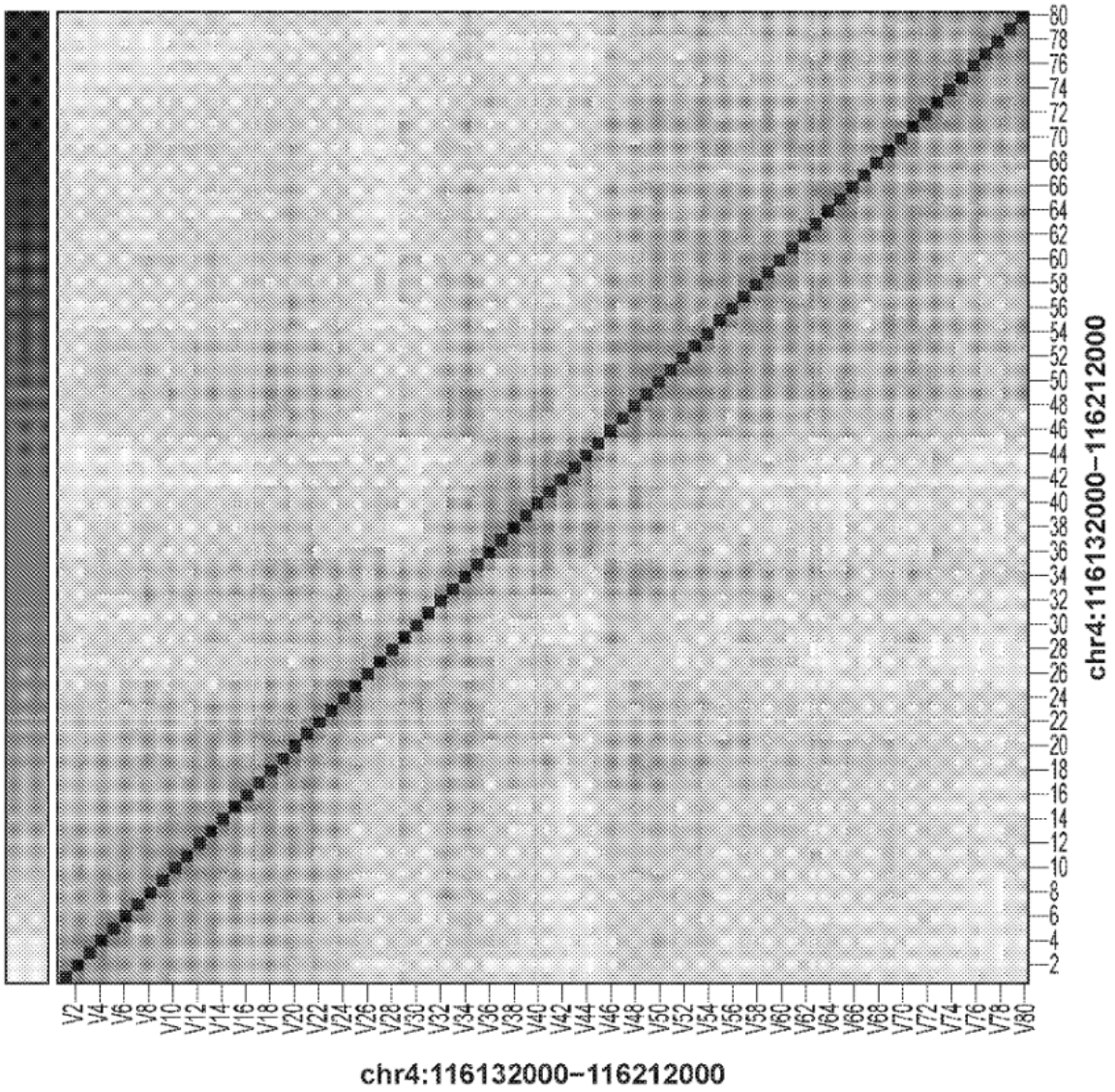


Figura 8C

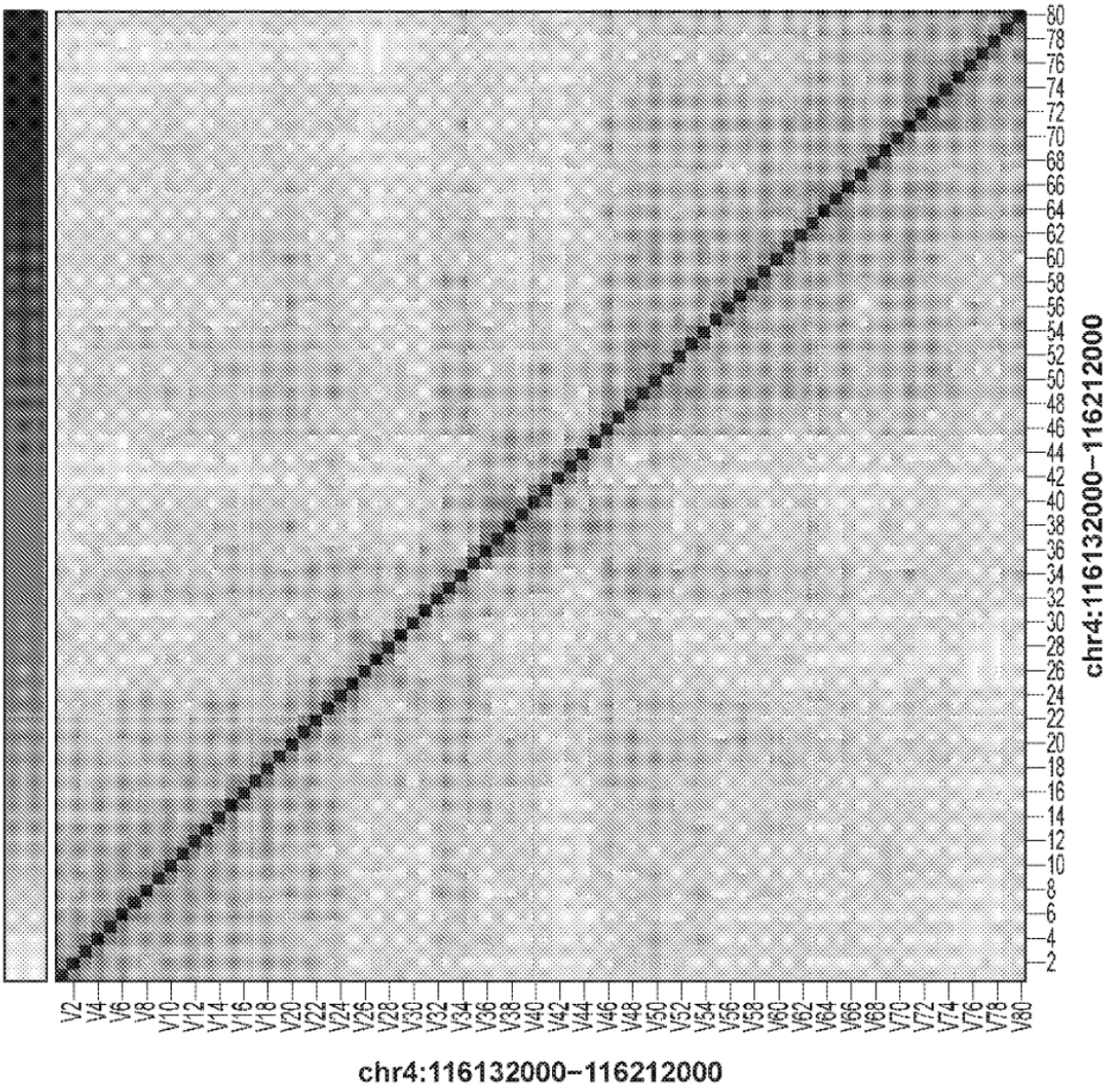


Figura 8D

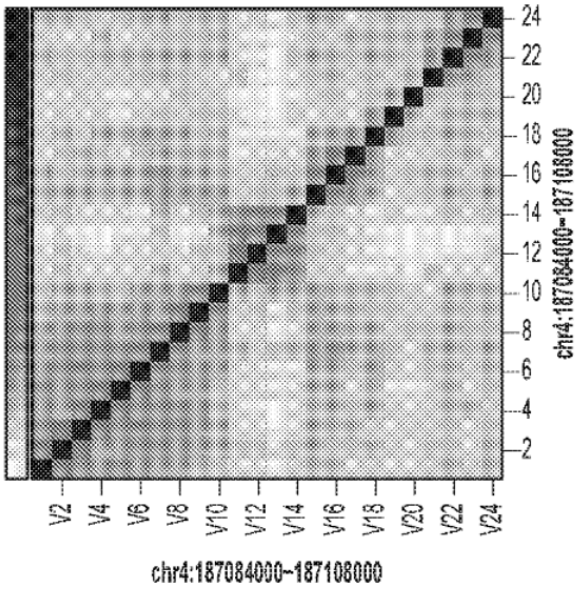


Figura 8E

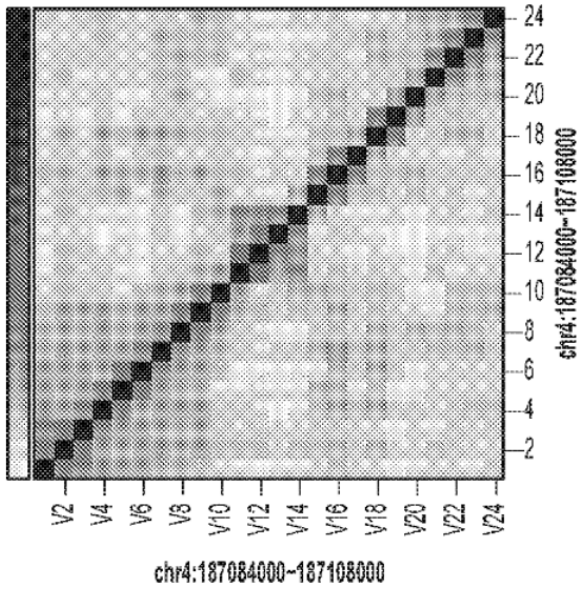


Figura 8F

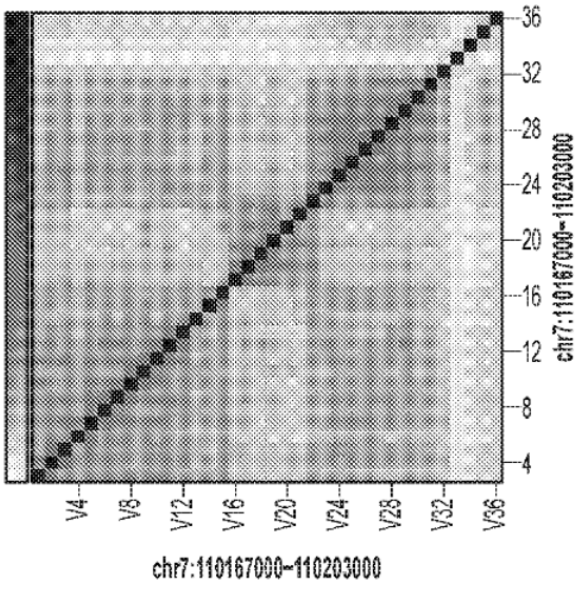


Figura 8G

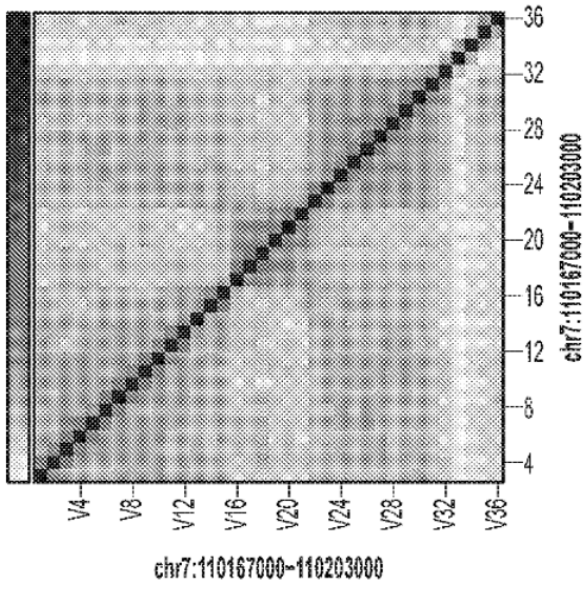


Figura 8H

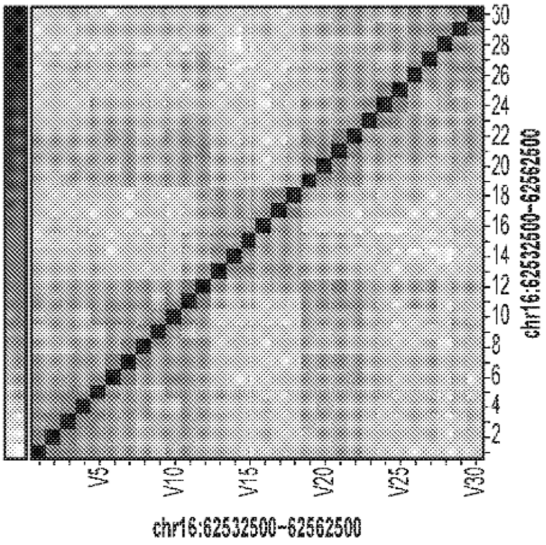


Figura 8I

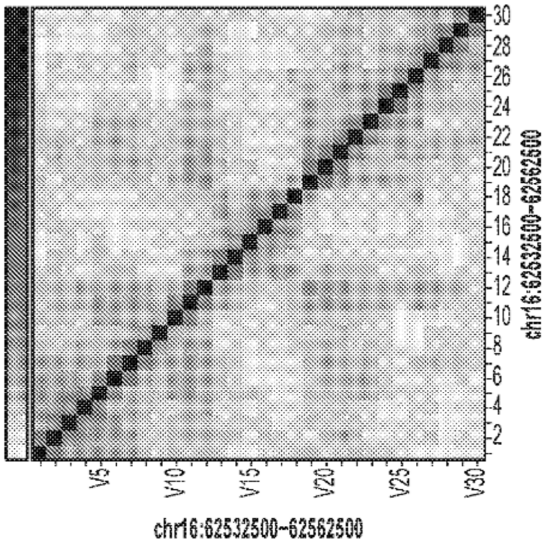


Figura 8J

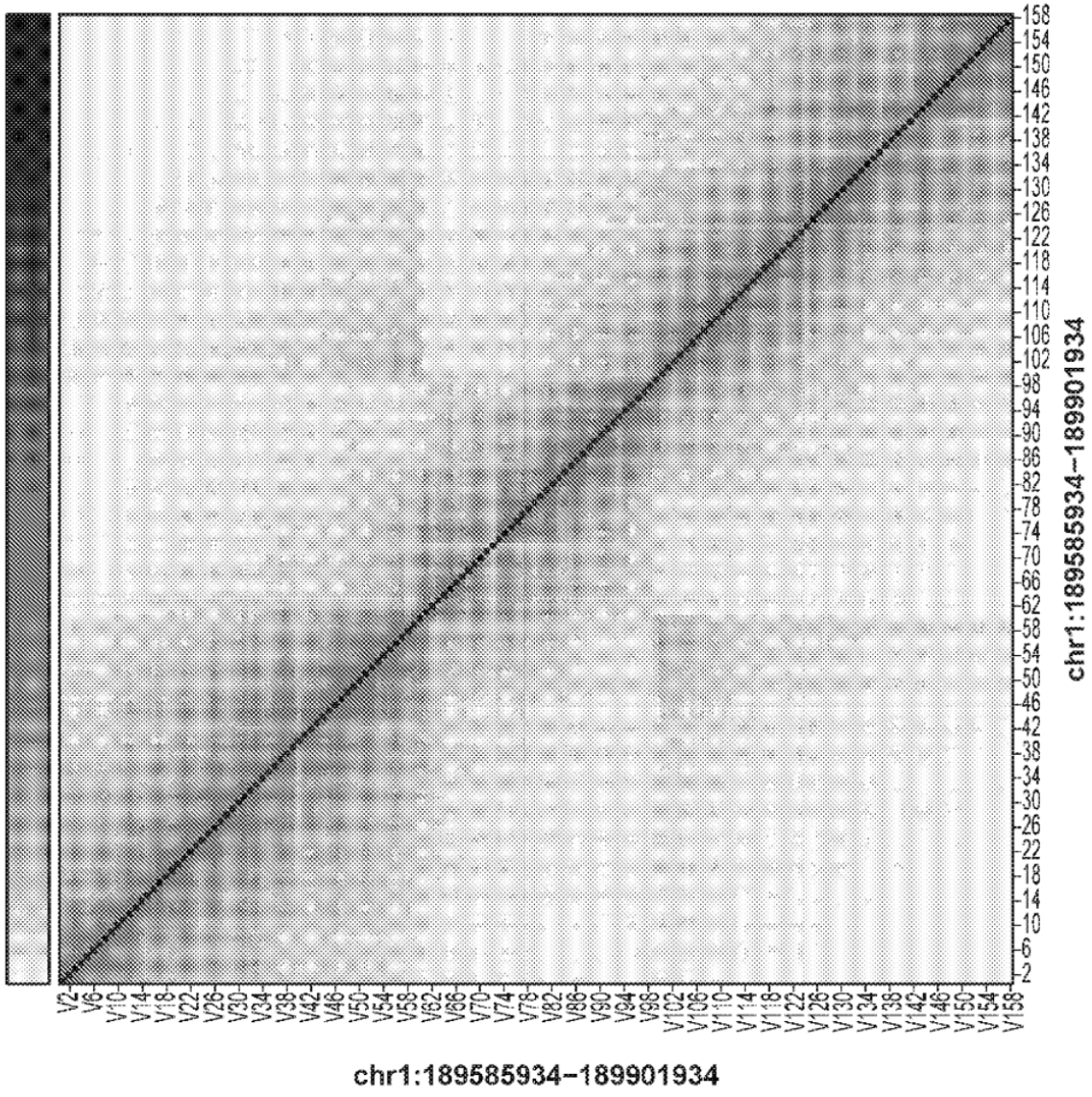


Figura 8K

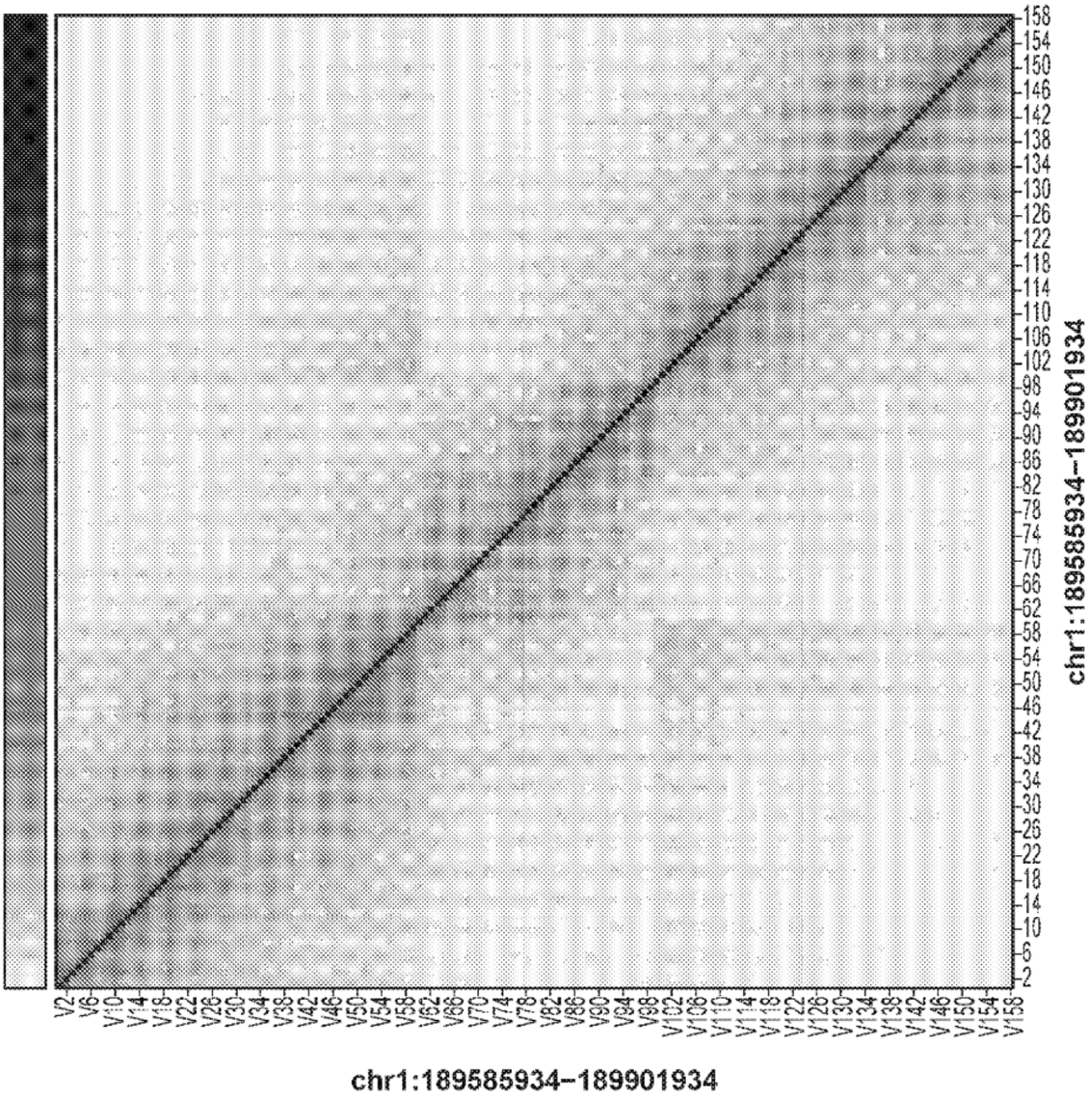


Figura 8L

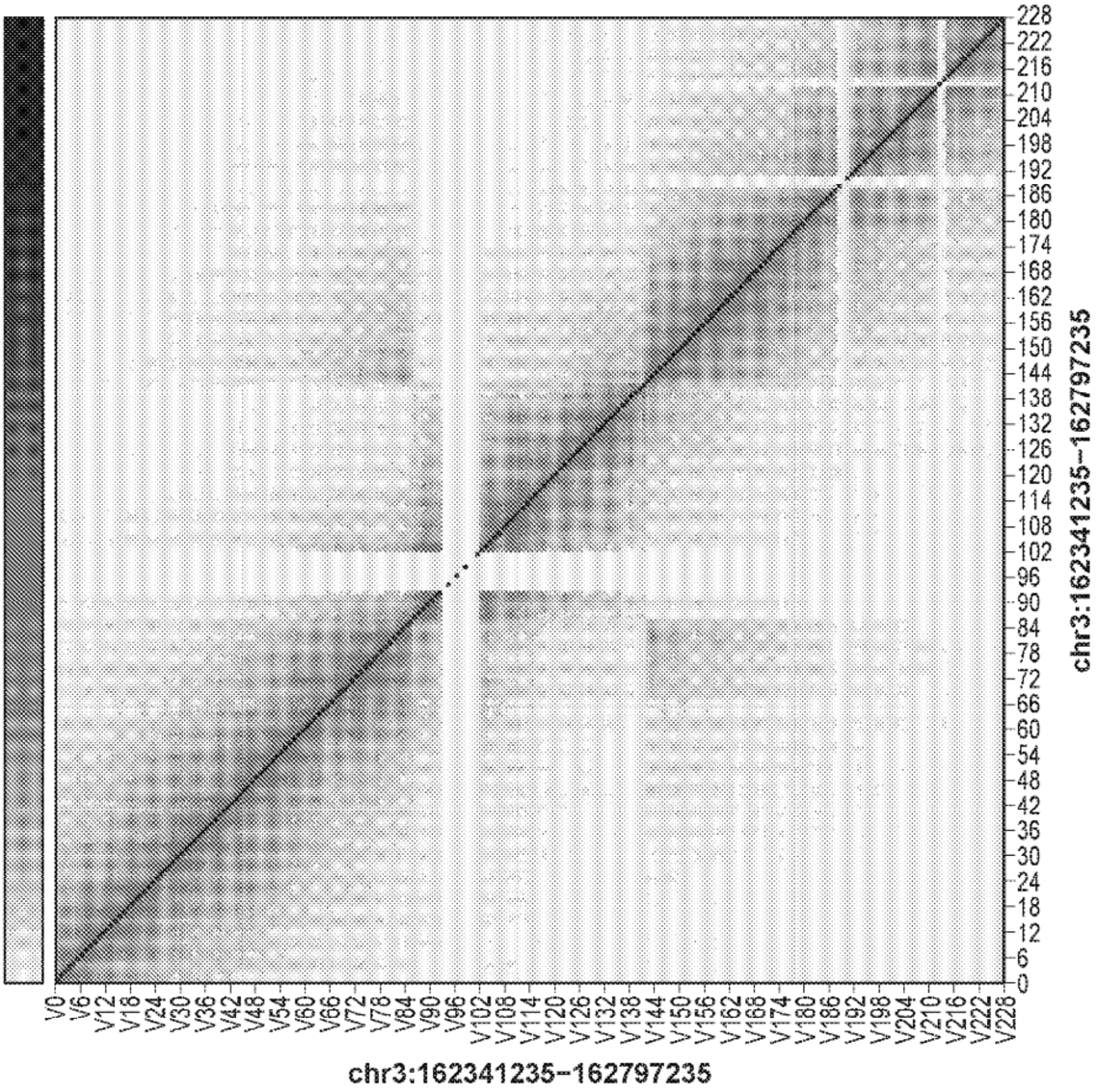


Figura 8M

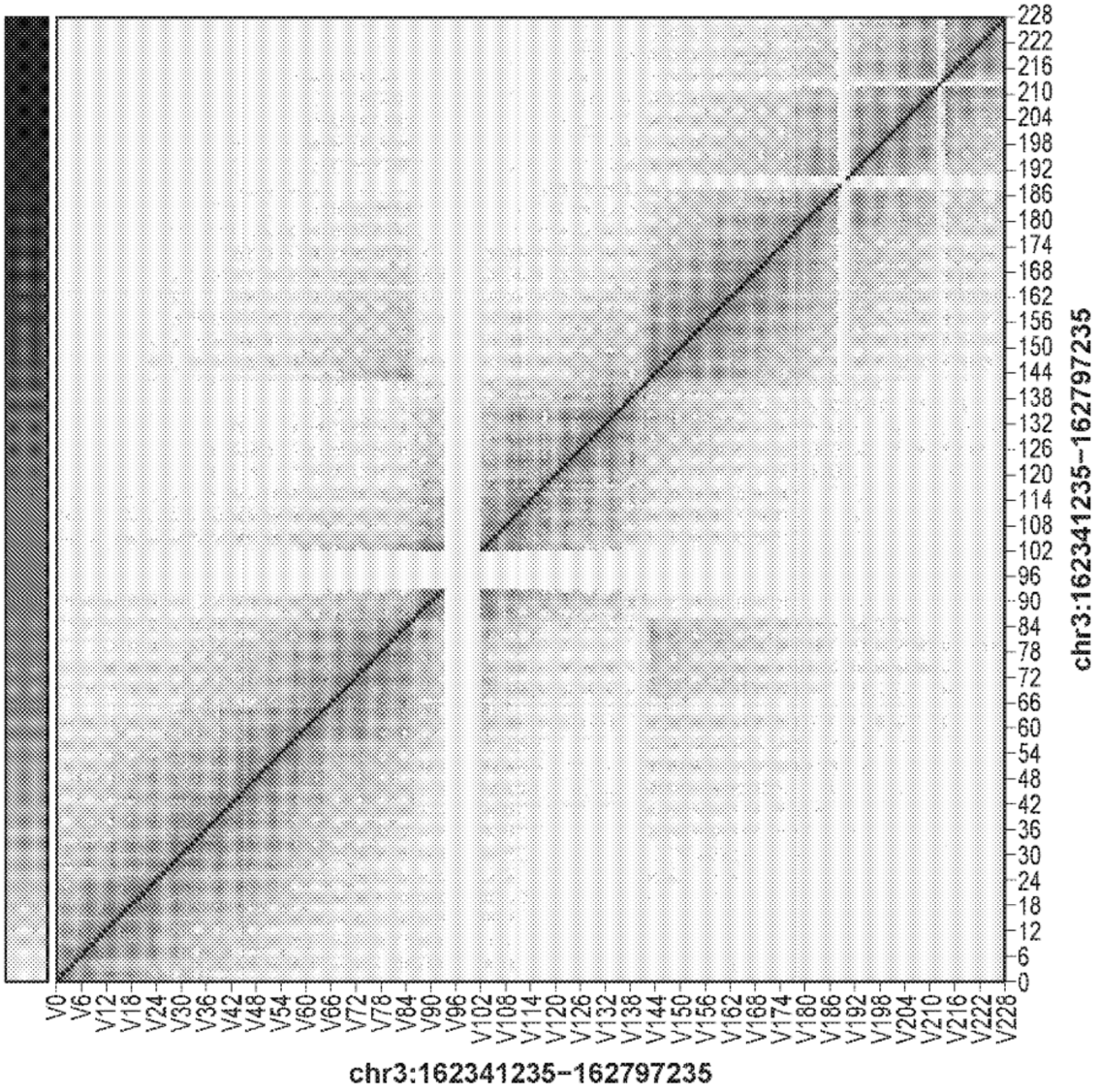


Figura 8N

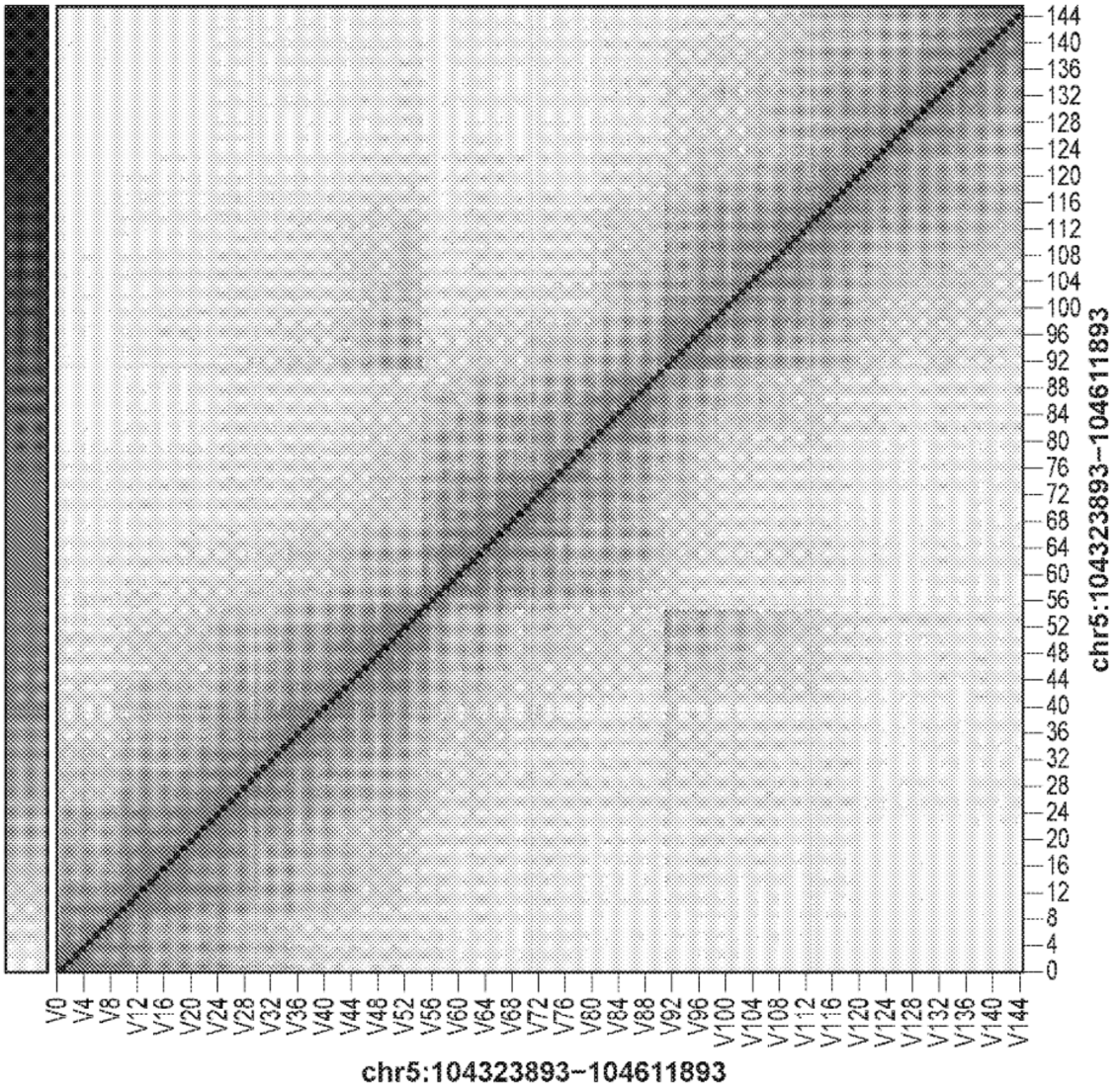


Figura 8O

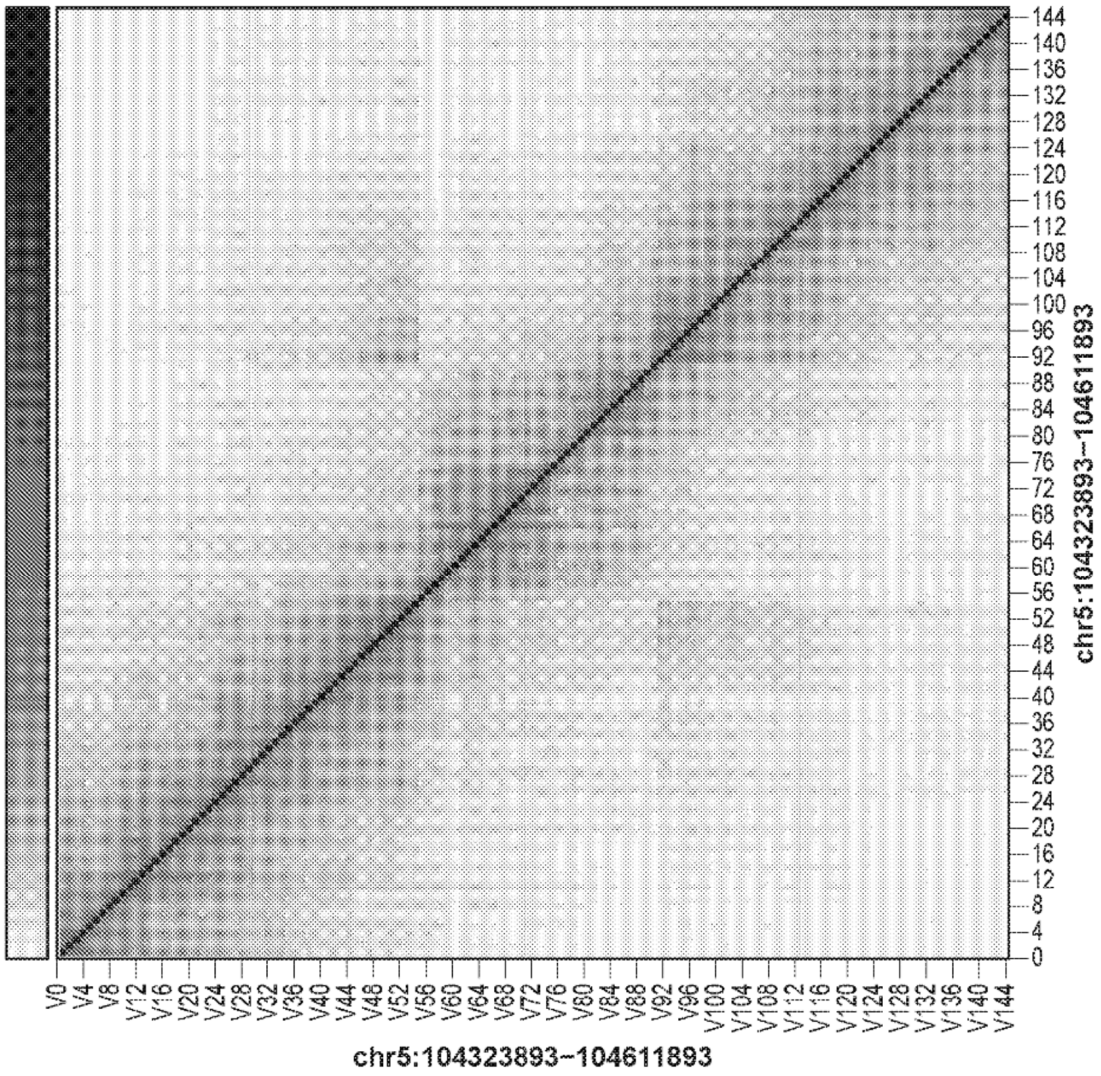


Figura 8P

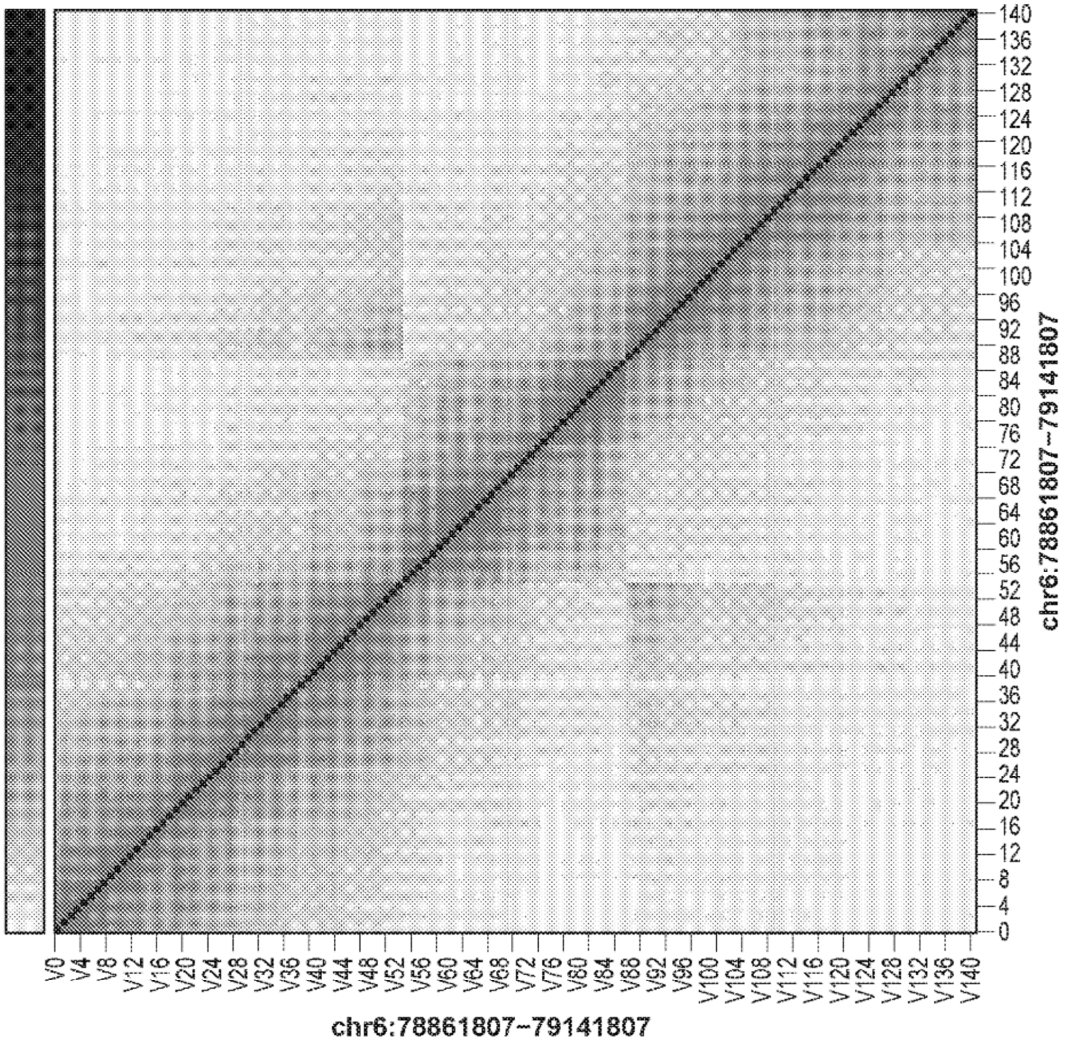


Figura 8Q

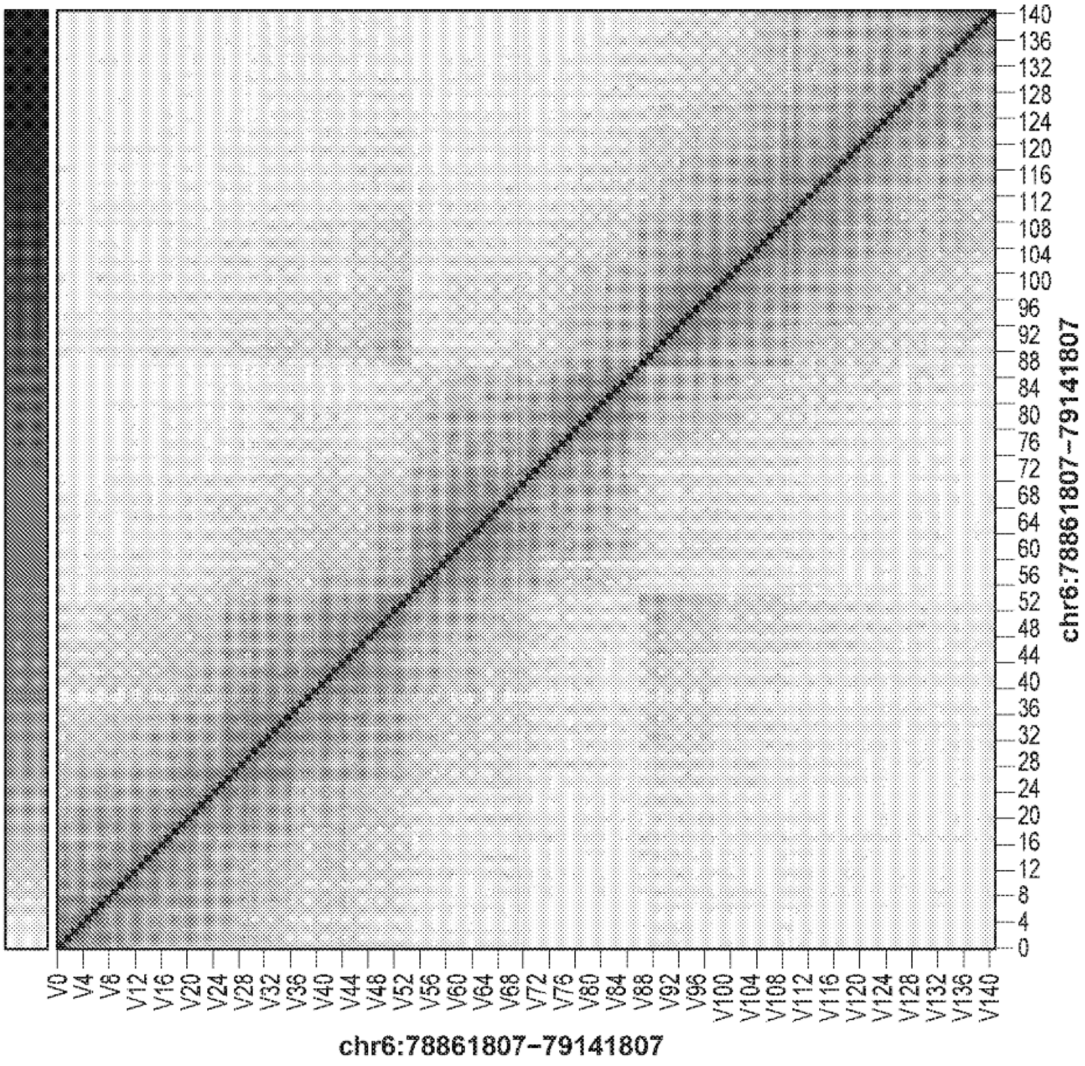


Figura 8R

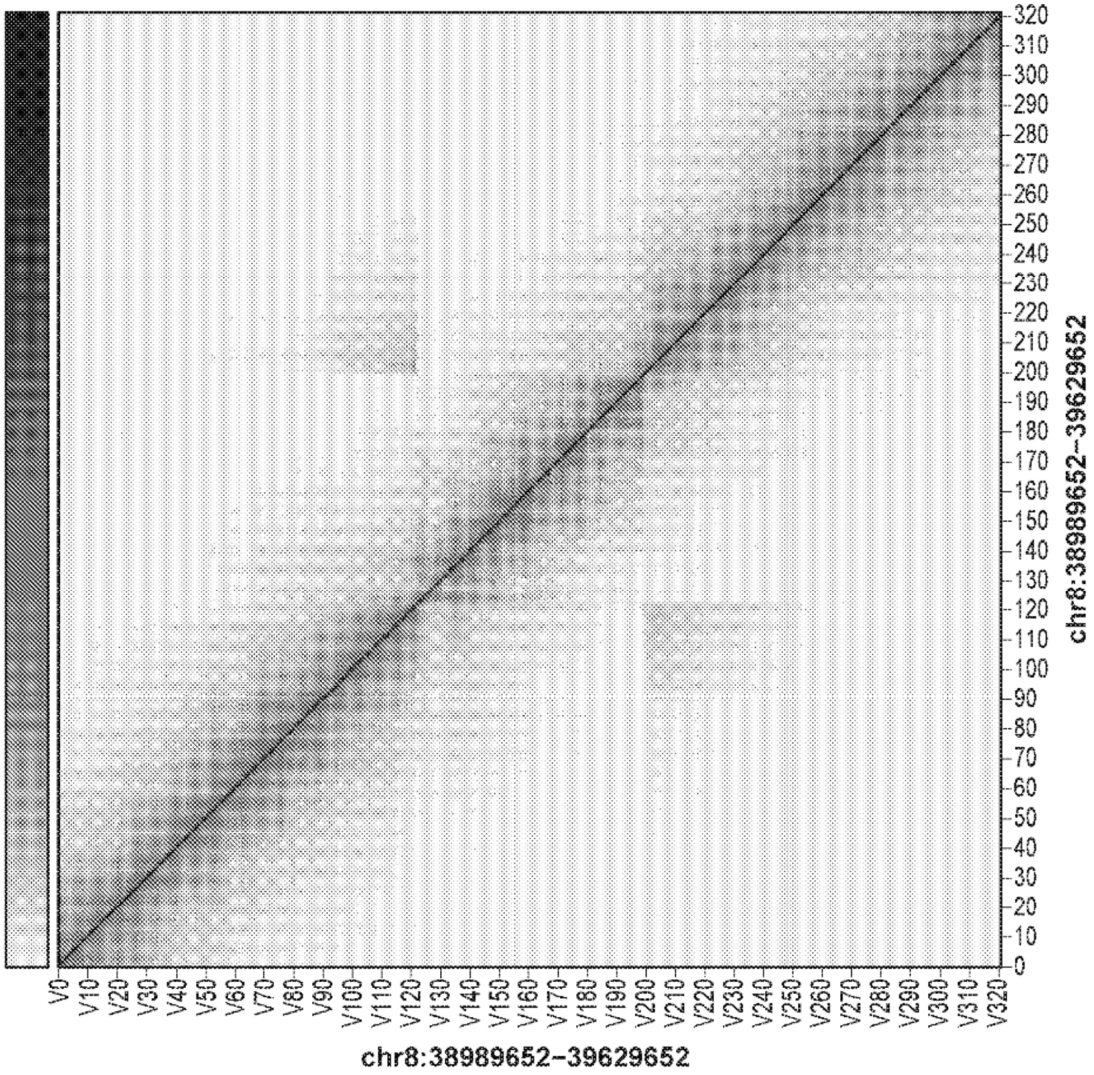


Figura 8S

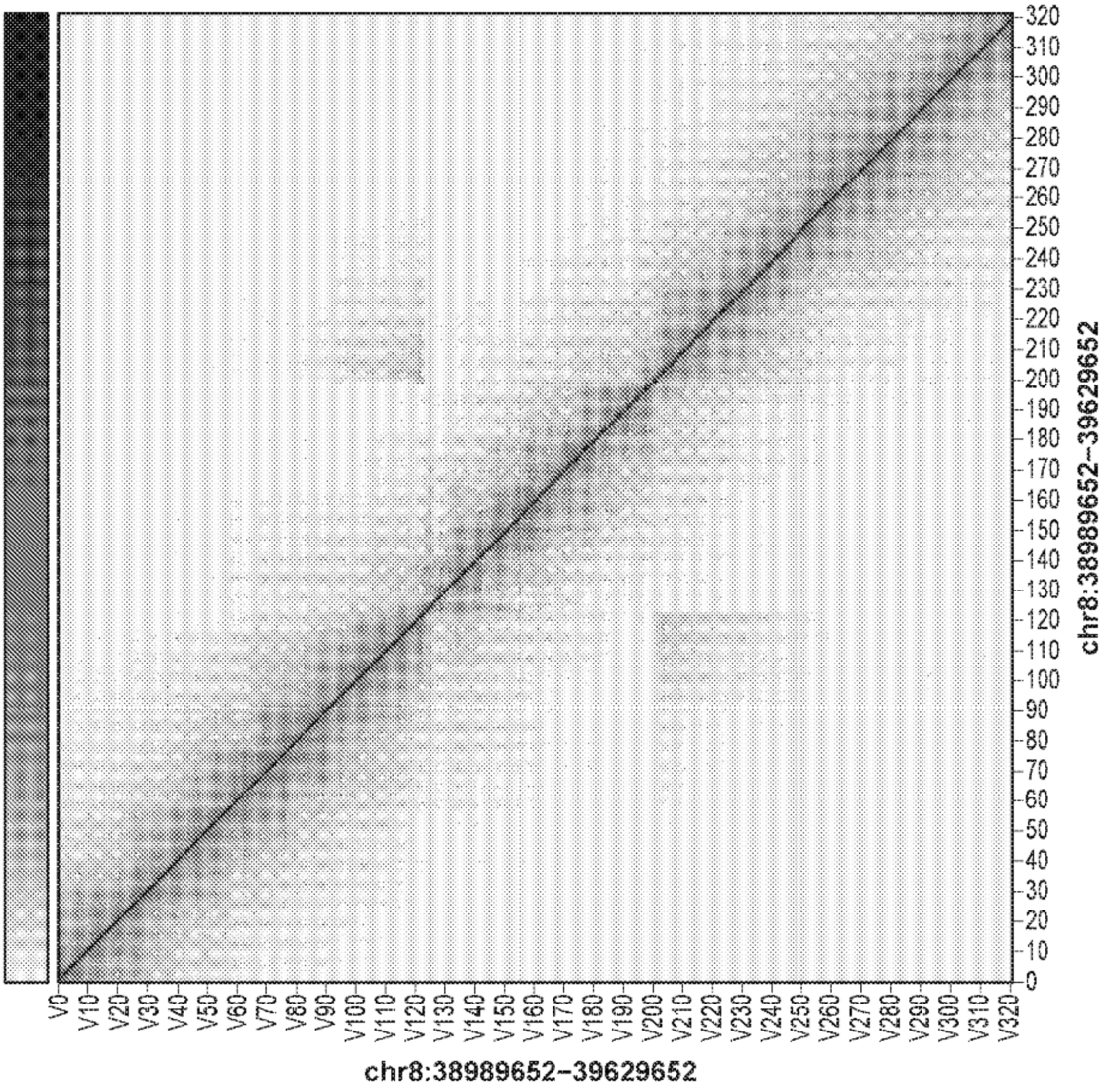


Figura 8T

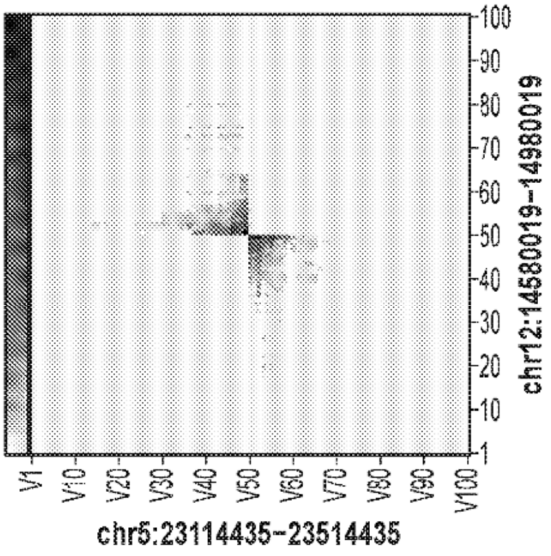


Figura 9A

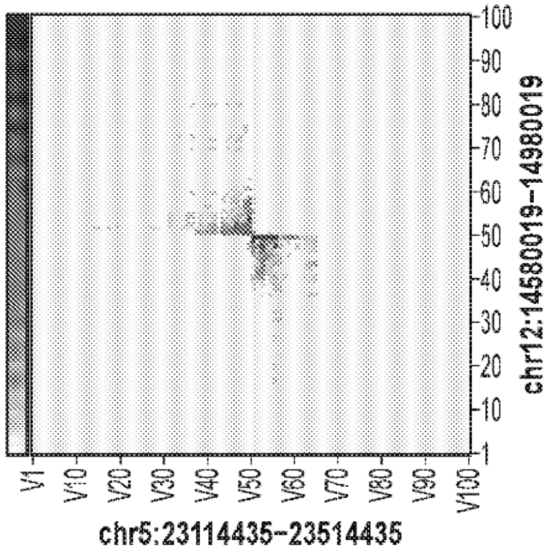


