(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 938 896**

(51) Int. CI.:

C12Q 1/6855 (2008.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **21.10.2019 E 19204418 (8)**

(97) Fecha y número de publicación de la concesión europea: **23.11.2022 EP 3812472**

(54) Título: **Un ensayo in vitro verdaderamente imparcial para perfilar la actividad fuera de objetivo de una o más nucleasas programables específicas de objetivo en células (ABNOBA-SEQ)**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
17.04.2023

(73) Titular/es:

ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG (100.0%)
Friedrichstrasse 39
79098 Freiburg, DE

(72) Inventor/es:

CATHOMEN, TONI;
HAAS, SIMONE ALEXANDRA;
HILDENBEUTEL, MARKUS;
MUSSOLINO, CLAUDIO;
BÖRRIES, MELANIE y
ANDRIEUX, GEOFFROY

(74) Agente/Representante:

GONZÁLEZ PECES, Gustavo Adolfo

Observaciones:

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

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

DESCRIPCIÓN

Un ensayo in vitro verdaderamente imparcial para perfilar la actividad fuera de objetivo de una o más nucleasas programables específicas de objetivo en células (ABNOBA-SEQ)

CAMPO DE LA INVENCIÓN

La presente invención se refiere a un procedimiento para detectar sitios fuera de objetivo de una o más nucleasas programables específicas de objetivo en un genoma *in vitro,* permitiendo la detección de sitios fuera de objetivo en una muestra de entrada de menos de 1 µg de ADN genómico. La presente invención también se refiere a un procedimiento para detectar *in vivo* sitios fuera de objetivo utilizando el procedimiento *in vitro* para detectar sitios fuera de objetivo de una o más nucleasas programables específicas de objetivo en un genoma.

ANTECEDENTES

La edición del genoma con nucleasas del tipo de las repeticiones palindrómicas cortas agrupadas y regularmente espaciadas (CRISPR)/Cas, pero también con nucleasas efectoras similares al activador de la transcripción (TALEN) y nucleasas de dedos de zinc (ZFN), ya se está utilizando terapéuticamente en estudios clínicos. Estas nucleasas programables específicas de objetivos hechas a medida pueden dirigirse a una región específica del genoma, donde introducen una rotura de ADN de doble cadena (ARNdc), que desencadena mecanismos de reparación de ADN celular. Las roturas de ARNdc introducidas se reparan mediante una de las dos principales vías de reparación celular, la unión de extremos no homólogos (NHEJ) o la reparación dirigida por homología (HDR) (Hustedt y Durocher, 2016, Nat Cell Biol 19(1): 1-9).

Mientras que las nucleasas programables específicas de objetivo como ZFN, TALEN y meganucleasas se basan en el reconocimiento proteína-ADN, la especificidad de las nucleasas CRISPR/Cas guiadas por ARN (CRISPR/Cas RGN) depende del reconocimiento ARN-ADN. En la CRISPR/Cas RGN, un ARN guía (ARNg) se hibrida con una secuencia de aproximadamente 20 nucleótidos de longitud (secuencia objetivo) junto a un motivo adyacente al protoespaciador (PAM) y permite a la endonucleasa Cas escindir el sitio objetivo (Wiedenheft et al., 2012, Nature 482, 331-338; Terns et al., 2011, Curr Opin Microbiol 14, 321-327). El cambio de la secuencia de nucleótidos del extremo 5' del ARNg modifica el sitio al que se dirige la CRISPR/Cas RGN. En su forma activa, la CRISPR/Cas RGN es una ribonucleoproteína (RNP), en la que la endonucleasa Cas está unida a un ARN guía (ARNg). Las endonucleasas Cas más utilizadas incluyen Cas9, Cpf1 y variantes de las mismas. Por ejemplo, se han utilizado diferentes endonucleasas Cas9 derivadas de *Streptococcus pyogenes, Neisseria meningitidis, Staphylococcus aureus, Streptococcus thermophilus* y *Treponema denticola* para desarrollar herramientas de edición del genoma.

Además de la secuencia objetivo real (sitio dentro del objetivo), las nucleasas programables específicas de objetivo pueden ser activas en sitios del genoma que difieren en secuencia hasta cierto punto de la secuencia objetivo real, los llamados sitios fuera del objetivo, generando así mutaciones de rotura de ARNdc en sitios fuera del objetivo que difieren hasta en varios nucleótidos de los sitios dentro del objetivo. Por ejemplo, se ha demostrado que las nucleasas CRISPR-Cas9 pueden tolerar hasta seis desajustes entre el ARNg y el ADN objetivo, especialmente en la región distal PAM del sitio objetivo (por ejemplo, Tsai et al., 2015, Nat Biotechnol 33(2): 187-197).

La actividad fuera del objetivo en células humanas puede tener consecuencias potencialmente peligrosas. La generación de roturas de ARNdc en sitios del genoma fuera del objetivo puede producir inserciones y deleciones (indels), así como translocaciones que deben vigilarse cuidadosamente. Si un sitio fuera del objetivo se encuentra, por ejemplo, en una región codificante o en un elemento regulador, en el peor de los casos se puede inducir una degeneración cancerígena de la célula. Por lo tanto, la actividad fuera del objetivo de las nucleasas programables específicas del objetivo debe determinarse y, si es necesario, la especificidad de las nucleasas programables específicas del objetivo utilizadas debe optimizarse antes de que las herramientas de edición del genoma puedan usarse en entornos clínicos. Así pues, la comprensión y prevención de los efectos fuera de objetivo es fundamental en el contexto del uso terapéutico de nucleasas programables específicas de objetivo.

Es importante destacar que las diferencias individuales en el genoma, tal como los polimorfismos de un solo nucleótido (SNP), así como el estado de la cromatina específico del tipo de célula, pueden cambiar los perfiles dentro y fuera del objetivo de cualquier nucleasa programable específica de objetivo (Lessard et al., 2017, Proc Natl Acad Sci U S A 114(52): E11257-E11266, Verkuijl y Rots, 2019, Curr Opin Biotechnol 55: 68-73.). Esto subraya aún más la necesidad de identificar los posibles sitios fuera del objetivo de forma específica e individualizada para cada tipo celular antes de la aplicación terapéutica de una nucleasa programable específica.

Se han desarrollado diferentes procedimientos, con sus correspondientes ventajas e inconvenientes, para determinar la actividad fuera del objetivo de nucleasas programables específicas de objetivo (resumidos en Kim et al., 2019, Annu Rev Biochem 88: 191-220). Por un lado, los sitios potencialmente fuera del objetivo pueden predecirse *in silico* mediante algoritmos bioinformáticos basándose en la similitud de secuencia de dichos sitios potencialmente fuera del objetivo con la secuencia objetivo real y, a continuación, pueden investigarse mediante secuenciación de amplicones objetivo, por ejemplo, mediante secuenciación de próxima generación (NGS). Los efectos de un solo

desajuste con respecto a la secuencia dentro del objetivo no siempre son predecibles. Además, el contexto genómico o epigenómico, o ambos, también podrían afectar a la frecuencia de escisión. Estos factores dificultan enormemente el desarrollo de un algoritmo capaz de identificar posibles sitios fuera del objetivo. Estos procedimientos *in silico* son rápidos y relativamente baratos. Sin embargo, estos procedimientos *in silico* suelen tener una baja tasa de verdaderos positivos (baja sensibilidad). No se detecta la actividad fuera del objetivo en los sitios pasados por alto por dichos algoritmos.

Como alternativa, la actividad fuera de objetivo puede determinarse experimentalmente, lo que significa que la detección no se inicia mediante predicciones asistidas por ordenador. Los procedimientos experimentales requieren más mano de obra y son más caros. También están sujetos a limitaciones técnicas y bioinformáticas que deben superarse. Se conocen diferentes enfoques experimentales para determinar la especificidad, por ejemplo, GUIDE-Seq, DISCOVER-Seq, Digenome-Seq, CIRCLE-Seq y SITE-Seq. A grandes rasgos, pueden dividirse en procedimientos basados en células y procedimientos *in vitro*.

El documento US 2018/0016572 divulga una versión optimizada de GUIDE-seq y representa Target-EnrichedGUIDE-seq que también se designa abreviadamente como TEG-seq. El procedimiento se basa en la detección y el mapeo de etiquetas de ADN cortas de doble cadena que se integran en las roturas de doble cadena de ADN generadas por una nucleasa CRISPR-Cas, por lo que el procedimiento está basado en células que detectan la actividad fuera del objetivo en líneas celulares.

Wienert et al. "Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq", Science (2019), 364, pp 286-289 divulgan el denominado DISCOVER-Seq, que es un procedimiento para detectar la actividad fuera de objetivo de las nucleasas CRISPR-Cas. Es un procedimiento que utiliza líneas celulares tal como las células humanas K562 o las células murinas B16-F10. El procedimiento se basa en la detección de la actividad fuera del objetivo en las células mediante ChIP-Seq. Presenta una especificidad y una sensibilidad bajas. Dado que el procedimiento requiere, por ejemplo, de 2 a 10 millones de células como entrada experimental, este procedimiento es poco adecuado para un análisis individualizado de la especificidad de nucleasas programables específicas de objetivo, en particular de células con un tamaño de muestra limitado y una disponibilidad limitada.

El documento WO 2016/081798 divulga procedimientos relacionados con la detección de roturas de doble cadena recurrentes y no específicas en el genoma. El procedimiento se basa en células y debe detectar translocaciones. Este procedimiento no es muy sensible, ya que requiere 20-100 µg de ADN genómico de entrada, lo que corresponde aproximadamente a un equivalente de 3-15 millones de células.

Los ensayos basados en células GUIDE-seq (Tsai et al., 2015, Nat Biotechnol 33(2): 187-197) y DISCOVER-seq (Wienert et al., 2019, Science 364(6437): 286-289) en células vivas aprovechando los mecanismos de reparación celular desencadenados por la actividad nucleasa programable específica del objetivo. GUIDE-Seq es un procedimiento basado en células en el que se introducen en la célula oligodesoxinucleótidos cortos de doble cadena (ODNds) además de una nucleasa programable específica del objetivo. Si la nucleasa escinde el genoma, los ODNds cortos se integran en el ADN en el lugar de las roturas de doble cadena, que luego sirven como punto de partida para la secuenciación de alto rendimiento. Este procedimiento funciona bien, pero sólo en determinadas líneas celulares que pueden diferir de células procedentes de distintos individuos en sitios clave del genoma.

DISCOVER-Seq es otro procedimiento celular basado en el procedimiento ChIP-seq. Si el material genético es cortado por nucleasas programables específicas, se activan los mecanismos de reparación celular y las proteínas de reparación del ADN, tal como la MRE11, se dirigen a la rotura de la doble cadena de ADN. A continuación, el ADN se reticula/fija con las proteínas unidas a él. El "ADN cromatina" resultante se fragmenta y todos los fragmentos de ADN con proteína MRE11 unida se precipitan utilizando un anticuerpo dirigido contra MRE11. Por último, mediante NGS de los fragmentos de ADN aislados, se pueden identificar los sitios con actividad fuera del objetivo. Este procedimiento mide las roturas reales de ARNdc en la célula, pero también las no inducidas por nucleasas programables específicas del objetivo. Además, este procedimiento sólo puede detectar las roturas de doble cadena del ADN que aún no se hayan reparado en el momento de la fijación celular. Por lo tanto, cabe suponer que tanto el número de falsos positivos como el reducido número de resultados correctamente positivos limitan la aplicabilidad de este procedimiento.

En general, estos procedimientos basados en células funcionan muy bien en determinadas líneas celulares. Sin embargo, debido a la toxicidad observada de ODNds, por ejemplo, en células primarias, GUIDE-Seq no puede transferirse fácilmente o no puede transferirse en absoluto a determinadas células, mientras que DISCOVER-Seq mostrará una especificidad y sensibilidad bajas. Por lo tanto, estos procedimientos son poco adecuados para un análisis individualizado de la especificidad de las nucleasas programables específicas de objetivo, en particular de células con un tamaño de muestra y una disponibilidad limitados.

Los procedimientos *in vitro* Digenome-Seq (Kim et al, 2015, Nat Methods 12(3): 237-243), CIRCLE-Seq (Tsai et al., 2017, Nat Methods 14(6): 607-614) y SITE-Seq (Cameron et al., 2017, Nat Methods 14(6): 600-606) descubren sitios fuera del objetivo marcando los sitios que son escindidos en ADN genómico aislado por CRISPR-Cas RGN *in vitro*.

En Digenome-Seq, el ADN genómico se escinde en el tubo de ensayo utilizando nucleasas programables específicas

del objetivo. Los adaptadores de ADN se ligan a los puntos de fractura, a los que se pueden unir los cebadores, que se utilizan para la secuenciación de alto rendimiento. Una de las principales desventajas de este enfoque es que también debe secuenciarse un genoma de referencia y que, debido a la falta de un paso de enriquecimiento, la secuenciación sólo puede llevarse a cabo en costosos secuenciadores de alto rendimiento (por ejemplo, HiSeq), lo que hace que este enfoque sea muy caro. Además, hay que utilizar cantidades relativamente grandes de ADN genómico de entrada. Este procedimiento corresponde al objeto de la solicitud de patente europea publicada como EP 3 219 810 A1.

En CIRCLE-Seq, el ADN genómico se fragmenta primero por ultrasonidos y las piezas individuales se ligan después en moléculas con forma de anillo. En un siguiente paso, el ADN circular se corta con CRISPR-Cas RGN, dando lugar a moléculas linealizadas. A los extremos de estas moléculas linealizadas se ligan adaptadores, que pueden utilizarse para secuenciar e identificar los sitios fuera del objetivo. La principal desventaja de este procedimiento es que hay que utilizar grandes cantidades de ADN genómico de entrada. Este procedimiento corresponde al objeto de la patente de EE. UU con el no. de patente US 9,850,484.

En SITE-Seq, el ADN genómico se escinde en el tubo de ensayo utilizando nucleasas programables específicas de objetivo. Los adaptadores biotinilados se ligan a los sitios escindidos, a través de los cuales se pueden enriquecer las moléculas escindidas tras la fragmentación enzimática. A continuación, se ligan adaptadores en sus extremos, que pueden utilizarse para la secuenciación y la identificación de los fuera del objetivo. Además, este procedimiento requiere cantidades relativamente grandes de ADN genómico de entrada.

Una de las principales desventajas de estos procedimientos *in vitro* es que deben utilizarse cantidades relativamente grandes de ADN genómico de entrada (aproximadamente 8 a 25 µg, dependiendo del procedimiento), lo que dificulta o incluso imposibilita el análisis fuera del objetivo en muestras celulares valiosas de tamaño y disponibilidad de muestra limitados. Además, los procedimientos *in vitro* se caracterizan generalmente por una alta tasa de falsos positivos (baja especificidad), es decir, muchos de los sitios fuera del objetivo cortados *in vitro* no se cortan en la célula, por ejemplo, porque la concentración de nucleasas programables específicas de objetivo en la célula es menor o porque estos sitios no son accesibles debido a la configuración de la cromatina (Kim et al., 2019, Annu Rev Biochem 88: 191-220).

En resumen, el uso de los procedimientos anteriores de forma individualizada en células con un tamaño de muestra limitado está limitado principalmente debido a su requerimiento de una gran cantidad de ADN de entrada de los ensayos *in vitro* actuales o al hecho de que los procedimientos basados en células no son fácilmente transferibles a ciertas células, por ejemplo, células primarias.

Existe la necesidad de un procedimiento para detectar sitios fuera de objetivo de una o más nucleasas programables específicas de objetivo en un genoma que permita la detección de forma individualizada de sitios fuera del objetivo en el ADN genómico de muestras celulares valiosas con un tamaño de muestra y una disponibilidad limitados.

SUMARIO DE LA INVENCIÓN

Los presentes inventores han encontrado un procedimiento para detectar sitios fuera del objetivo de una o más nucleasas programables específicas de objetivo en un genoma *in vitro* (Abnoba-Seq), que permite la detección de sitios fuera del objetivo en el campo de la edición del genoma en muestras celulares valiosas de tamaño de muestra limitado. A diferencia de los procedimientos *in vitro* establecidos como Digenome-Seq, CIRCLE-Seq y SITE-Seq, el procedimiento inventivo sólo requiere cantidades mínimas de ADN genómico de entrada (< 1 µg), lo que permite probar una o más nucleasas programables específicas de objetivo en muestras celulares muy pequeñas de suministro limitado, por ejemplo, en el ADN genómico derivado de células primarias, material de biopsia o un injerto editado genéticamente. Además, el procedimiento inventivo puede llevarse a cabo con secuenciadores NGS básicos (de bajo coste), sin necesidad de un costoso secuenciador de alto rendimiento como los sistemas de secuenciación HiSeq. Por último, el procedimiento puede identificar sitios fuera del objetivo con una sensibilidad comparable a la de los procedimientos *in vitro* ya establecidos, pero con una tasa de falsos positivos inferior, lo que significa con una especificidad significativamente mejor que la de procedimientos comparables.

En un segundo aspecto, la presente invención, además, proporciona un procedimiento para detectar *in vivo* sitios fuera de objetivo utilizando el procedimiento para detectar sitios fuera del objetivo de una o más nucleasas programables específicas del objetivo en un genoma *in vitro* (Dual-Abnoba-Seq). Así pues, este procedimiento permite determinar la especificidad *in vivo* de una o más nucleasas programables específicas de objetivo en muestras celulares valiosas de tamaño y disponibilidad de muestra limitados, en otras palabras, la capacidad de determinar la actividad fuera del objetivo real de una o más nucleasas programables específicas de objetivo en células específicas de pequeño tamaño de muestra. Esto significa que el procedimiento de acuerdo con el segundo aspecto de la presente invención puede utilizarse para ensayar la actividad de nucleasas programables específicas de objetivo *in vivo* para demostrar la actividad *in vivo* real de una o más nucleasas programables específicas de objetivo en un tejido o tipo celular concreto de tamaño de muestra y disponibilidad limitados.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Figura 1: Visión general del flujo de trabajo de Abnoba-Seq
El ADN genómico se cortó aleatoriamente hasta un tamaño de fragmento medio de ~ 400 pb, los extremos se protegieron mediante la ligadura de un adaptador de bloqueo (gris) y los fragmentos no protegidos se eliminaron mediante tratamiento con exonucleasas. A continuación, los fragmentos de ADN bloqueados en los extremos se escindieron con CRISPR-Cas9, y el éxito de la escisión del genoma se controló mediante la amplificación por PCR del sitio dentro del objetivo como control en proceso (IPC). Los fragmentos escindidos se marcaron mediante la ligadura de un adaptador de biotina (gris claro), se enriquecieron en perlas de estreptavidina (StA) y posteriormente se amplificaron mediante PCR: en primer lugar, una PCR lineal en las perlas de StA con un único cebador que se une al adaptador de biotina y, en segundo lugar, mediante PCR exponencial tras la adición de un segundo cebador que se une al adaptador de bloqueo. Por último, se preparó una biblioteca y se sometió a secuenciación de próxima generación (NGS).

Figura 2: Prueba de bloqueo final de fragmentos de ADN
En la parte superior se muestra la ligadura del adaptador de bloqueo a fragmentos de ADN (amplicones de PCR), seguida de un tratamiento con exonucleasa y electroforesis capilar posterior, con los porcentajes de fragmentos de ADN bloqueados en los extremos (amplicones de PCR) en las tablas debajo de los gráficos. A: Tratamiento con exonucleasas: Ligación del adaptador de bloqueo (75 pmol) a un amplicón de PCR de 189 pb (400 ng) y posterior digestión con una combinación de exonucleasas durante 30 min en una reacción de dos pasos con la exonucleasa III (200 U), la lambda exonucleasa (10 U) y la exonucleasa I (20 U). B: Valoración de la concentración del adaptador de bloqueo: Ligación del adaptador de bloqueo en diferentes concentraciones a un amplicón PCR de 189 pb (400 ng). El amplicón PCR solo, con adaptador de bloqueo ligado a un extremo o con adaptador de bloqueo ligado a ambos extremos está representado por el pico de 189 pb, el pico de 215 pb o el pico de 241 pb, respectivamente, y un intermedio por el pico de 225 pb. picos de 15 pb y 5.000 pb (primer y último pico), fragmentos del marcador de alineación; los porcentajes se calcularon mediante AUC.

Figura 3: Controles en curso
A: Células U2OS. PCR para amplificar el sitio genómico dentro del objetivo tras la escisión *in vitro* de ADN genómico bloqueado en el extremo con nucleasas CRISPR-Cas dirigidas a RNF2, FANCF, VEGFA y HEKs4. B: HSC. PCR para amplificar el sitio genómico dentro del objetivo después de la escisión in vitro del ADN genómico bloqueado en el extremo con una nucleasa CRISPR-Cas dirigida a VEGFA. +RNP, muestras escindidas; -RNP, muestras no tratadas; ◄, posiciones de los amplicones PCR esperados.