Figura 9B

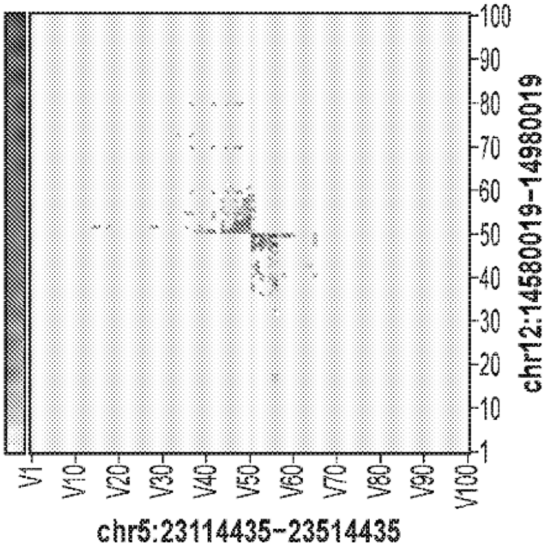


Figura 9C

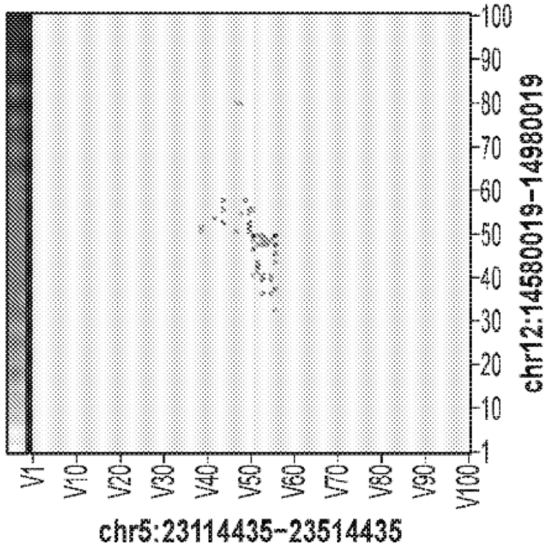


Figura 9D

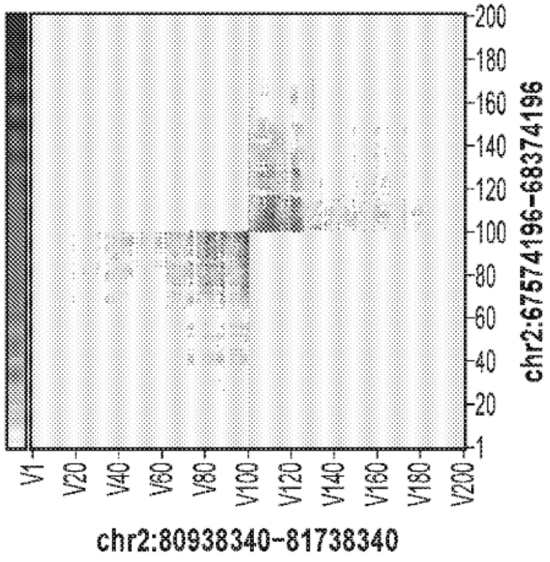


Figura 9E

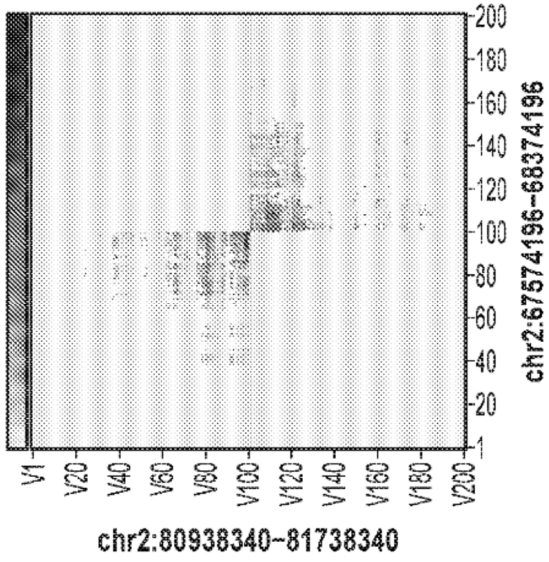


Figura 9F

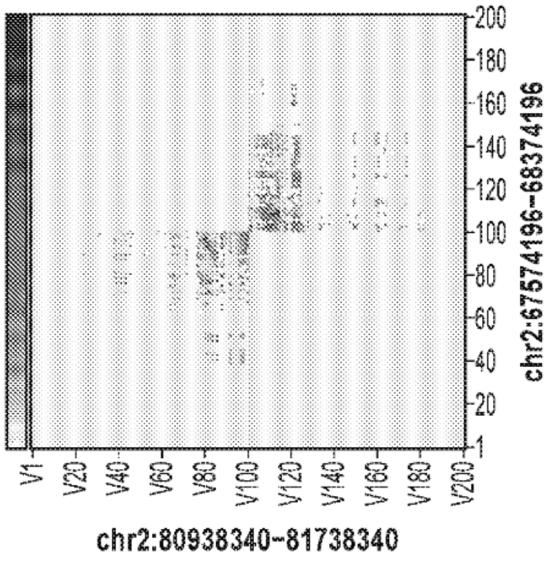


Figura 9G

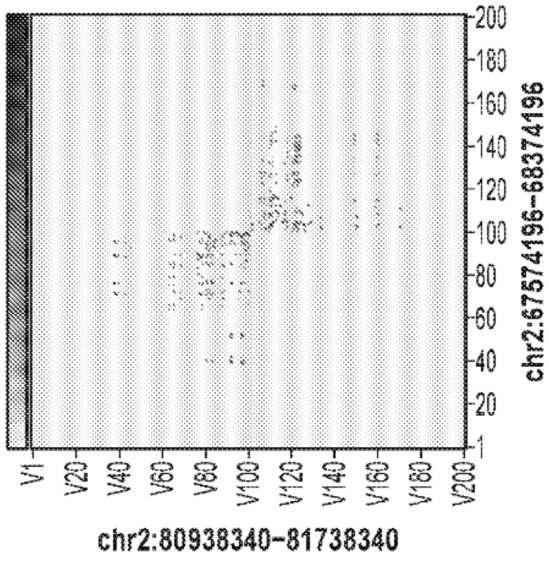


Figura 9H

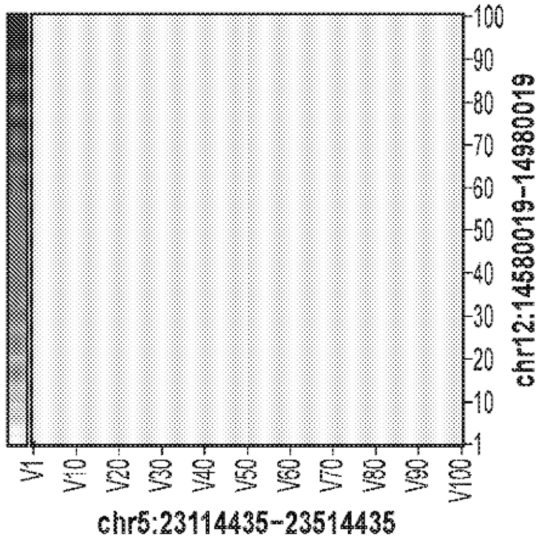


Figura 9I

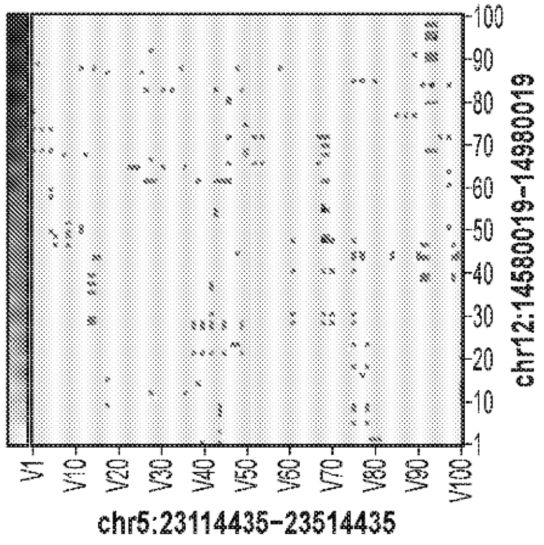


Figura 9J

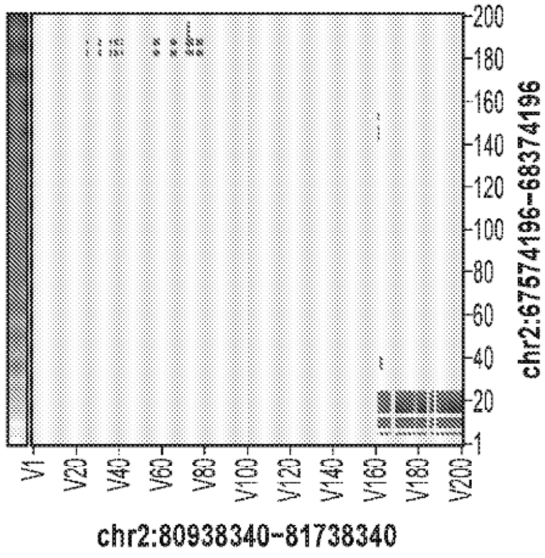


Figura 9K

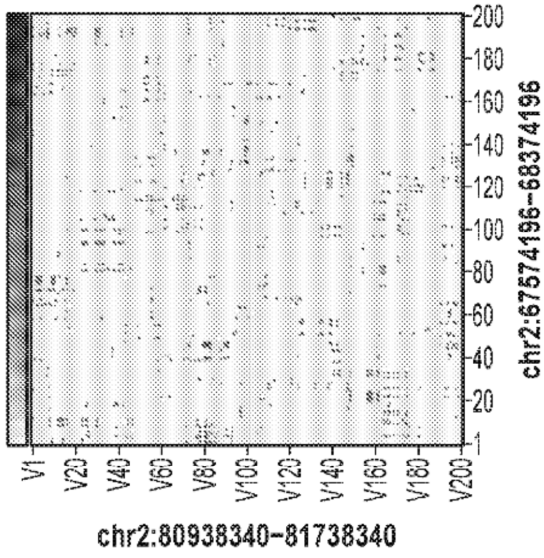


Figura 9L

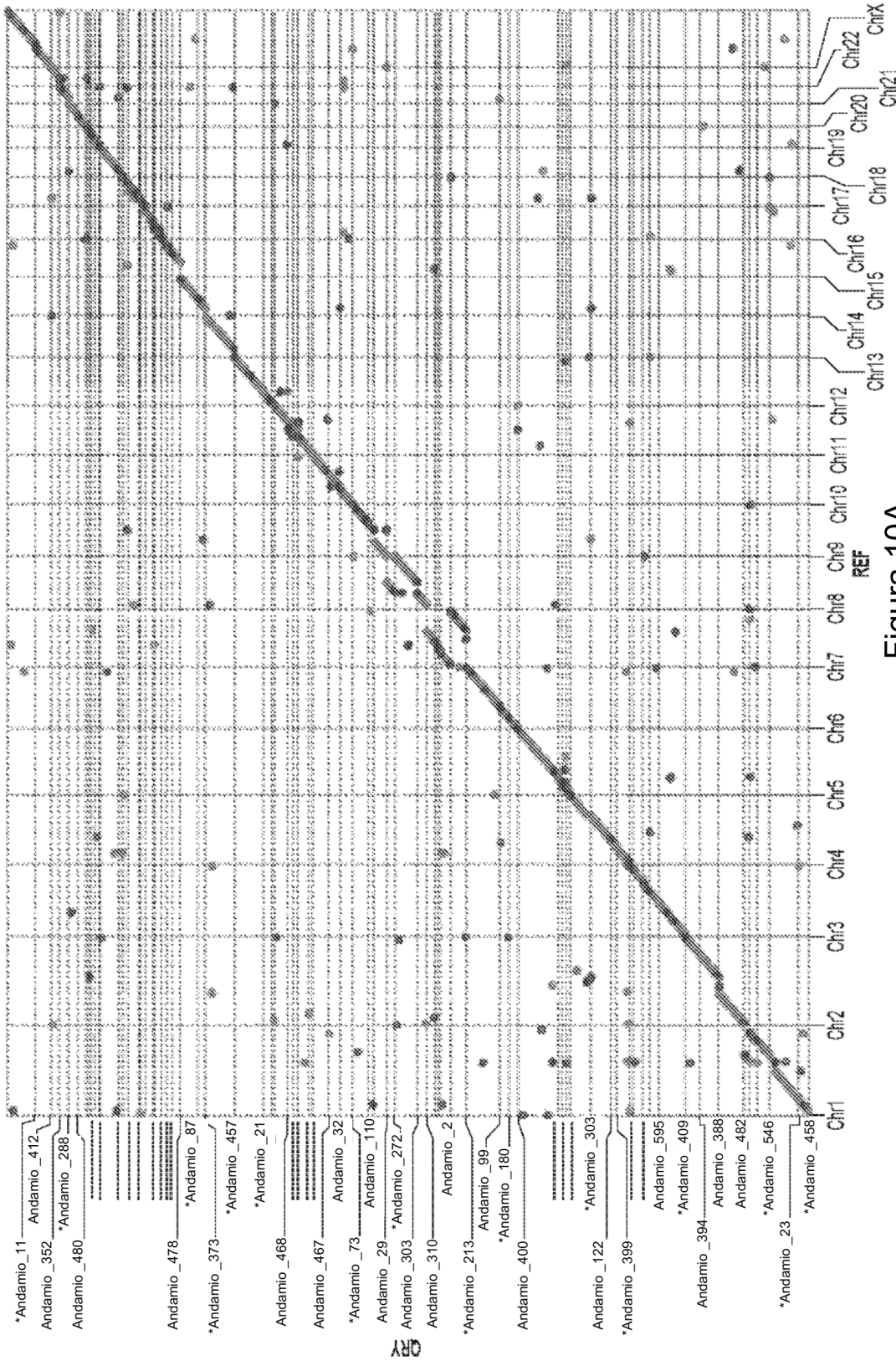

*Andamio_11
Andamio_412
Andamio_352
*Andamio_288
Andamio_480
Andamio_478
*Andamio_87
*Andamio_373
*Andamio_457
*Andamio_21
Andamio_468
Andamio_467
Andamio_32
*Andamio_73
Andamio_110
Andamio_29
*Andamio_272
Andamio_303
Andamio_310
Andamio_2
*Andamio_213
Andamio_99
*Andamio_180
Andamio_400
Andamio_122
*Andamio_303
*Andamio_399
Andamio_595
*Andamio_409
Andamio_394
Andamio_388
Andamio_482
*Andamio_546
*Andamio_23
*Andamio_458
QRY
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
ChrX
REF

Figura 10A

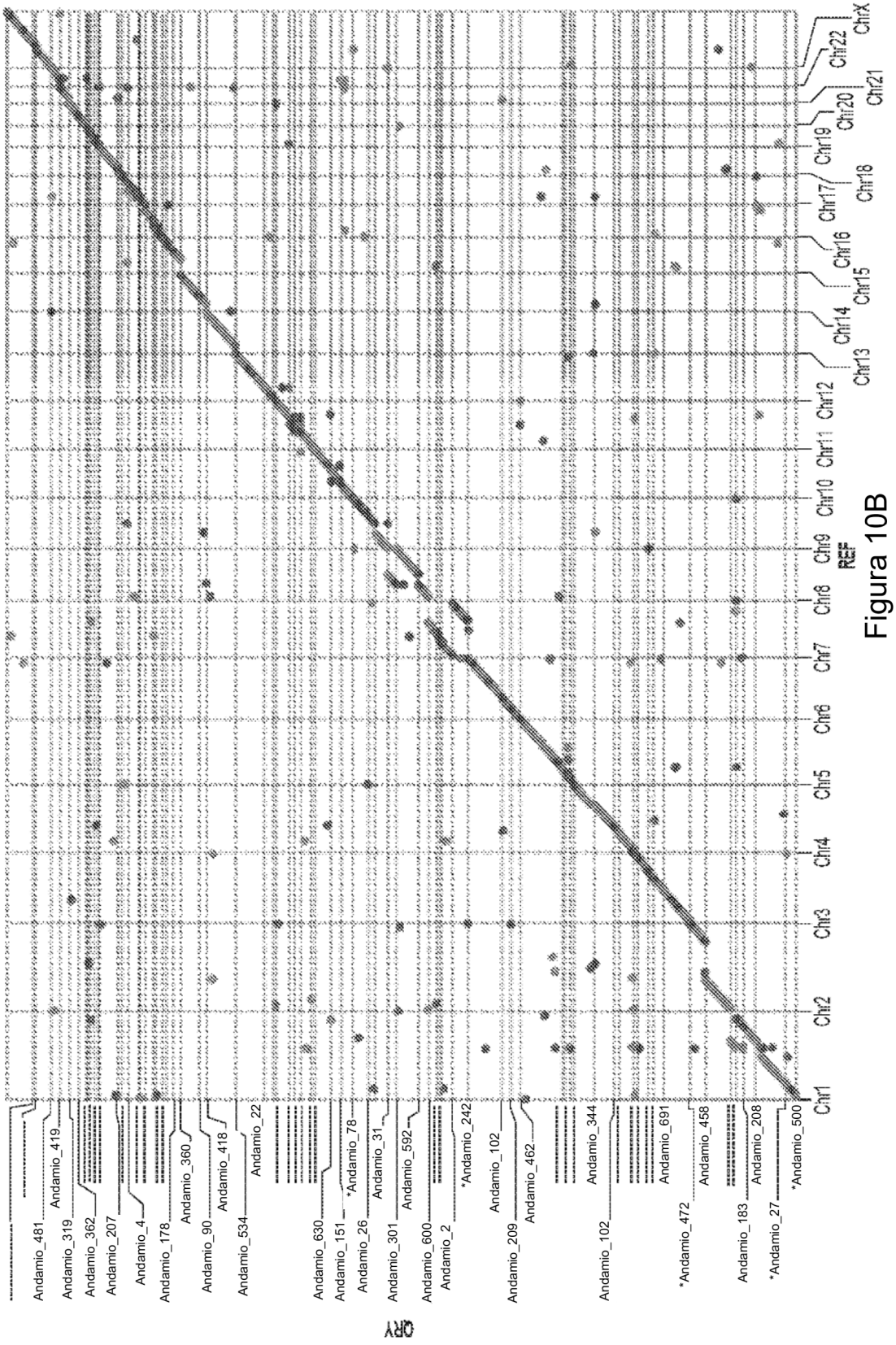

QRY
REF
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
ChrX
Andamio_481
Andamio_419
Andamio_319
Andamio_362
Andamio_207
Andamio_4
Andamio_178
Andamio_360
Andamio_90
Andamio_418
Andamio_534
Andamio_22
Andamio_630
Andamio_151
*Andamio_78
Andamio_26
Andamio_31
Andamio_301
Andamio_592
Andamio_600
Andamio_2
*Andamio_242
Andamio_102
Andamio_209
Andamio_462
Andamio_344
Andamio_102
Andamio_691
*Andamio_472
Andamio_458
Andamio_183
Andamio_208
*Andamio_27
*Andamio_500

Figura 10B

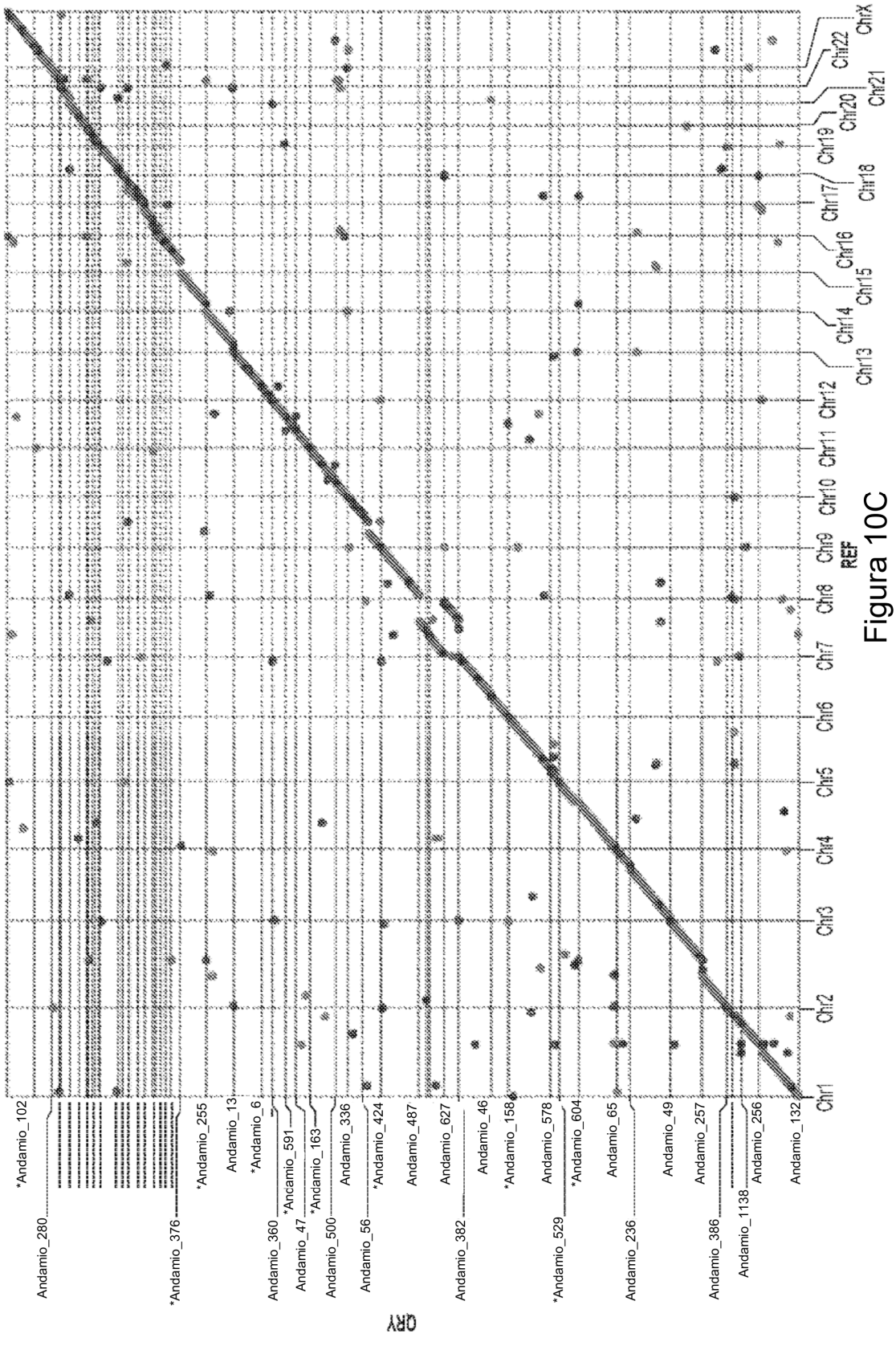

QRY
REF
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
ChrX
*Andamio_102
Andamio_280
*Andamio_376
*Andamio_255
Andamio_13
*Andamio_6
Andamio_360
*Ancamio_591
Andamio_47
*Andamio_163
Andamio_500
Andamio_336
Andamio_56
*Andamio_424
Andamio_487
Andamio_627
Andamio_382
Andamio_46
*Andamio_158
Andamio_578
*Andamio_529
*Andamio_604
Andamio_65
Andamio_236
Andamio_49
Andamio_257
Andamio_386
Andamio_1138
Andamio_256
Andamio_132