Figura 4: Proceso bioinformático
Para la identificación de los sitios objetivo, el proceso bioinformático siguió los siguientes pasos: (1) cálculo de la cobertura de intervalos, (2) descarte de intervalos vacíos, (3) división de intervalos en ventanas deslizantes de 4 nt con pasos de 2 nt, (4) identificación de sitios de escisión putativos con lecturas que comienzan o terminan en el mismo sitio, (5) eliminación de sitios falsos positivos aplicando la relación de filtros entre cobertura de fondo y cobertura de señal ≤ 0,5, relación entre cobertura inicial y cobertura final ≥ 1/5 y ≤ 5, cobertura de sitio (suma de lecturas iniciales y finales) ≥ 6 y cobertura de intervalos ≥ 10.

Figura 5: Patrón de alineación de lecturas NGS en sitios dentro del objetivo
Las lecturas de NGS se alinearon con el genoma de referencia humano (hg19) y se visualizaron con Integrative Genomics Viewer (IGV). Las localizaciones cromosómicas de los respectivos sitios dentro del objetivo se indican con una barra. Las lecturas individuales se muestran como flechas, indicando su dirección. En la parte inferior se indican los genes anotados (exones como recuadros, intrones como líneas). A: ADN genómico de células U2OS. B: ADN genómico de las CMH.

Figura 6: Los 20 principales sitios de las dianas analizadas objetivo de los objetivos analizados
Gráficos de alineación que muestran las secuencias de los 20 principales sitios objetivo identificados, inlcuyendo el sitio dentro del objetivo (•) identificado para las nucleasas CRISPR-Cas dirigidas a FANCF, RNF2, VEGFA y HEKs4. La cobertura (número de lecturas) se muestra junto a la secuencia. Las filas superiores muestran la secuencia del sitio dentro del objetivo, incluyendo la PAM (NGG). En las filas subyacentes, las diferencias con el respectivo sitio dentro del objetivo se resaltan en gris; los números positivos y negativos indican protuberancias en el ADN o en el ARNg, respectivamente. N, cualquier nucleótido.

Figura 7: Comparación de Abnoba-Seq con GUIDE-Seq y CIRCLE-Seq
Diagrama de Venn que muestra la superposición de los sitios objetivo identificados entre Abnoba-Seq, GUIDE-Seq y CIRCLE-Seq.

Figura 8: Abnoba-Seq aplicado a células madre hematopoyéticas
humanas A: Top 20 (fuera-) sitios objetivo. Diagrama de alineación que muestra las secuencias de los 20 principales sitios objetivo identificados, incluyendo el sitio dentro del objetivo (•) identificado para una nucleasa CRISPR-Cas dirigida a VEGFA. La cobertura (número de lecturas) se muestra junto a la secuencia. La fila superior muestra la secuencia del sitio dentro del objetivo, incluyendo PAM. En las filas subyacentes, las

diferencias con el sitio dentro del objetivo se resaltan en gris. Los números positivos o negativos indican protuberancias en el ADN o en el ARNg, respectivamente. N, cualquier nucleótido. B: Sitios fuera del objetivo compartidos y distintos. Diagrama de Venn que muestra la superposición de los sitios fuera del objetivo identificados entre las células U2OS y las HSC. C, D: Comparación cualitativa. Diagrama de Venn que muestra la superposición de los 20 o 50 principales sitios objetivo identificados entre U2OS y HSC, respectivamente.

Figura 9: Secuenciación de amplicones dirigidos de los 20 principales sitios fuera del objetivo
Las células U2OS y las HSC se transfectaron con RNP dirigidas a VEGFA o HEKs4.La edición se evaluó mediante la secuenciación de amplicones del sitio dentro del objetivo y de los 20 principales sitios fuera del objetivo. A: Eficacia de la edición con una RNP dirigida a VEGFA. Para cada (fuera-) sitio objetivo (eje y) la barra muestra el porcentaje de alelos editados en una escala logarítmica (eje x) ya sea en células U2OS (negro) o en HSCs (gris). Se indica la localización cromosómica de cada sitio. ON, sitio dentro del objetivo; OT, sitios fuera del objetivo. B: Eficacia de la edición con una RNP dirigida a HEKs4. Para cada (fuera-) sitio objetivo (eje y) la barra muestra el porcentaje de alelos editados en una escala logarítmica (eje x) en HSCs (gris). La localización cromosómica de cada sitio se indica en el sitio correcto. ON: sitio dentro del objetivo; OT: sitios fuera del objetivo; n.d.: no determinado por fallo de secuenciación.

DESCRIPCIÓN DETALLADA

Abnoba-Seq

En un primer aspecto, la presente invención se refiere a un procedimiento para detectar sitios fuera del objetivo de una o más nucleasas programables específicas de objetivo en un genoma *in vitro.* Este procedimiento de detección de sitios fuera del objetivo de una o más nucleasas programables específicas de objetivo permite detectar todos los sitios de escisión de una o más nucleasas programables específicas de objetivo en un genoma *in vitro.* A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que comúnmente entiende un experto.

El procedimiento del primer aspecto de la presente invención tiene varias ventajas en comparación con el procedimiento conocido para determinar sitios fuera del objetivo de nucleasas programables específicas de objetivo. El procedimiento no utiliza la homología con el sitio/secuencia dentro del objetivo para identificar los sitios fuera del objetivo, como, por ejemplo, utilizan los procedimientos *in vitro* conocidos; en particular, las similitudes con la secuencia objetivo real no desempeñan ningún papel en la detección de los sitios fuera del objetivo y, a este respecto, el procedimiento es completamente imparcial. El resultado es un procedimiento con una sensibilidad comparable, pero con una mayor selectividad en comparación con otros procedimientos *in vitro* conocidos.

Ventajosamente, el procedimiento inventivo puede realizarse con pequeñas cantidades de ADN genómico de entrada, lo que permite probar una o más nucleasas programables específicas de objetivo en muestras celulares valiosas de tamaño y disponibilidad de muestra limitados, por ejemplo, mediante la determinación de la especificidad de nucleasas programables específicas de objetivo en células primarias, en material de biopsia o en un injerto editado genéticamente. Además, puede llevarse a cabo con secuenciadores NGS básicos (de bajo coste), sin necesidad de un costoso secuenciador de alto rendimiento como HiSeq.

El término "nucleasa programable específica de objetivo" es conocido por el experto. Preferentemente, el término "nucleasa programable específica de objetivo", tal como se utiliza en el contexto de la presente invención, se refiere a todas las formas de nucleasas que son capaces de reconocer y escindir un sitio específico en un genoma deseado. En particular, puede incluir, entre otras, ZFN, TALEN, meganucleasas, nucleasas de ADN guiadas por ADN basadas en proteínas de la familia Argonauta y CRISPR/Cas RGN. Términos como "CRISPR/Cas RGN", "ZFN" y "TALEN" pueden referirse a una nucleasa o a más de una (dos, tres, etc.) nucleasas de ese tipo, por ejemplo, CRISPR/Cas RGN puede referirse a una nucleasa guiada por ARN CRISPR/Cas o a más de una nucleasa guiada por ARN CRISPR)/Cas (dos, tres, etc.), de forma similar ZFN puede referirse a una nucleasa de dedos de zinc o a más de una nucleasa de dedos de zinc, etc. En una realización preferida, una nucleasa programable específica de objetivo es una nucleasa que escinde ADN de doble cadena (ARNdc) específico de objetivo, es decir, se dirige a un sitio o secuencia objetivo específico (sitio dentro del objetivo) para la escisión, para introducir preferentemente una rotura de ARNdc en el ADN.

El término "sitio objetivo", tal como se utiliza en el contexto de la presente invención, es conocido por el experto y puede referirse preferentemente a sitios dentro del objetivo y sitios fuera del objetivo. El término "sitio dentro del objetivo", tal como se utiliza en el contexto de la presente invención, es conocido por el experto. Se refiere preferentemente a un sitio o secuencia específica, que es el objetivo de una nucleasa programable específica de objetivo para la escisión de ácido nucleico, preferentemente la escisión de ADN, y puede seleccionarse libremente dentro de una secuencia de nucleótidos, preferentemente una secuencia de ADN como una secuencia de ADN genómico, dependiendo de la finalidad de la misma.

Por supuesto, el experto entiende que una nucleasa programable específica de objetivo puede dirigirse a ácidos

nucleicos, preferentemente ADN como el ADN genómico, para la escisión, además del sitio dentro del objetivo, en uno o más sitios distintos del sitio dentro del objetivo; estos otros sitios son sitios fuera del objetivo. El término "sitio fuera del objetivo", tal como se utiliza en el contexto de la presente invención, es conocido por el experto. Preferentemente, el término "sitio fuera de objetivo" se refiere a un sitio que no es idéntico en secuencia de ácido nucleico al sitio o secuencia objetivo (sitio dentro del objetivo) de la nucleasa programable específica de objetivo. Es decir, puede referirse a un sitio distinto del sitio dentro del objetivo que es el objetivo de escisión de la nucleasa programable específica de objetivo.

Preferentemente, el sitio fuera del objetivo es un sitio que no es idéntico en secuencia de ácido nucleico al sitio dentro del objetivo y que es el objetivo para la escisión por la nucleasa programable específica de objetivo *in vitro.* En una realización específica, el sitio fuera del objetivo es un sitio que no es idéntico en secuencia de ácido nucleico al sitio dentro del objetivo y que es objetivo para la escisión por la nucleasa programable específica de objetivo *in vitro* e *in vivo* (es decir, dentro de una célula). Por selección de objetivo para la escisión, se entiende preferentemente que una nucleasa programable específica de objetivo utiliza un sitio o secuencia específicos para unirse directa o indirectamente a un ácido nucleico, por ejemplo, a un ADN, para a continuación escindir el ácido nucleico en, o muy cerca de, este sitio o secuencia específicos.

Paso a) - Preparación de fragmentos purificados de ARNdc con extremos bloqueados

El procedimiento del primer aspecto de la invención comprende el paso (a) de proporcionar una muestra de entrada que comprende ADN genómico, fragmentar aleatoriamente el ADN genómico en fragmentos de ARNdc, y ligar un adaptador de bloqueo a los extremos de los fragmentos de ARNdc y eliminar los fragmentos de ARNdc que comprenden extremos, que no son extremos bloqueados, para preparar fragmentos de ARNdc purificados con extremos bloqueados.

El ADN genómico de la muestra de entrada no está particularmente limitado siempre que sea adecuado para su uso en el procedimiento de la presente invención. Preferentemente, el ADN genómico es ADN genómico aislado, lo que significa que el ADN genómico se extrae de las células y, preferentemente, se purifica, es decir, se separa hasta cierto punto, preferentemente tanto como sea posible, de las proteínas y otros componentes celulares y/o reactivos utilizados durante la extracción. La fuente del ADN genómico no está especialmente limitada. El ADN genómico puede proceder de células de un organismo específico, por ejemplo, de un organismo unicelular o multicelular, y para un organismo multicelular de células de un tejido específico, o de un tipo celular específico. En una realización, el ADN genómico procede de células eucariotas (por ejemplo, de un animal, un mamífero, un insecto, una planta, un hongo, un insecto, un ave, un pez, un anfibio, un reptil o un cnidario).

En una realización preferida, el ADN genómico comprendido en la muestra de entrada proporcionada en el paso (a) es ADN genómico humano, por ejemplo, de células primarias o material de biopsia. Dichas células pueden o no ser células de un tejido específico, o de un tipo celular específico, por ejemplo, de un tejido específico.

La cantidad de ADN genómico en la muestra de entrada no está particularmente limitada. En una realización preferida de la presente invención, la muestra de entrada de ADN genómico comprende menos de 1 µg, preferentemente menos de 900 ng, más preferentemente menos de 800 ng, aún más preferentemente menos de 700 ng, aún más preferentemente aproximadamente 650 ng o menos, aún más preferentemente aproximadamente 600 ng o menos, alternativamente preferentemente aproximadamente 500 ng, alternativamente preferentemente aproximadamente 400 ng, de ADN genómico.

El término "aproximadamente", tal como se utiliza en el presente documento en combinación con un valor o intervalo, indica que el valor o intervalo de una cantidad dada puede incluir cantidades que oscilan dentro del 10% del valor o intervalo establecido, u opcionalmente dentro del 5% del valor o intervalo, o en algunas realizaciones dentro del 1% del valor o intervalo. Por supuesto, el término "aproximadamente" también incluye el valor o el intervalo exactos del valor o intervalo con el que se combina el término.

La forma en que el ADN genómico se fragmenta aleatoriamente en fragmentos de ARNdc, de acuerdo con la presente invención, no está particularmente limitada. Se conocen formas de fragmentar aleatoriamente el ADN genómico. El ADN genómico puede, por ejemplo, fragmentarse aleatoriamente por cizallamiento, es decir, por rotura mecánica del ADN, o por fragmentación enzimática. De acuerdo con una realización de la presente invención, el ADN genómico se fragmenta por sonicación (ultrasonicación), cizallamiento acústico, cizallamiento centrífugo, cizallamiento por hundimiento puntual, cizallamiento por aguja o fragmentación basada en enzimas, preferentemente por sonicación. Preferentemente, el ADN genómico se fragmenta hasta una longitud definida (mediana o promedio). La longitud particular del ADN genómico en fragmentos de ARNdc no está limitada. En una realización específica del primer aspecto de la presente invención, el ADN genómico se fragmenta aleatoriamente en fragmentos de ARNdc con un tamaño de fragmento promedio o mediano, preferentemente mediano, de 100 a 2.000 pb, preferentemente de 200 a 800 pb, más preferentemente de 300 a 600 pb, aún más preferentemente de 300 a 600 pb, aún más preferentemente de 400 a 500 pb, aún más preferentemente de 400 pb. En una realización preferida de la presente invención, el ADN genómico se fragmenta aleatoriamente en fragmentos de ARNdc con un tamaño mediano de fragmento de aproximadamente 300 a 600 pb, más preferentemente de 300 a 600 pb, aún más preferentemente de

aproximadamente 400 pb, aún más preferentemente de 400 pb. El término "pb" se refiere a uno o más pares de bases, es decir, uno o más pares de nucleótidos complementarios en un ácido nucleico de doble cadena (dc), como un ARNdc o un ARNdc. Tras la fragmentación aleatoria del ADN genómico de la muestra de entrada para obtener fragmentos de ADNdc, se liga un adaptador de bloqueo a los extremos de los fragmentos de ADNdc y, a continuación, se eliminan los fragmentos de ADNdc que comprende extremos no bloqueados para preparar fragmentos purificados de ADNdc con extremos bloqueados.

El término "adaptador de bloqueo", tal como se utiliza en el presente documento, se refiere a una molécula, preferentemente compuesta por ARNdc, que, tras ligar el adaptador de bloqueo a un extremo de un fragmento de ARNdc, bloquea la ligación posterior, así como la escisión (digestión) por exonucleasas de ese extremo. El término "extremo bloqueado" en relación con un fragmento de ARNdc se refiere a un fragmento de ARNdc al que se ha ligado un adaptador de bloqueo en ambos extremos.

El adaptador de bloqueo no está particularmente limitado siempre que cumpla las funciones descritas anteriormente. Se conocen en la técnica formas de preparar un adaptador de bloqueo que, tras la ligadura, bloquee la ligadura posterior, así como la escisión por exonucleasa del extremo de un fragmento de ADN al que esté ligado. Preferentemente, el adaptador de bloqueo es un adaptador basado en ADN, más preferentemente un adaptador basado en ARNdc (es decir, adaptador de bloqueo basado en ARNdc), lo que significa que comprende ARNdc. Dicho adaptador de bloqueo basado en ARNdc puede comprender modificaciones adicionales que incluyan nucleótidos modificados y otras moléculas que no sean nucleótidos. Dicho adaptador de bloqueo basado en ARNdc tiene preferentemente una longitud de secuencia que permite la incorporación de un sitio de cebador en el adaptador de bloqueo para la amplificación por PCR.

De acuerdo con una realización preferida de la presente invención, el adaptador de bloqueo, preferentemente un adaptador de bloqueo basado en ARNdc, comprende una o más modificaciones, que tras la ligadura del adaptador de bloqueo a un fragmento de ARNdc, impiden la ligadura del extremo al que se liga el adaptador de bloqueo al fragmento de ARNdc y una o más de las modificaciones, que tras la ligadura del adaptador de bloqueo a un fragmento de ARNdc, impiden la escisión (digestión) por exonucleasas del fragmento de ARNdc del extremo al que se liga el adaptador de bloqueo al fragmento de ARNdc.

Las modificaciones en un adaptador de bloqueo, que después de la ligadura del adaptador de bloqueo a un fragmento de ARNdc, impiden la ligadura, del extremo al que el adaptador de bloqueo está ligado al fragmento de ARNdc son conocidas en la técnica, y en un adaptador de bloqueo basado en ARNdc incluyen uno o más espaciadores C3, uno o más Dideoxy-G/Dideoxy-C, u otros espaciadores, preferentemente uno o más espaciadores C3, preferentemente en cada cadena del adaptador de bloqueo basado en ADNdc, preferentemente en el extremo del adaptador de bloqueo basado en ADNdc, que no se utiliza para ligar el adaptador de bloqueo basado en ADNdc al fragmento de ADNdc (es decir, el extremo del adaptador de bloqueo que, después de la ligadura, no está directamente adyacente al extremo del fragmento de ADNdc al que se ligó el adaptador de bloqueo).

Las modificaciones en un adaptador de bloqueo, que después de la ligadura del adaptador de bloqueo a un fragmento de ARNdc, impiden la escisión por exonucleasa del fragmento de ARNdc desde el extremo al que el adaptador de bloqueo está ligado al fragmento de ARNdc también son conocidas en la técnica, y en un adaptador de bloqueo basado en ARNdc incluyen uno o más enlaces fosforotioato, y uno o más espaciadores C3, y combinaciones de los mismos. En una realización preferida, el adaptador de bloqueo es un adaptador de bloqueo basado en ARNdc que comprende uno o más, preferentemente dos o más, enlaces fosforotioato, preferentemente al menos dos (por ejemplo, 2, 3, 4, etc.) en cada cadena del adaptador de bloqueo basado en ARNdc, preferentemente cerca del extremo, por ejemplo, entre los últimos 3-5 nucleótidos, del adaptador de bloqueo basado en ARNdc que no se utiliza para ligar el adaptador de bloqueo basado en ARNdc al fragmento de ARNdc, (es decir, el extremo del adaptador de bloqueo que, tras la ligadura, no está directamente adyacente al extremo del fragmento de ADNdc al que se ligó el adaptador de bloqueo); y uno o más espaciadores C3, preferentemente al menos en cada cadena del adaptador de bloqueo basado en ADNdc, preferentemente en el extremo del adaptador de bloqueo basado en ADNdc que no se utiliza para ligar el adaptador de bloqueo basado en ADNdc al fragmento de ADNdc.

El adaptador de bloqueo en forma de adaptador de bloqueo basado en ARNdc puede comprender modificaciones adicionales que incluyan, por ejemplo, una o más desoxiuridinas internas, preferentemente al menos dos en cada cadena, preferentemente cerca del extremo del adaptador que no se utiliza para ligar el adaptador de bloqueo basado en ARNdc al fragmento de ARNdc. Un adaptador de bloqueo preferido es un adaptador de bloqueo basado en ARNdc que comprende una combinación de uno o más espaciadores C3 y uno o más enlaces fosforotioato, al menos uno en cada cadena del adaptador de bloqueo basado en ARNdc y preferentemente en el extremo del adaptador de bloqueo basado en ARNdc, que no se utiliza para ligar el adaptador de bloqueo basado en ARNdc al fragmento de ARNdc, y opcionalmente una o más desoxiuridinas internas.

Además, el adaptador de bloqueo comprende preferentemente un sitio de cebado compatible para su uso en cebado por PCR, lo que significa una secuencia de nucleótidos a la que un cebador de PCR puede fusionarse para amplificar el fragmento de ARNdc al que se liga el adaptador de bloqueo. Preferentemente, la secuencia de un adaptador de bloqueo basado en ARNdc y/o el sitio del cebador en el adaptador de bloqueo no está presente en el genoma objetivo

(es decir, la secuencia del ADN genómico en la muestra de entrada y/o el genoma de referencia). En una realización específica de la presente invención, el adaptador de bloqueo es un adaptador de bloqueo basado en ARNdc que comprende la secuencia directa definida en SEQ ID NO:01 y la secuencia inversa de SEQ ID NO02. En otra realización específica de la presente invención, el adaptador de bloqueo es un adaptador de bloqueo basado en ARNdc que consiste en las secuencias SEQ ID NO:01 y SEQ ID NO:02.

El adaptador de bloqueo se liga a los extremos de los fragmentos de ARNdc por procedimientos conocidos en la técnica. Por ejemplo, el adaptador de bloqueo puede ligarse a los extremos del fragmento de ARNdc mediante ligación enzimática o química utilizando un protocolo de ligación conocido en la técnica. La ligadura enzimática puede llevarse a cabo utilizando una ADN ligasa conocida, que facilita la unión de cadenas de ADN catalizando la formación de un enlace fosfodiéster, e incluye la ligadura de extremo romo y de extremo pegajoso. En una realización preferida, el adaptador de bloqueo se liga a los extremos del fragmento de ARNdc mediante ligadura enzimática.

Preferentemente, la ligadura del adaptador de bloqueo a los extremos de los fragmentos de ARNdc, comprende 1. preparar los fragmentos de ARNdc para la ligadura del adaptador de bloqueo; y 2. ligar un adaptador de bloqueo a los extremos de los fragmentos de ARNdc, para obtener fragmentos de ARNdc con extremos bloqueados. El experto sabe cómo preparar un fragmento de ARNdc para bloquear la ligadura del adaptador. La preparación de los fragmentos de ARNdc para la ligadura del adaptador de bloqueo puede comprender una o ambas reparaciones de los extremos y formación de cola dA (incorporación de un desoxiadenosina 5'-monofosfato (dAMP) no templado en el extremo 3' de los fragmentos de ADN romos) del ARNdc fragmentado.

En una realización específica de la presente invención, los fragmentos de ADNdc se preparan para bloquear la ligadura del adaptador mediante la reparación del extremo, para preparar fragmentos de ADNdc, que están libres de salientes, y contienen grupos fosfato 5' e hidroxilo 3', y formación de cola dA de los fragmentos de ADN de extremo romo así preparados. A los extremos de un fragmento de ARNdc así preparado puede ligarse un adaptador de bloqueo basado en ARNdc con un saliente dT en el extremo 3' de un extremo del adaptador.

Después de ligar el adaptador de bloqueo a los extremos de los fragmentos de ARNdc, los fragmentos de ARNdc que comprenden extremos no bloqueados se eliminan para preparar fragmentos purificados de ARNdc con extremos bloqueados. En estos fragmentos de ADNdc con extremos no bloqueados, uno o ambos extremos no están ligados a un adaptador de bloqueo. El experto sabe cómo eliminar estos fragmentos de ADNdc con extremos no bloqueados.

En una realización preferida, la eliminación de fragmentos de ARNdc que comprenden extremos no están bloqueados, comprende el contacto de la muestra que comprende los fragmentos de ARNdc con extremos bloqueados (después de la ligadura del adaptador de bloqueo a los fragmentos de ARNdc) con una o más exonucleasas para eliminar los fragmentos de ARNdc, para obtener fragmentos purificados de ADN con extremos bloqueados Las exonucleasas adecuadas son conocidas en la técnica e incluyen, por ejemplo, una o más de las exonucleasas III, exonucleasa V, exonucleasa VIII, exonucleasa lambda, exonucleasa T5, exonucleasa T7, nucleasa BAL-31 y combinaciones de las mismas, preferentemente exonucleasa III, exonucleasa lambda o combinaciones de las mismas. Opcionalmente se utilizan una o más exonucleasas de ADN monocatenario (ssADN), por ejemplo, seleccionadas de la exonucleasa I, la exonucleasa termoestable I, la exonucleasa T, la exonucleasa VII, la RecJf, la nucleasa de judía mungo, la nucleasa P1 y combinaciones de las mismas.

En una realización específica, la una o más exonucleasas consiste en una mezcla de exonucleasa I, exonucleasa III y exonucleasa lambda, por ejemplo, en una relación de unidades enzimáticas de 2 : 20 : 5.

En una realización preferida del procedimiento del primer aspecto de la presente invención, el paso a) comprende los pasos de a1) fragmentar aleatoriamente el ADN genómico a una longitud definida para proporcionar fragmentos de ARNdc; a2) proteger los extremos de los fragmentos de ARNdc ligando un adaptador de bloqueo para obtener fragmentos de ARNdc con extremos bloqueados; y a3) poner en contacto la muestra que comprende los fragmentos de ARNdc con extremos bloqueados con una o más exonucleasas para eliminar los fragmentos de ARNdc, que no tienen extremos bloqueados, para obtener fragmentos purificados de ADN con extremos bloqueados.

En otra realización específica del procedimiento del primer aspecto de la presente invención, en el paso a) los extremos libres de los fragmentos de ADN restantes con extremos no bloqueados se desfosforilan con un reactivo de desfosforilación, preferentemente después de contactar la muestra que comprende los fragmentos de ADNdc con extremos bloqueados con una o más exonucleasas para eliminar los fragmentos de ADNdc. En esta realización preferentemente en el paso a) el reactivo de desfosforilación consiste en una o más fosfatasas, la una o más fosfatasas preferentemente comprenden, o consisten en, una o más de Fosfatasa Antártica, Fosfatasa Alcalina Intestinal de Ternera, Quick-CIP, Fosfatasa Alcalina de Camarón, y combinaciones de las mismas.

Paso b) - Preparación de una biblioteca de ADN escindido y amplificado de fragmentos de ADNdc

El procedimiento del primer aspecto de la invención comprende como segundo paso, el paso (b) de preparar una biblioteca de ADN escindido y amplificado de fragmentos de ADNdc, mediante b1) escindir los fragmentos de ADNdc con extremos bloqueados purificados con una o más nucleasas programables específicas de objetivo para obtener

fragmentos de ADNdc escindidos; b2) ligar un adaptador de afinidad a los sitios de escisión de los fragmentos de ADNdc escindidos para obtener fragmentos de ADNdc modificados por adaptador de afinidad; b3) enriquecer los fragmentos de ADNdc modificados por adaptador de afinidad utilizando el adaptador de afinidad para obtener fragmentos de ADNdc modificados por adaptador de afinidad enriquecidos; y b4) amplificar los fragmentos de ADNdc modificados por adaptador de afinidad enriquecidos mediante amplificación por PCR para obtener una biblioteca de ADN escindida y amplificada de fragmentos de ADNdc.

De acuerdo con la presente invención, la cantidad de fragmentos purificados de ARNdc con extremos bloqueados utilizados en el paso b1) no está particularmente limitada. De acuerdo con una realización específica, en el paso b1) se escinden menos de 400 ng, preferentemente menos de 300 ng, más preferentemente menos de 200 ng, aún más preferentemente menos de 100 ng, por ejemplo, aproximadamente 90 ng, aproximadamente 80 ng, aproximadamente 70 ng, aproximadamente 60 ng, menos de 60 ng, aproximadamente 50 ng, o 50 ng, de fragmentos purificados de ADNdc con extremos bloqueados, alternativamente menos de 50 ng, por ejemplo, aproximadamente 40 ng de fragmentos purificados de ADNdc con extremos bloqueados con una o más nucleasas programables específicas de objetivo, para obtener fragmentos de ADNdc escindidos. En una realización preferida, en el paso b1) se utilizan menos de 100 ng, preferentemente por ejemplo aproximadamente 50 ng, de fragmentos purificados de ARNdc con extremos bloqueados.

De acuerdo con otra realización específica de la presente invención, en el paso b1) se purifican fragmentos de ARNdc bloqueados por el extremo correspondientes a menos de 1.500.000 genomas haploides, preferentemente menos de 150.000 genomas haploides, más preferentemente menos de 80.000 genomas haploides, aún más preferentemente menos de 60.000 genomas haploides, por ejemplo menos de 50.000 genomas haploides, aún más preferentemente menos de 50.000 genomas haploides, por ejemplo aproximadamente 45.000, aproximadamente 30.000, aproximadamente 20.000 o aproximadamente 15.000 genomas haploides.

Los fragmentos de ARNdc con extremos bloqueados purificados pueden escindirse con una o más nucleasas programables específicas de objetivo para obtener fragmentos de ARNdc escindidos poniendo en contacto una muestra que comprende los fragmentos purificados de ARNdc con extremos bloqueados con una o más nucleasas programables específicas de objetivo.

La una o más nucleasas programables específicas de objetivo que se van a utilizar en el paso b1) no están particularmente limitadas siempre que puedan utilizarse en los procedimientos de la presente invención, lo que significa siempre que sean adecuadas para su uso como nucleasa programable específica de objetivo para la escisión específica de objetivo de ADN genómico, por ejemplo para la edición de genoma, y sean adecuadas para su uso en la detección de sitios fuera del objetivo utilizando los procedimientos de la presente invención, preferentemente *in vitro* e *in vivo.* Las nucleasas programables específicas de objetivo adecuadas incluyen TALEN, ZFN, meganucleasas, megaTAL, FokI-dCas9, nucleasas de ADN guiadas por ADN basadas en proteínas de la familia Argonauta, CRISPR/Cas RGN, o una variante de las mismas, o una combinación de las mismas.

Los TALEN que pueden utilizarse en los procedimientos de la presente invención no están particularmente limitados, siempre que puedan utilizarse en los procedimientos de la presente invención. TALEN se refiere a una o más proteínas de fusión que comprenden un dominio de unión al ADN derivado de las proteínas efectoras similares al activador de la transcripción (TALE) y un dominio de escisión de nucleótidos, preferentemente el dominio de escisión FokI. Las proteínas TALE contienen múltiples dominios repetidos de 33-35 aminoácidos (conjuntos de repeticiones TALE) que reconocen cada uno un único par de bases (véase, por ejemplo, Gaj et al., 2013, Trends Biotechnol., 31(7):397-405). Los procedimientos para generar conjuntos de repeticiones TALE modificados y fusionarlos con una nucleasa para preparar TALEN son conocidos en la técnica (véase, por ejemplo, Reyon et al., 2012, Nature Biotechnology 30, 460-465; Bogdanove y Voytas, 2011, Science 333, 1843-1846; Bogdanove et al., 2010, Curr Opin Plant Biol 13, 394-401; Scholze y Boch, 2011, J. Curr Opin Microbiol 2011; Boch et al.,2009, Science 326, 1509-1512; Cermak et al., 2011, Nucleic Acids Res 39, e82; Tesson et al., 2011, Nat Biotechnol 29, 695-696; Sander et al., 2011, Nat Biotechnol 29, 697-698; y Zhang et al., 2011, Nat Biotechnol 29, 149-153).

Las meganucleasas que pueden utilizarse en los procedimientos de la presente invención no están particularmente limitadas, siempre que puedan utilizarse en los procedimientos de la presente invención. Las meganucleasas son endonucleasas de secuencia específica procedentes de diversos organismos, tales como bacterias, levaduras, algas y orgánulos vegetales. Las meganucleasas endógenas tienen sitios de reconocimiento de aproximadamente 12 a 30 pb; se han descrito sitios de unión al ADN personalizados con sitios de reconocimiento de meganucleasas de 18 pb y 24 pb de longitud, que pueden utilizarse en los presentes procedimientos (véase, por ejemplo. Silva et al., 2011, Current Gene Therapy, 11:11-27; Arnould et al., 2011, Protein Engineering Design & Selection, 24:27-31;). La meganucleasa puede seleccionarse, por ejemplo, de las meganucleasas de la familia LAGLIDADG, la familia GIY-YIG, la familia His-Cyst box y la familia HNH. Algunas meganucleasas ejemplares son I-SceI, I-CeuI, PI-PspI, PI-SceI, I-SceIV, I-CsmI, I-PanI, I-SceII, I-PpoII, I-SceIII, I-CreI, I-TevI, I-TevII e I-TevIII.

Los ZFN que pueden usarse en los procedimientos de la presente invención no están particularmente limitados, siempre que sean adecuados para usarse en los procedimientos de la presente invención. Los ZFN son nucleasas artificiales modificadas que comprenden un dominio de unión al ADN con dedos de zinc y un dominio de escisión del

ADN. Es posible diseñar artificialmente las características de unión al ADN de los dominios individuales de unión al ADN de los dedos de zinc mediante la aleatorización de los aminoácidos en las posiciones alfa-helicoidales implicadas en la unión al ADN y el uso de metodologías de selección tal como la visualización de fagos para identificar las variantes deseadas capaces de unirse a los sitios objetivo de ADN de interés (Rebar et al., 1994, Science, 263:671; Choo et al., 1994 Proc. Natl. Acad. Sci. EE. UU., 91:11163; Jamieson et al., 1994, Biochemistry 33:5689; Wu et al., 1995 Proc. Natl. Acad. Sci. EE. UU., 92: 344). Estos dominios recombinantes de unión al ADN de dedos de zinc pueden fusionarse con dominios funcionales, tales como nucleasas, para introducir alteraciones selectivas en genomas de organismos modelo, plantas y células humanas (Carroll, 2008, Gene Ther, 15:1463-68; Cathomen, 2008, Mol. Ther., 16:1200-07; Wu et al., 2007, Cell. Mol. Life Sci., 64:2933-44).

La selección de sitios/secuencias objetivo específicos y el diseño y construcción de ZFN son bien conocidos por los expertos en la técnica, y se describen en detalle, por ejemplo, en Patente de EE. UU. No. 7,888,121; 8,409,861; 6,479,626 ; 6,903,185; and 7,153,949). El ZFN puede incluir cualquier combinación de enlazadores adecuados entre cada dedo de zinc de la proteína. El dominio de unión al ADN de los dedos de zinc puede diseñarse para que se una a la secuencia seleccionada (Beerli et al., 2002, Nat. Biotechnol, 20:135-141; Segal et al., 2003, Biochemistry, 42:2137-48; Mandell et al., 2006, Nucleic Acids Res., 34:W516-523; Carroll et al., 2006, Nat. Protoc. 1:1329-41; Bae et al., 2003, Nat. Biotechnol, 21:275-280; Wright et al., 2006, Nat. Protoc., 1:1637-52; Maeder et al., 2008, Mol. Cell, 31:294-301; Joung et al., 2010, Nat. Methods, 7:91-92). En realizaciones preferidas, los ZFN se preparan como se describe en los documentos WO 2011/017293 y WO 2004/099366. Los dominios de unión a ADN con dedos de zinc adecuados adicionales se describen en la patente de EE. UU. No. 6.511.808; 6.013.453; 6.007.988; y 6.503.717).

Los megaTAL y FokI-dCas9 que pueden utilizarse en los procedimientos de la presente invención no están particularmente limitados, siempre que sean adecuados para su uso en los procedimientos de la presente invención. MegaTAL son proteínas artificiales de fusión de una meganucleasa con un efector TAL (véase, por ejemplo. Boissel et al., 2014, Nucl. Acids Res. 42(4):2591-2601; Boissel y Scharenberg, 2015, Methods Mol Biol. 2015; 1239:171-96). FokI-dCas9 son proteínas de fusión que comprenden Cas9 inactivada para la selección de objetivo de secuencias y la nucleasa FokI (Guilinger et al., 2014, Nat Biotechnol. Jun; 32(6): 577-582; documentos WO 2014/144288y WO 2014/204578).

Las nucleasas de ADN guiadas por ADN basadas en proteínas de la familia Argonauta que pueden utilizarse en los procedimientos de la presente invención no están particularmente limitadas, siempre que puedan utilizarse en los procedimientos de la presente invención. Se conocen en la técnica formas de utilizar nucleasas de ADN guiadas por ADN basadas en proteínas de la familia Argonauta para la edición del genoma (véase, por ejemplo, el documento WO2017139264A1; Wei et al., 2016, Genes & Diseases, Vol. 3, Iss. 3, p.169-170).

Las nucleasas CRISPR/Cas guiadas por ARN (CRISPR/Cas RGN) que pueden utilizarse en los procedimientos de la presente invención no están particularmente limitadas, siempre que puedan utilizarse en los procedimientos de la presente invención. Esto significa que son al menos adecuadas para su uso como nucleasas programables específicas de objetivo para la escisión específica de ADN genómico, por ejemplo, para la edición del genoma, y son adecuadas para su uso en la detección de sitios fuera de objetivo utilizando los procedimientos de la presente invención, preferentemente *in vitro* e *in vivo.* CRISPR/Cas RGN se refiere a la combinación de una endonucleasa Cas9 y un ARNg. La endonucleasa Cas9 puede unirse al ARNg para formar una ribonucleoproteína (RNP) Cas9-ARNg. Las formas de preparar CRISPR/Cas RGN específicos de la secuencia objetivo son bien conocidas en la técnica.

La endonucleasa Cas de la CRISPR/Cas RGN que se utilizará en los procedimientos de la presente invención no está particularmente limitada, siempre que sea adecuada para su uso en una CRISPR/Cas RGN para la edición del genoma y para su uso en la detección de sitios fuera del objetivo utilizando los procedimientos de la presente invención, preferentemente *in vitro* e *in vivo.* En una realización específica de la presente invención, la endonucleasa Cas de la CRISPR/Cas RGN es la endonucleasa Cas9 o una variante de la endonucleasa Cas9 o la endonucleasa Cpf1 o una variante de la endonucleasa Cpf1. La endonucleasa Cas puede guiarse a través de una complementariedad de nucleótidos (nt) simple de aproximadamente 20 nucleótidos entre un ARNg modificado y la cadena complementaria de una secuencia de ADN genómico objetivo de interés que se encuentra junto a una PAM. El término "nt", tal como se utiliza en el presente documento, se refiere a uno o más nucleótidos (por ejemplo, uno, dos, tres, cuatro, etc.) de una molécula de ADN o ARN.

Puede utilizarse la endonucleasa Cas9, por ejemplo, de S. *pyogenes* (en adelante SpyCas9) (Shen et al., 2013, Cell Res; Jinek et al., 2013, Elife 2, e00471; Hwang et al., 2013, Nat Biotechnol 31, 227-229 (2013); Cong et al., 2013, Science 339, 819-823; Mali et al., 2013, Science 339, 823-826; Cho et al., 2013, Nat Biotechnol 31, 230-232. También puede utilizarse la endonucleasa CRISPR de Prevotella y Francisella 1 Cpf1 (endonucleasa asociada a CRISPR en Prevotella y Francisella 1) (véase, por ejemplo, Zetsche et al., 2015, Cell 163, 759-771y Makarova et al., 2015, Nat Rev Microbiol 13, 722-736).

En algunas realizaciones, los procedimientos inventivos utilizan una endonucleasa Cas9 de tipo silvestre o variante del género *Streptococcus*, el género *Neisseria*, el género *Pasteurella*, el género *Francisella*, el género *Campylobacter,por ejemplo* de S. *pyogenes, Staphylococcus aureus, Neisseria meningitidis, Streptococcus*

thermophilus o Treponema denticola; o una endonucleasa Cpf1 de tipo silvestre o variante del género *Lachnospira*, el género *Butyrivibrio*, *Peregrinibacteria*, el género *Acidominococcus*, el género *Porphyromonas*, el género *Prevotella*, el género *Francisella*, *Candidatus methanoplasma*, y el género *Eubacterium*, por ejemplo de Candidatus Paceibacterium, el género *Acidominococcus*, el género *Porphyromonas*, el género *Prevotella*, el género *Francisella*, *Candidatus methanoplasma*, y el género *Eubacterium*, por ejemplo de *Candidatus Paceibacter*, *Acidaminococcus sp. BV3L6 o la bacteria Lachnospiraceae ND2006;* bien codificada en bacterias o, por ejemplo, optimizada en codones para su expresión en células de mamífero y/o modificada en su especificidad de reconocimiento de PAM y/o su especificidad en todo el genoma. Se han descrito diversas variantes (véase, por ejemplo, el documenot WO 2016/141224, PCT/US2016/049147,Kleinstiver et al., 2016, Nat Biotechnol., 34(8):869-74; Tsai and Joung, 2016, Nat Rev Genet., 17(5):300-12; Kleinstiver et al., 2016, Nature., 529(7587):490-5; Shmakov et al., 2015, Mol Cell., 5; 60(3):385-97; y Tsai et al., 2014, Nat Biotechnol., 32(6):569-76).

El término "ARNg" o "ARN guía", tal como se utiliza en el presente documento, se refiere a un ARN específico del ADN objetivo, que puede unirse a una endonucleasa Cas para formar la ribonucleoproteína (RNP) CRISPR/Cas RGN, y guía a una proteína Cas hasta un sitio/secuencia objetivo. Puede utilizarse cualquier tipo de ARNg en los procedimientos de la presente invención, siempre que sea adecuado para su uso en una CRISPR/Cas RGN para la edición del genoma y para su uso en la detección de sitios fuera del objetivo utilizando los procedimientos de la presente invención, preferentemente *in vitro* e *in vivo.* El ARNg puede ser, por ejemplo, un ARN guía único (ARNsg) o un ARNg dual (ARNdg) que comprenda un par de moléculas de ARN ARNcr (CRISPRARN) /ARNtracr (ARNcr transactivador). De acuerdo con una realización específica de la presente invención, el ARNg es un ARNsg, que comprende un ARN, o un ARNdg que comprende dos ARN, preferentemente un ARNcr y un ARNtracr, como componentes.

La una o más nucleasas programables específicas de objetivo se dirigen a un sitio o secuencia de nucleótidos específico (sitio dentro del objetivo). El sitio dentro del objetivo no está particularmente limitado y puede incluir uno o más sitios o secuencias específicas.

De acuerdo con una realización de la presente invención en el paso b1), la una o más nucleasas programables específicas de objetivo se seleccionan de nucleasas efectoras similares a activadores de transcripción (TALEN), nucleasas con dedos de zinc (ZFN), meganucleasas, megaTAL, FokI-dCas9, repeticiones palindrómicas cortas agrupadas regularmente interespaciadas (CRISPR)/nucleasas guiadas por ARN Cas (CRISPR/Cas RGN) o una variante de los mismos, o una combinación de los mismos.