Figura 10C

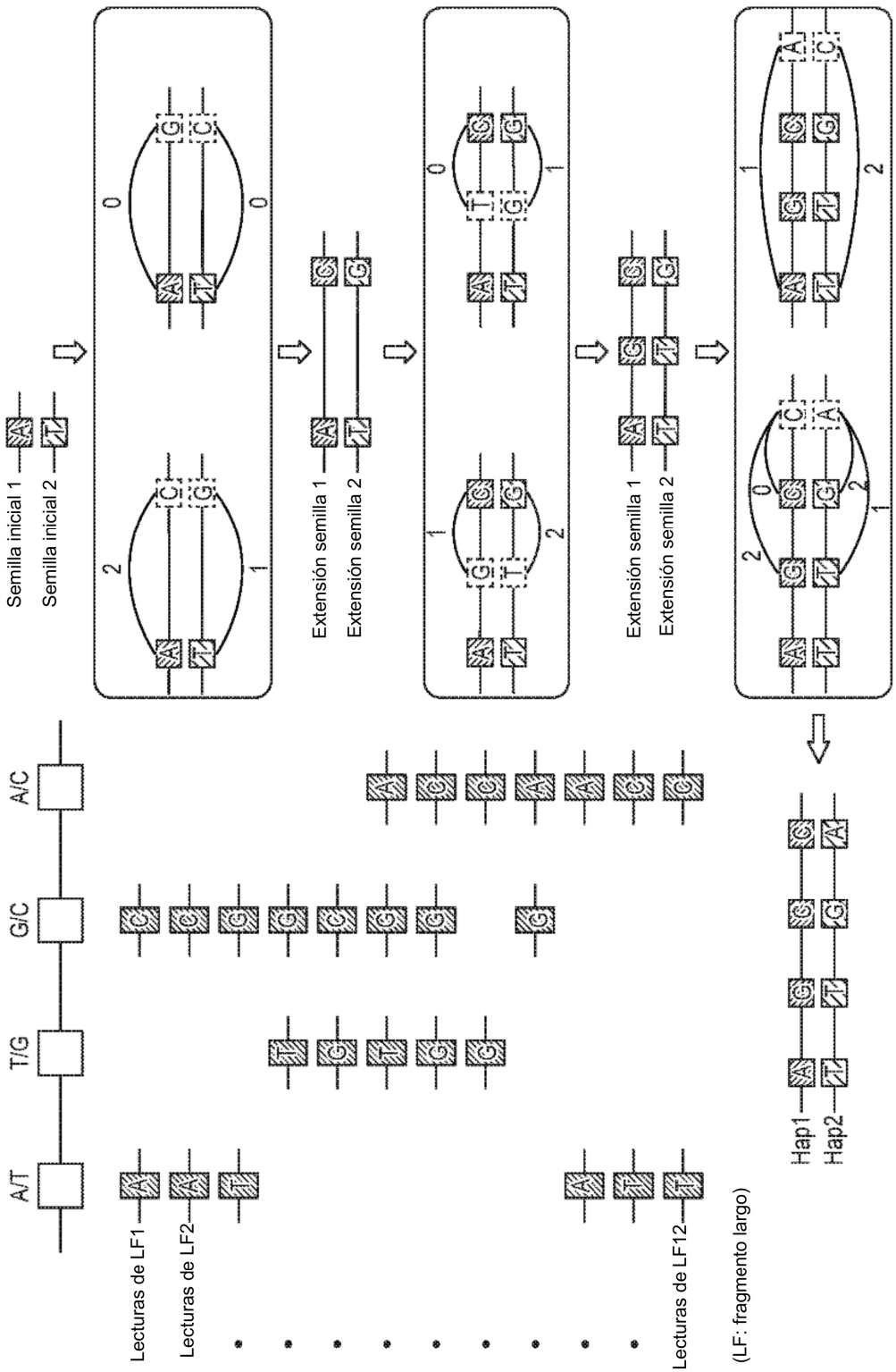

Semilla inicial 1
Semilla inicial 2
Extensión semilla 1
Extensión semilla 2
Extensión semilla 1
Extensión semilla 2
A/T
T/G
G/C
A/C
Lecturas de LF1
Lecturas de LF2
Lecturas de LF12
(LF: fragmento largo)
Hap1
Hap2

Figura 11

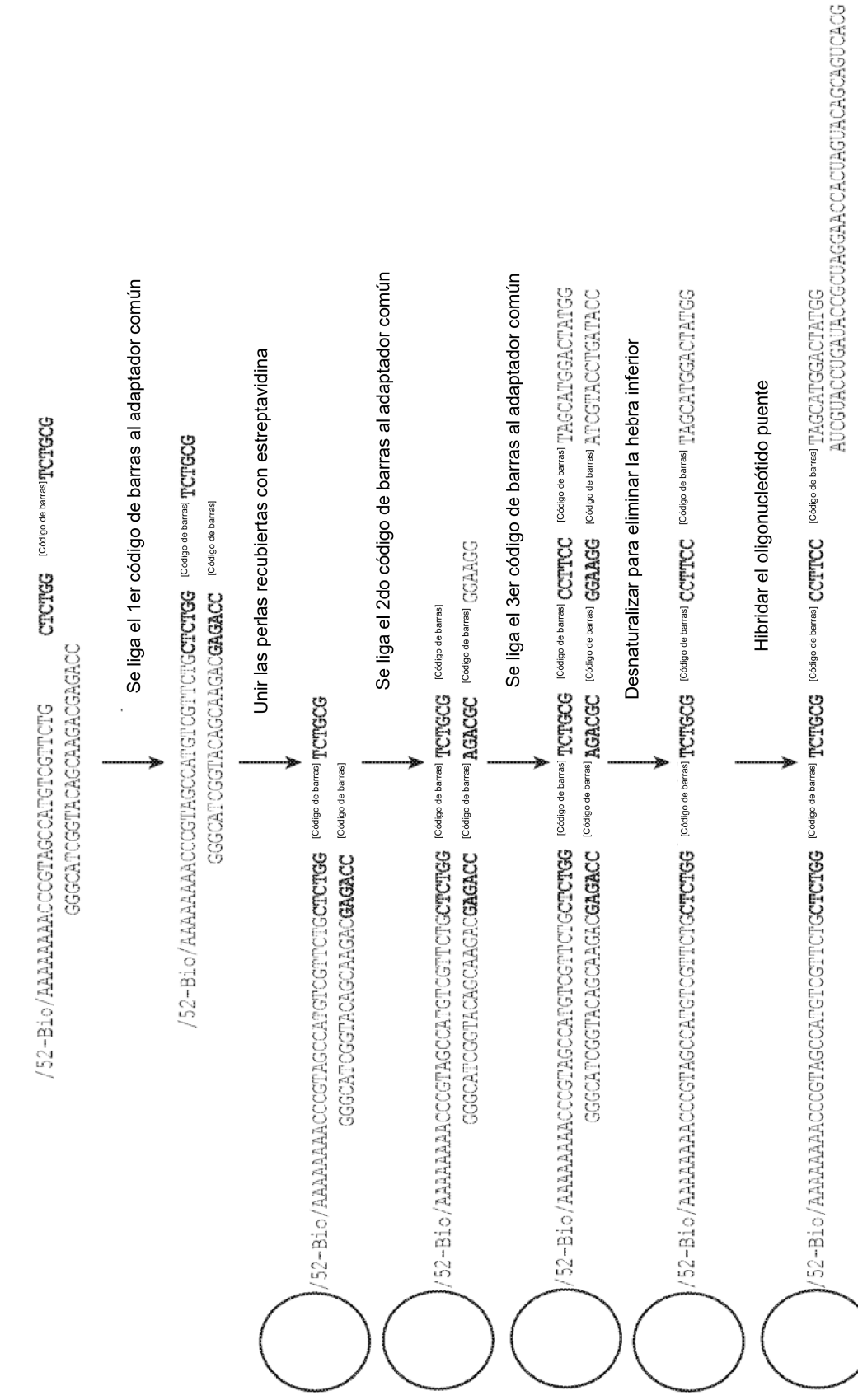

/52-Bio/AAAAAAAACCCGTAGCCATGTCGTTCTG [Código de barras] CTCTGG TCTGCG
GGGCATCGGTACAGCAGACGAGACC
Se liga el 1er código de barras al adaptador común
/52-Bio/AAAAAAAACCCGTAGCCATGTCGTTCTG [Código de barras] CTCTGG TCTGCG
GGGCATCGGTACAGCAAGAC [Código de barras] GAGACC
Unir las perlas recubiertas con estreptavidina
/52-Bio/AAAAAAAACCCGTAGCCATGTCGTTCTG [Código de barras] CTCTGG TCTGCG
GGGCATCGGTACAGCAAGAC [Código de barras] GAGACC
Se liga el 2do código de barras al adaptador común
/52-Bio/AAAAAAAACCCGTAGCCATGTCGTTCTG [Código de barras] CTCTGG TCTGCG [Código de barras]
GGGCATCGGTACAGCAAGAC [Código de barras] GAGACC [Código de barras] AGACGC
Se liga el 3er código de barras al adaptador común
/52-Bio/AAAAAAAACCCGTAGCCATGTCGTTCTG [Código de barras] CTCTGG TCTGCG [Código de barras] CCTTCC [Código de barras] TAGCATGGACTATGG
GGGCATCGGTACAGCAAGAC [Código de barras] GAGACC [Código de barras] AGACGC [Código de barras] GGAAGG [Código de barras] ATCGTACCTGATACC
Desnaturalizar para eliminar la hebra inferior
/52-Bio/AAAAAAAACCCGTAGCCATGTCGTTCTG [Código de barras] CTCTGG TCTGCG [Código de barras] CCTTCC [Código de barras] TAGCATGGACTATGG
Hibridar el oligonucleótido puente
/52-Bio/AAAAAAAACCCGTAGCCATGTCGTTCTG [Código de barras] CTCTGG TCTGCG [Código de barras] CCTTCC [Código de barras] TAGCATGGACTATGG
AUCGUACCUGAUACCGCUAGGAACCACUAGUACAGCAGUCACG

Figura 12

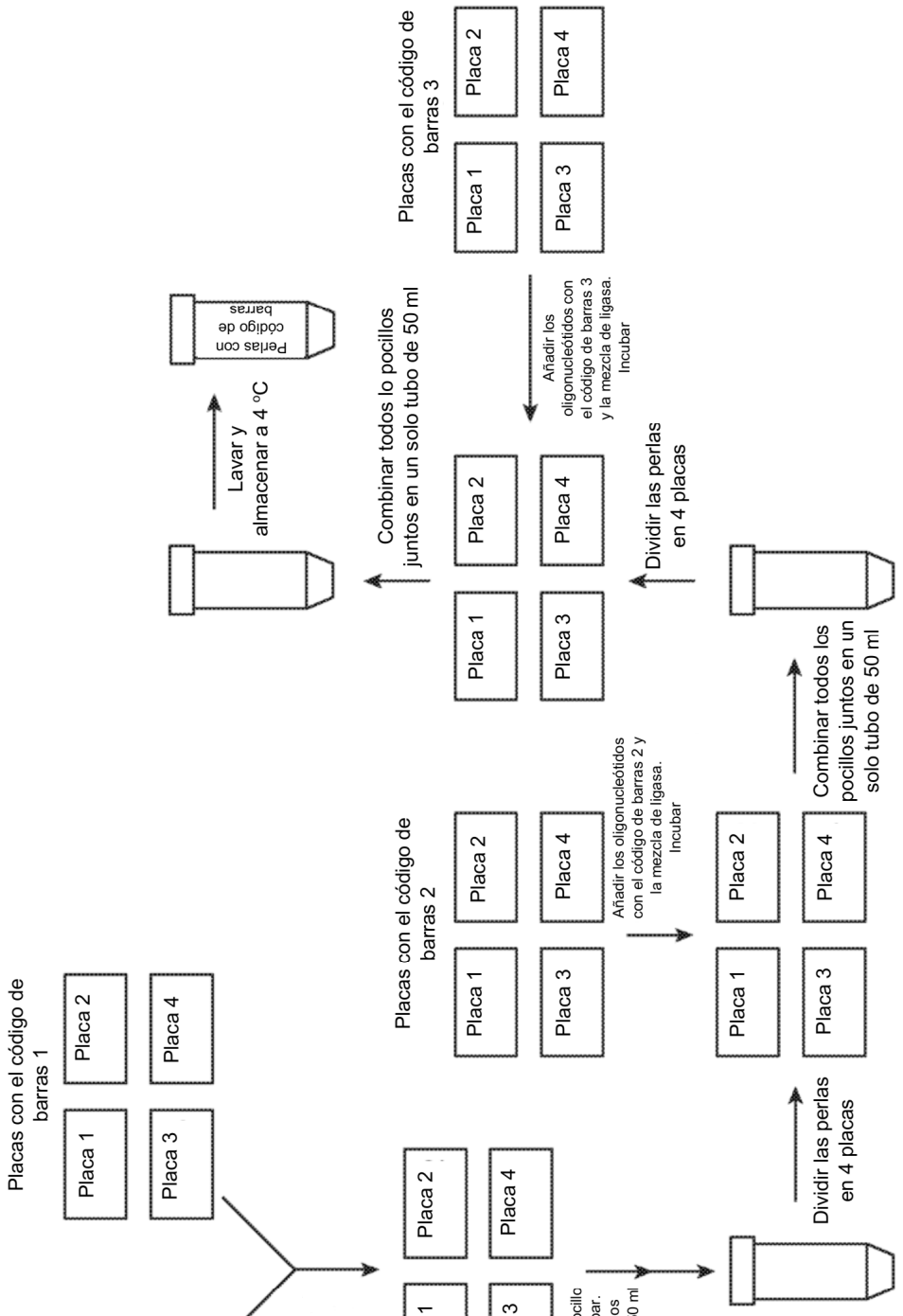

Oligonucleótido común de las perlas marcado con biotina hibridado
Añadir el oligonucleótido común de las perlas a cada pocillo, cada oligonucleótido con el código de barras 1 a su pocillo correspondiente y la mezcla de ligasa a las placas de ligazón. Incubar.
Placas con el código de barras 1
Placa 1
Placa 2
Placa 3
Placa 4
Añadir las perlas a cada pocillo de todas las placas. Incubar. Combinar todos los pocillos juntos en un solo tubo de 50 ml
Dividir las perlas en 4 placas
Añadir los oligonucleótidos con el código de barras 2 y la mezcla de ligasa. Incubar
Placas con el código de barras 2
Combinar todos los pocillos juntos en un solo tubo de 50 ml
Dividir las perlas en 4 placas
Añadir los oligonucleótidos con el código de barras 3 y la mezcla de ligasa. Incubar
Placas con el código de barras 3
Combinar todos lo pocillos juntos en un solo tubo de 50 ml
Lavar y almacenar a 4 ºC
Perlas con código de barras

Figura 13

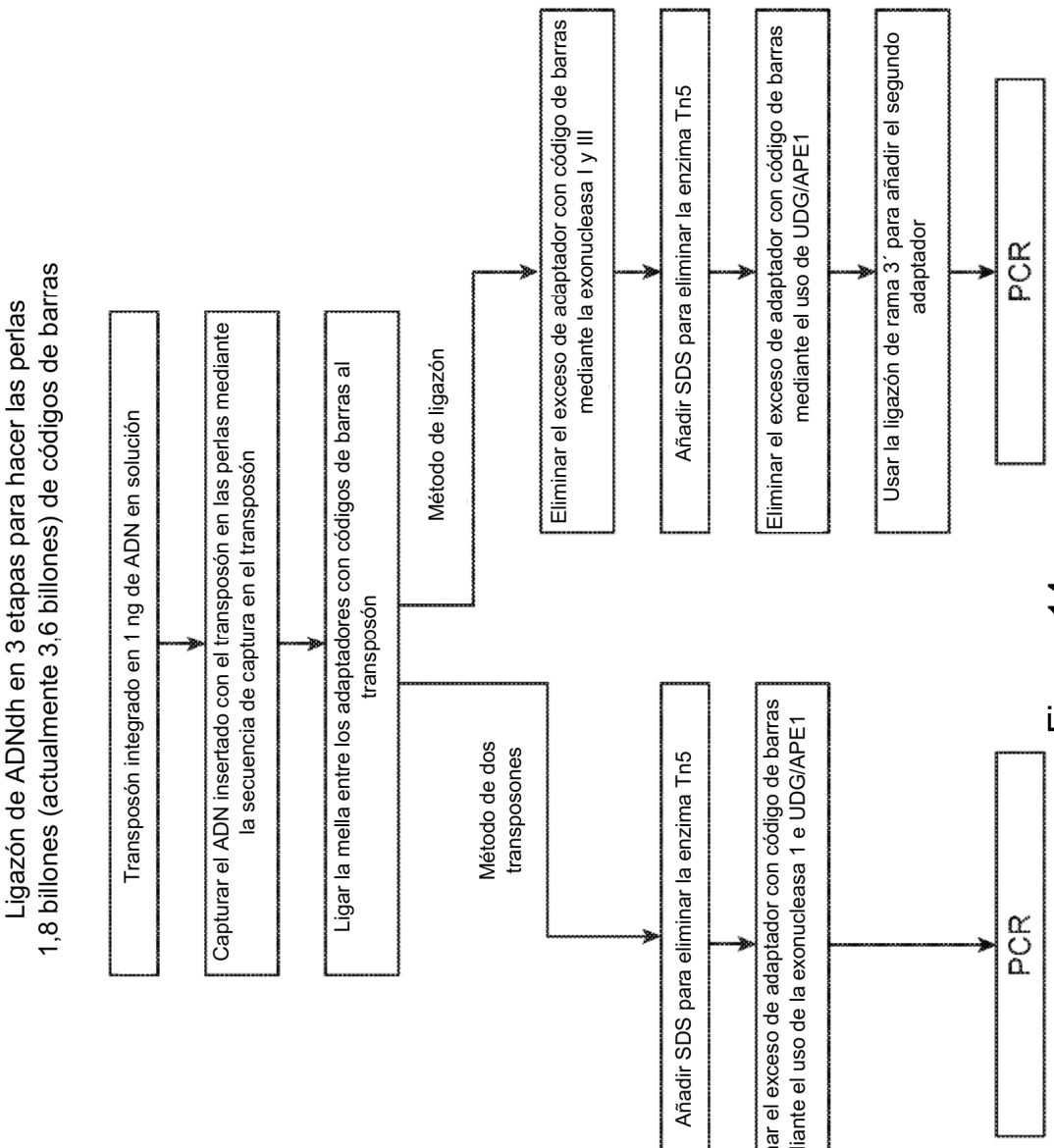

Ligazón de ADNdh en 3 etapas para hacer las perlas
1,8 billones (actualmente 3,6 billones) de códigos de barras
Transposón integrado en 1 ng de ADN en solución
Capturar el ADN insertado con el transposón en las perlas mediante la secuencia de captura en el transposón
Ligar la mella entre los adaptadores con códigos de barras al transposón
Método de ligazón
Eliminar el exceso de adaptador con código de barras mediante la exonucleasa I y III
Añadir SDS para eliminar la enzima Tn5
Eliminar el exceso de adaptador con código de barras mediante el uso de UDG/APE1
Usar la ligazón de rama 3´ para añadir el segundo adaptador
PCR
Método de dos transposones
Añadir SDS para eliminar la enzima Tn5
Eliminar el exceso de adaptador con código de barras mediante el uso de la exonucleasa 1 e UDG/APE1
PCR

Figura 14

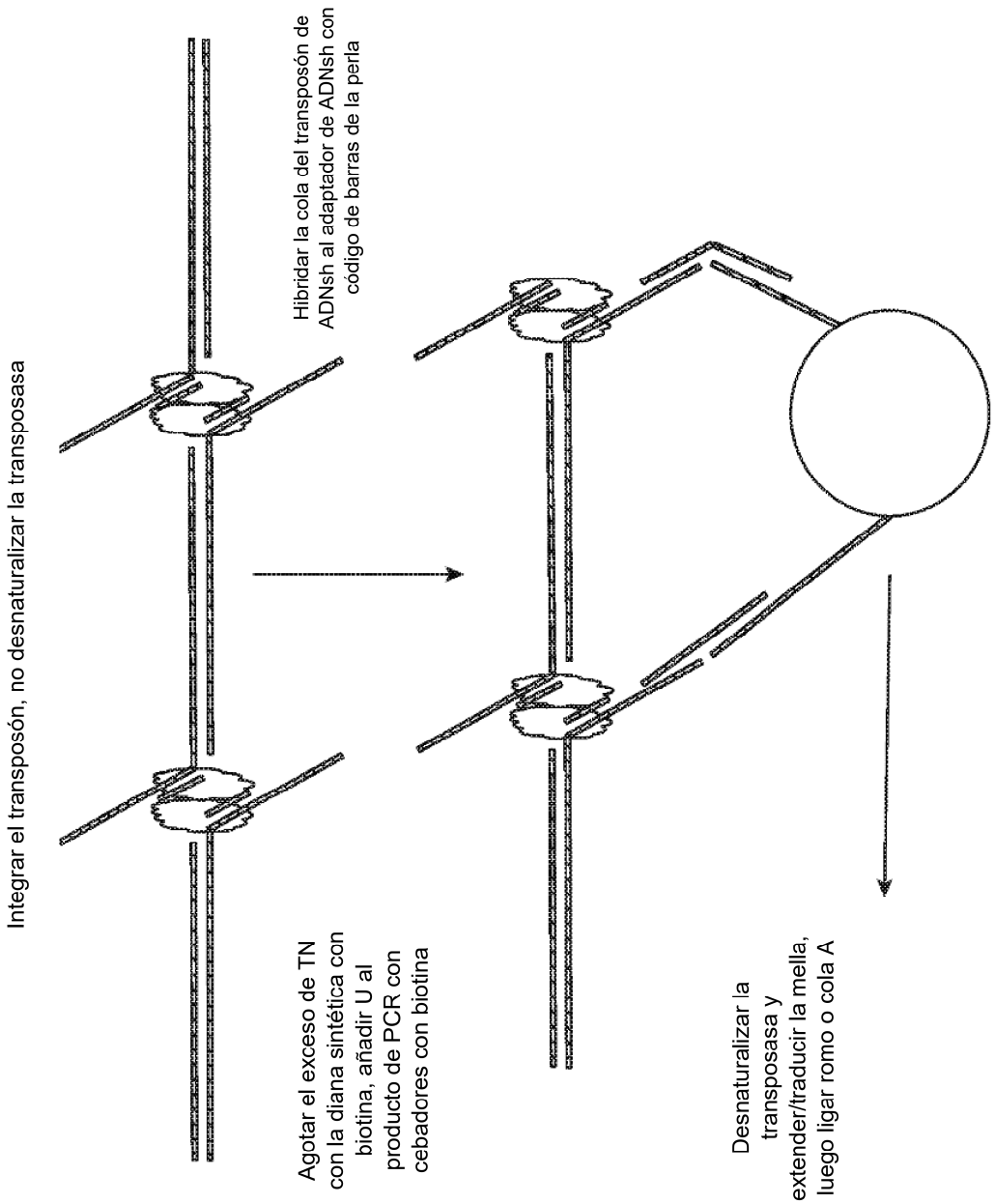

Integrar el transposón, no desnaturalizar la transposasa
Hibridar la cola del transposón de ADNsh al adaptador de ADNsh con código de barras de la perla
Agotar el exceso de TN con la diana sintética con biotina, añadir U al producto de PCR con cebadores con biotina
Desnaturalizar la transposasa y extender/traducir la mella, luego ligar romo o cola A