En una realización específica de la presente invención, en el paso b1), la una o más nucleasas programables específicas de objetivo comprende, preferentemente consiste en, una o más CRISPR/Cas RGN, cada una utilizada preferentemente como una ribonucleoproteína (RNP) preensamblada, en la que la endonucleasa Cas está unida a un ARNg, por ejemplo, en una relación molar de 1:3. La endonucleasa Cas puede ser, por ejemplo, una proteína recombinante. De acuerdo con una realización ejemplar, se utilizan aproximadamente 1,6 pmol de endonucleasa Cas y aproximadamente 4,8 pmol de ARNsg, preferentemente preensamblados en una RNP de endonucleasa Cas/ARNgc para escindir aproximadamente 50 ng de fragmentos purificados de ADNdc con extremos bloqueados; para una cantidad mayor o menor de fragmentos purificados de ADNdc con extremos bloqueados, la endonucleasa Cas y el ARNsg utilizados pueden, por ejemplo, adaptarse proporcionalmente.

Puede utilizarse un control en proceso para monitorizar la actividad dentro del objetivo de una o más nucleasas programables específicas de objetivo. De acuerdo con una realización de la presente invención, la amplificación por PCR utilizando cebadores específicos del sitio dentro del objetivo para los sitios dentro del objetivo de una o más nucleasas programables específicas de objetivo se utiliza como control durante el proceso.

El adaptador de afinidad que se utilizará en el paso b2) del procedimiento del primer aspecto de la invención no está particularmente limitado, siempre que permita, después de la ligadura del adaptador de afinidad a un sitio de escisión de un fragmento purificado de ADNdc con extremo bloqueado, que ha sido escindido por una o más nucleasas programables específicas del objetivo (es decir, un fragmento de ADNdc escindido), el enriquecimiento específico del fragmento de ADNdc modificado por adaptador de afinidad así preparado utilizando el adaptador de afinidad. Tales adaptadores de afinidad son conocidos en la técnica. Preferentemente, el adaptador de afinidad comprende una o más modificaciones que, tras la ligadura del adaptador de afinidad al sitio de escisión de un fragmento de ADNdc escindido, permiten el enriquecimiento específico del fragmento de ADNdc modificado por adaptador de afinidad.

Preferentemente, el adaptador de afinidad que se utilizará en los procedimientos inventivos es un adaptador basado en ADN, preferentemente un adaptador basado en ARNdc (es decir, un adaptador de afinidad basado en ARNdc) que comprende una o más modificaciones adecuadas para la purificación por afinidad. Dicho adaptador de afinidad basado en ARNdc tiene preferentemente una longitud de secuencia que permite la incorporación de un sitio de cebado para la amplificación por PCR en el adaptador de afinidad.

Las modificaciones adecuadas para la purificación por afinidad son conocidas en la técnica e incluyen, por ejemplo, la biotina y las modificaciones basadas en biotina, en las que la biotina o la modificación basada en biotina se une al extremo 5' o 3' del adaptador de afinidad basado en ADN, o internamente, o a un nucleótido específico (o derivado de

nucleótido) del adaptador de afinidad basado en ADN, por ejemplo, a una desoxitimidina. Las modificaciones adecuadas de biotina y basadas en biotina incluyen, por ejemplo, biotina, biotina dT (base de desoximidina biotinilada), biotina TEG (espaciador de 15 átomos de trietilenglicol); destiobiotina TEG; biotina dual; biotina fotocontraíble (PC); biotina azida; nitrilotriacetato (NTA); DNP TEG (2, 4-dinitrofenil) y combinaciones de las mismas. Otras modificaciones adecuadas son la digoxigenina, las etiquetas proteínicas y combinaciones de las mismas. Preferentemente, la una o más modificaciones se seleccionan de una o más, preferentemente al menos dos, modificaciones basadas en biotina y biotina, y combinaciones de las mismas. En una realización específica, las modificaciones de biotina se añaden sólo a una cadena del adaptador de afinidad basado en ARNdc.

Además, el adaptador de afinidad comprende preferentemente un sitio de cebado compatible para su uso en cebado PCR, lo que significa una secuencia de nucleótidos a la que un cebador PCR puede unirse para amplificar el fragmento de ARNdc escindido al que se liga el adaptador de afinidad. Preferentemente, la secuencia de un adaptador de afinidad, preferentemente basado en ARNdc, y/o el sitio del cebador en el adaptador de afinidad no están presentes en el genoma objetivo (es decir, la secuencia del ADN genómico en la muestra de entrada y/o el genoma de referencia).

En una realización preferida, el adaptador de afinidad es un adaptador de afinidad basado en ARNdc, que comprende una o más modificaciones de biotina o basadas en biotina. En una realización preferida específica, el adaptador de afinidad es un adaptador basado en ADN que comprende uno o más, preferentemente al menos dos, dT de biotina internos. En otra realización específica de la presente invención, el adaptador de afinidad es un adaptador de afinidad basado en ARNdc que comprende la secuencia definida en SEQ ID NO:03 y la secuencia de SEQ ID NO04. En otra realización específica de la presente invención, el adaptador de afinidad está formado por las secuencias SEQ ID NO:03 y SEQ ID NO:04.

El adaptador de afinidad se liga a los sitios de escisión del ARNdc escindido por procedimientos conocidos en la técnica. El sitio de escisión de un fragmento de ARNdc escindido se refiere a un sitio de escisión en un fragmento de ARNdc, que ha sido escindido por una o más nucleasas programables específicas de objetivo. Por ejemplo, el adaptador de afinidad puede ligarse a los sitios de escisión de los fragmentos de ARNdc escindidos mediante ligadura enzimática o química utilizando un protocolo de ligadura conocido en la técnica. La ligadura enzimática puede realizarse utilizando una ADN ligasa conocida. En una realización preferida, el adaptador de afinidad se liga a los sitios de escisión del ARNdc escindido mediante ligadura enzimática.

Preferentemente, la ligadura del adaptador de afinidad a los sitios de escisión de los fragmentos de ADNdc escindidos, comprende 1. preparar los fragmentos de ADNdc escindidos para la ligadura del adaptador de afinidad; y 2. ligar un adaptador de afinidad a los sitios de escisión de los fragmentos de ADNdc escindidos, para obtener fragmentos de ADNdc modificados por adaptador de afinidad. La forma de preparar un fragmento de ARNdc escindido para la ligadura del adaptador de afinidad es conocida en la técnica. La preparación de los fragmentos de ARNdc escindidos para la ligadura del adaptador de afinidad puede comprender una o ambas reparaciones de los extremos y la formación de cola dA del ARNdc escindido.

En una realización específica de la presente invención, los sitios de escisión de los fragmentos de ADNdc escindidos se preparan para la ligadura de adaptador de afinidad mediante reparación de extremos, para preparar fragmentos de ADNdc escindidos, que están libres de salientes, y contienen grupos fosfato 5' e hidroxilo 3', y formación de cola dA. A los extremos de un fragmento de ARNdc escindido así preparado puede ligarse un adaptador de afinidad basado en ARNdc con un saliente dT en el extremo 3' de un extremo del adaptador.

Después de la ligadura, el adaptador de afinidad se utiliza para enriquecer fragmentos de ARNdc modificados por adaptador de afinidad, para obtener fragmentos de ARNdc modificados por adaptador de afinidad enriquecidos. En este contexto, "enriquecer con fragmentos de ARNdc modificados por adaptador de afinidad" se refiere a la unión específica, transitoria o permanente, a los fragmentos de ARNdc modificados por adaptador de afinidad a través del adaptador de afinidad y la separación de los fragmentos de ARNdc, que no están modificados por adaptador de afinidad, y opcionalmente otros componentes. Por lo tanto, la forma en que se enriquecen los fragmentos de ARNdc modificados por adaptador de afinidad depende del adaptador de afinidad utilizado (es decir, del tipo de etiqueta(s) de afinidad utilizada(s) en el adaptador de afinidad). Los protocolos para enriquecer los fragmentos de ARNdc modificados por adaptador de afinidad específico son conocidos en la técnica.

De acuerdo con una realización preferida, se utiliza una molécula de unión a adaptador de afinidad, que es preferentemente inmovilizable sobre un soporte sólido, para enriquecer fragmentos de ARNdc modificados por adaptador de afinidad. Una molécula de unión al adaptador de afinidad se refiere a una molécula, por ejemplo, proteína, anticuerpo, etc., que se une específicamente (de forma reversible o irreversible) al adaptador de afinidad. La molécula de unión al adaptador de afinidad puede inmovilizarse en un soporte sólido antes o después de la unión al adaptador de afinidad. La naturaleza de la molécula de unión del adaptador de afinidad depende del adaptador de afinidad utilizado. La molécula de unión del adaptador de afinidad puede ser, por ejemplo, una proteína. El soporte sólido no está particularmente limitado e incluye, por ejemplo, perlas, por ejemplo, perlas magnéticas. En una realización preferida se utilizan perlas magnéticas como soporte sólido, preferentemente perlas magnéticas recubiertas de polímero, como las Dynabeads.

De acuerdo con una realización, la molécula de unión al adaptador de afinidad es una proteína, preferentemente seleccionada de estreptavidina; avidina; anticuerpos o fragmentos funcionales de los mismos específicos de una parte del adaptador de afinidad, por ejemplo, anticuerpos anti-digoxigenina; proteína de unión a digoxigenina; o una variante de los mismos; más preferentemente estreptavidina, avidina o una variante de las mismas. Los anticuerpos o fragmentos funcionales de los mismos específicos de una parte del adaptador de afinidad se unen preferentemente de forma específica a una o más modificaciones adecuadas para la purificación por afinidad en el adaptador de afinidad, por ejemplo, una digoxigenina o una etiqueta proteica.

De acuerdo con una realización preferida, el adaptador de afinidad comprende una o más biotinas o modificaciones basadas en biotina, y la molécula de unión del adaptador de afinidad es estreptavidina, avidina o una variante de la misma, preferentemente estreptavidina o una variante de la misma. Los protocolos para enriquecer ADN modificado con una o más biotina o modificaciones basadas en biotina son conocidos en la técnica. La estreptavidina o una variante de la misma puede utilizarse inmovilizada en perlas magnéticas recubiertas con polímero. Por ejemplo, se pueden utilizar las dynabeads MyOne Streptavidin C1, Dynabeads M-270, Dynabeads M-280 o Dynabeads MyOne Streptavidin T1.

Los fragmentos de ARNdc modificados por adaptador de afinidad enriquecidos se amplifican mediante amplificación por PCR para obtener una biblioteca de ADN escindida y amplificada de fragmentos de ARNdc. La amplificación por PCR de fragmentos de ARNdc, incluyendo los fragmentos de ARNdc modificados por adaptador de afinidad descritos en el presente documento, es conocida en la técnica. En una realización preferida, para la amplificación por PCR se utilizan cebadores PCR específicos de partes del adaptador de bloqueo y/o del adaptador de afinidad (es decir, complementarios de parte de la secuencia del adaptador de bloqueo y/o del adaptador de afinidad).

Si los fragmentos de ARNdc modificados por adaptador de afinidad se han enriquecido utilizando una molécula de unión a adaptador de afinidad inmovilizada en un soporte sólido, los fragmentos de ARNdc modificados por adaptador de afinidad unidos a través de la molécula de unión a adaptador de afinidad al soporte sólido pueden servir como molde para la amplificación por PCR. Alternativamente, los fragmentos de ARNdc modificados por adaptador de afinidad enriquecidos se separan (por ejemplo, se liberan) de la molécula de unión a adaptador de afinidad y/o del soporte sólido antes de la amplificación por PCR. En una realización preferida, la amplificación por PCR se lleva a cabo utilizando, al menos inicialmente, el fragmento de ARNdc modificado por adaptador de afinidad unido mediante la molécula de unión a adaptador de afinidad al soporte sólido como molde. En este contexto, el término "molde" se refiere a la función de un fragmento de ARNdc modificado por adaptador de afinidad enriquecido para servir como molde de ADN para dirigir la síntesis de otro ADN o ARN durante la amplificación por PCR.

De acuerdo con una realización preferida de la presente invención, en el paso b4), obtener la biblioteca de ADN escindido y amplificado de fragmentos de ADNdc comprende una amplificación por PCR en dos pasos, preferentemente con 1. un primer paso de amplificación por PCR para amplificación por PCR lineal utilizando fragmentos de ADNdc modificados por adaptador de afinidad enriquecidos unidos mediante una molécula de unión a adaptador a un soporte sólido como un molde y un cebador específico de adaptador de afinidad; y 2. después del primer paso, un segundo paso de amplificación por PCR 2. Después del primer paso, un segundo paso de amplificación por PCR, después de la eliminación de fragmentos de ADNdc modificados con adaptador de afinidad enriquecidos unidos a través de una molécula de unión a adaptador al soporte sólido de la mezcla de reacción y la adición de un cebador específico de adaptador de bloqueo como segundo cebador.

Paso c) - Secuenciación

El procedimiento del primer aspecto de la presente invención comprende el paso adicional (c) de realizar una secuenciación de la biblioteca de ADN escindida y amplificada de fragmentos de ARNdc para obtener lecturas de secuencia. La forma de realizar la secuenciación de una biblioteca de ADN de fragmentos de ARNdc es conocida en la técnica.

Tal como se utiliza en el presente documento, "secuenciación" incluye cualquier procedimiento de determinación de la secuencia de un ácido nucleico. El procedimiento de secuenciación que se utilizará en los presentes procedimientos no está particularmente limitado e incluye, por ejemplo, la secuenciación con terminador de cadena (Sanger) y la secuenciación con terminador de colorante. En las realizaciones preferidas, se utiliza la secuenciación de próxima generación (NGS). Aunque las distintas plataformas de NGS utilizan distintas químicas de ensayo, por lo general generan datos de secuencias a partir de un gran número de reacciones de secuenciación ejecutadas simultáneamente en un gran número de plantillas. Normalmente, los datos de la secuencia se recogen mediante un escáner y, a continuación, se ensamblan y analizan bioinformáticamente. Así, las reacciones de secuenciación se realizan, leen, ensamblan y analizan en paralelo (véase, por ejemplo, el documento US 2014/0162897y MacLean et al., 2009 Nature Rev. Microbiol. 7: 287-296).

Algunos procedimientos NGS requieren amplificación de la plantilla y otros no. Los procedimientos que requieren amplificación incluyen la pirosecuenciación; la plataforma Solexa/Illumina; o la ligadura y detección de oligonucleótidos soportada (SOLiD) (véase, por ejemplo, los documentos EP 0 946 752; EP 1634 963; EP 0 777 749; EP 1 117 827y EP 0 777 749y la Pante de EE. UU. No. 6,969,488; y 6,130,073). Los procedimientos que no requieren amplificación

incluyen, por ejemplo, procedimientos de secuenciación de una sola molécula como la secuenciación por nanoporos; la secuenciación de una sola molécula HeliScope; y la secuenciación en tiempo real mediante procedimientos de secuenciación de ADN en tiempo real de una sola molécula de síntesis (SMRT) utilizando guías de onda de modo cero (véase, por ejemplo los documentos EP 2 007 908; EP 1 848 829; EP 2 100 971y la Patente de EE.UU No. 7,169,560; 7,501,245; 6,818,395; 7,501,245; y 7,170,050). En una realización preferida se utiliza la secuenciación Illumina.

En una realización preferida, en el paso c) para la secuenciación de la biblioteca de ADN escindido y amplificado de fragmentos de ADNdc, en el paso c1) la biblioteca de ADN escindido y amplificado de fragmentos de ADNdc se prepara para la secuenciación, opcionalmente ligando adaptadores de secuenciación a ambos extremos de los fragmentos de ADNdc; y en el paso c2) la biblioteca de ADN escindido y amplificado de fragmentos de ADNdc preparada se secuencia, preferentemente por NGS, para obtener las lecturas de secuencia. En una realización específica, en el paso c1), la preparación de la biblioteca de ADN escindido y amplificado de fragmentos de ADNdc para secuenciación comprende la preparación de la biblioteca de ADN escindido y amplificado de fragmentos de ADNdc para la ligadura de adaptadores de secuenciación a ambos extremos, preferentemente mediante una o ambas reparaciones de extremos y formación de cola dA de la biblioteca de ADN escindido y amplificado de fragmentos de ADNdc; a continuación, se ligan los adaptadores de secuenciación, preferentemente a ambos extremos de los fragmentos de ADNdc, y los adaptadores de secuenciación son, por ejemplo, adaptadores de secuenciación de Illumina. En el paso c2) puede utilizarse preferentemente la PCR digital en gotas (PCRdd), por ejemplo, con el kit de cuantificación de bibliotecas PCRdd para Illumina TruSeq.

La secuenciación en el paso (c) no requiere un secuenciador de alto rendimiento como por ejemplo HiSeq. En su lugar, se puede utilizar un secuenciador NGS básico como un instrumento MiSeq (Illumina), por ejemplo, utilizando MiSeq Reagent Kit v2.

Paso d) - Alineación de las lecturas de secuencias con un genoma de referencia

El procedimiento del primer aspecto de la invención comprende el paso adicional (d) alinear las lecturas de secuencia con un genoma de referencia para obtener lecturas de secuencia alineadas. En este contexto, por alineación se entiende la correspondencia entre las lecturas de secuencias y las secuencias correspondientes del genoma de referencia, de modo que, para una lectura de secuencia específica, la secuencia correspondiente del genoma de referencia coincida lo mejor posible (es decir, que varía desde una coincidencia incompleta hasta la identidad de la secuencia) con la secuencia de la lectura de secuencia. Son conocidas en la técnica las formas de alinear las lecturas de secuencias con un genoma de referencia. Por ejemplo, se puede utilizar cualquier programa informático siempre que las lecturas de secuencia se puedan alinear como se ha descrito anteriormente, que puede ser un programa ya conocido en la técnica pertinente, o un programa adaptado al propósito. En una realización específica de la presente divulgación, las lecturas de secuencia se alinean con un genoma de referencia utilizando Bowtie2.

De acuerdo con una realización preferida de la presente invención, el ADN genómico es ADN genómico humano y el genoma de referencia es un genoma humano de referencia, por ejemplo, seleccionado de hg16 (NCBI Build 34), hg17(NCBI Build 35), hg18 (NCBI Build 36.1), hg19 (GRCh37) y hg38 (GRCh38), preferentemente seleccionado de hg19 y hg38.

En una realización preferida específica, en el paso d) las lecturas de secuencia se alinean con un genoma de referencia para obtener lecturas de secuencia alineadas, primero en el paso d1) recortando la secuencia del adaptador de bloqueo, la secuencia del adaptador de afinidad y la secuencia del adaptador de secuenciación opcional de las lecturas de secuencia para obtener lecturas de secuencia recortadas; y segundo en el paso d2) mapeando las lecturas de secuencia recortadas con un genoma de referencia para obtener lecturas de secuencia alineadas.

Paso e) - Identificación de los puntos de escisión

El procedimiento del primer aspecto de la invención comprende el paso adicional (e) de identificar los sitios de escisión, determinando los sitios de escisión putativos y excluyendo los sitios de escisión falsos positivos, en el que los sitios de escisión putativos se determinan localizando regiones de secuencia en el genoma de referencia donde coinciden los inicios y los finales de las lecturas de secuencia alineadas, e identificando los sitios de escisión putativos como los sitios en estas regiones de secuencia, donde coinciden los inicios y los finales de las lecturas de secuencia.

Así pues, la determinación de los sitios de escisión no se basa en la homología de secuencia con los sitios de escisión dentro del objetivo de una o más nucleasas programables específicas de objetivo.

El término "sitio de escisión putativo" puede incluir cualquier sitio de escisión escindido por una o más nucleasas programables específicas de objetivo *in vitro*, así como sitios de escisión falsos positivos. Los sitios de escisión falsos positivos son sitios que durante la detección de sitios de escisión se determinan como sitios de escisión putativos sin ser un sitio escindido por la una o más nucleasas programables específicas de objetivo *in vitro* y, por lo tanto, se identificaron incorrectamente, es decir, falsos positivos. El término "sitio de escisión putativo", tal y como se utiliza en el presente documento, se refiere a una secuencia corta de nt (es decir, una secuencia corta de ADN en el genoma

de referencia), preferentemente de longitud fija (es decir, que no varía de un sitio de escisión putativo a otro). En una realización preferida, un sitio de escisión putativo determinado en el paso e) consiste en la secuencia nt de una ventana específica de anchura de secuencia establecida (es decir, una ventana que cubre una longitud de secuencia definida del genoma de referencia, que no varía de una ventana a otra), preferentemente de una ventana de aproximadamente 6 nt o menos, más preferentemente de una ventana de (aproximadamente) 4 nt. En otras palabras, en esta realización preferida, el sitio de escisión putativo corresponde a la secuencia de ADN en el genoma de referencia cubierto por la ventana específica.

Las regiones de secuencia en el genoma de referencia en las que coinciden tanto los inicios como los finales de las lecturas de secuencia alineadas son regiones en el genoma de referencia de longitud de secuencia variable o establecida, que comprenden lecturas de secuencia alineadas, que comprenden tanto los inicios de las lecturas de secuencia como los finales de las lecturas de secuencia, es decir, regiones de secuencia en las que, de todas las lecturas de secuencia alineadas en dicha región que están alineadas allí con el genoma de referencia, una o más de las lecturas de secuencia alineadas comienzan y una o más lecturas de secuencia terminan en esa región de secuencia. En una realización preferida, cada región de secuencia es una zona de longitud de secuencia definida, preferentemente de aproximadamente 20 a 1.000 nt, más preferentemente de aproximadamente 50 a 500 nt, aún más preferentemente de al menos aproximadamente 50 nt, aún más preferentemente de al menos aproximadamente 80 nt, más preferentemente de al menos (aproximadamente) 100 nt. En otra realización preferida, en el paso e) cada región de secuencia es un intervalo de longitud de secuencia definida, preferentemente de aproximadamente 100 nt o un múltiplo de los mismos, más preferentemente de 100 nt o un múltiplo de los mismos (es decir, 100 nt, 200 nt, 300 nt, 400 nt, etc.).

Un inicio de lectura de secuencia se refiere a un extremo de una lectura de secuencia alineada que está en el lado opuesto del extremo de lectura de secuencia. Para las lecturas de secuencias alineadas con la misma cadena de nucleótidos del genoma de referencia, todos los inicios de las lecturas de secuencias apuntan en la misma dirección y todos los finales de las lecturas de secuencias apuntan en dirección opuesta a la cadena de nucleótidos de la secuencia del genoma de referencia. Esto significa, preferentemente, que si para una primera lectura de secuencia alineada, que es complementaria a una primera secuencia monocatenaria del genoma de referencia, el inicio de la lectura de secuencia está en el extremo 3' de esta primera lectura de secuencia alineada, para una hipotética segunda lectura de secuencia alineada, que es exactamente de la misma longitud y exactamente complementaria a la primera lectura de secuencia alineada, el inicio de la lectura de secuencia estaría en el extremo 5' de la segunda lectura de secuencia alineada, de manera que ambas lecturas de secuencia alineadas comienzan en la misma ubicación espacial en el genoma de referencia. Así, de acuerdo con una realización específica de la presente invención, cualquier inicio de lectura de secuencia de lecturas de secuencia alineadas, alineadas con la misma secuencia continua del genoma de referencia apunta en dirección opuesta a cualquier final de lectura de secuencia de lecturas de secuencia alineadas, alineadas con esa misma secuencia continua del genoma de referencia; es decir todos los inicios de lecturas de secuencias de lecturas de secuencias alineadas, alineadas con la misma secuencia continua del genoma de referencia, están situados espacialmente en secuencias de lecturas de secuencias alineadas en un lado de los respectivos centros de secuencias de lecturas de secuencias alineadas individuales, y todos los finales de lecturas de secuencias de lecturas de secuencias alineadas, alineadas con la misma secuencia continua del genoma de referencia, están situados espacialmente en las secuencias de lecturas de secuencias alineadas en el lado opuesto de los respectivos centros de secuencias de lecturas de secuencias alineadas individuales (en comparación con los inicios de lecturas de secuencias).

En las regiones de secuencia así identificadas, los sitios de escisión putativos se identifican como los sitios en estas regiones de secuencia, donde se encuentran los inicios de las lecturas de secuencia y los finales de las lecturas de secuencia. El término "encontrarse" en este contexto significa que los inicios de las lecturas de secuencia y los finales de las lecturas de secuencias que apuntan la una a la otra están muy próximos, es decir, a unos pocos nt, o se tocan, es decir, se alinean limpiamente de forma que ningún nt los separa en la secuencia del genoma de referencia a la que se alinean las lecturas, o se solapan unos pocos nt. Los posibles puntos de escisión en estas regiones de la secuencia pueden identificarse, por ejemplo, encontrando los máximos locales de la suma de inicios de lectura y finales de lectura de lecturas de secuencia alineadas en estas regiones de secuencia.

En una realización específica de la presente invención, en el paso e) en cada región de secuencia con una cobertura de inicio y final de lectura distinta de cero, se determinan uno o más sitios de escisión putativos (por ejemplo, exactamente un sitio de escisión putativo por región de secuencia, o alternativamente dos sitios de escisión putativos por región de secuencia, etc.) como los sitios donde los inicios de lectura de secuencia y los finales de lectura de secuencia de las lecturas de secuencia alineadas coinciden y donde la suma del número de inicios de lectura de secuencia y finales de lectura de secuencia tiene un máximo local. Una cobertura de inicio de lectura y final de lectura distinta de cero significa que dicha región de secuencia es una región donde se alinean una o más lecturas de inicio y uno o más finales de lectura de secuencia, de lecturas de secuencia alineadas.

De acuerdo con una realización preferida de la presente invención, en el paso e) se determinan sitios de escisión putativos dividiendo en primer lugar (la secuencia de ADN del) genoma de referencia que comprende lecturas de secuencia alineadas en intervalos de una longitud definida; y en segundo lugar determinando para cada intervalo con una cobertura de inicio de lectura y final de lectura distinta de cero, uno o más sitios de escisión putativos como sitios

con un máximo local para la suma de inicios de lectura de secuencia y finales de lectura de secuencia, y opcionalmente minimizando la distancia al centro del intervalo; donde un inicio de lectura de secuencia consiste en los primeros 1 a 5 nt, preferentemente los primeros 2 a 4 nt, más preferentemente los primeros 3 nt, y un final de lectura de secuencia consiste en los últimos 1 a 5 nt, preferentemente los últimos 2 a 4 nt, más preferentemente los últimos 3 nt, de una lectura de secuencia alineada. En una realización específica de la presente invención, se determina un sitio putativo de escisión para cada intervalo.

En una realización preferida de la presente invención, un sitio de escisión putativo determinado en el paso e) consiste en el nt de una ventana específica de anchura de secuencia establecida (es decir, una ventana que cubre una longitud de secuencia definida del genoma de referencia, que no varía de un sitio de escisión putativo a otro), preferentemente de una ventana de 4 nt. En otras palabras, en esta realización preferida, el sitio de escisión putativo corresponde a la secuencia de ADN en el genoma de referencia cubierto por la ventana específica. Preferentemente dicha ventana es una ventana de anchura definida, por ejemplo 4 nt, en un intervalo, y tiene un máximo local para la suma de inicios de lectura de secuencia y finales de lectura de secuencia (alineados en dicha ventana). Opcionalmente, la determinación de dicha ventana incluye la minimización de la distancia al centro del intervalo como paso adicional.

Un intervalo con una cobertura de inicio de lectura y final de lectura distinta de cero se refiere a un intervalo donde se alinean uno o más inicios de lectura de secuencia y uno o más finales de lectura de secuencia, de lectura de secuencia alineadas En la realización anterior, la anchura del intervalo (es decir, la longitud de la secuencia de nucleótidos cubierta por el intervalo) no está particularmente limitada y puede ser, por ejemplo, aproximadamente 20 a 1000 nt, preferentemente aproximadamente 50 a 500 nt, más preferentemente de al menos aproximadamente 50 nt, incluso más preferentemente de al menos aproximadamente 80 nt. En una realización preferida, la anchura del intervalo es de al menos aproximadamente 100 nt, preferentemente de al menos 100 nt. En otra realización, la anchura del intervalo es de aproximadamente 100 nt o un múltiplo del mismo, más preferentemente 100 nt o un múltiplo del mismo (es decir, 100 nt, 200 nt, 300 nt, 400 nt, etc.).

De acuerdo con una realización específica de la presente invención, en el paso e) los sitios de escisión se identifican e1) dividiendo (la secuencia de ADN de) el genoma de referencia que comprende las lecturas de secuencia alineadas en intervalos de longitud de secuencia establecida, preferentemente de aproximadamente 100 nt de longitud; e2) determinando preferentemente para cada intervalo una cobertura de lectura, descartando los intervalos con una cobertura de lectura de 0 y combinando preferentemente los intervalos con una cobertura de lectura de >0 consecutivamente, basándose en el orden de secuencia en el genoma de referencia, en un conjunto continuo de intervalos; e3) dividir cada intervalo en ventanas de anchura (de secuencia) establecida, preferentemente de 4 nt, con un tamaño de paso establecido, que es igual o menor, preferentemente menor, que la anchura de las ventanas, preferentemente de 2 nt; e4) determinar el inicio, el final y la cobertura de fondo de cada ventana, agrupando,

- los inicios de lectura de la secuencia que caen en la región de la ventana en un grupo de cobertura de inicio;
- secuenciar los finales de lectura que caen en la región ventana en un grupo de cobertura final; y
- las lecturas de la secuencia que caen en la región de la ventana, pero no pertenecen al grupo de cobertura inicial y al grupo de cobertura final en un grupo de cobertura de fondo;

y sumar para cada ventana el número de lecturas en la cobertura inicial, cobertura final y cobertura de fondo, respectivamente; e5) determinar para cada ubicación que comprende al menos una lectura de secuencia inicial y al menos una lectura de secuencia final, uno o más sitios de escisión putativos maximizando para la suma de cobertura inicial y cobertura final y minimizando la distancia al centro de la ubicación; y e6) excluir los sitios de escisión falsos positivos; en los que el inicio de una lectura de secuencia consiste en los primeros 1 a 5 nt, preferentemente los primeros 2 a 4 nt, más preferentemente los primeros 3 nt, y el final de una lectura de secuencia consiste en los últimos 1 a 5 nt, preferentemente los últimos 2 a 4 nt, más preferentemente los últimos 3 nt, de una lectura de secuencia alineada.

Preferentemente, cada uno de los uno o más sitios de escisión putativos determinados en el paso e5) consiste en el nt de una ventana específica de anchura de secuencia establecida (es decir, una ventana que cubre una longitud de secuencia definida del genoma de referencia, que no varía de un sitio de escisión putativo a otro), preferentemente de una ventana de 4 nt, establecida en el paso e3). El tamaño del paso, tal como se utiliza en el presente documento, se refiere preferentemente a la diferencia en nt del primer nt de una primera ventana y el primer nt de una segunda ventana, que es la ventana situada justo al lado de la primera ventana a lo largo de la secuencia del genoma de referencia (es decir, la ventana situada más cerca de la primera ventana cuando se desplaza desde la primera ventana a lo largo de la secuencia del genoma de referencia en una dirección). Un tamaño de paso fijo se refiere a un tamaño de paso que no varía entre diferentes ventanas posicionadas lo más cerca posible unas de otras, por ejemplo, un tamaño de paso de 2 nt.

En una realización preferida específica de la presente invención, el paso e1) comprende dividir (la secuencia de ADN de) el genoma de referencia que comprende lecturas de secuencia alineadas en intervalos de longitud de secuencia establecida (intervalos preliminares), preferentemente de aproximadamente 100 nt de longitud; y el paso e2)

comprende determinar para cada intervalo (preliminar) de e1) una cobertura de lectura, descartar intervalos con una cobertura de lectura de 0 y preferiblemente combinar intervalos con una cobertura de lectura >0 consecutivamente, con base en el orden de secuencia en el genoma de referencia, en un conjunto continuo de intervalos (intervalos finales); en el que en el conjunto continuo de intervalos del paso e2), preferiblemente, cada intervalo (final) consiste en uno o un múltiplo de uno de los intervalos del paso e1) (es decir, 1, 2, 3, 4, 5, etc. de intervalos preliminares).

Preferentemente, en esta realización preferida específica, en el conjunto continuo de intervalos del paso e2), cada intervalo (final) tiene una anchura de intervalo, que se define por la suma de intervalos consecutivos del paso e1) entre dos intervalos descartados del paso e2) más cercanos entre sí en la secuencia del genoma de referencia, de tal manera que cada intervalo (final) del paso e2) inicia en la secuencia del genoma de referencia con el intervalo (es decir, con el inicio de este intervalo) del paso e1) (intervalo preliminar) que sigue inmediatamente después de un primer intervalo descartado en el paso e2) y finaliza con el intervalo (es decir, con el final de este intervalo) del paso e1) que está inmediatamente antes del siguiente intervalo (es decir, el siguiente en la secuencia del genoma de referencia) descartado en el paso e2). que está inmediatamente antes de la siguiente (es decir, la siguiente en la secuencia del genoma de referencia) casilla descartada en la etapa e2). Esto significa que, en esta realización, los intervalos individuales (finales) del paso e2) pueden tener una longitud diferente (es decir, cubrir una longitud diferente en la secuencia del genoma de referencia), que sin embargo es siempre uno o un múltiplo de uno de los intervalos del paso e1) (es decir, 1, 2, 3, 4, 5, etc. intervalos del paso e1)).

En otra realización específica de la presente invención, el paso e4) comprende además determinar la cobertura inicial, final y de fondo de cada intervalo, agrupando,

- los inicios de lectura de la secuencia que caen en la región de intervalo en un grupo de cobertura de inicio;
- secuenciar los finales de lectura que caen en la región de intervalo en un grupo de cobertura final; y
- las lecturas de secuencias que caen en la región de intervalo pero que no pertenecen al grupo de cobertura inicial y al grupo de cobertura final en un grupo de cobertura de fondo;

y sumando para cada intervalo el número de lecturas en la cobertura inicial, la cobertura final y la cobertura de fondo, respectivamente.

En el paso e3), la anchura establecida de las ventanas (es decir, la longitud establecida y constante de la secuencia de nucleótidos del genoma de referencia cubierta por cada ventana) y el tamaño de paso establecido no están particularmente limitados. En una realización preferida, la anchura establecida de las ventanas, que es constante para todas las ventanas, es por ejemplo 8 nt o menos, preferentemente 6 nt o menos, más preferentemente 4 nt; y el tamaño de paso establecido es igual o menor que la anchura de las ventanas, preferentemente menor que la anchura de las ventanas, y es por ejemplo 6 nt o menos, preferentemente 4 nt o menos (por ejemplo 4, 3, 2, o 1 nt), más preferentemente 2 nt. En una realización específica, se determina un sitio de escisión putativo para cada intervalo en el paso e5). En una realización preferida de la presente invención, un sitio de escisión putativo determinado en el paso e) consiste en el nt de una ventana específica (es decir, la secuencia cubierta por una ventana específica) definida en el paso e3), preferentemente una ventana de 4 nt.

En un sitio de escisión putativo específico, los inicios de las lecturas de secuencia y los finales de las lecturas de secuencia, respectivamente, de las lecturas de secuencia alineadas de ese sitio de escisión putativo, pueden no alinearse perfectamente en una posición específica en la secuencia del genoma de referencia. De acuerdo con una realización, los inicios de lectura de secuencia de las lecturas de secuencia alineadas en un inicio de sitio de escisión putativo dentro de +/- 2 nt o menos, preferentemente +/- 1 nt o menos, de una primera posición específica en la secuencia del genoma de referencia; y los finales de lectura de secuencia de las lecturas de secuencia alineadas en un final de sitio de escisión putativo dentro de +/- 2 nt o menos, preferentemente +/- 1 nt o menos, de una segunda posición específica en la secuencia del genoma de referencia; las dos posiciones específicas preferentemente están a sólo unos pocos nt (por ejemplo, 1. 2, o 3 nt de distancia entre sí o son directamente adyacentes entre sí en la secuencia del genoma de referencia, 2, o 3 nt) entre sí o directamente adyacentes entre sí en la secuencia del genoma de referencia.

Tras la determinación de los sitios de escisión putativos, los sitios de escisión falsos positivos (en el presente documento también denominados "sitios falsos positivos") se excluyen de la lista de sitios de escisión putativos para identificar los sitios de escisión. El término "sitio de escisión", tal como se utiliza en el presente documento, se refiere a una secuencia corta de nt (es decir, una secuencia corta en el genoma de referencia), preferentemente de longitud fija (es decir, que no varía de un sitio de escisión putativo a otro). En una realización preferida, un sitio de escisión tiene una longitud de secuencia de (aproximadamente) 4 nt. En una realización específica, un sitio de escisión identificado en el paso e) consiste en el nt de una ventana específica de anchura de secuencia establecida (es decir, una ventana que cubre una longitud de secuencia definida del genoma de referencia, que no varía de una ventana a otra), preferentemente de una ventana de aproximadamente 6 nt o menos, más preferentemente de una ventana de (aproximadamente) 4 nt. En una realización preferida de la presente invención, los sitios de escisión falsos positivos se excluyen basándose en la relación entre la cobertura de fondo y la cobertura de la señal en un sitio de escisión

putativo; y/o la relación de la cobertura inicial a la cobertura final en un sitio de escisión putativo; y/o la cobertura del sitio; y/o la cobertura del intervalo. Preferentemente, la escisión real de un sitio de escisión del paso e) por una de las una o más nucleasas programables específicas del objetivo se produce dentro de la secuencia de ese sitio de escisión específico, es decir, para un sitio de escisión, que consiste en el nt de una ventana específica de anchura de secuencia establecida, en algún lugar de la secuencia cubierta por esta ventana específica.

La cobertura de inicio en un sitio de escisión putativo es la suma de inicios de lecturas de secuencia alineadas en la región de secuencia del sitio de escisión putativo, es decir, la suma de lecturas de secuencia alineadas, que tienen su inicio de lectura de secuencia alineado en la región de secuencia del genoma de referencia definida por el sitio de escisión putativo. En una realización preferida, la cobertura de inicio en un sitio de escisión putativo es la suma de inicios de lectura de secuencia alineados en la región de secuencia de un sitio de escisión putativo, estando el sitio de escisión putativo definido por una ventana específica de anchura de secuencia establecida (es decir, la anchura no varía para diferentes ventanas correspondientes a diferentes sitios de escisión putativos), preferentemente una ventana con una anchura de secuencia de 4 nt. En otras palabras, en esta realización preferida, la cobertura de inicio en un sitio de escisión putativo es la suma de lecturas de secuencia alineadas, que tienen su inicio de lectura de secuencia alineado en la región de secuencia del genoma de referencia definida por la ventana específica, de por ejemplo 4 nt, que representa un sitio de escisión putativo (es decir, preferentemente una ventana establecida en el paso e3) descrito anteriormente).

De forma similar, la cobertura final en un sitio de escisión putativo es la suma de los extremos de lectura de secuencia alineados en la región de secuencia del sitio de corte putativo, es decir, la suma de lecturas de secuencia alineadas, que tienen su extremo de lectura de secuencia alineado en la región de secuencia del genoma de referencia definida por el sitio de escisión putativo. En una realización preferida, la cobertura final en un sitio de escisión putativo es la suma de los finales de lectura de secuencia alineados en la región de secuencia de un sitio de escisión putativo, estando el sitio de escisión putativo definido por una ventana específica de anchura de secuencia establecida (es decir, la anchura no varía para diferentes ventanas correspondientes a diferentes sitios de escisión putativos), preferentemente una ventana con una anchura de secuencia de 4 nt. En otras palabras, en esta realización preferida, la cobertura final en un sitio de escisión putativo es la suma de lecturas de secuencia alineadas, que tienen su final de lectura de secuencia alineado en la región de secuencia del genoma de referencia definida por la ventana específica, de por ejemplo 4 nt, que representa un sitio de escisión putativo (es decir, preferentemente una ventana establecida en el paso e3) descrito anteriormente).

Preferentemente, la cobertura de inicio de una ventana específica que define un sitio de escisión putativo específico, incluye también todos los inicios de lectura de secuencia, que se alinean sólo con parte de la secuencia del inicio de lectura de secuencia en la región de secuencia del genoma de referencia definido por la ventana específica, es decir, es suficiente que por ejemplo 1 nt, o 2 nt, etc. de un inicio de lectura de secuencia (dependiendo de la longitud del inicio de lectura de secuencia) se alinee en la región de secuencia de la ventana específica para que ese inicio de lectura de secuencia se incluya en el grupo de cobertura de inicio de una ventana específica. Del mismo modo, preferentemente, la cobertura final de una ventana específica que define un sitio de escisión putativo específico, incluye también todos los extremos de lectura de secuencia, que se alinean sólo con parte de la secuencia del final de lectura de secuencia en la región de secuencia del genoma de referencia definido por la ventana específica, es decir, es suficiente que, por ejemplo, 1 nt, o 2 nt, etc. de un extremo de lectura de secuencia (dependiendo de la longitud del extremo de lectura de secuencia) se alinee en la región de secuencia de la ventana específica para que ese final de lectura de secuencia se incluya en el grupo de cobertura final de una ventana específica.

La cobertura de fondo en un sitio de escisión putativo es la suma de lecturas de secuencia alineadas en la región de secuencia del sitio de escisión putativo, pero que no pertenecen al grupo de cobertura inicial y al grupo de cobertura final, es decir la suma de lecturas de secuencia alineadas en la región de secuencia del genoma de referencia definida por el sitio de escisión putativo, sin que su inicio de lectura de secuencia o final de lectura de secuencia esté alineada en la región de secuencia del genoma de referencia definida por el sitio de escisión putativo; estando el sitio de escisión putativo definido opcionalmente por una ventana específica de anchura de secuencia establecida (es decir, la anchura no varía para diferentes ventanas correspondientes a diferentes sitios de escisión putativos), preferentemente una ventana con una anchura de secuencia de 4 nt (es decir, preferentemente una ventana establecida en el paso e3) descrito anteriormente).