Figura 15

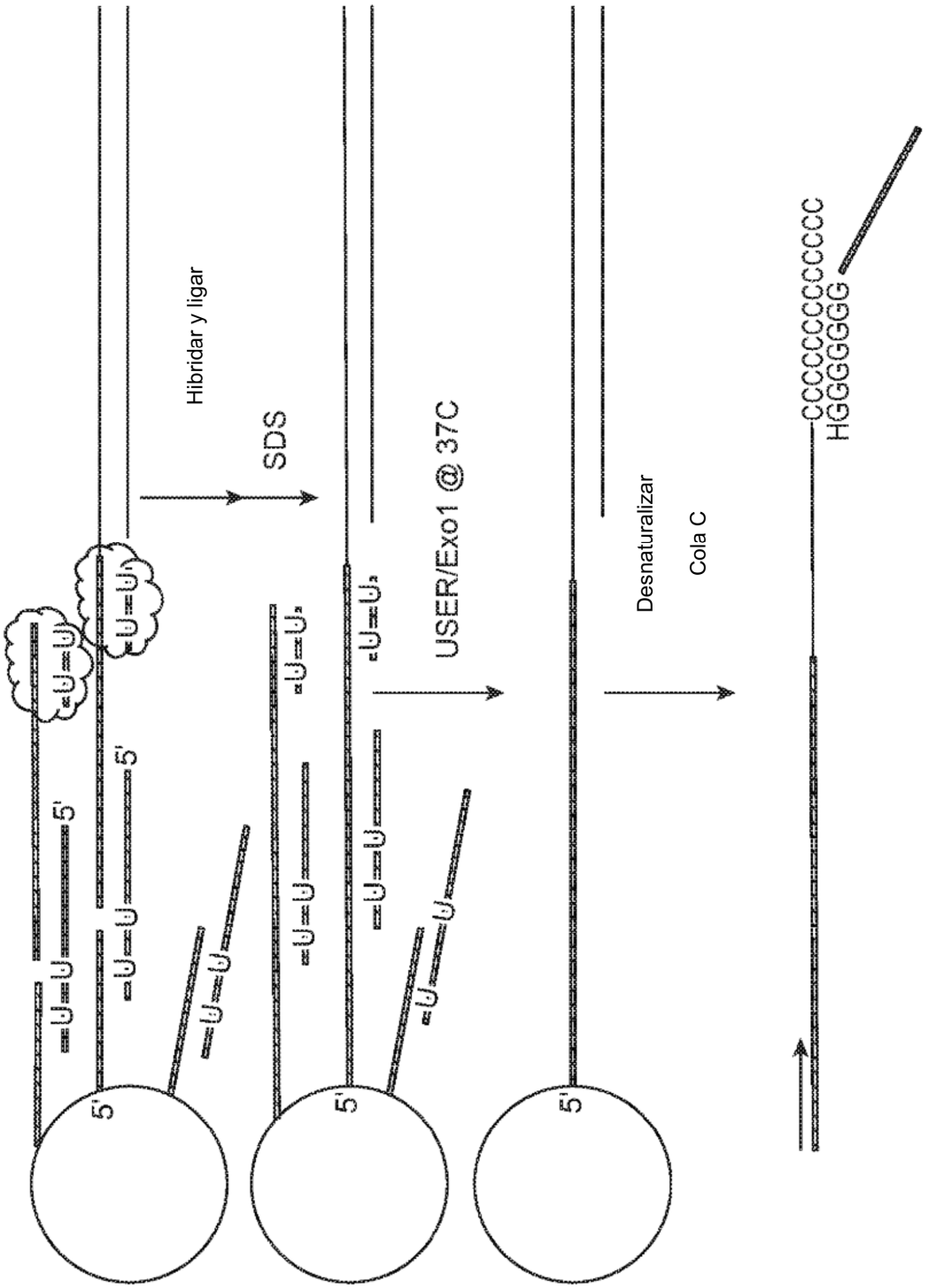

Hibridar y ligar
SDS
USER/Exo1 @ 37C
Desnaturalizar
Cola C
CCCCCCCCCCC
HGGGGGGG
5'

Figura 16

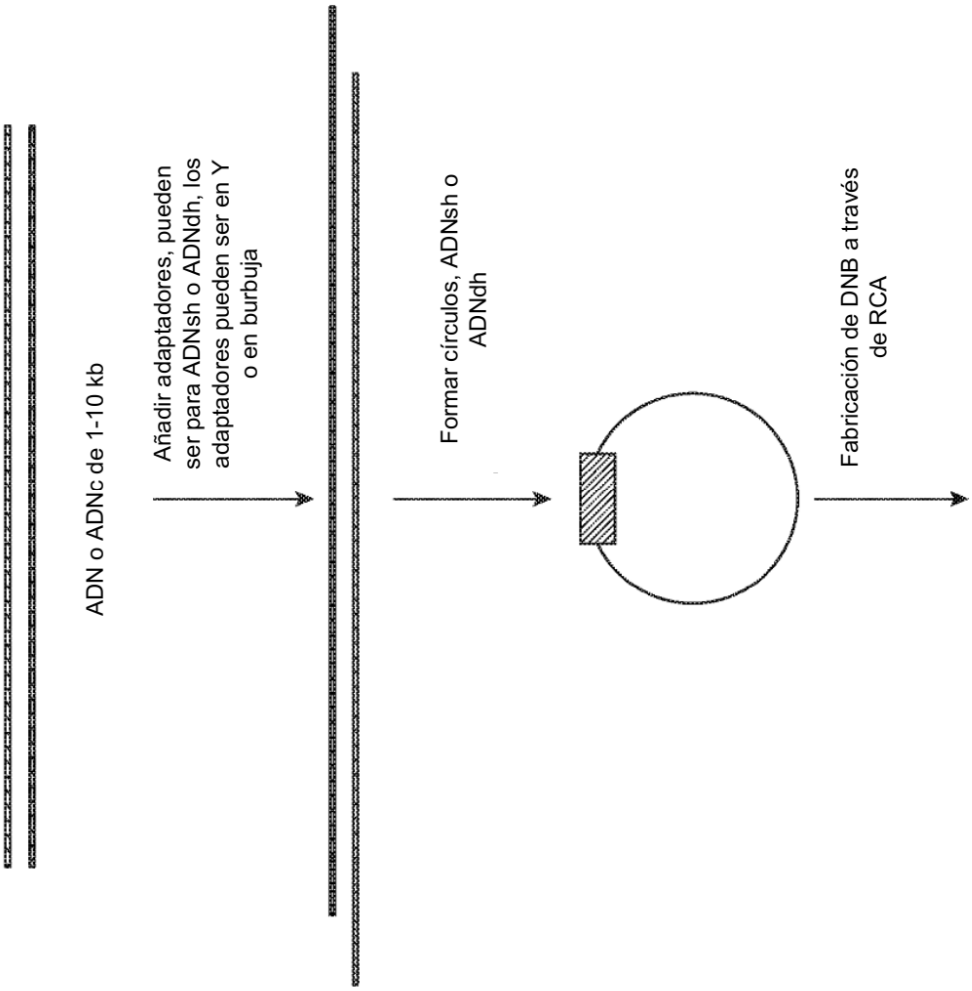

ADN o ADNc de 1-10 kb
Añadir adaptadores, pueden ser para ADNsh o ADNdh, los adaptadores pueden ser en Y o en burbuja
Formar círculos, ADNsh o ADNdh
Fabricación de DNB a través de RCA

Figura 17

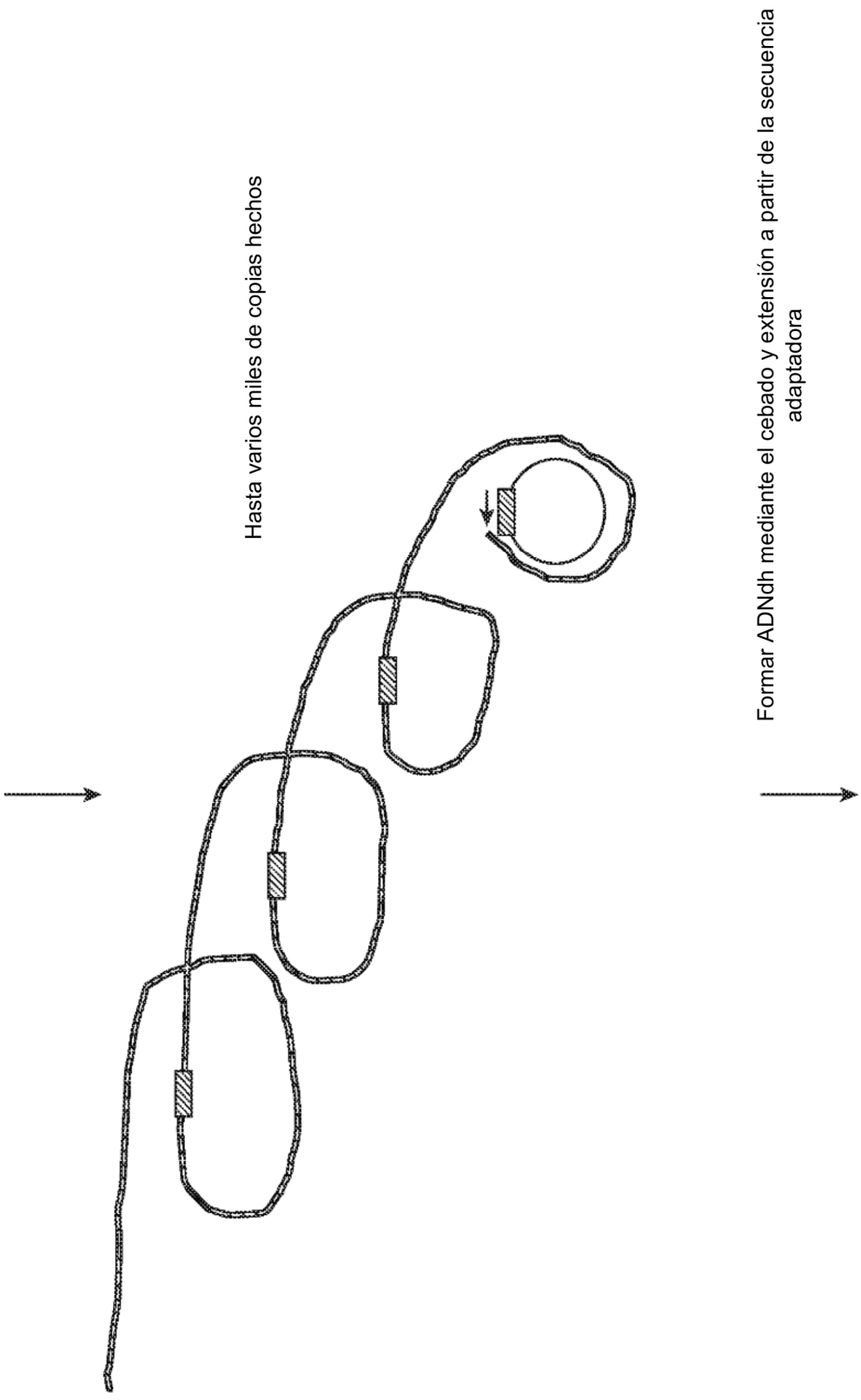
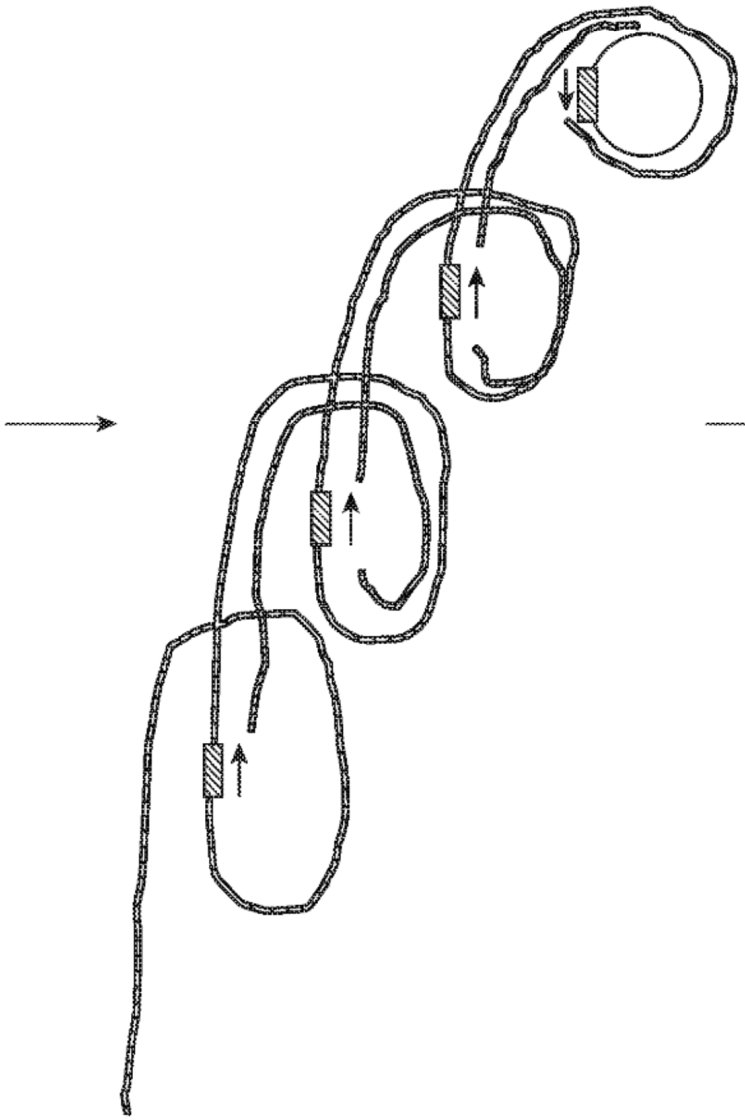


Figura 17 (cont.1)



La DNB de doble hebra puede ahora insertarse con el transposón y capturarse en las perlas de stLFR, posiblemente se hicieron miles de copias de la molécula original y todas en la misma hebra de ADN, esto permite una alta cobertura de la molécula original con la secuenciación de stLFR

Figura 17 (cont. 2)

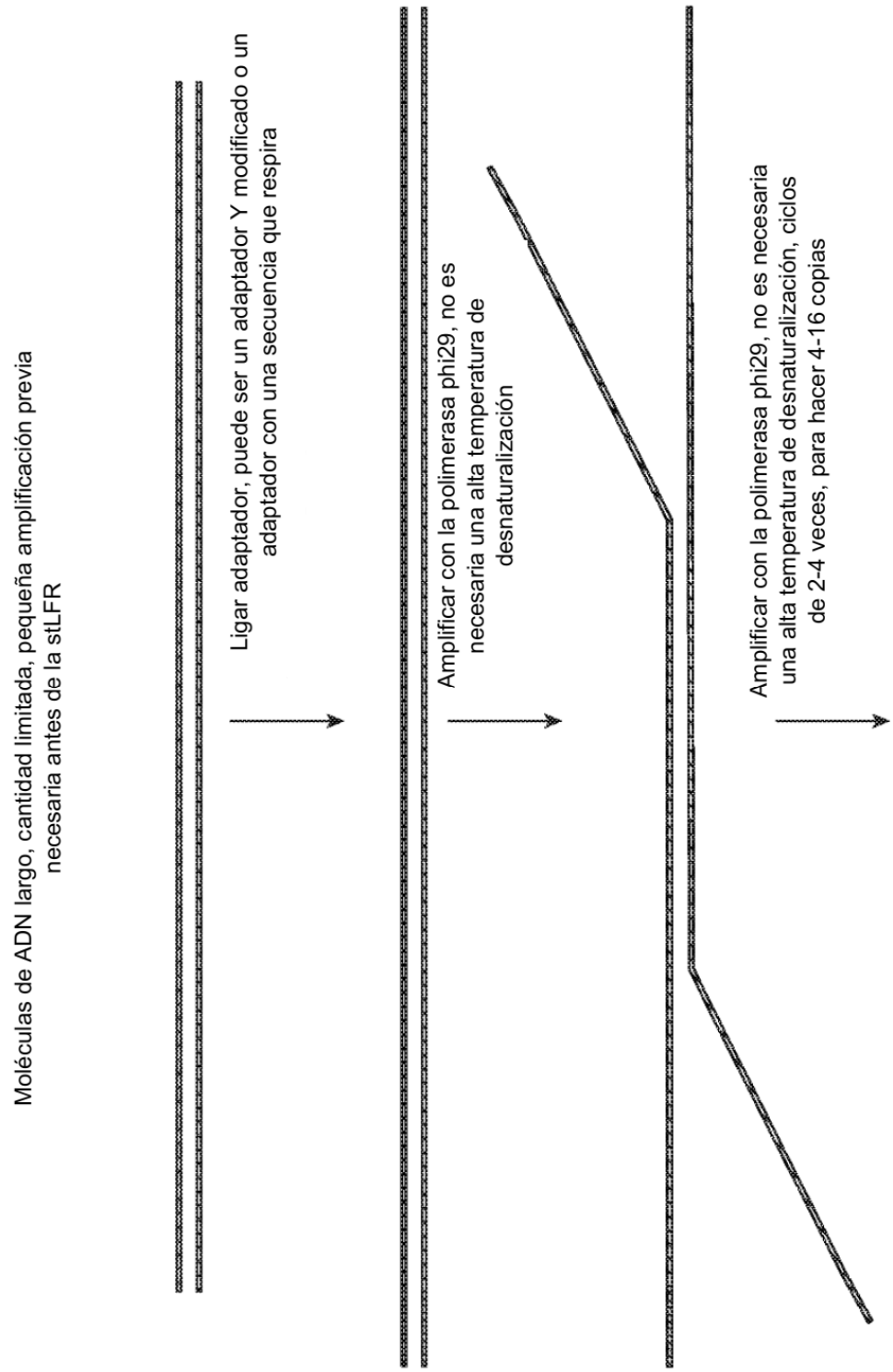

Moléculas de ADN largo, cantidad limitada, pequeña amplificación previa necesaria antes de la stLFR
Ligar adaptador, puede ser un adaptador Y modificado o un adaptador con una secuencia que respira
Amplificar con la polimerasa phi29, no es necesaria una alta temperatura de desnaturalización
Amplificar con la polimerasa phi29, no es necesaria una alta temperatura de desnaturalización, ciclos de 2-4 veces, para hacer 4-16 copias

Figura 18

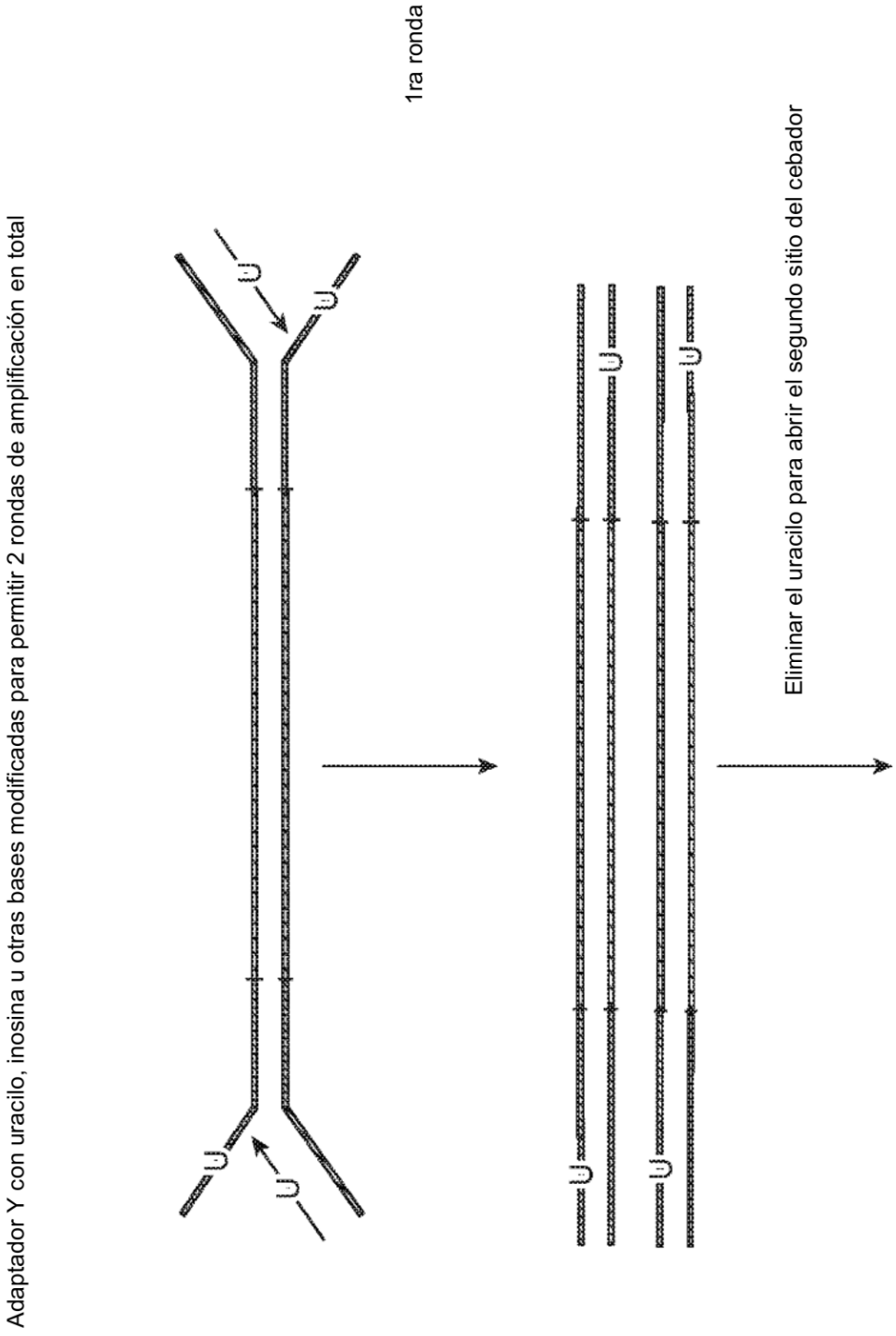


Figura 18 (cont. 1)

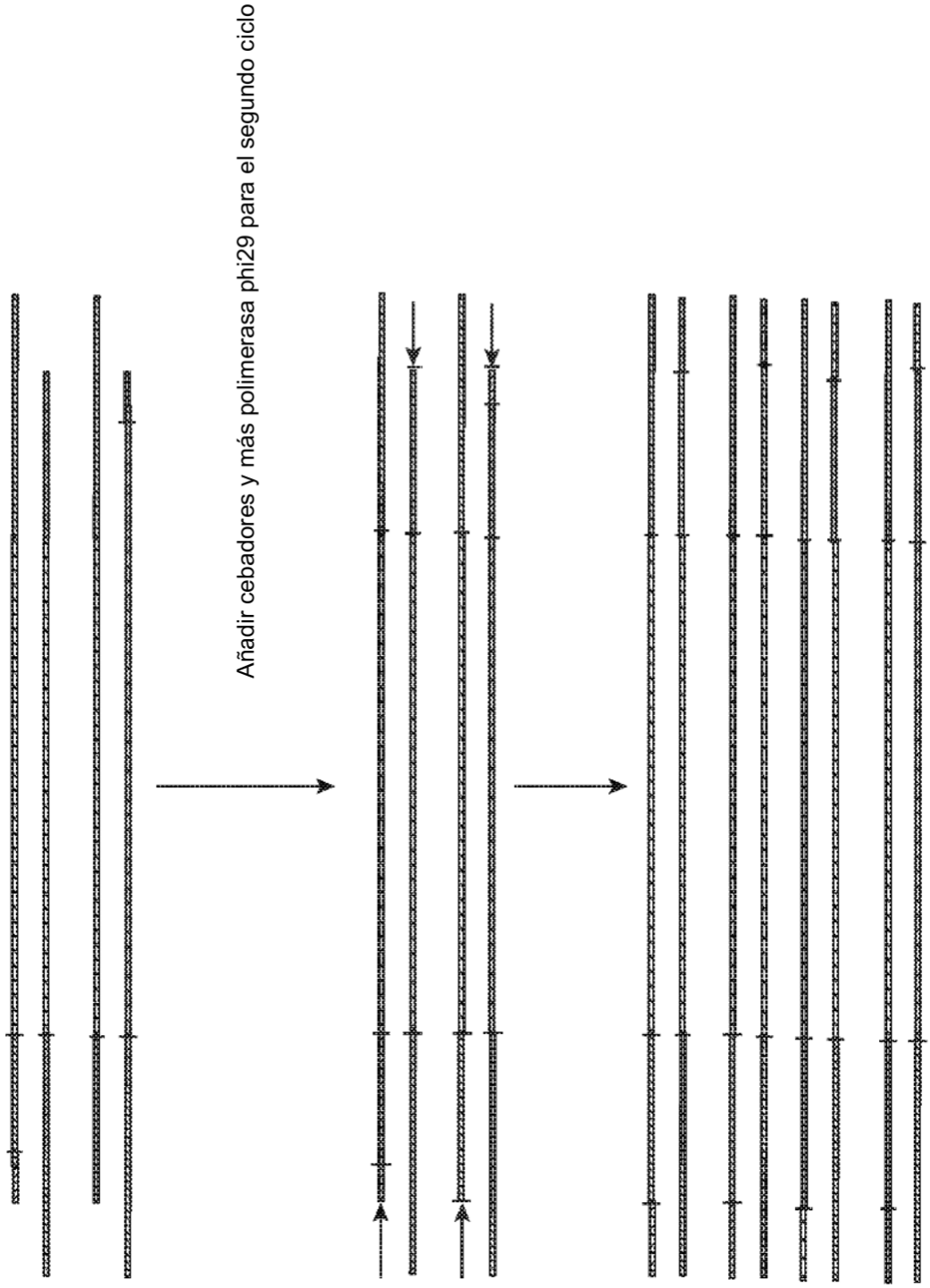


Figura 18 (cont. 2)

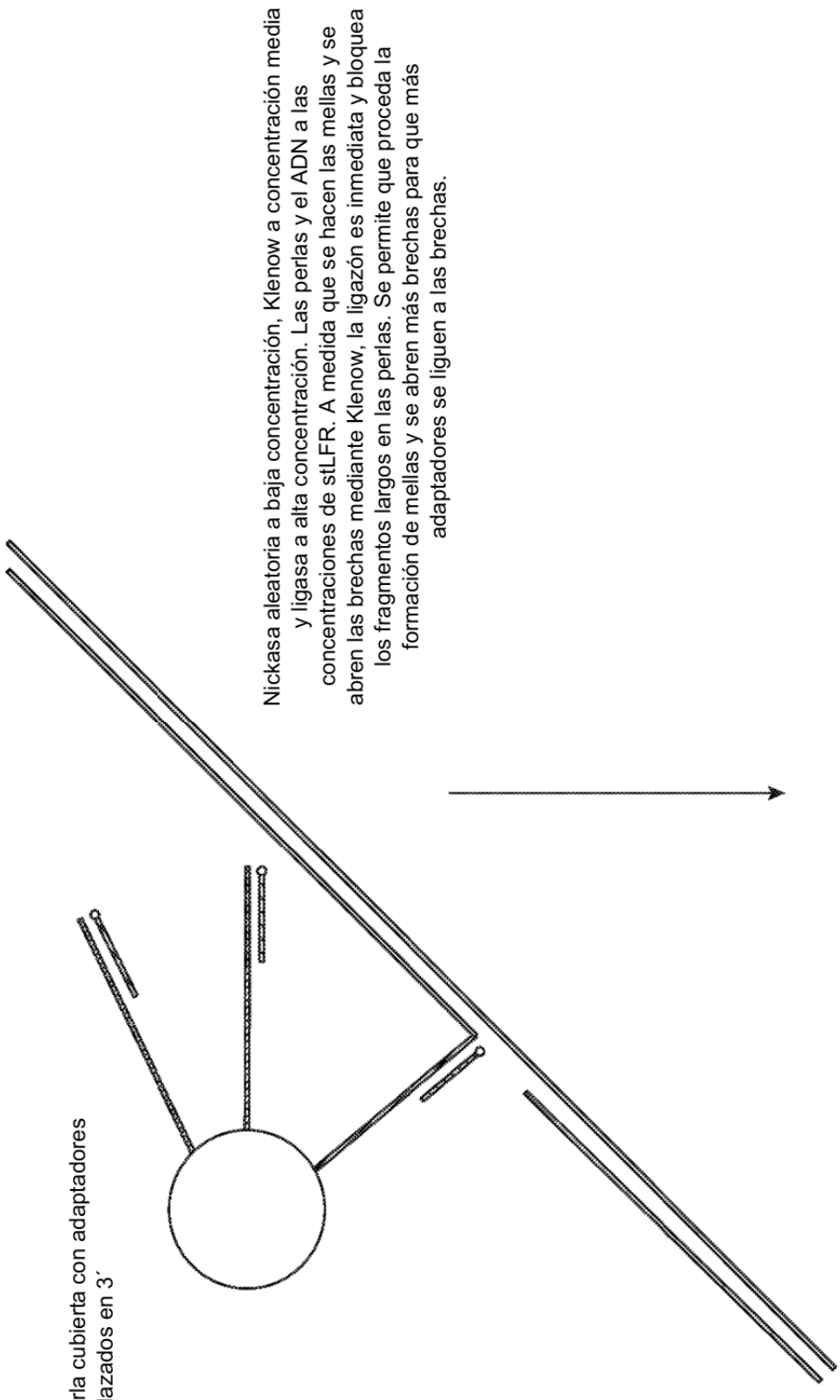

Perla cubierta con adaptadores enlazados en 3´
Nickasa aleatoria a baja concentración, Klenow a concentración media y ligasa a alta concentración. Las perlas y el ADN a las concentraciones de stLFR. A medida que se hacen las mellas y se abren las brechas mediante Klenow, la ligazón es inmediata y bloquea los fragmentos largos en las perlas. Se permite que proceda la formación de mellas y se abren más brechas para que más adaptadores se liguen a las brechas.