El término "cobertura del sitio", tal como se utiliza en el presente documento, se refiere a la suma de la cobertura del inicio y la cobertura del final en un sitio de escisión putativo, es decir, la suma de los inicios de lectura de secuencia y los finales de lectura de secuencia de las lecturas de secuencias alineadas, que coinciden en un sitio de escisión putativo. En una realización preferida, el término "cobertura del sitio" se refiere a la suma de la cobertura del inicio y la cobertura del final en un sitio de escisión putativo, estando el sitio de escisión putativo definido por una ventana específica de anchura de secuencia establecida (es decir, la anchura no varía para diferentes ventanas correspondientes a diferentes sitios de escisión putativos), preferentemente una ventana con una anchura de secuencia de 4 nt; en otras palabras, el término "cobertura del sitio" se refiere a la suma de inicios de lectura de secuencia y finales de lectura de secuencia de lecturas de secuencia alineadas, alineadas con el genoma de referencia en la región de secuencia de una ventana específica de anchura de secuencia establecida, por ejemplo 4 nt, que representa un sitio de escisión putativo específico (es decir, preferentemente un conjunto de ventanas en el paso e3)

descrito anteriormente). El término "cobertura del intervalo" como se usa en el presente documento se refiere a la suma de todas las lecturas de secuencia alineadas en un intervalo (es decir, la suma de todas las lecturas de secuencia alineadas que se alinean en la región de secuencia del genoma de referencia de un intervalo específico).

En otra realización preferida de la presente invención, los sitios de escisión falsos positivos se excluyen seleccionando solamente sitios de escisión putativos con 1.) una relación de cobertura de fondo a cobertura de señal, que no excede un cierto umbral, preferentemente aproximadamente 0,5; y/o 2.) una relación de cobertura de inicio a cobertura de final en el sitio de escisión putativo, que cae dentro de un cierto intervalo numérico, preferentemente entre aproximadamente 1/5 y 5/1; y/o 3.) una cobertura de sitio mínima, de preferentemente aproximadamente 6 o más; y/o 4.) una cobertura de intervalo mínima, de preferentemente 10 o más.

En una realización específica de la presente invención, los sitios de escisión falsos positivos se excluyen seleccionando únicamente sitios de escisión putativos con 1. una relación de cobertura de fondo a cobertura de señal de 0,5 o menos en el sitio de escisión putativo; y 2. una relación de cobertura inicial a cobertura final de entre 1/5 y 5/1 en el sitio de escisión putativo; y 3. una cobertura de sitio de 6 o más; y 4. una cobertura de intervalo de 10 o más.

De acuerdo con otra realización alternativa específica de la presente invención, los sitios de escisión falsos positivos se excluyen o se excluyen adicionalmente utilizando un modelo de aprendizaje automático, preferentemente entrenado en sitios falsos positivos excluidos por el procedimiento descrito en las realizaciones anteriores, preferiblemente entrenado usando la relación de cobertura de fondo a cobertura de señal; relación de cobertura inicial a cobertura final; anchura de intervalo; distancia relativa entre sitios falsos positivos; y/o relación de sitios falsos positivos a intervalos como características. La forma de utilizar un modelo de aprendizaje automático, incluyendo la selección de un modelo adecuado, la selección de características y el entrenamiento, es conocida en la técnica. Por ejemplo, se puede construir un modelo de bosque aleatorio, por ejemplo, utilizando una validación cruzada de 3 veces 10 repeticiones y obteniendo un área bajo la curva de 1.

En una realización de la presente invención, cada sitio de escisión del paso e) constituye un sitio objetivo (es decir, un sitio dentro del objetivo o un sitio fuera del objetivo) de una o más nucleasas programables específicas de objetivo, de modo que preferentemente cada sitio objetivo (es decir, sitio dentro del objetivo o sitio fuera del objetivo) consiste en un sitio de escisión específico, que es diferente para cada sitio objetivo (es decir, uno de los sitios de escisión individuales identificados en el paso e)). En una realización específica, donde un sitio de escisión específico del paso e) consiste en el nt de una ventana específica de anchura de secuencia establecida, cada sitio objetivo (es decir, sitio dentro del objetivo o sitio fuera del objetivo) consiste en la secuencia de una ventana específica de anchura de secuencia establecida que define un sitio de escisión específico, de tal manera que cada ventana específica, que define un sitio de escisión específico, identificado en el paso e), da lugar a un sitio objetivo.

De acuerdo con otra realización de la presente invención, los sitios de escisión del paso e) están comprendidos en los sitios objetivo (es decir, sitio dentro del objetivo o sitio fuera de objetivo) de una o más nucleasas programables específicas de objetivo, de tal manera que cada sitio objetivo (es decir, sitio dentro del objetivo o sitio fuera del objetivo) comprende un sitio de escisión. El experto entiende que, en esta realización, un sitio de escisión determinado en el paso e), que puede consistir en el nt de una ventana específica de anchura de secuencia establecida en el paso e), no constituye necesariamente toda la secuencia del sitio objetivo (es decir, la secuencia del sitio dentro del objetivo o del sitio fuera del objetivo), ya que, por ejemplo, un sitio dentro del objetivo puede comprender, además, por ejemplo, una secuencia objetivo (es decir. la secuencia que una nucleasa programable específica de objetivo está diseñada para reconocer) y/o una secuencia de unión a nucleasas (para CRISPR/Cas RGN, por ejemplo, una secuencia PAM); del mismo modo, un sitio fuera del objetivo puede, en esta realización, comprender además, por ejemplo, una secuencia que es, en cierto grado (incluso en un grado muy pequeño, por ejemplo, 1 nt, o 2 nt, o 3 nt, etc. de homología de secuencia), homóloga a una secuencia objetivo y/o a una secuencia de unión a nucleasas.

Un sitio objetivo CRISPR/Cas RGN, en esta otra realización, puede comprender además del sitio de escisión una secuencia objetivo y una secuencia PAM. Por ejemplo, si una o más nucleasas programables específicas de objetivo utilizadas en el paso b) consisten en una o más CRISPR/Cas RGN, un sitio objetivo, en esta otra realización, puede comprender o consistir en una secuencia objetivo, por ejemplo, de aproximadamente 18-23 nt, preferentemente aproximadamente 20 nt, que es idéntica, su homóloga al menos en un grado muy bajo (es decir, por ejemplo, 1 nt, o 2 nt, o 3 nt, o 4 nt, o 5 nt, o 6 nt, o 7 nt, u 8 nt, etc. homología de secuencia), a la secuencia objetivo de una de las una o más nucleasas programables específicas de objetivo utilizadas en el paso b); y una secuencia PAM, por ejemplo de 2-6 nt, preferentemente de 3 nt, directamente adyacente a la secuencia objetivo. Así, en esta otra realización, en una realización preferida específica en la que la una o más nucleasas programables específicas de objetivo consisten en una o más CRISPR/Cas RGN, los sitios objetivo, preferentemente los sitios fuera del objetivo comprenden, o consisten en, una secuencia de, por ejemplo, aproximadamente 20-29 nt de longitud, preferentemente aproximadamente 23 nt. Tales secuencias de, por ejemplo, aproximadamente 23 nt pueden identificarse, por ejemplo, identificando los sitios de escisión del paso e) que tienen en estrecha proximidad o superposición con su secuencia con 1) una secuencia PAM de 2-6 nt preferentemente 3 nt, y 2) una secuencia objetivo, que es al menos hasta cierto grado, preferentemente muy bajo, homóloga (por ejemplo al menos 1 nt, o al menos 2 nt, o al menos 3 nt, o al menos 4 nt, o al menos 5 nt, o al menos 6 nt, o al menos 7 nt, o al menos 8 nt, o al menos 9 nt, o al menos 10 nt, o al menos 11 nt, o al menos 12 nt, o al menos 13 nt, etc. homología de secuencia) a una secuencia objetivo de uno de los de uno o más CRISPR/Cas

RGN utilizados en el paso b). Preferentemente, el sitio de escisión está situado en un sitio objetivo de este tipo (es decir, sitio dentro del objetivo o sitio fuera del objetivo) de, por ejemplo, 23 nt de longitud, cerca del extremo de la secuencia objetivo directamente adyacente a la secuencia PAM, por ejemplo, dentro de un intervalo de secuencias consistente en los primeros aproximadamente 1-7 nt del extremo de la secuencia objetivo directamente adyacente a la secuencia PAM y los aproximadamente primeros 1-2 nt del extremo de la secuencia PAM directamente adyacente a la secuencia objetivo.

De acuerdo con una realización específica del procedimiento del primer aspecto de la presente invención, los sitios de escisión fuera del objetivo de una o más nucleasas programables específicas de objetivo se identifican como sitios de escisión del paso (e) que no son sitios de escisión dentro del objetivo de la una o más nucleasas programables específicas de objetivo. De acuerdo con una realización específica alternativa del procedimiento del primer aspecto de la presente invención, los sitios de escisión fuera del objetivo de la una o más nucleasas programables específicas de objetivo se identifican como sitios objetivo que comprenden un sitio de escisión del paso (e) que no son sitios dentro del objetivo de la una o más nucleasas programables específicas de objetivo. En la práctica, esto puede hacerse identificando (por ejemplo, etiquetando) en la lista de sitios de escisión identificados en el paso e) los sitios dentro del objetivo de la una o más nucleasas programables específicas de objetivo, que se conocen como resultado del diseño de la una o más nucleasas programables específicas de objetivo. Esto puede hacerse por medios conocidos en la técnica.

Doble-Abnoba-Seq

En un segundo aspecto, la presente invención se refiere a un procedimiento para detectar sitios fuera del objetivo *in vivo* utilizando el procedimiento para detectar sitios fuera del objetivo de una o más nucleasas programables específicas de objetivo en un genoma *in vitro* del primer aspecto (Dual-Abnoba-Seq). En particular, el segundo aspecto de la presente invención se refiere a un procedimiento para detectar sitios fuera del objetivo en un genoma *in vivo* que comprende los pasos de i) proporcionar una primera muestra, que comprende ADN genómico de células tratadas con una o más nucleasas programables específicas de objetivo, y una segunda muestra, que comprende ADN genómico de células de fondo genético relacionado, preferentemente células del mismo fondo genético, no tratadas con la una o más nucleasas programables específicas de objetivo; ii) para ambas muestras, detectar independientemente los sitios fuera de objetivo *in vitro* utilizando el procedimiento del primer aspecto de la presente invención tal como se define en una cualquiera de las realizaciones descritas anteriormente y utilizando la una o más nucleasas programables específicas de objetivo del paso i); y iii) comparar los sitios fuera del objetivo entre las dos muestras para identificar los sitios fuera del objetivo solo identificados en la muestra que comprende ADN genómico de células no tratadas con la una o más nucleasas programables específicas de objetivo, pero no en la muestra que comprende ADN genómico de células tratadas con una o más nucleasas programables específicas de objetivo como sitios fuera del objetivo *in vivo* de la una o más nucleasas programables específicas de objetivo.

El procedimiento de acuerdo con el segundo aspecto de la invención permite la determinación individualizada de la especificidad *in vivo* de una o más nucleasas programables específicas de objetivo, en otras palabras, la capacidad de determinar la actividad fuera de objetivo real de la una o más nucleasas programables específicas de objetivo en células específicas. Esto significa que el procedimiento de acuerdo con el segundo aspecto de la presente invención puede utilizarse para demostrar la actividad de nucleasas programables específicas de objetivo *in vivo*, por ejemplo, en un tipo celular concreto.

El presente inventor descubrió que al aislar ADN genómico de células editadas genéticamente y células no tratadas y para ambas muestras de ADN genómico llevar a cabo el procedimiento para detectar sitios fuera de objetivo de una o más nucleasas programables específicas de objetivo en un genoma *in vitro* de acuerdo con el primer aspecto de la invención con las nucleasas programables específicas de objetivo utilizadas para la edición génica en las células editadas genéticamente, permite la determinación de la actividad fuera de objetivo real en las células.

Por comparación de los sitios fuera del objetivo identificados en la muestra de ADN genómico editado genéticamente con los sitios fuera del objetivo en la muestra de ADN genómico de las células no tratadas, puede determinarse la actividad *in* vivo real de una o más nucleasas programables específicas de objetivo, por ejemplo, en células de un tipo celular o tejido particular, ya que esos sitios ya no se identifican en la muestra de ADN genómico derivada de las células editadas genéticamente, utilizando cantidades mínimas de ADN genómico, por ejemplo de menos de 1 µg por muestra, permitiendo así el ensayo de los efectos *in* vivo fuera del objetivo en muestras celulares valiosas de tamaño de muestra muy pequeño y disponibilidad limitada.

Se conocen en la técnica formas de tratar células con una o más nucleasas programables específicas de objetivo (véase, por ejemplo, Yin et al., 2017, Delivery technologies for genome editing. Nat Rev Drug Discov 16, 387-399). Para ello, se introducen en las células una o más nucleasas programables específicas de objetivo por medios conocidos en la técnica y en concentraciones y forma adecuadas para la edición del genoma *in vivo.* Por ejemplo, si la una o más nucleasas programables específicas de objetivo comprende una o más CRISPR/Cas RGN, por ejemplo, el ARNg, la endonucleasa Cas, o ambos, pueden introducirse como proteína purificada y/o ácido nucleico en las células, o pueden expresarse de forma transitoria o estable en la célula, utilizando procedimientos conocidos en la

técnica. La una o más nucleasas programables específicas de objetivo puede ser una de las especificadas anteriormente en la primera realización de la presente invención.

En particular, este segundo aspecto de la invención se refiere a un procedimiento para detectar sitios fuera del objetivo en un genoma *in vivo* que comprende los pasos de i) proporcionar una primera muestra, que comprende ADN genómico de células tratadas con una o más nucleasas programables específicas de objetivo, y una segunda muestra, que comprende ADN genómico de células de fondo genético relacionado, preferentemente células del mismo fondo genético, no tratadas con la una o más nucleasas programables específicas de objetivo. Preferentemente, un fondo genético relacionado significa que dichas células tienen el mismo fondo genético, lo que significa que tienen las mínimas diferencias posibles, idealmente un código genético idéntico, es decir, el ADN genómico de estas células es el mismo excepto por pequeñas variaciones que pueden producirse de una célula a otra debido, por ejemplo, a mutaciones aleatorias. Preferentemente, las células relacionadas con el fondo genético son preferentemente células originarias de la misma región anatómica en un organismo multicelular, por ejemplo, un cuerpo humano, por ejemplo, del mismo tejido y/o del mismo tipo celular.

En una realización preferida del procedimiento del segundo aspecto de la presente invención, las células de fondo genético relacionado son células del mismo fondo genético, preferentemente con el mismo código genético, y preferentemente con el mismo origen anatómico, preferentemente células procedentes del mismo tejido y/o tipo celular. En otra realización, las células de origen genético relacionado son células del mismo linaje celular. Las células del mismo linaje celular, tal como se utilizan en el presente documento, se refieren muy generalmente a células derivadas de una célula ancestro común que son el resultado de la división celular mitótica y, opcionalmente, de la diferenciación celular. Preferentemente, las células del mismo linaje celular son células originarias del mismo tejido, y preferentemente células del mismo tipo celular.

Preferentemente, las células de fondo genético relacionado son células derivadas originalmente del mismo organismo individual. El tipo de organismo del que proceden las células de la primera y segunda muestra no está particularmente limitado siempre que las células sean adecuadas para su uso en el procedimiento del segundo aspecto de la invención. En una realización, las células son células eucariotas (por ejemplo, de un animal, un mamífero, un insecto, una planta, un hongo, un insecto, un ave, un pez, un anfibio, un reptil o un cnidario). En una realización específica, las células son humanas. Por ejemplo, las células pueden ser células primarias.

Para ambas muestras se detectan independientemente sitios fuera de objetivo *in vitro* utilizando el procedimiento del primer aspecto de la presente invención y utilizando la una o más nucleasas programables específicas de objetivo del paso i). Por último, se comparan los sitios fuera del objetivo entre las dos muestras para identificar los sitios fuera del objetivo solo identificados en la muestra que comprende ADN genómico de células no tratadas con la una o más nucleasas programables específicas de objetivo pero no en la muestra que comprende ADN genómico de células tratadas con la una o más nucleasas programables específicas de objetivo como sitios fuera del objetivo *in vivo* de la una o más nucleasas programables específicas de objetivo. En la técnica se conocen formas de comparar los sitios fuera del objetivo entre las dos muestras. Preferentemente, esto se hace utilizando un programa informático, existente o escrito específicamente para este fin. Por ejemplo, tras la normalización de la cobertura de los intervalos por la anchura de los intervalos, los sitios de escisión pueden clasificarse por su ubicación en el genoma y los resultados de las células editadas pueden escribirse junto a los resultados de las células no tratadas.

En otro aspecto, la presente invención se refiere a un kit para llevar a cabo el procedimiento de acuerdo con el primer aspecto de la presente invención, tal como se define en cualquiera de las realizaciones descritas anteriormente, y un kit para llevar a cabo el procedimiento de acuerdo con el segundo aspecto de la presente invención, tal como se define en cualquiera de las realizaciones descritas anteriormente. Las formas de preparar estos kits son conocidas en la técnica

Los siguientes ejemplos explican mejor la invención.

EJEMPLOS

Materiales y procedimientos aplicados en los ejemplos

Formación del complejo ribonucleoproteico

Para los experimentos *in vitro*, los ARNg se transcribieron *in vitro* a partir de amplicones de PCR utilizando el kit HiScribe T7 High Yield RNA Syntesis (NEB) seguido de una limpieza con el kit MEGAclear (Thermo Fisher Scientific). Para los experimentos basados en células, se adquirieron ARNg modificados químicamente (Synthego). La proteína Cas9 se produjo internamente siguiendo un protocolo publicado anteriormente (Jinek et al, 2012, Science 337(6096): 816-821). Para permitir la formación del complejo RNP, Cas9 y ARNg se incubaron a temperatura ambiente durante 10 min en una relación molar de 1:3 proteína Cas9 por ARNg.

Cultivo celular y transfección

Las células U2OS se cultivaron en DMEM con bajo contenido en glucosa y piruvato (Gibco) suplementado con 10% de FBS y 1% de penicilina/estreptomicina (Sigma-Aldrich). Las células CD34+ se purificaron a partir de sangre de cordón umbilical humano obtenida del Department of Obstetrics and Gynecology con consentimiento informado, mediante centrifugación en gradiente de densidad basado en Ficoll y posterior selección positiva utilizando el kit MicroBead UltraPure de CD34 humano (Miltenyi Biotec). Las células CD34+ se cultivaron en CellGro SCGM (CellGenix®) suplementado con 20 ng/µl de TPO, 60 ng/µl de SCF y 60 ng/µl de Flt-3L (CellGenix®). Antes de la electroporación, se añadieron 12 ng/µl de IL-3 al medio de cultivo de células CD34+. Las células se transfectaron con RNPs utilizando un Nucleofector 4D (Lonza Group AG) usando el kit SE y el programa DN-100 para células U2OS o el kit P3 y el programa CA-136 para células CD34+ de acuerdo con las instrucciones del fabricante.

Aislamiento y fragmentación del ADN genómico

El ADN genómico de las células se aisló utilizando el QIAamp DNA Blood Mini Kit (Qiagen). El cizallamiento del ADN genómico hasta una longitud media de 400 pb se realizó con un ultrasonido focalizado M220 (Covaris) siguiendo las instrucciones del fabricante.

Flujo de trabajo de Abnoba-Seq

Todos los cebadores utilizados se indican en la Tabla 1. Durante todo el flujo de trabajo, se utilizó el kit NEBNext Ultra II (NEB) y perlas AMPure XP (Beckman Coulter). Creación de lotes de fragmentos de ADN con extremos bloqueados: El ADN genómico esquilado (600 ng por reacción) se preparó en el extremo y el adaptador de bloqueo (112,5 pmol) se ligó utilizando el kit NEBNext Ultra II (NEB) durante 30 min a 20°C. Los fragmentos con extremos no bloqueados se eliminaron mediante tratamiento con exonucleasas: Exonucleasa III (200 U, NEB), Exonucleasa I (20 U, NEB) y Exonucleasa Lambda (50 U, NEB).

Escisión *in vitro* : Se escindieron *in vitro* fragmentos de ADN genómico de 50 ng con extremos bloqueados con 1,6 pmol de proteína Cas9 (fabricada en laboratorio) y 4,8 pmol de ARNg en un volumen de reacción de 25 µl con 1x NEB Buffer 3 y 1x BSA durante 1 h a 37 °C. La reacción se detuvo añadiendo 4 µg de RNasa y una solución STOP que contenía 50 % de EDTA 0,5 M pH 8,0, 30 % de glicerol y 1 % de SDS, incubándose durante 15 min a 37 °C cada una. Se realizó una PCR de control de calidad utilizando la polimerasa Q5 (NEB) y se amplificaron los sitios dentro del objetivo.

Biotinilación de fragmentos de ADN escindidos: Se prepararon los extremos de los fragmentos de ADN escindidos y se ligó el adaptador de biotina (75 pmol) utilizando el kit NEBNext Ultra II durante 30 min a 20 °C. Los fragmentos biotinilados se enriquecieron mediante Dynabeads MyOne Streptavidin C1 (Thermo Fisher Scientific) de acuerdo con el protocolo del fabricante. A continuación, los fragmentos de ADN biotinilados unidos a estreptavidina se aplicaron como molde para la amplificación lineal por PCR utilizando únicamente un cebador específico para el adaptador de biotina. Tras 10 ciclos, la reacción de PCR se detuvo a 4 °C y la reacción de PCR con los productos lineales de PCR se transfirió a un nuevo tubo, se añadió el cebador específico del adaptador de bloqueo y se continuó la PCR durante otros 10 ciclos. Los amplicones se aplicaron a la preparación de bibliotecas de ADN para prepararlas para la secuenciación de próxima generación.