Figura 19

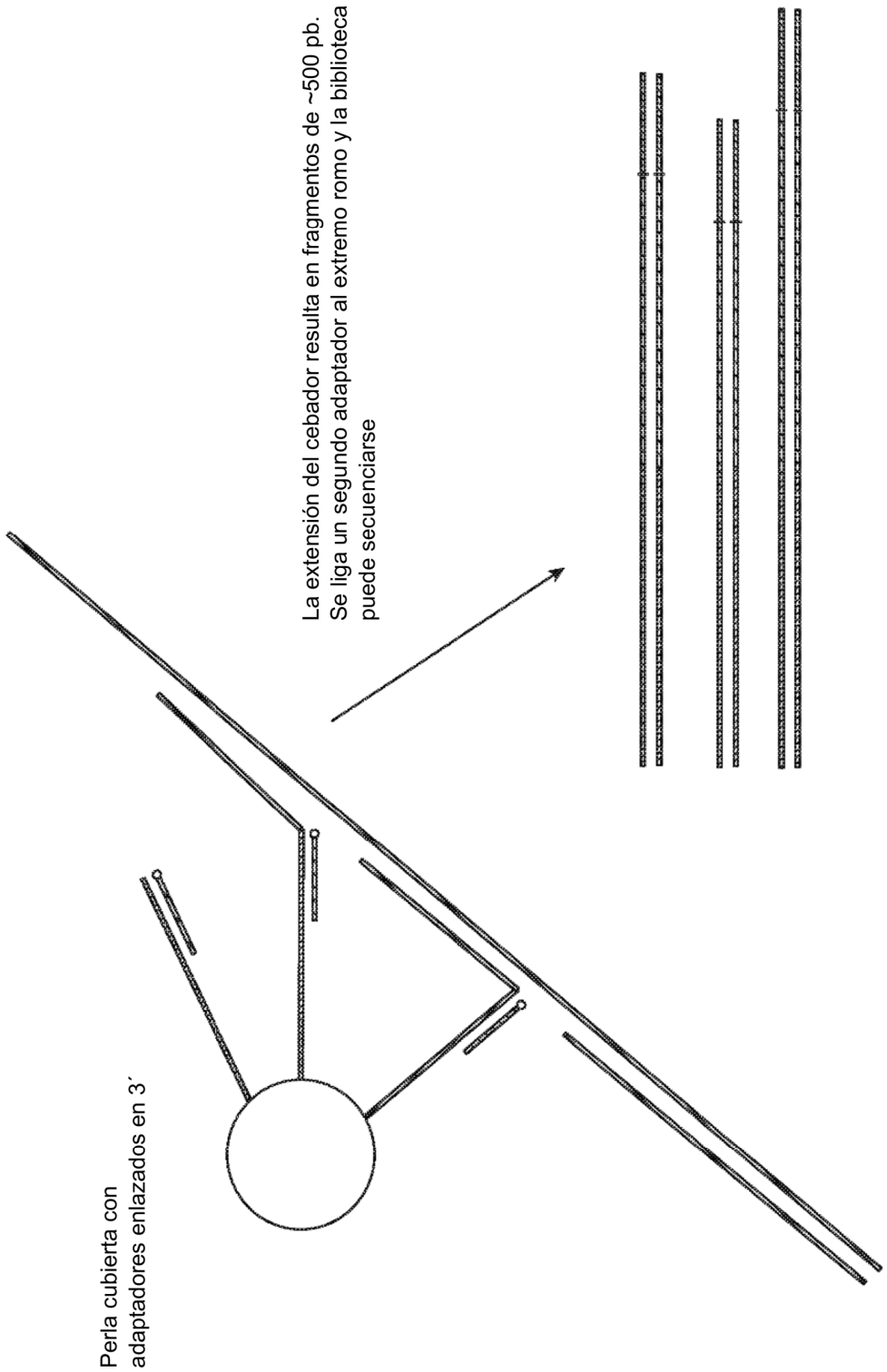


Figura 19 (cont. 1)

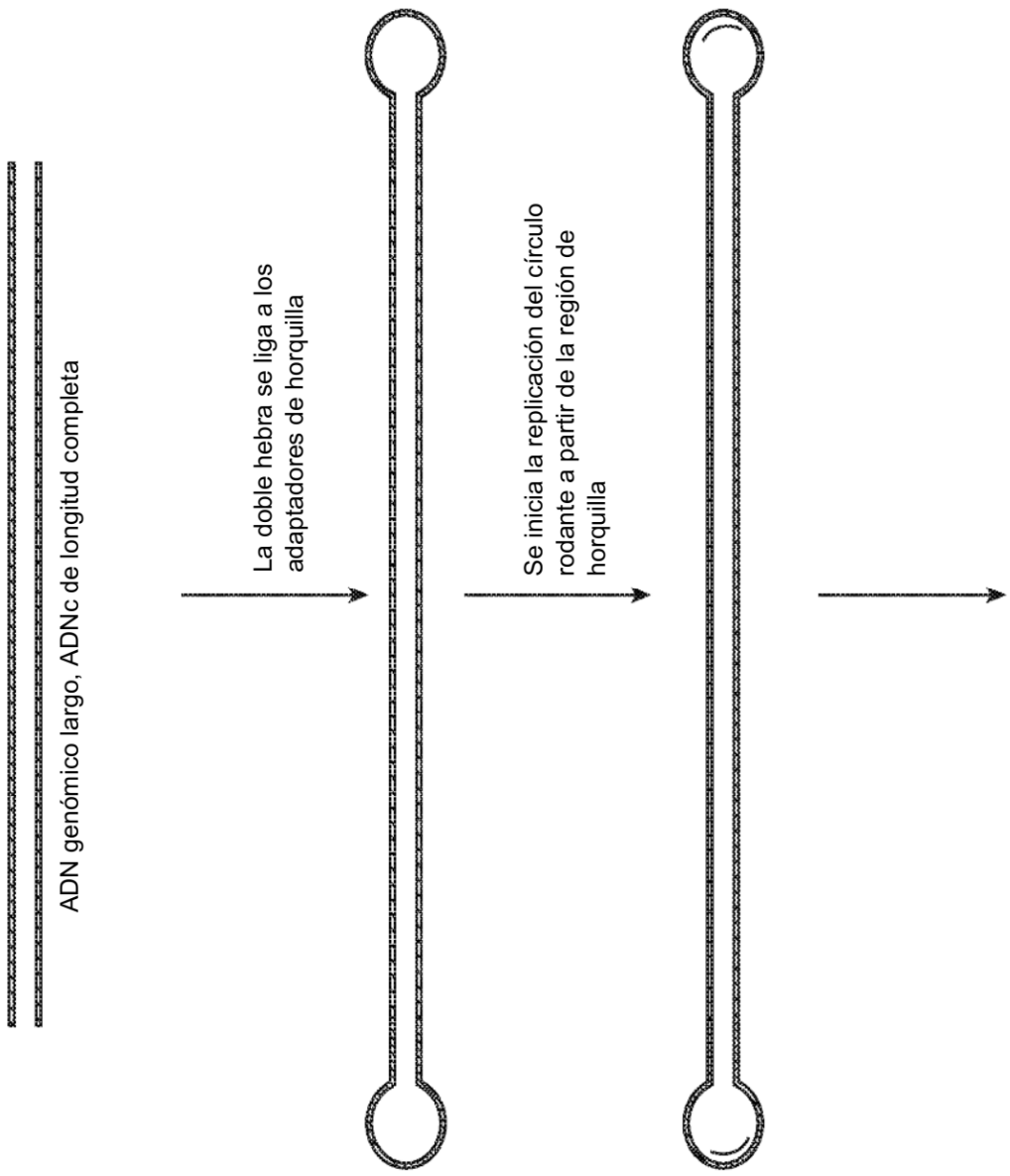

ADN genómico largo, ADNc de longitud completa
La doble hebra se liga a los adaptadores de horquilla
Se inicia la replicación del círculo rodante a partir de la región de horquilla

Figura 20

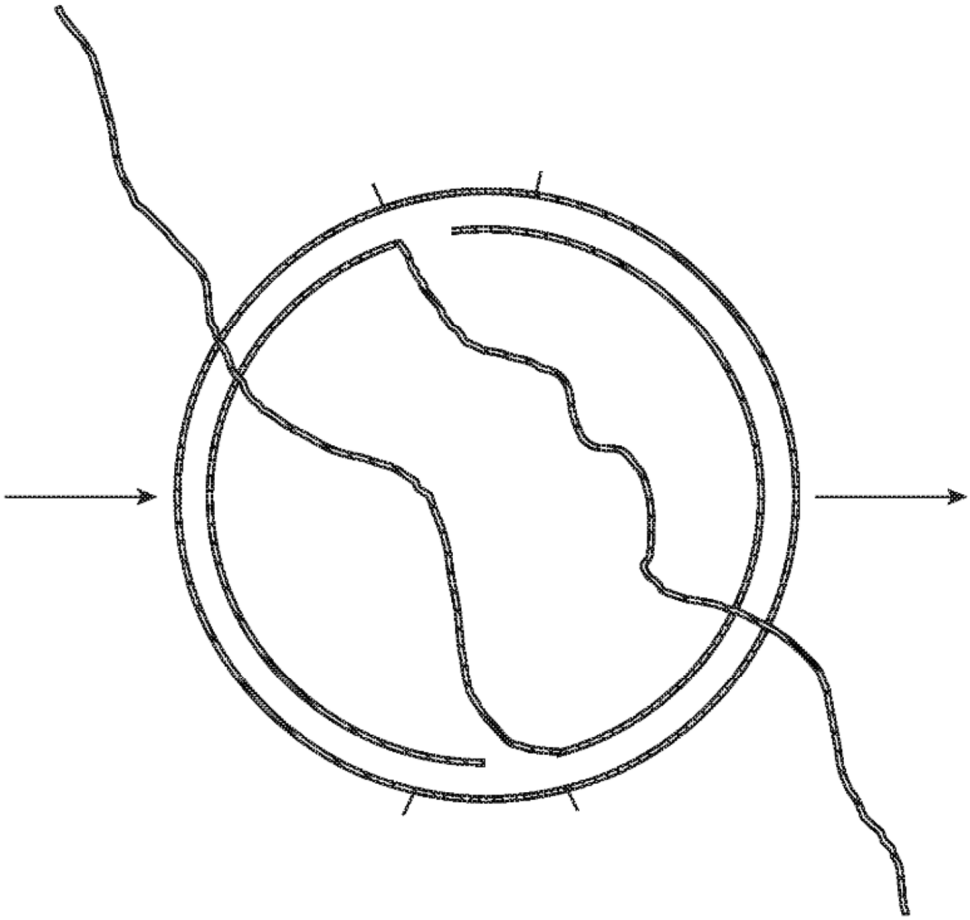
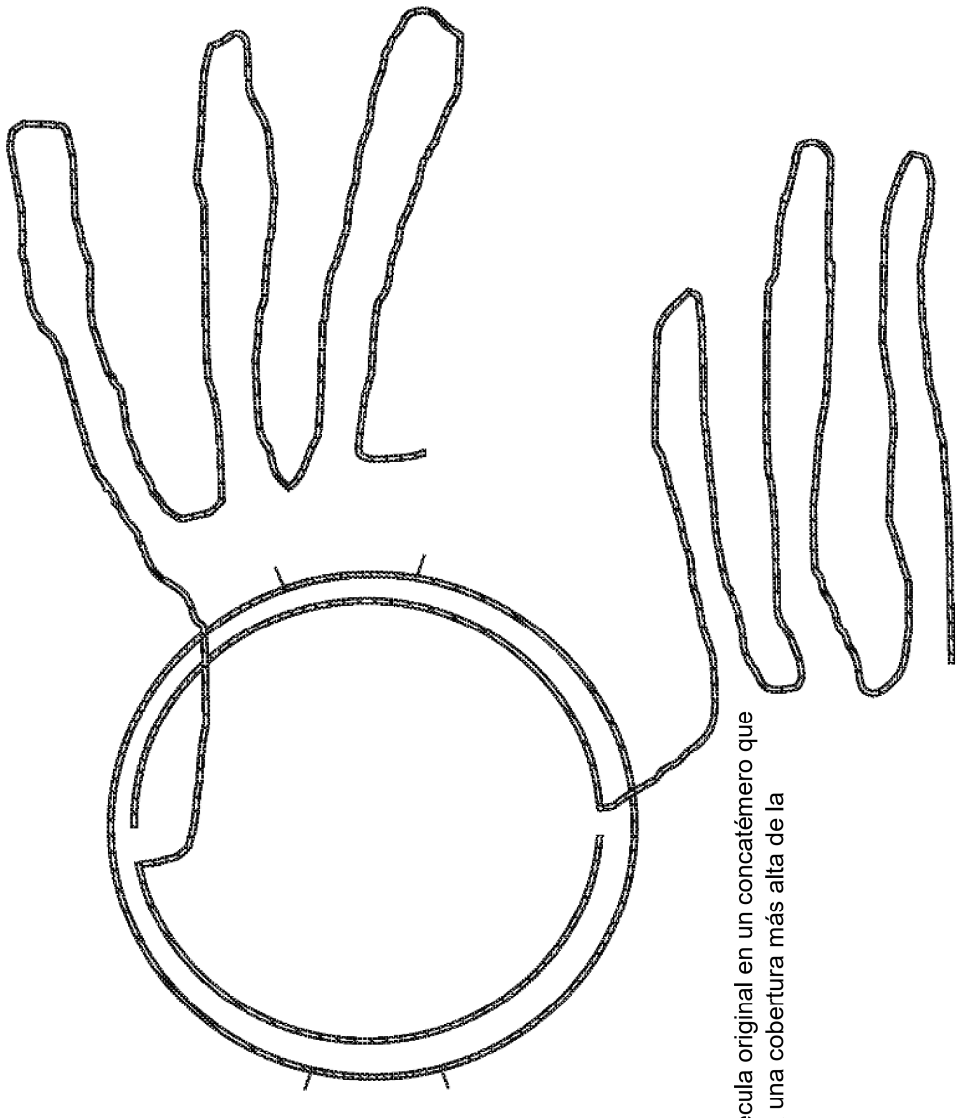


Figura 20 (cont. 1)



Múltiples copias de ADNdh de la molécula original en un concatémero que está listo para la stLFR, esto permitirá una cobertura más alta de la molécula original

Figura 20 (cont. 2)