Secuenciación de próxima generación: Las bibliotecas de ADN se cuantificaron mediante PCRdd con el kit de cuantificación de bibliotecas PCRdd para Illumina TruSeq (Bio-Rad) y la secuenciación se realizó en un instrumento MiSeq (Illumina) utilizando el kit de reactivos MiSeq v2 (300 ciclos).

Análisis fuera del objetivo: Los primeros pasos del análisis de los datos de secuenciación y algunos de los pasos posteriores se realizaron utilizando Galaxy en usegalaxy.eu (Afgan et al., 2018, Nucleic Acids Res 46(W1): W537-W544). En primer lugar, se recortaron las secuencias adaptadoras del adaptador Illumina, el adaptador de bloqueo y el adaptador de biotina de las lecturas utilizando Trim Galore con la configuración predeterminada. Las lecturas recortadas se asignaron al genoma humano de referencia (hg19) con Bowtie2 con la configuración predeterminada. A continuación, se analizaron adicionalmente las lecturas de secuencias alineadas cuantificando en primer lugar la cobertura total de lecturas dentro de una ventana deslizante de 100 nt sobre el genoma humano utilizando bedtools (Quinlan y Hall, 2010, Bioinformatics 26(6): 841-842). Las ventanas consecutivas no vacías con una cobertura >0 se fusionaron en intervalos, mientras que las ventanas vacías se descartaron. A continuación, los intervalos se dividieron en ventanas superpuestas de 4 nt con un paso de 2 nt. Se consideraron el primer y el último nucleótido (nt) de cada lectura para calcular las coberturas inicial y final, respectivamente. Las lecturas que no pertenecían ni al grupo de cobertura inicial ni al grupo de cobertura final se tuvieron en cuenta para la cobertura de fondo. Para cada uno de los intervalos con una cobertura inicial y final no nula, se obtuvo un sitio de escisión putativo maximizando la cobertura inicial y final y minimizando la distancia al centro del intervalo. Los sitios falsos positivos se filtraron aplicando los criterios relación fondo/señal ≤ 0,5, relación inicio/final ≥ 1/5 y ≤ 5, cobertura del sitio (cobertura inicial y final) ≥ 6 y cobertura del intervalo ≥ 10. Con el fin de eliminar aún más los sitios falsos positivos restantes, se entrenó un modelo de aprendizaje automático en los sitios anotados RNF2 basado en las características de la relación de inicio a fin, la relación de fondo a señal, la anchura del intervalo, la distancia relativa y la relación de sitio a intervalo. Los sitios RNF2 identificados por el proceso utilizado se clasificaron como verdaderos o falsos positivos basándose en la literatura y el conocimiento previo. Se construyó un modelo de bosque aleatorio en dos tercios de estos sitios utilizando una

validación cruzada triple de 10 repeticiones y se obtuvo un área bajo la curva de 1, tanto en los conjuntos de entrenamiento como en los de prueba. A continuación, se utilizó el modelo entrenado para clasificar los demás sitios objetivo y se filtraron los sitios predichos como falsos positivos.

Visualización: Para la visualización de los sitios objetivo identificados, se utilizó el Integrative Genomics Viewer (IGV) (Robinson et al., 2011, Nat Biotechnol 29(1): 24-26; Thorvaldsdottir et al., 2013, Brief Bioinform 14(2): 178-192).

Secuenciación selectiva de amplicones

Las células U2OS y las células CD34+ aisladas de la sangre del cordón umbilical se nucleofectaron con RNPs dirigidas a VEGFA y HEKs4, y el ADN genómico de las células tratadas y no tratadas se cosechó el día 7 después de la transfección. El sitio dentro del objetivo, así como los 20 principales sitios fuera del objetivo de las nucleasas CRISPR-Cas9 se amplificaron por PCR con tamaños de amplicón de alrededor de 300 pb utilizando Q5 Polymerase (NEB). Los amplicones de PCR se purificaron con el kit de purificación de PCR QIAquick (Qiagen) y la concentración se midió con el fluorómetro Qubit (Thermo Fisher Scientific). Se mezclaron 20 ng de cada uno de los 21 amplicones en grupos (no tratados/genes editados, U2OS/HSCs) y se aplicaron a la preparación de bibliotecas de ADN utilizando el kit NEBNext Ultra II (NEB). Las bibliotecas de ADN se cuantificaron mediante PCRdd con el kit de cuantificación de bibliotecas PCRdd para Illumina TruSeq (Bio-Rad) y la secuenciación se realizó en un MiSeq (Illumina) utilizando el kit de reactivos MiSeq v2 (500 ciclos). Tras la NGS, las lecturas se analizaron utilizando CRISPResso (Pinello et al., 2016, Nat Biotechnol 34(7): 695-697) con la siguiente configuración: filtrar las lecturas con base en la puntuación de calidad de phred33 con la puntuación de calidad media mínima fijada en 30 (q 30); recortar la secuencia adaptadora de Illumina de las lecturas; cuantificar las indels dentro de una ventana de 40 nt alrededor del sitio de escisión previsto (w 40); ignorar los eventos de sustitución para la cuantificación con el fin de no considerar los SNP. La eficacia de la edición se calculó mediante la fórmula: % editado = % de lecturas indel en células editadas genéticamente - % de lecturas indel en células no tratadas

Tabla 1: Oligonucleótidos

Las secuencias se muestran en orientación 5'-3'; modificaciones: * = enlace fosforotioato, /5Fos/ = fosforilación 5', /5SpC3/ = espaciador 5' C3, /3SpC3/ = espaciador 3' C3, /iBiodT/ = deoxitimidina de biotina interna, /ideoxiU/ = deoxiuridina interna

SEQ-ID NO:	Nombre de oligo	Secuencia (5'-3')	Propósito
1	Bloqueando_fw	/5 Fos /CCTTGGTATCTGCTGTC/ ideoxi U/ATGAC*T*C/3Sp C3/	adaptador de bloqueo
2	Bloqueando_rv	/5SpC3/GA*G*TCA/ ideoxiU/AGACAGCAGATACCAAGGT	adaptador de bloqueo
3	Purific_Adaptador +	/5 Fos /GCATGTCATCGGATCCAG/iBiodT/AC /iBiodT/CATCTGCA	adaptador de biotina
4	Purific_Adaptador -	TGCAGATGAGTACTGGATCCGATGACATGCT	adaptador de biotina
5	pCAGneo-R	AGTCGTGTCTTACCGGGTTG	secuenciación
6	U6P_Fw	TGGATCCGGTACCAAGGTCG	PCR
7	VEGFA-3F	acaccGGTGAGTGAGTGTGTGCGTGg	clonación
8	VEGFA-3R	aaaacCACGCACACACTCACTCACCg	clonación
9	T7gRNArv	AAAAGCACCGACTCGGTGCCAC	síntesis de ARN

10	ARNg-FANCF_For	acaccGGAATCCCTTCTGCAGCACCg	clonación
11	ARNg-FANCF_Rev	aaaacGGTGCTGCAGAAGGGATTCCg	clonación
12	ARNg-HEK4_For	acaccGGCACTGCGGCTGGAGGTGGg	clonación
13	ARNg-HEK4_Rev	aaaacCCACCTCCAGCCGCAGTGCCg	clonación
14	T7-FANCF ARNg_fw	ttaatacgactcactataGGAATCCCTTCT	síntesis de ARN
15	T7-HEK4gARNg_fw	ttaatacgactcactataGGCACTGCGGCT	síntesis de ARN
16	T7-VEGFA3 ARNg_fw	ttaatacgactcactataGGTGAGTGAGTG	síntesis de ARN
17	VEGFA-sitio3sp-RP	CGAGGAGGGAGCAGGAAAGT	PCR
18	= no biot. adaptador	TGCAGATGAGTACTGGAT	PCR
19	VEGFA-sitio3sp-Fu-FP	TCCAGATGGCACATTGTCAG	PCR
20	FP-FANCF	GCCCTACATCTGCTCTCCCTCCA	PCR
21	RP-FANCF	GGGCCGGGAAAGAGTTGCTG	PCR
22	FP-RNF2	TCCTAAAAATCAGTTTTGAGATTTACTTCC	PCR
23	RP-RNF2	AAAGTGTTAGCCAACATACAGAAGTCAGGA	PCR
24	FP-HEK sitio 4	AGGCAGAGAGGGGTTAAGGT	PCR
25	RP-HEK sitio 4	GGGTCAGACGTCCAAAACCA	PCR
26	= propio adaptador de bloqueo	GACAGCAGATACCAAGG	PCR
27	VEGFA3 ON_for-nuevo	GTGTGCAGACGGCAGTCACTA	Amplicón-seq
28	VEGFA3 ON_rev-nuevo	GAAAGTGAGGTTACGTGCGGAC	Amplicón-seq
29	VEGFA3-FP OT9(nuevo)	ACCTCGGAGCTCACCTTCCA	Amplicón-seq
30	VEGFA3-RP OT9(nuevo)	GCCGAGAGTGAGTGAGTGAG	Amplicón-seq
31	VEGFA OT2 Abnoba_for2	ATGTGAGTGTGTCTGGGTGTGC	Amplicón-seq
32	VEGFA OT2 Abnoba_rev	CCCATAGCTCAGCGGGTGAA	Amplicón-seq
33	VEGFA OT3 Abnoba_for	TGAGGATGTATGTGAGACTGAATGTGTGTG	Amplicón-seq
34	VEGFA OT3 Abnoba_rev	TGTACTTCAAAATGCAGCCGAGCC	Amplicón-seq
35	VEGFA OT4 Abnoba_for	GTGTGCCTATGGGCATATGTGG	Amplicón-seq
36	VEGFA OT4 Abnoba_rev	GAGACACACATACGCACACGC	Amplicón-seq
37	VEGFA OT5 Abnoba_for2	GATGCATGTGTACGTGTGACGTG	Amplicón-seq
38	VEGFA OT5 Abnoba_rev	ACACATGTACACACACCCTCGC	Amplicón-seq
39	VEGFA OT6 Abnoba_for2	GCTCTGCATAGGGGGAAAGAATC	Amplicón-seq
40	VEGFA OT6 Abnoba_rev2	ACCACACCCATCCACATGCAC	Amplicón-seq
41	VEGFA sitio 3-FP OT7	AGAAGCAGCTTTGCAGAAGCAGG	Amplicón-seq
42	VEGFA sitio 3-RP OT7	CCATTTGAGGCACTGGAAAGCTAC	Amplicón-seq
43	VEGFA OT8 Abnoba_for2	CAAGAACTACAAGTCTCGGCTGC	Amplicón-seq
44	VEGFA OT8 Abnoba_rev2	TCGCTCTCCACCTCGATGTC	Amplicón-seq
45	VEGFA OT9 Abnoba_for	TCAGTCTAGGACAGCTGGGG	Amplicón-seq
46	VEGFA OT9 Abnoba_rev	GGTCGCTGCCCATTCTGCAA	Amplicón-seq
47	VEGFA OT10 Abnoba_for	GTCCTGAGAGTTGGTGACATGTG	Amplicón-seq
48	VEGFA OT10 Abnoba_rev	ACACTGGCTGGCACAATCCATC	Amplicón-seq
49	VEGFA OT11 Abnoba_for	ATGGATCCACATGGGGGCAC	Amplicón-seq
50	VEGFA OT11 Abnoba_rev	GCACACACATATGCACATCCGC	Amplicón-seq
51	VEGFA OT12 Abnoba_for	CCTATTCCCCAGCGTGTGCA	Amplicón-seq

52	VEGFA OT12 Abnoba_rev	CGCACTCTTCATAATAGGGGTCG	Amplicón-seq
53	VEGFA sitio 3-FP OT4	TCCTTTGAGGTTCATCCCCC	Amplicón-seq
54	VEGFA sitio 3-RP OT4	CCAATCCAGGATGATTCCGC	Amplicón-seq
55	VEGFA OT14 Abnoba_for	ATAGGCTGTCTCCCATTAACGGGC	Amplicón-seq
56	VEGFA OT14 Abnoba_rev	CTGTTTGGAAAACGAAGCCCAGAG	Amplicón-seq
57	VEGFA OT15 Abnoba_for	GTGCTCCTGTGTTTATGTGTGTGC	Amplicón-seq
58	VEGFA OT15 Abnoba_rev	CACGGAGAAGGCACACTCAC	Amplicón-seq
59	VEGFA OT16 Abnoba_for	ACAGGAGGAGCCAGGTGAGA	Amplicón-seq
60	VEGFA OT16 Abnoba_rev	TTGCAGGGGACAGCAGTGGT	Amplicón-seq
61	VEGFA OT17 Abnoba_for2	GCAGAGGTTGCAGTGAGCCTA	Amplicón-seq
62	VEGFA OT17 Abnoba_rev2	AGCTGCATTTTTCACCTCTCCCTG	Amplicón-seq
63	VEGFA OT18 Abnoba_for	GGTCGTAATGCACCATAGATGCTC	Amplicón-seq
64	VEGFA OT18 Abnoba_rev	CCCACATTTCTCCCTGTTGACATC	Amplicón-seq
65	VEGFA OT19 Abnoba_for	TGGGGTGGCAAGAGACAGTG	Amplicón-seq
66	VEGFA OT19 Abnoba_rev	CAGTGCCTCCACTTCTGCAG	Amplicón-seq
67	VEGFA OT20 Abnoba_for	ACAGAACTCAGCTGTCCCGG	Amplicón-seq
68	VEGFA OT20 Abnoba_rev	CCTCCAACTTCAAATGGCCCC	Amplicón-seq
69	HEK4 ON_for-new	GACTCCTTCTGGGGCCTTTTTC	Amplicón-seq
70	HEK4 ON_rev-new	AGACCCGCTGGTCTTCTTTCC	Amplicón-seq
71	HEKs4 OT1 Abnoba_for	CATCAACACATAACAGGCGTGCAG	Amplicón-seq
72	HEKs4 OT1 Abnoba_rev	GGCCCCTCCCAGTTTATAGC	Amplicón-seq
73	HEKs4 OT2 Abnoba_for2	GAAGGTGGGAGTGGCTGGAGT	Amplicón-seq
74	HEKs4 OT2 Abnoba_rev2	ACCTGTCAGCATCCAGTGGC	Amplicón-seq
75	HEKs4 OT3 Abnoba_for	CACCTCCCGACTGACTGAGTGT	Amplicón-seq
76	HEKs4 OT3 Abnoba_rev	TGGATCAGGGGGCTCCACAT	Amplicón-seq
77	HEKs4 OT4 Abnoba_for	AAAAATCAGCTGGGCGTGGTGG	Amplicón-seq
78	HEKs4 OT4 Abnoba_rev	GAGCCATGATACCACACACAATCC	Amplicón-seq
79	HEKs4 OT5 Abnoba_for	GCTCCAGGCAAGTAAAACACCAG	Amplicón-seq
80	HEKs4 OT5 Abnoba_rev	CCAAGGAGCCTCCAGGTATAC	Amplicón-seq
81	HEKs4 OT6 Abnoba_for	CCCTGTGGGAAAGTGCTTACC	Amplicón-seq
82	HEKs4 OT6 Abnoba_rev	GCAGGAGAGGTTGCAGGTGA	Amplicón-seq
83	HEKs4 OT7 Abnoba_for2	TCAGGCAATGGCCAGTGAAGG	Amplicón-seq
84	HEKs4 OT7 Abnoba_rev2	AGAGAATCCCCTGTCGCCGT	Amplicón-seq
85	HEKs4 OT8 Abnoba_for	GGACAAGTGACTTAGCCTCTCAG	Amplicón-seq
86	HEKs4 OT8 Abnoba_rev	CTTGGTCAAGTCTCCTCAACTCTC	Amplicón-seq
87	HEKs4 OT9 Abnoba_for	CTCTGACTGTTACCGGGGCA	Amplicón-seq
88	HEKs4 OT9 Abnoba_rev	TCCAAACGAGGGCGGTGGACA	Amplicón-seq

89	HEKs4 OT10 Abnoba_for	CAGTAAGTCAAGGTCAGCTAGGG	Amplicón-seq
90	HEKs4 OT10 Abnoba_rev	CCAGAGGCACCACTGTGCTT	Amplicón-seq
91	HEKs4 OT11 Abnoba_for	CCTCTTAATTGGAGGGGAGAATGC	Amplicón-seq
92	HEKs4 OT11 Abnoba_rev	ACTGCAGGGTCCCCATGGTA	Amplicón-seq
93	HEKs4 OT12 Abnoba_for	GATGCCTCATCTAAGCTGCTGC	Amplicón-seq
94	HEKs4 OT12 Abnoba_rev	GTTCTTCCTCCACCCTGGCT	Amplicón-seq
95	HEKs4 OT13 Abnoba_for	CTATACCCCTACCCACCTGAG	Amplicón-seq
96	HEKs4 OT13 Abnoba_rev	GTAGTCAGGCATCGGGGCAT	Amplicón-seq
97	HEKs4 OT14 Abnoba_for	AGGCATGTATCCCGCTCTGG	Amplicón-seq
98	HEKs4 OT14 Abnoba_rev	AACTGGTGCTGGGGAAAAGGG	Amplicón-seq
99	HEKs4 OT15 Abnoba_for2	GATGTCTGGGGCAGTTGAGTC	Amplicón-seq
100	HEKs4 OT15 Abnoba_rev2	GTTTTACCCACATCCTCACTGTGC	Amplicón-seq
101	HEK sitio 4-FP OT7	GAACACAATGGCTCCGGTGG	Amplicón-seq
102	HEK sitio 4-RP OT7	AGCACCAGGTTGTCGCTCAG	Amplicón-seq
103	HEKs4 OT17 Abnoba_for	TTGGCTGTGGTGATGGTCGC	Amplicón-seq
104	HEKs4 OT17 Abnoba_rev	ATTTGGGGTGGGACAGGCCT	Amplicón-seq
105	HEKs4 OT18 Abnoba_for	AGAGAAGCTGCGCAAGGAGG	Amplicón-seq
106	HEKs4 OT18 Abnoba_rev	ATCTGCACGCGCACCGAGAA	Amplicón-seq
107	HEKs4 OT19 Abnoba_for	AGCAGGCTTGACATCACTTTCTCC	Amplicón-seq
108	HEKs4 OT19 Abnoba_rev	CAGCTTCATGTGTCCCGGTG	Amplicón-seq
109	HEKs4 OT20 Abnoba_for	CTGTCCTGGAAAGCCTAGTCTC	Amplicón-seq
110	HEKs4 OT20 Abnoba_rev	GGTACCATCCTCTTAATTCACCCC	Amplicón-seq

Resultados

Ejemplo 1: Flujo de trabajo de Abnoba-Seq

Los presentes inventores plantearon la hipótesis de que la escisión de ADN genómico fragmentado y con extremos bloqueados con CRISPR-Cas *in vitro* permitirá el posterior marcado de los sitios de corte con adaptadores biotinilados, lo que a su vez permitirá un fácil enriquecimiento de los fragmentos de ADN escindidos y la identificación final de los sitios objetivo tras la secuenciación de alto rendimiento (Figura 1). Para generar una biblioteca de ADN genómico fragmentado y con extremos bloqueados, el ADN genómico se cortó aleatoriamente hasta un tamaño mediano de ~400 pb, y se ligó un adaptador de ADN de doble cadena (dc) con espaciadores C3 en ambas cadenas a los fragmentos de ADN con extremos reparados y cola dA. Los fragmentos de ADN con adaptador de bloqueo ligado a ambos extremos, denominados aquí fragmentos como con extremos bloqueados, representaron más del 99 % tras el tratamiento de los fragmentos con una combinación de Exonucleasa III, Exonucleasa I y Exonucleasa Lambda (Figura 2). A continuación, 50 ng de fragmentos de ADN con extremos bloqueados, correspondientes a ~15.000 genomas haploides, se escindieron con 1,6 pmol de RNPs (proteína SpyCas9 complejada con ARNg en una relación molar de 1:3). La reacción de escisión creó extremos de ADN libres en todos los sitios objetivo, a los que se ligó un adaptador de afinidad basado en ARNdc con dos bases dT biotiniladas internas tras la reparación de los extremos y la formación de cola dA. Como control durante el proceso, la actividad de CRISPR-Cas9 en el objetivo se monitorizó mediante amplificación por PCR utilizando cebadores específicos del sitio dentro del objetivo. Si el sitio dentro del objetivo se ha escindido eficazmente, la PCR sólo debería producir bandas débiles (Figura 3). Si las muestras pasaban el control durante el proceso, se utilizaban perlas cargadas con estreptavidina, que tiene una gran afinidad por la biotina, para enriquecer los fragmentos de ADN escindidos por CRISPR-Cas. Estos adaptadores también añadieron secuencias objetivo para los pasos posteriores de la PCR que se utilizaron en primer lugar para liberar fragmentos de ADN biotinilados de las perlas de estreptavidina mediante amplificación lineal y, en segundo lugar, para crear de nuevo ARNdc. Por último, se generó una biblioteca de secuenciación con adaptadores Illumina y se sometió a secuenciación de próxima generación (NGS) de extremos pareados. Es importante destacar que el proceso Abnoba-Seq no se basa en un algoritmo basado en la homología de secuencias. En consecuencia, se mantiene totalmente independiente de las secuencias ARNg y PAM. El proceso bioinformático se basa en un algoritmo de reconocimiento de patrones que

busca una disposición específica de las lecturas de secuencias alineadas, es decir, lecturas con inicios (3 primeras bases) y finales (3 últimas bases) superpuestos. Brevemente, se escaneó todo el genoma para identificar regiones genómicas, que se denominaron intervalos, y que contienen sitios putativos de escisión. Estos sitios se definen por la presencia de múltiples lecturas que comienzan o terminan en la misma posición (+/- 3 nt). Finalmente se introdujo un algoritmo de aprendizaje automático, basado en bosques aleatorios, para eliminar los sitios de escisión falsos positivos. Por último, los intervalos seleccionados se clasificaron de acuerdo con su cobertura (para más detalles, véase la sección sobre el procedimiento y la Figura 4).