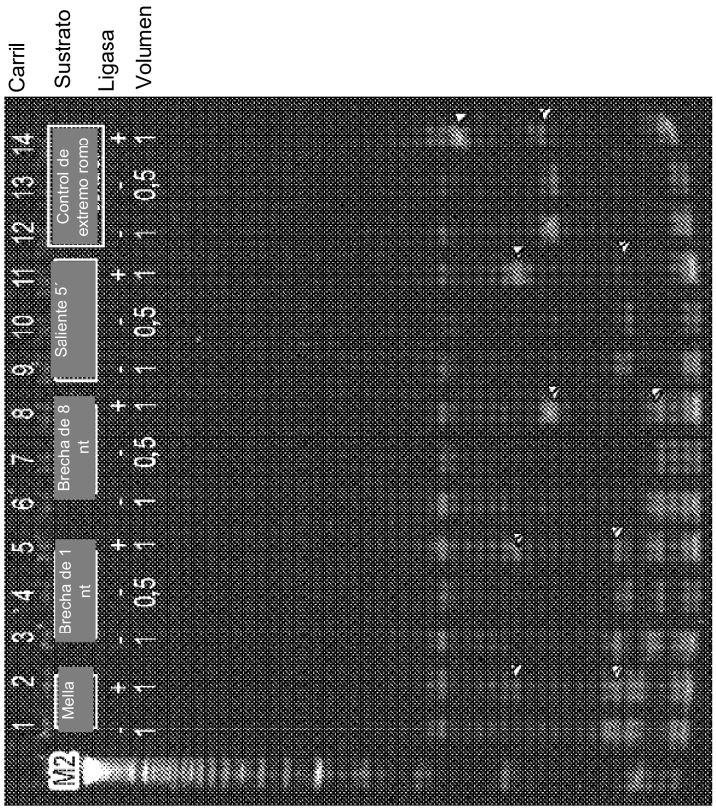


Figura 21B

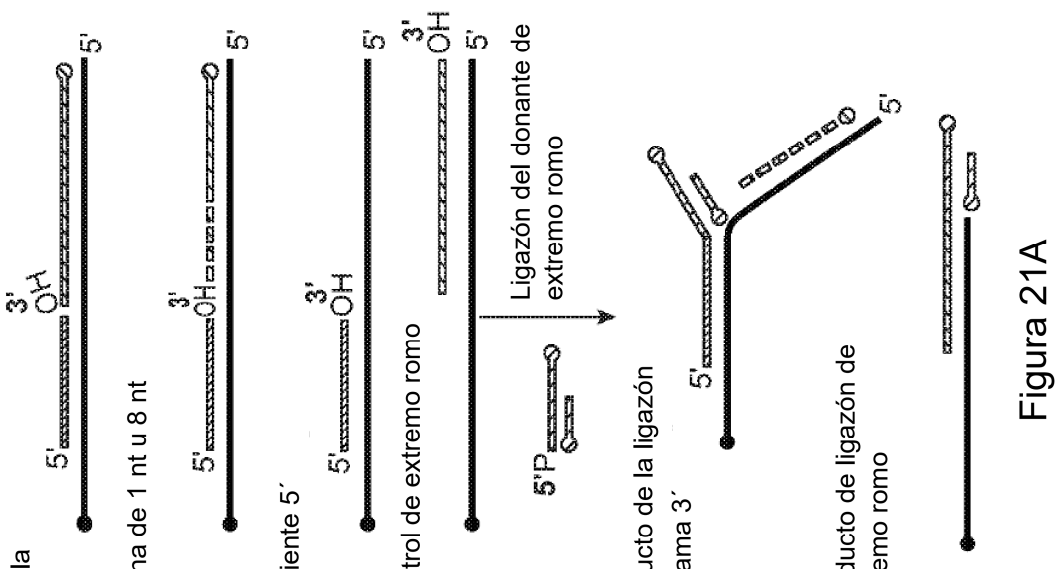


Figura 21A

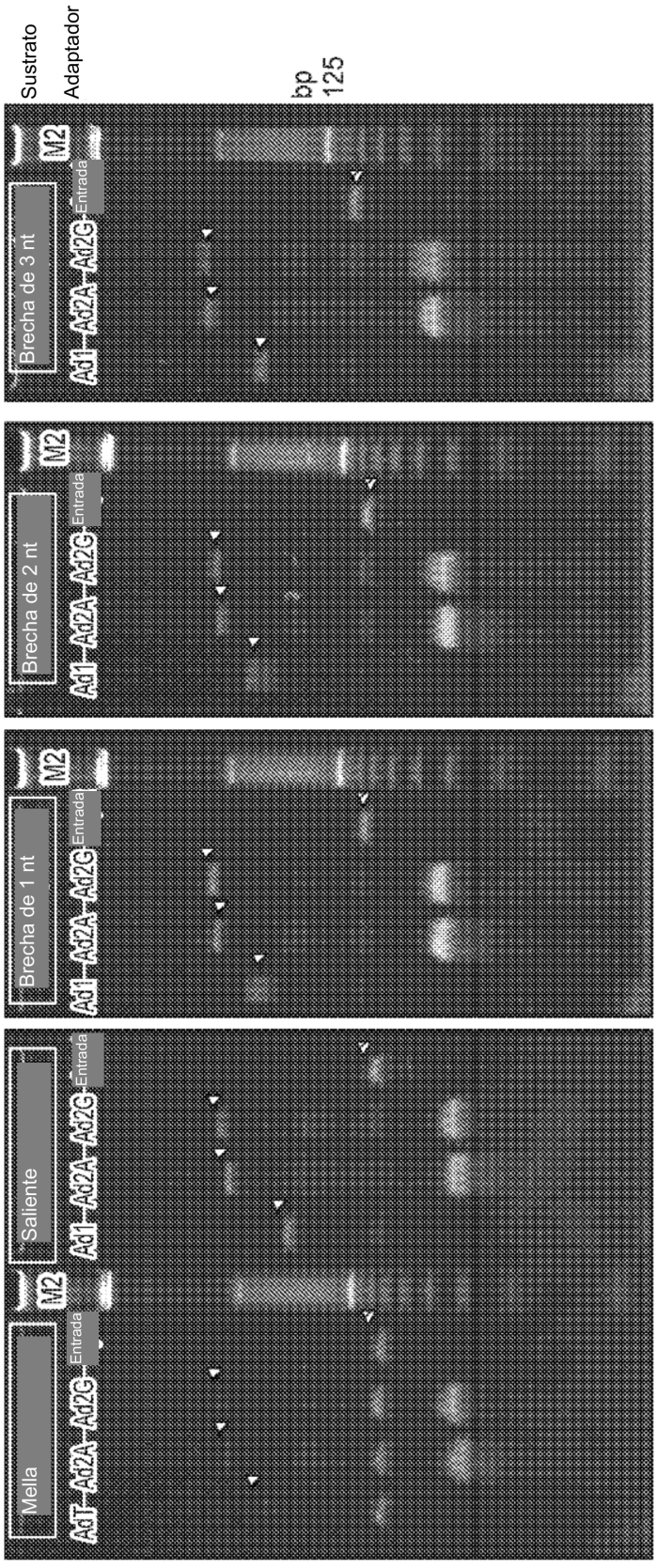


Figura 22A Figura 22B Figura 22C Figura 22D

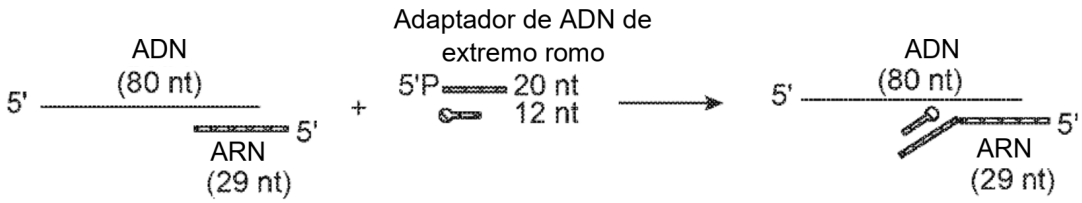


Figura 23A

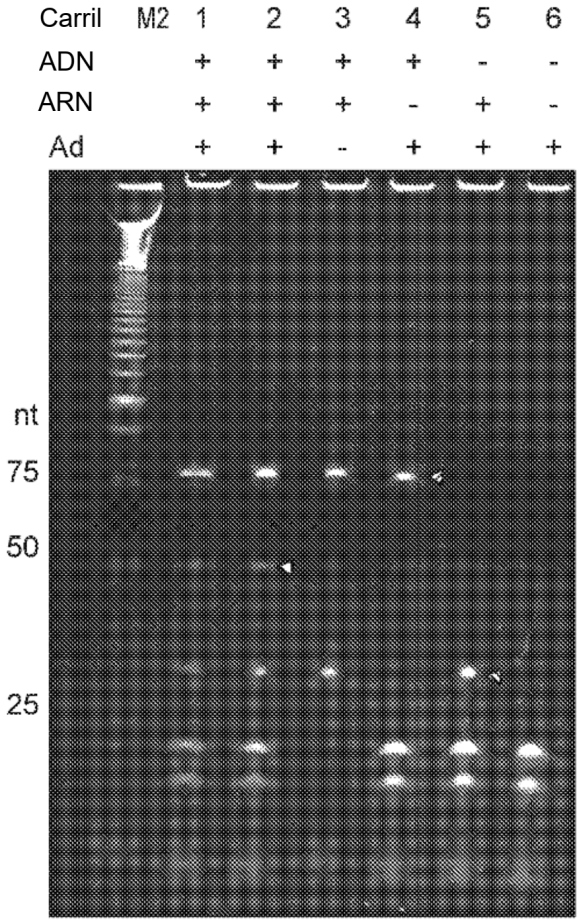


Figura 23B

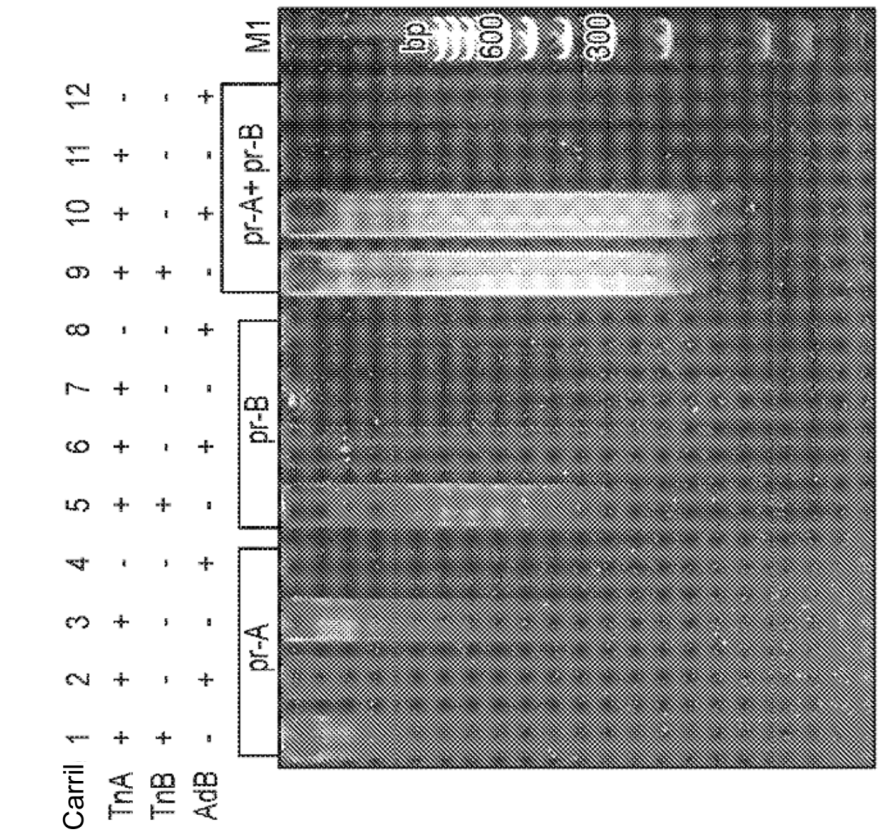


Figura 24B

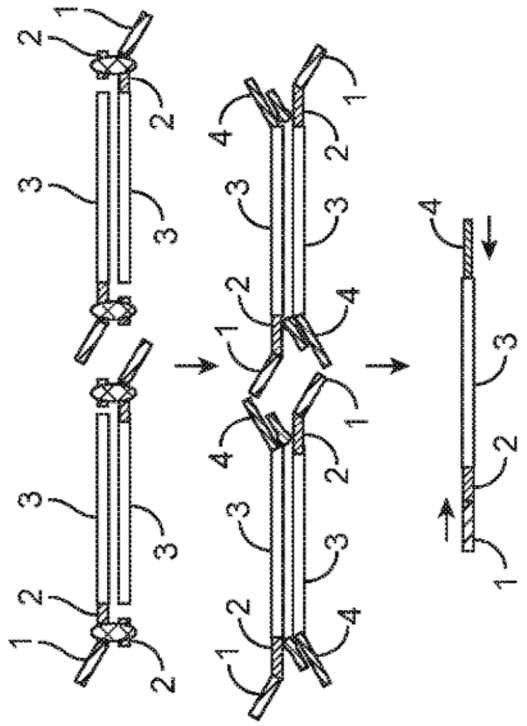


Figura 24A

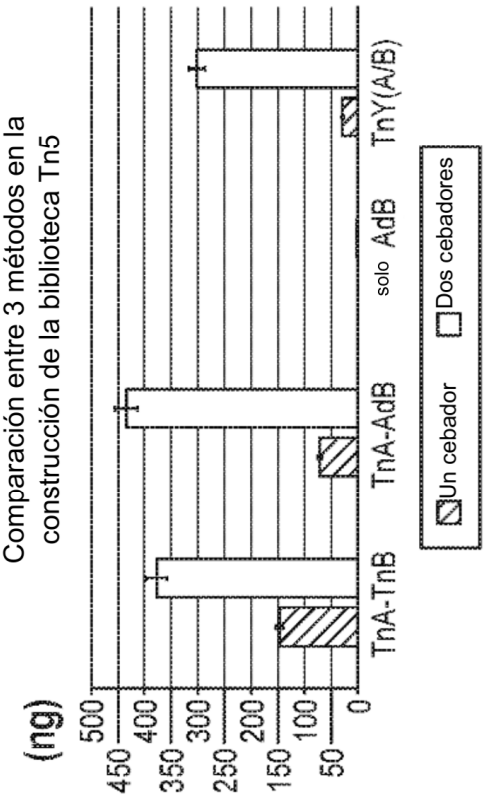


Figura 24C

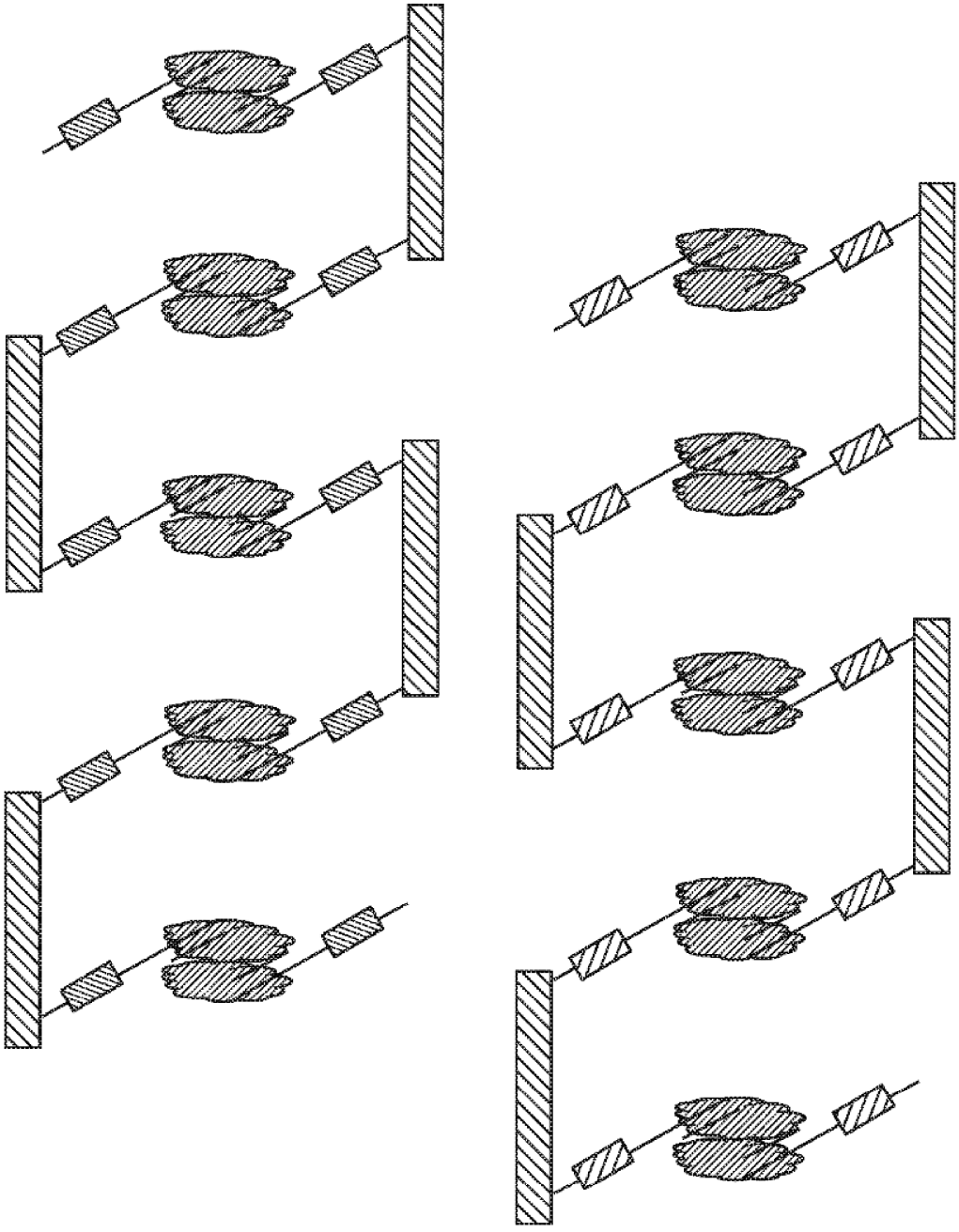


Figura 25

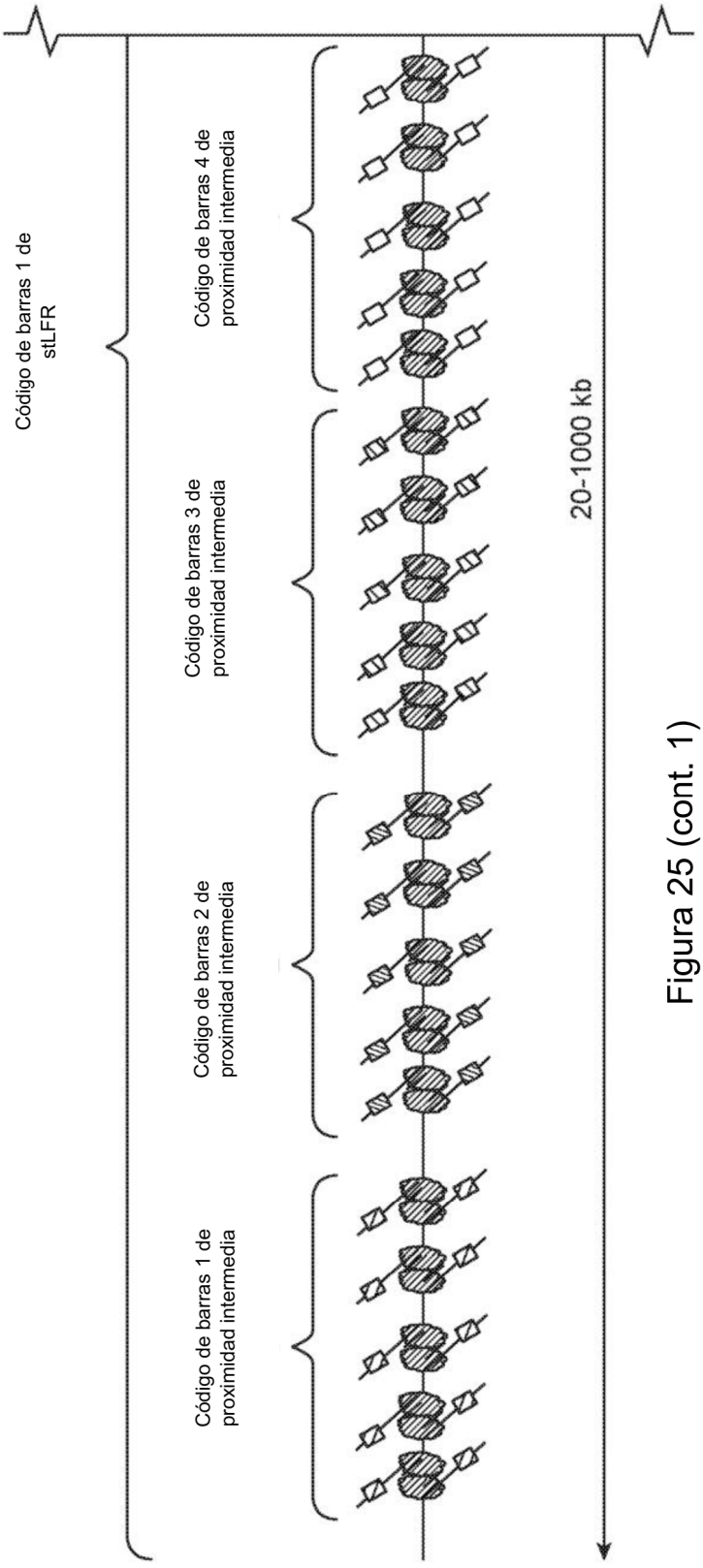


Figura 25 (cont. 1)

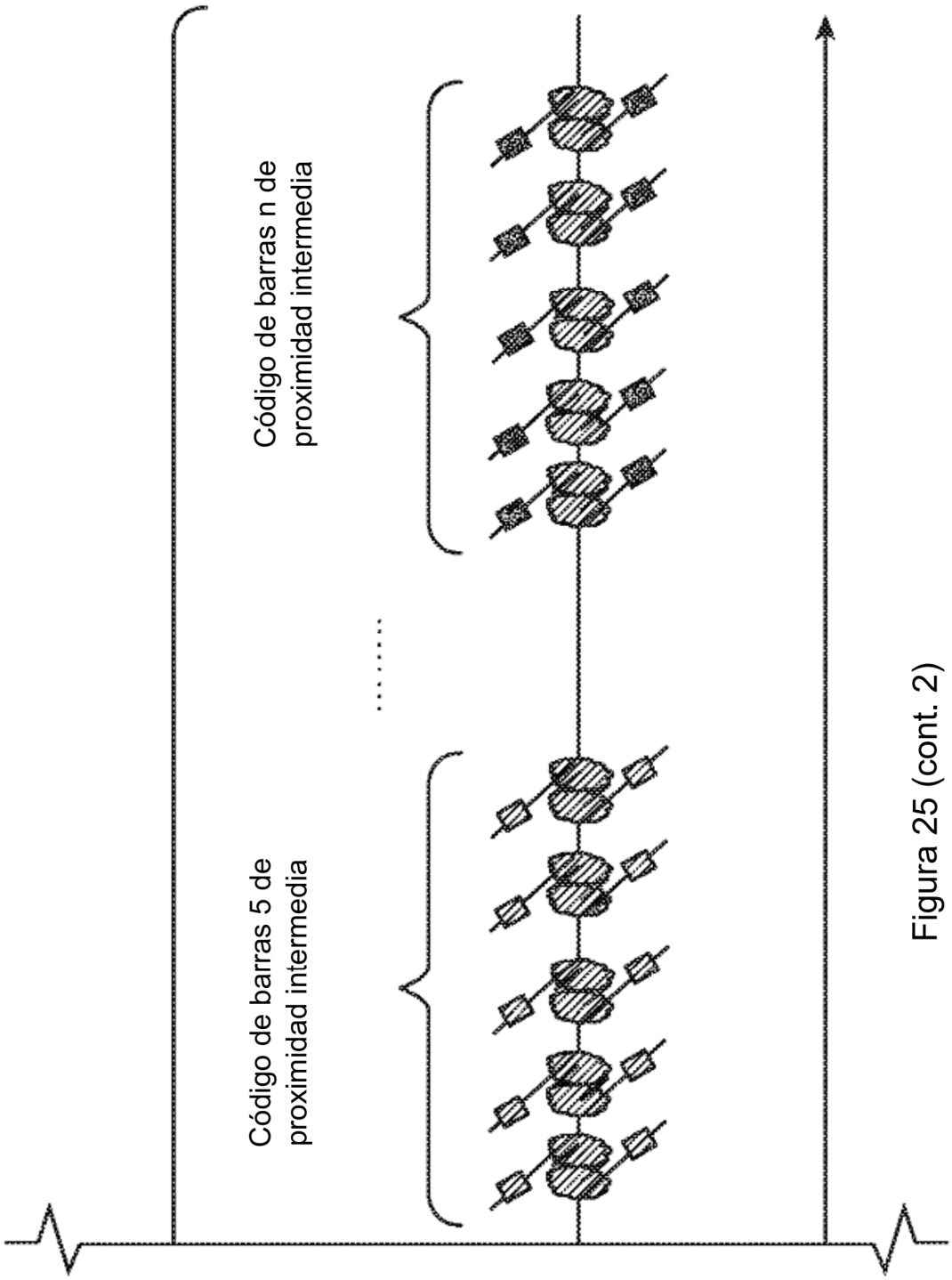


Figura 25 (cont. 2)

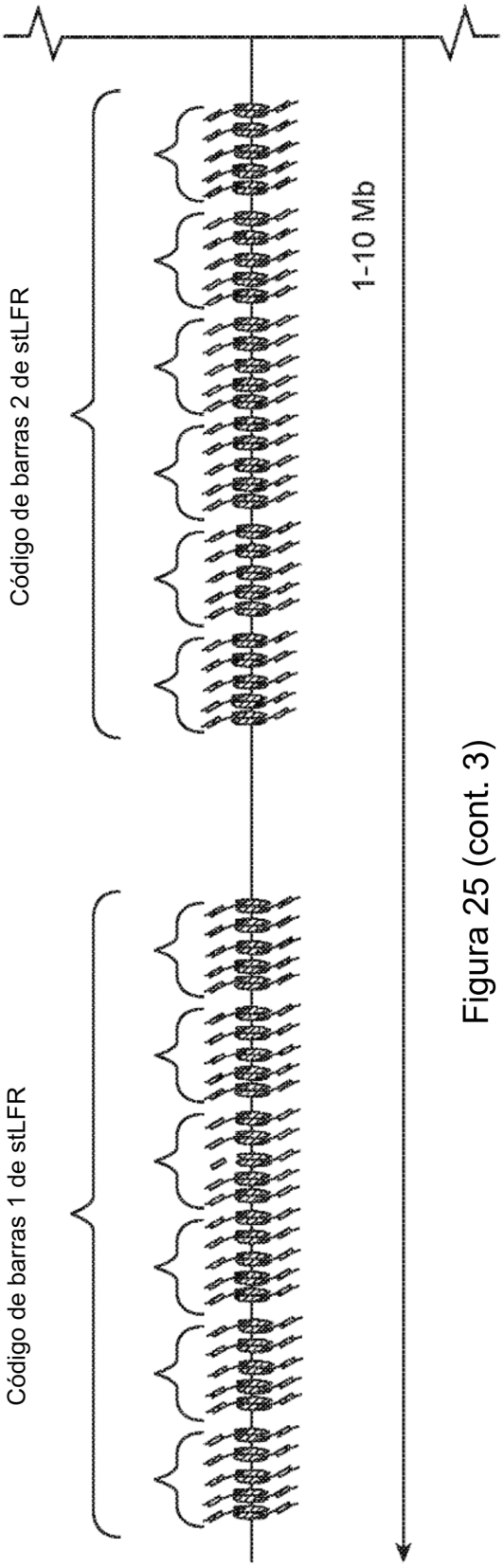


Figura 25 (cont. 3)

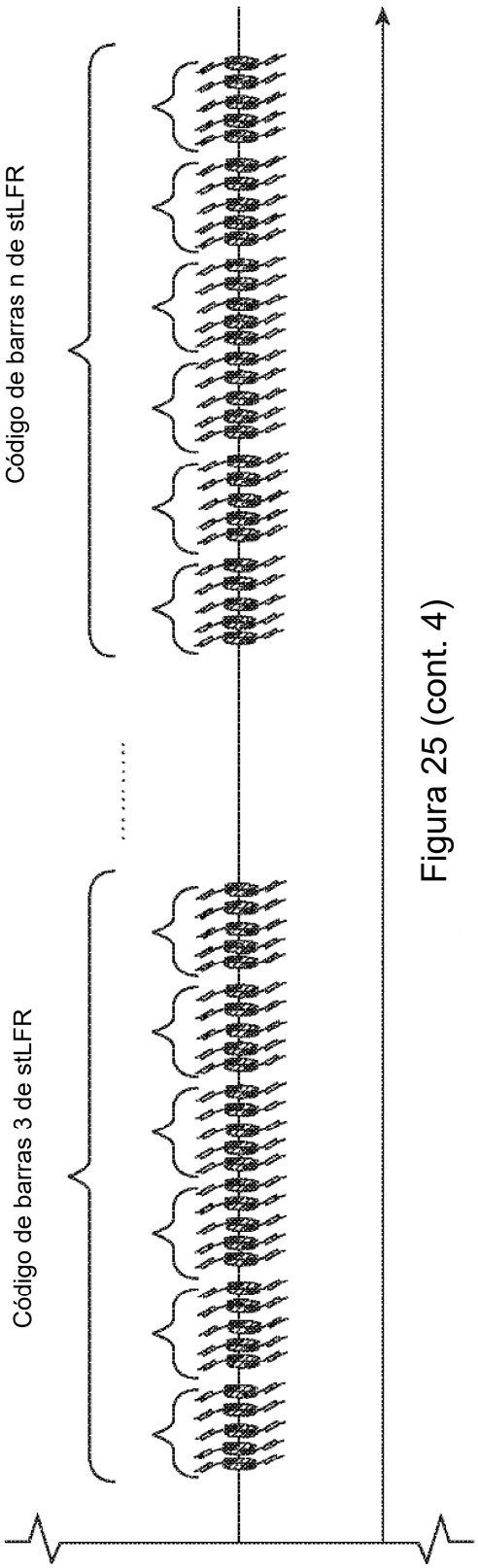


Figura 25 (cont. 4)

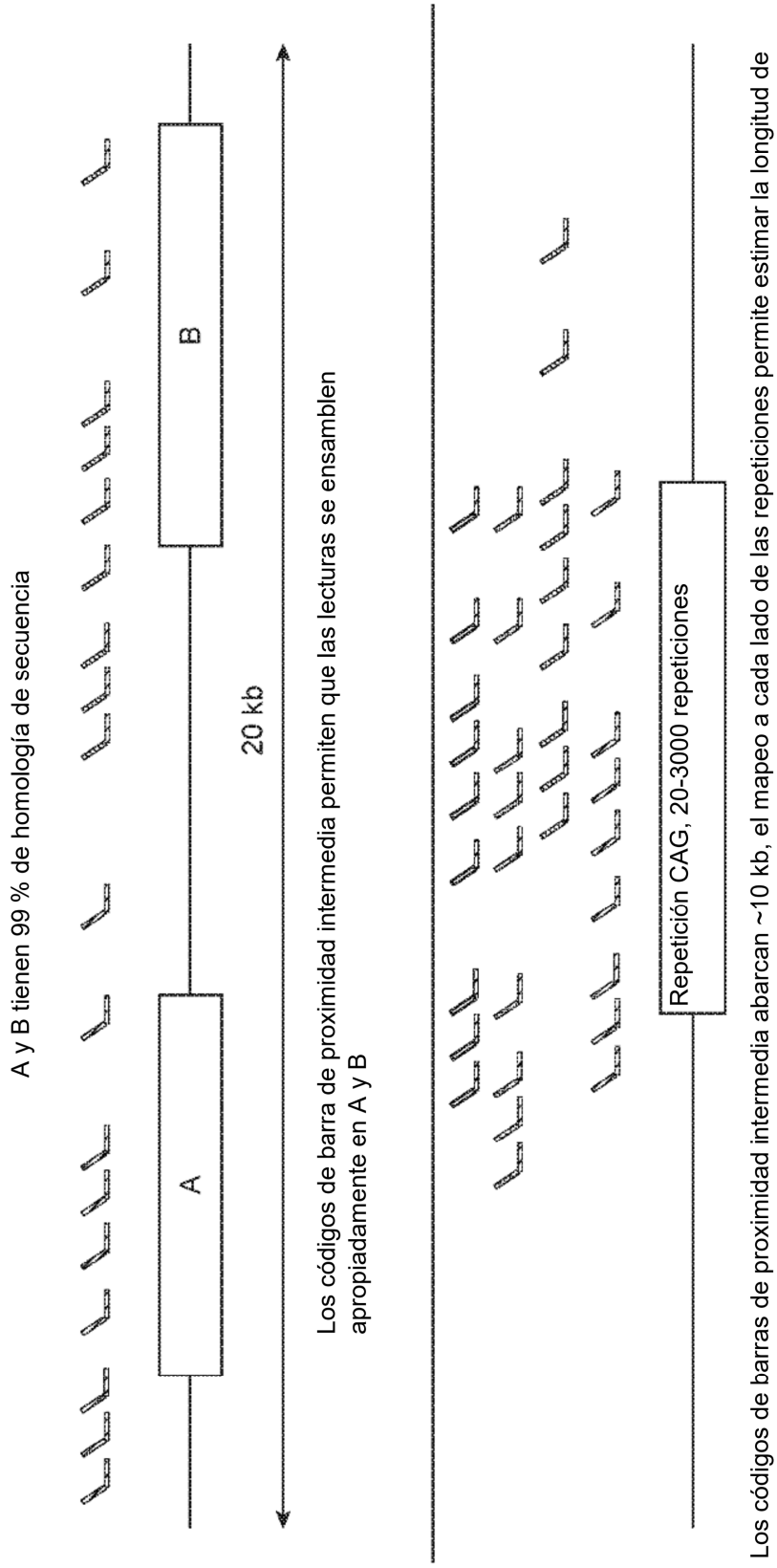


Figura 25 (cont. 5)

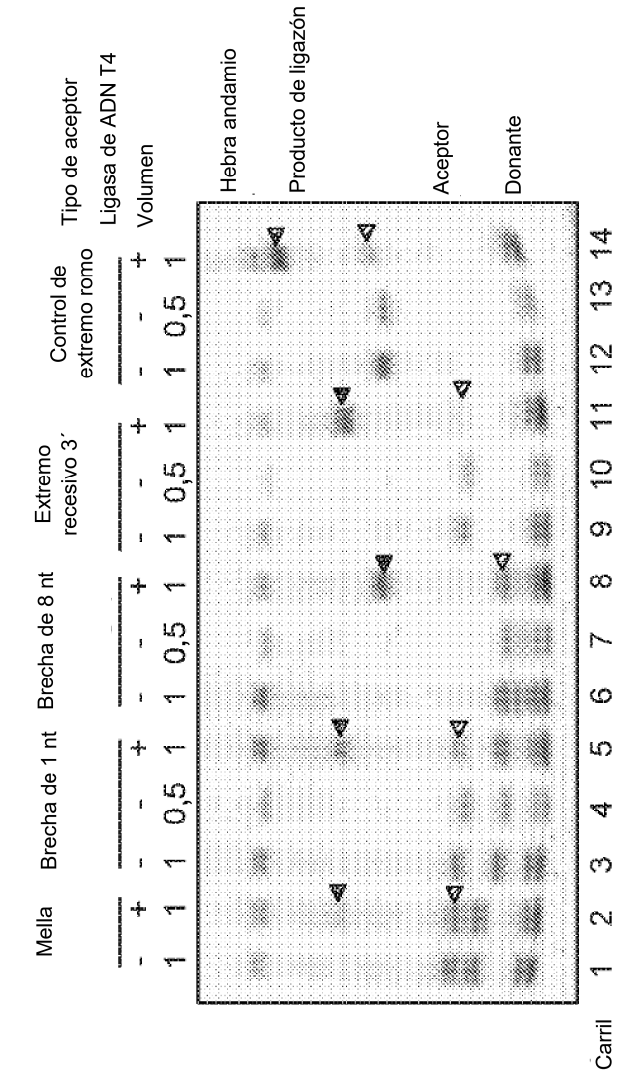


Figura 26B

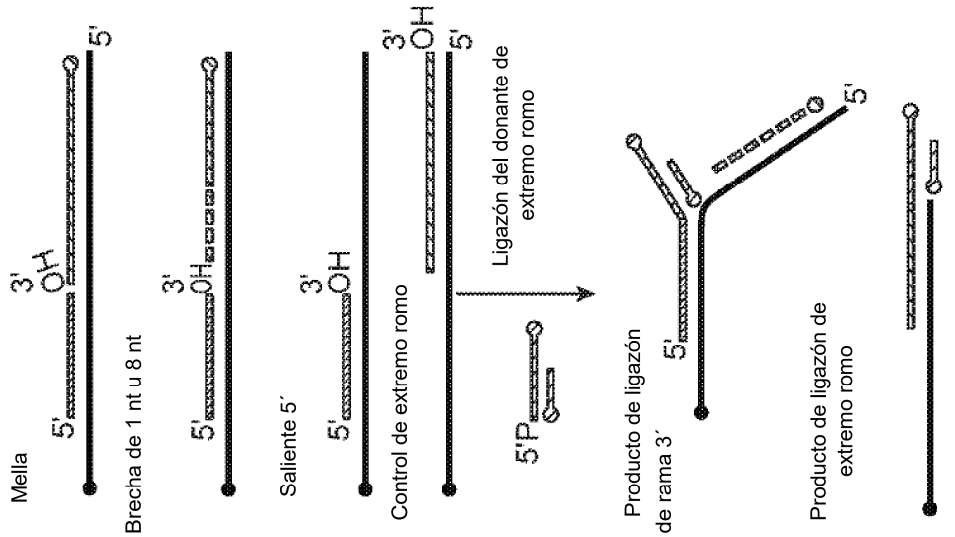


Figura 26A

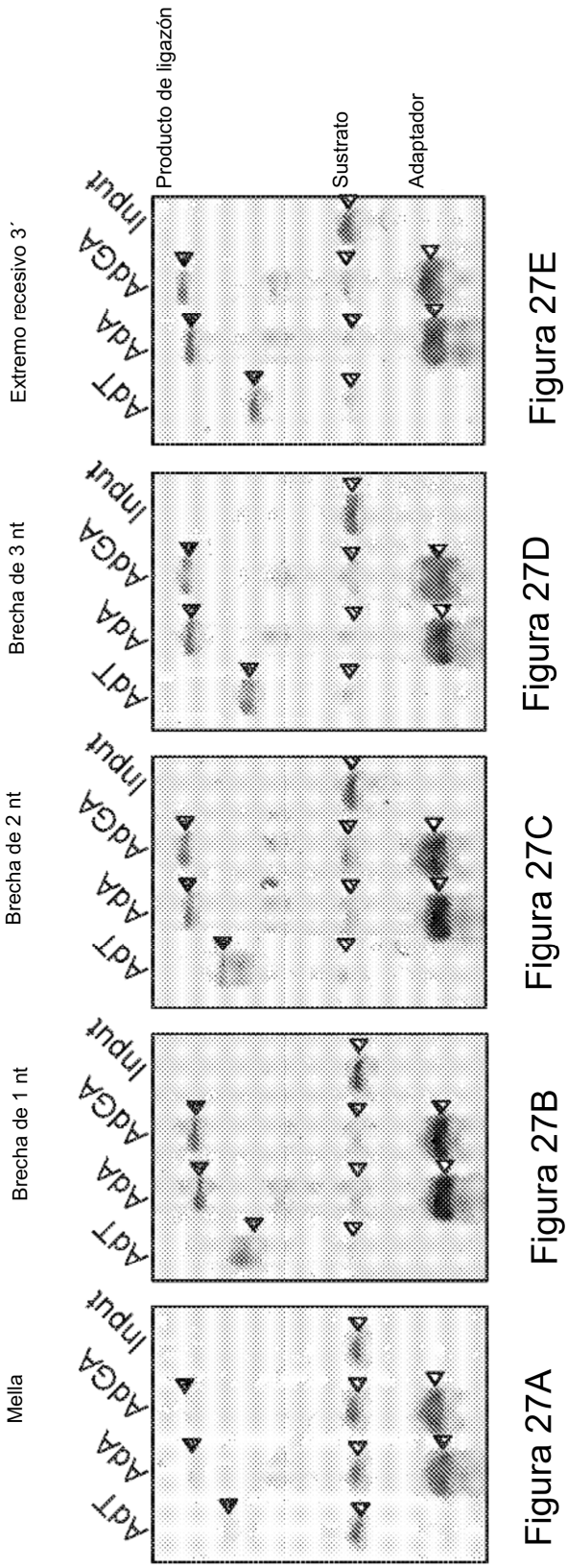


Figura 27A Figura 27B Figura 27C Figura 27D Figura 27E

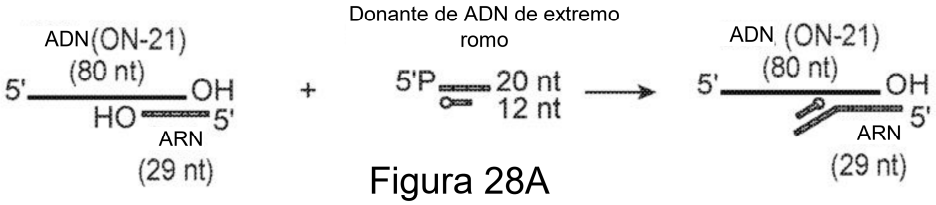


Figura 28A

Figura 28B

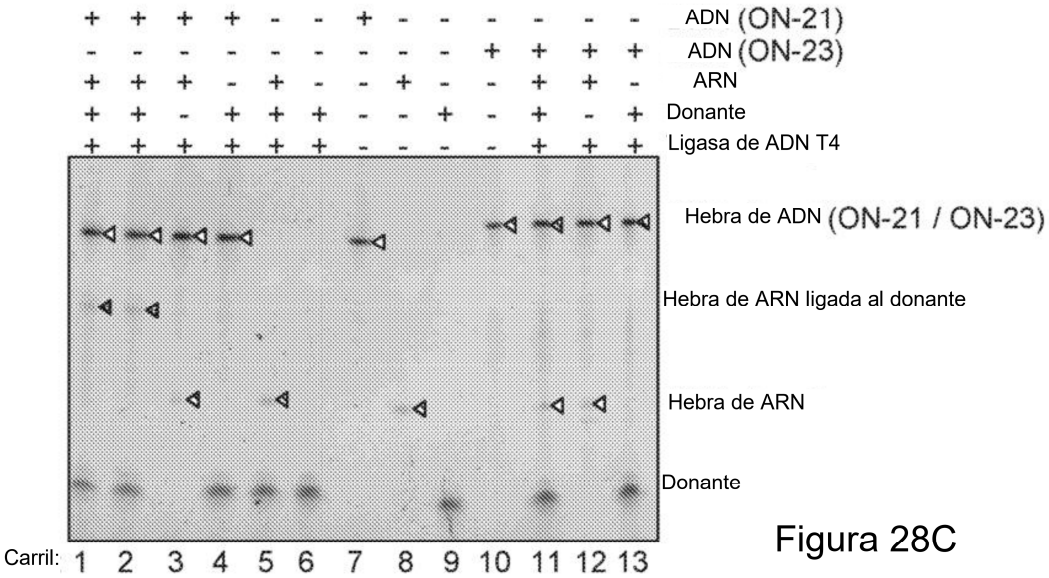


Figura 28C

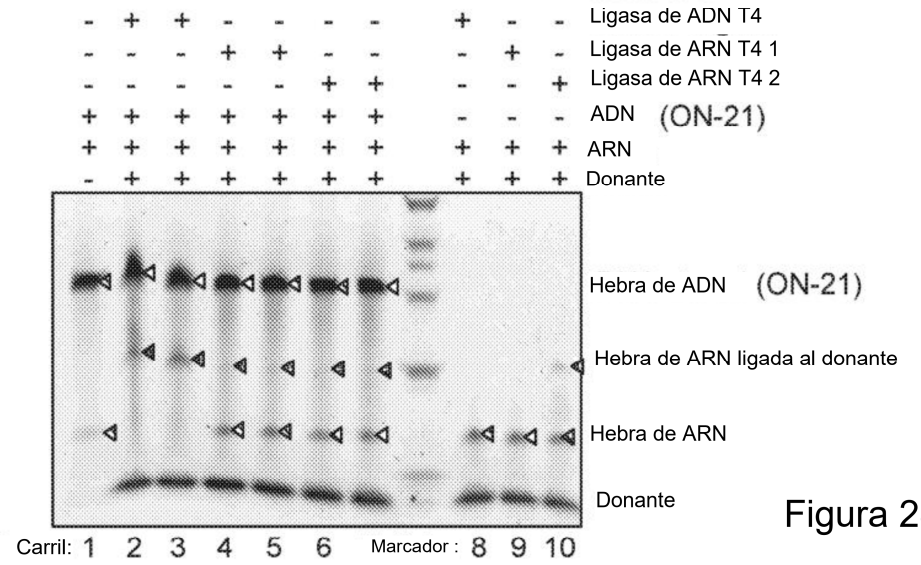
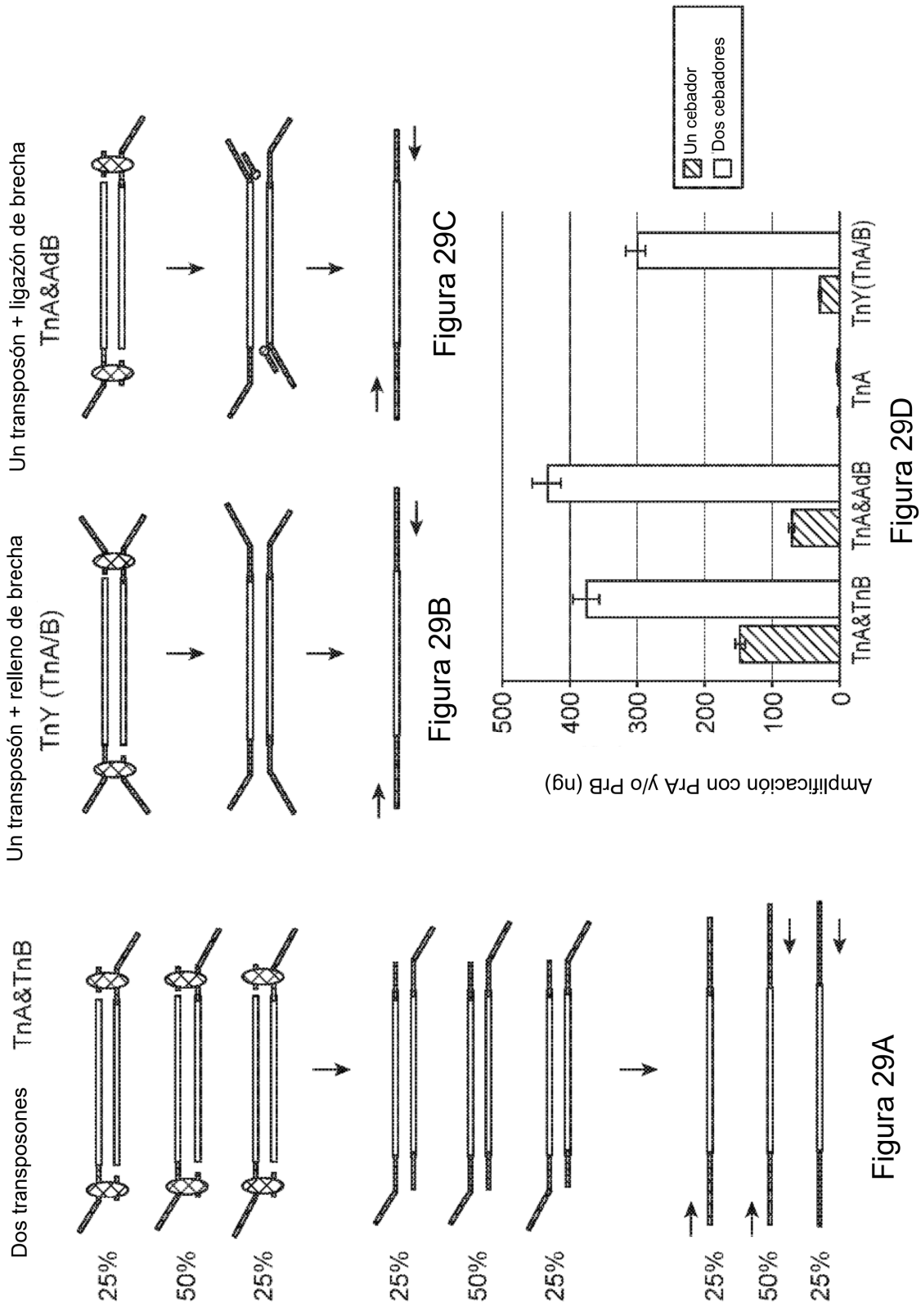


Figura 28D


Dos transposones
TnA&TnB
25%
50%
25%
25%
50%
25%
25%
50%
25%
Figura 29A
Un transposón + relleno de brecha
TnY (TnA/B)
Figura 29B
Un transposón + ligazón de brecha
TnA&AdB
Figura 29C
Amplificación con PrA y/o PrB (ng)
500
400
300
200
100
0
TnA&TnB
TnA&AdB
TnA
TnY(TnA/B)
Un cebador
Dos cebadores
Figura 29D

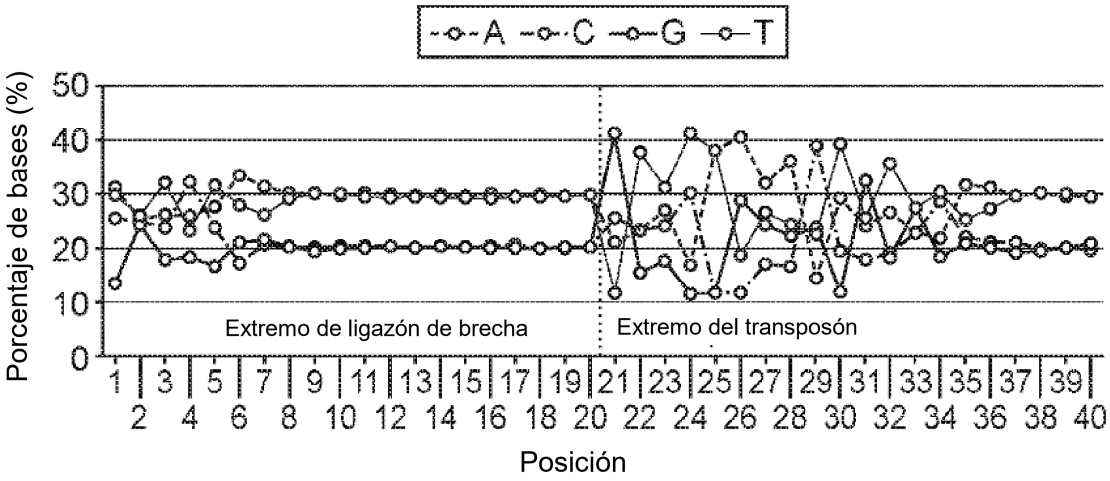


Figura 30A

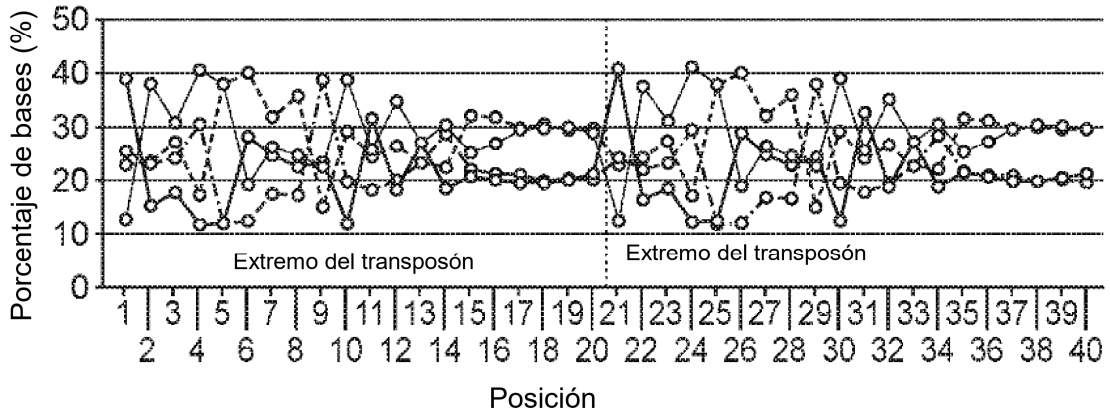


Figura 30B

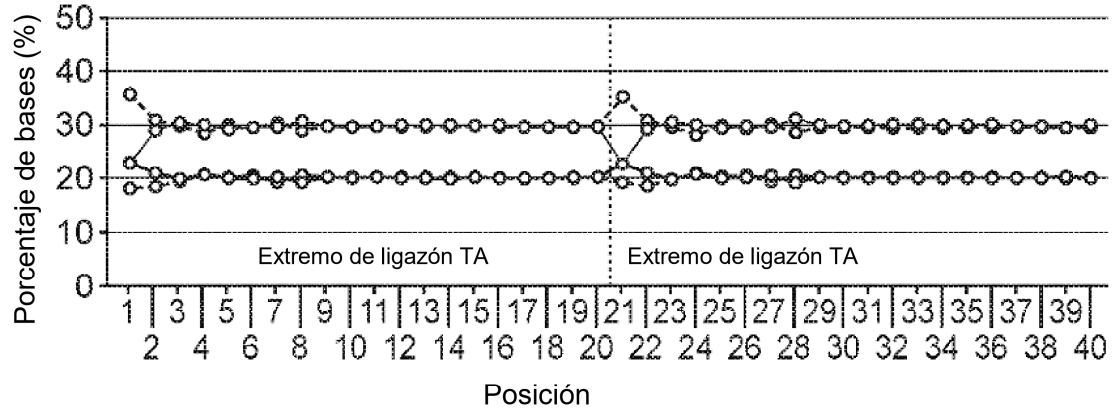


Figura 30C

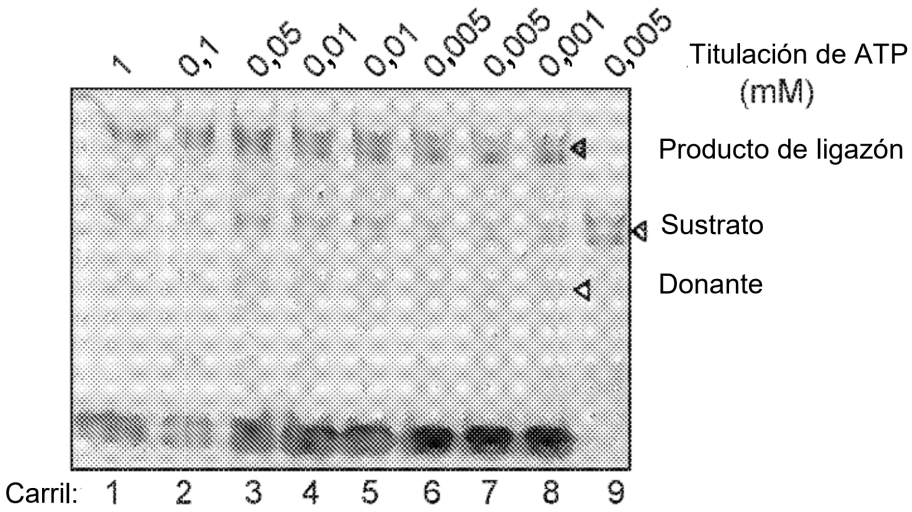


Figura 31A

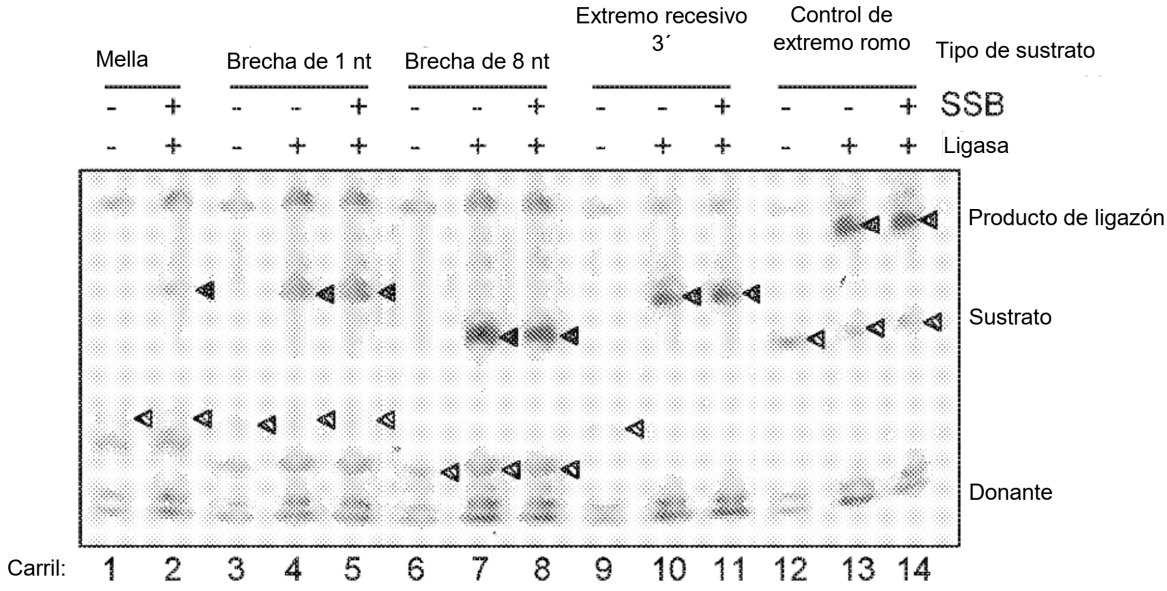


Figura 31B