Ejemplo 2: Validación del flujo de trabajo Abnoba-Seq

Para validar este procedimiento, se aplicó Abnoba-Seq al ADN genómico de la línea celular humana U2OS y se utilizaron cuatro nucleasas CRISPR-Cas diferentes que se evaluaron previamente mediante GUIDE-Seq y CIRCLE-Seq (Tsai et al. 2015, Tsai et al. 2017). Como era de esperar, la alineación de las lecturas NGS con el genoma humano de referencia indicó la posición del sitio de escisión en el genoma humano, con la mayoría de las lecturas comenzando o terminando tres nucleótidos corriente arriba de la PAM (Figura 5). Se observaron lecturas adicionales que no empezaban ni terminaban en el sitio de escisión, sino que formaban parte del grupo de lecturas que abarcaban los sitios objetivo. Estos fragmentos de ADN probablemente empezaban o terminaban en el lugar de la escisión, pero no se secuenciaron en toda su longitud debido a que la longitud máxima de lectura de la secuencia era de 150 nt.

El análisis bioinformático reveló 42 sitios objetivo para la nucleasa CRISPR-Cas9 dirigida a FANCF, 51 sitios para la nucleasa dirigida a RNF2, 307 sitios objetivo para la nucleasa dirigida a VEGFA y 233 sitios objetivo para la nucleasa CRISPR-Cas9 dirigida al sitio 4 de HEK (HEKs4) (Figura 6). Cabe destacar que, en el caso de Abnoba-Seq, al igual que en GUIDE-Seq y CIRCLE-Seq, el sitio dentro del objetivo a menudo no devolvía el mayor número de lecturas. La mayoría de los sitios fuera del objetivo identificados mostraron un alto grado de homología de secuencia con el sitio dentro del objetivo. Sin embargo, además del simple desajuste de nucleótidos, Abnoba-Seq identificó un gran número de sitios fuera del objetivo con protuberancias en el ARNg o en el sitio objetivo del ADN.

Una comparación lado a lado de los sitios fuera del objetivo identificados por Abnoba-Seq con los sitios fuera del objetivo encontrados previamente por GUIDE-Seq y CIRCLE-Seq reveló un grado razonable de superposición para las nucleasas de orientación FANCF, VEGFA y HEKs4, pero no RNF2, respectivamente (Figura 3). Entre 58 % y 100 % de los sitios fuera del objetivo descubiertos por GUIDE-Seq y 11-21 % de los descubiertos por CIRLCE-Seq, respectivamente, fueron identificados por Abnoba-Seq (Tabla 2). Los tres procedimientos permitieron identificar las zonas dentro del objetivo. Por otro lado, una comparación más cualitativa puso de manifiesto que la mayoría de los 20 sitios fuera del objetivo más identificados fueron detectados por los tres procedimientos (Tabla 2), lo que sugiere que estos tres ensayos son capaces de detectar de forma fiable sitios fuera del objetivo de corte frecuente.

En resumen, Abnoba-Seq fue capaz de identificar sitios fuera del objetivo comúnmente cortados con una cantidad sustancialmente menor de ADN genómico de entrada, lo que lo convierte en un procedimiento adecuado para evaluar la especificidad de nucleasas programables específicas de objetivo en muestras celulares valiosas de tamaño de muestra y disponibilidad limitados.

Ejemplo 3: ***Análisis Abnoba-Seq en muestras celulares de tamaño y disponibilidad limitados***

Tabla 2

	En general con GUIDE-Seq			**En general con CIRCLE-Seq**		
	General	**Top 20**	**Top 10**	**General**	**Top 20**	**Top 10**
FANCF	67%	67%	67%	21%	60%	80%
RNF2	100%	100%	100%	14%	15%	30%
VEGFA	75%	90%	100%	19%	90%	90%
HEKs4	58%	100%	100%	11%	85%	80%

Una desventaja importante de los procedimientos de análisis fuera del objetivo ya existentes es que no pueden aplicarse fácilmente a muestras celulares de tamaño y disponibilidad de muestra limitados. Como prueba de concepto de que Abnoba-Seq puede realizarse en ADN genómico de tipos celulares clínicamente relevantes con disponibilidad limitada, Abnoba-Seq se aplicó a ADN genómico aislado de células madre y progenitoras hematopoyéticas (HSC) CD34-positivas derivadas de sangre de cordón umbilical. Como se visualizó en el control durante el proceso (Figura 3B), la nucleasa CRISPR-Cas9 dirigida a VEGFA mostró una elevada actividad de escisión dentro del objetivo y el resultado de Abnoba-Seq para las HSC (Figura 8A) fue comparable al anterior realizado en ADN genómico de células U2OS (Figura 6). Una comparación más detallada reveló que la mayoría de los sitios identificados eran comunes entre las dos muestras (Figura 8B). Al considerar los 20 o 50 sitios objetivo más importantes identificados en el ADN genómico de las HSC, 95 % (19/20) o 92 % (46/50) se identificaron previamente en el ADN genómico de las células U2OS, respectivamente (Figura 8C,D), lo que indica una alta reproducibilidad del ensayo. En resumen, Abnoba-Seq

es un procedimiento robusto que puede aplicarse con éxito al ADN genómico aislado de muestras de células valiosas de tamaño y disponibilidad de muestra limitados.

Ejemplo 4: Verificación de la actividad CRISPR-Cas fuera del objetivo en células

El análisis de la actividad *in vitro* de una nucleasa programable específica de objetivo devuelve un escenario del peor de los casos porque (i) *in vitro* todos los sitios genómicos potenciales fuera del objetivo son fácilmente accesibles a CRISPR-Cas9 y (ii) la concentración de nucleasa podría ser mucho mayor *in vitro* que en el núcleo de las células transfectadas. Por lo tanto, se analizaron los 20 sitios fuera del objetivo predichos por Abnoba-Seq para las nucleasas CIRPSR-Cas orientadas a VEGFA y HEKs4 mediante secuenciación de amplicones orientada siete días después de la transfección de células U2OS o HSC primarias. La eficiencia de edición general, según lo indicado por las frecuencias indel, fue 96 % en el sitio de objetivo de VEGFA en celdas U2OS, 61 % en el sitio de objetivo de VEGFA en HSC y 69 % en el sitio de objetivo de HEKs4 en HSC (Figura 9). La actividad global fuera del objetivo en estas células fue bastante baja. La mayor actividad fuera del objetivo se observó en el sitio fuera del objetico de VEGFA (OT) 8, con 20 % de alelos editados en células U2OS y 10 % de alelos editados en HSCs, y en VEGFA OT-13 con 5 % de alelos editados en la HSC. La actividad fuera del objetivo en los demás sitios fuera del objetivo previstos fue baja o inexistente. Algunos de los sitios fuera del objetivo sólo se escindieron en las células U2OS, pero no en las HSC primarias (OT-12, OT-14, OT-16), lo que sugiere efectos específicos del tipo celular. En resumen, sólo alrededor de la mitad de los sitios fuera del objetivo identificados *in vitro* se escindieron realmente en las células. Cabe destacar que CIRLCE-Seq y GUIDE-Seq también predijeron que la mayoría de estos sitios investigados estaban fuera del objetivo. En conjunto, es importante tener en cuenta que los procedimientos *in vitro* para detectar sitios fuera del objetivo en todo el genoma y de forma imparcial son sofisticados, pero siguen siendo meras herramientas de predicción.

Mientras que muchos procedimientos genómicos para identificar la actividad fuera del objetivo asociada a CRISPR-Cas son aparentemente imparciales, el proceso bioinformático para analizar los datos NGS introduce algún sesgo, en particular con respecto al número máximo de desajustes permitidos o el requisito de una PAM. Es importante destacar que Abnoba-Seq para identificar sitios fuera del objetivo no hace ninguna suposición sobre la naturaleza de un sitio fuera del objetivo, ni el grado de homología con el sitio dentro del objetivo ni la presencia de una PAM específica. Por lo tanto, Abnoba-Seq identificará sitios fuera del objetivo que son excluidos por otros procedimientos, incluyendo RNF2 OT7 con 7 desajustes y 2 protuberancias, RNF2 OT37 con 9 desajustes, VEGFA OT2 con 3 desajustes y un TGA PAM, VEGFA OT26 con 3 desajustes y un TGA PAM, o VEGFA OT75 con 2 desajustes y un TTG PAM. En conjunto, Abnoba-Seq es un procedimiento *in vitro* verdaderamente imparcial para perfilar la actividad CRISPR-Cas fuera del objetivo *in vitro* e *in vivo.* Dado que sólo se necesitan cantidades mínimas de ADN genómico para realizar el ensayo, Abnoba-Seq puede aplicarse de forma específica para cada tipo celular a muestras de células valiosas de tamaño y disponibilidad de muestra limitados.

REIVINDICACIONES

1. Un procedimiento para detectar sitios fuera del objetivo de una o más nucleasas programables específicas de objetivo en un genoma *in vitro* que comprende los pasos de:

(a) proporcionar una muestra de entrada que comprenda ADN genómico, fragmentar aleatoriamente el ADN genómico en fragmentos de ADNdc, y ligar un adaptador de bloqueo a los extremos de los fragmentos de ADNdc y eliminar los fragmentos de ADNdc que comprendan extremos no bloqueados, para preparar fragmentos purificados de ADNdc con extremos bloqueados;
(b) preparar una biblioteca de ADN escindido y amplificado de fragmentos de ADNdc, mediante
b1) escindir los fragmentos purificados de ARNdc con extremos bloqueados con una o más nucleasas programables específicas de objetivo para obtener fragmentos de ARNdc escindidos;
b2) ligar un adaptador de afinidad a los sitios de escisión de los fragmentos de ARNdc escindidos para obtener fragmentos de ARNdc modificados por adaptador de afinidad;
b3) enriquecer fragmentos de ADNdc modificados por adaptador de afinidad utilizando el adaptador de afinidad para obtener fragmentos de ADNdc modificados por adaptador de afinidad enriquecidos; y
b4) amplificar los fragmentos de ADNdc modificados por adaptador de afinidad enriquecidos mediante amplificación por PCR para obtener una biblioteca de ADN escindida y amplificada de fragmentos de ADNdc;
(c) realizar una secuenciación de la biblioteca de ADN escindido y amplificado de fragmentos de ARNdc para obtener lecturas de secuencias;
(d) alinear las lecturas de secuencias con un genoma de referencia para obtener lecturas de secuencias alineadas;
(e) identificar los sitios de escisión, determinando los sitios de escisión putativos y excluyendo los sitios de escisión falsos positivos, en los que los sitios de escisión putativos se determinan localizando regiones de secuencia en el genoma de referencia en las que coinciden los inicios y los finales de las lecturas de secuencia alineadas, e identificar los sitios de escisión putativos como los sitios de estas regiones de secuencia en los que coinciden los inicios y los finales de las lecturas de secuencia.

2. El procedimiento de acuerdo con la reivindicación 1, en el que la muestra de entrada comprende menos de 1 µg, preferentemente menos de 800 ng, más preferentemente menos de 700 ng, incluso más preferentemente aproximadamente 600 ng, de ADN genómico.

3. El procedimiento de acuerdo con cualquiera de las reivindicaciones anteriores, en el que el paso a) comprende

a1) fragmentación aleatoria del ADN genómico hasta una longitud definida para obtener fragmentos de ADNdc;
a2) proteger los extremos de los fragmentos de ARNdc ligando un adaptador de bloqueo para obtener fragmentos de ARNdc con extremos bloqueados, preferentemente comprendiendo los pasos de
1) preparar los fragmentos de ARNdc para la ligadura del adaptador de bloqueo; y
2) ligar un adaptador de bloqueo a los extremos de los fragmentos de ARNdc, para obtener fragmentos de ARNdc con extremos bloqueados; y
a3) poner en contacto la muestra que comprende los fragmentos de ADNdc con extremos bloqueados con una o más exonucleasas para eliminar los fragmentos de ADNdc con extremos no bloqueados, para obtener fragmentos purificados de ADN con extremos bloqueados.

4. El procedimiento de acuerdo con cualquiera de las reivindicaciones anteriores, en el que el adaptador de bloqueo comprende una o más modificaciones, que tras la ligadura del adaptador de bloqueo a un fragmento de ADNdc, impiden la ligadura, del extremo al que el adaptador de bloqueo está ligado al fragmento de ADNdc y una o más de las modificaciones, que tras la ligadura del adaptador de bloqueo a un fragmento de ADNdc, impiden la escisión por exonucleasa del fragmento de ADNdc del extremo al que el adaptador de bloqueo está ligado al fragmento de ADNdc.

5. El procedimiento de acuerdo con cualquiera de las reivindicaciones anteriores, en el que en el paso b1) la una o más nucleasas programables específicas de objetivo se selecciona de nucleasas efectoras similares a activadores de la transcripción (TALEN), nucleasas de dedos de zinc (ZFN), meganucleasas, megaTAL, FokI-dCas9, nucleasas guiadas por ARN de repeticiones palindrómicas cortas agrupadas regularmente espaciadas (CRISPR)/Cas (CRISPR/Cas RGN) o una variante de las mismas, nucleasas de ADN guiadas por ADN basadas en proteínas de la familia Argonauta o una variante de las mismas, o una combinación de las mismas, preferentemente una o más CRISPR/Cas RGN, en las que la endonucleasa Cas de la CRISPR/Cas RGN es preferentemente la endonucleasa Cas9 o una variante de la endonucleasa Cas9 o la endonucleasa Cpf1 o una variante de la endonucleasa Cpf1.

6. El procedimiento de acuerdo con cualquiera de las reivindicaciones anteriores, en el que el adaptador de bloqueo y/o el adaptador de afinidad son adaptadores basados en ARNdc, y en el que el adaptador de bloqueo y/o el adaptador de afinidad comprenden un sitio cebador compatible para su uso en cebado por PCR.

7. El procedimiento de acuerdo con cualquiera de las reivindicaciones anteriores, en el que el adaptador de afinidad es un adaptador basado en ARNdc que comprende una o más modificaciones basadas en biotina o biotina y en el que en el paso b3) se utiliza estreptavidina o una variante de la misma, preferentemente inmovilizada en perlas magnéticas, como molécula de unión del adaptador de afinidad para enriquecer los fragmentos de ARNdc escindidos y modificados por adaptador de afinidad.

8. El procedimiento de acuerdo con cualquiera de las reivindicaciones anteriores, en el que el paso b4) para obtener la biblioteca de ADN escindido y amplificado de fragmentos de ARNdc comprende una amplificación por PCR en dos pasos, con

1. un primer paso de amplificación PCR para la amplificación PCR lineal utilizando fragmentos de ARNdc modificados por adaptador de afinidad enriquecidos y unidos mediante una molécula de unión de adaptador a un soporte sólido como molde y un cebador específico de adaptador de afinidad; y
2. después del primer paso, un segundo paso de amplificación por PCR, tras eliminar de la mezcla de reacción los fragmentos de ARNdc modificados por el adaptador de afinidad enriquecido y unidos mediante la molécula de unión del adaptador al soporte sólido, y añadir un cebador específico del adaptador de bloqueo como segundo cebador.

9. El procedimiento de acuerdo con cualquiera de las reivindicaciones anteriores, en el que en el paso d) las lecturas de secuencia se alinean con un genoma de referencia para obtener lecturas de secuencia alineadas, mediante primero

d1) recortar el adaptador de bloqueo, el adaptador de afinidad y la secuencia del adaptador de secuenciación opcional de las lecturas de secuencia para obtener lecturas de secuencia recortadas; y segundo

d2) mapear las lecturas de secuencia recortadas a un genoma de referencia para obtener lecturas de secuencia alineadas.

10. El procedimiento de acuerdo con cualquiera de las reivindicaciones anteriores, en el que el ADN genómico es ADN genómico humano y en el que el genoma de referencia es un genoma de referencia humano, preferentemente hg19 o hg38.

11. El procedimiento de acuerdo con cualquiera de las reivindicaciones anteriores, en el que en el paso e) los sitios putativos de escisión se determinan mediante

- dividir el genoma de referencia que comprende las lecturas de secuencias alineadas en intervalos de una longitud definida, preferentemente de al menos aproximadamente100 nt; y
- determinar, para cada ubicación que comprenda al menos un inicio de lectura de secuencia y al menos un final de lectura de secuencia, uno o más sitios de escisión putativos como sitios con un máximo local para la suma de inicios de lectura de secuencia y finales de lectura de secuencia, y opcionalmente minimizar la distancia al centro del intervalo;

en el que un inicio de lectura de secuencia consiste en los primeros 1 a 5 nt, preferentemente los primeros 2 a 4 nt, más preferentemente los primeros 3 nt, y un final de lectura de secuencia consiste en los últimos 1 a 5 nt, preferentemente los últimos 2 a 4 nt, más preferentemente los últimos 3 nt, de una lectura de secuencia alineada.

12. El procedimiento de acuerdo con cualquiera de las reivindicaciones anteriores, en el que en el paso e) los sitios de escisión se identifican mediante

e1) dividir el genoma de referencia que comprende las lecturas de secuencias alineadas en intervalos de longitud de secuencia establecida, preferentemente de aproximadamente 100 nt de longitud,
e2) determinar preferentemente para cada intervalo una cobertura de lectura, descartando los intervalo con una cobertura de lectura de 0 y combinando preferentemente los intervalos con una cobertura de lectura de >0 consecutivamente, basándose en el orden de secuencia en el genoma de referencia, en un conjunto continuo de intervalos;
e3) dividir cada intervalo en ventanas de una anchura determinada, preferentemente de 4 nt, con un tamaño de paso establecido, que es igual o menor, preferentemente menor, que la anchura de las ventanas, preferentemente de 2 nt;
e4) determinar la cobertura inicial, final y de fondo de cada ventana, agrupando,
- inicios de lectura de secuencia que caen en la región de la ventana en un grupo de cobertura de inicio;
- finales de lectura de secuencia que caen en la región ventana en un grupo de cobertura final; y
- lecturas de secuencias que caen en la región de la ventana, pero no pertenecen al grupo de cobertura inicial y al grupo de cobertura final en un grupo de cobertura de fondo;

y sumando para cada ventana el número de lecturas en el grupo de cobertura inicial, cobertura final y cobertura de fondo, respectivamente;

e5) determinar, para cada intervalo que comprenda al menos un inicio de lectura de secuencia y al menos un final de lectura de secuencia, uno o más sitios de escisión putativos maximizando la suma de la cobertura inicial y la cobertura final y minimizando la distancia al centro del intervalo; y
e6) excluyendo los sitios de escisión falsos positivos;

en el que un inicio de lectura de secuencia consiste en los primeros 1 a 5 nt, preferentemente los primeros 2 a 4 nt, más preferentemente los primeros 3 nt, y un final de lectura de secuencia consiste en los últimos 1 a 5 nt, preferentemente los últimos 2 a 4 nt, más preferentemente los últimos 3 nt, de una lectura de secuencia alineada.

13. El procedimiento de acuerdo con cualquiera de las reivindicaciones anteriores, en el que los sitios de escisión falsos positivos se excluyen seleccionando únicamente sitios de escisión putativos con

1. una relación de la cobertura de fondo a la cobertura de la señal, que no supere un determinado umbral, preferentemente aproximadamente 0,5; y/o
2. una relación entre la cobertura inicial y la cobertura final en el sitio de escisión putativo, dentro de un cierto intervalo numérico, preferentemente entre 1/5 y 5/1 aproximadamente; y/o
3. una cobertura mínima del sitio, preferentemente de 6 o más; y/o
4. una cobertura mínima de intervalos, preferentemente de 10 o más.

14. Un procedimiento para detectar sitios fuera del objetivo en un genoma *in* vivo que comprende los pasos de

i) proporcionar una primera muestra, que comprende ADN genómico de células tratadas con una o más nucleasas programables específicas de objetivo, y una segunda muestra, que comprende ADN genómico de células de fondo genético relacionado, preferentemente células del mismo fondo genético, no tratadas con la una o más nucleasas programables específicas de objetivo;
ii) para ambas muestras, detectar independientemente *in vitro* los sitios fuera del objetivo mediante el procedimiento de las reivindicaciones 1 a 13 y utilizando la una o más nucleasas programables específicas del objetivo en el paso i); y
iii) comparar los sitios fuera de objetivo entre las dos muestras para identificar los sitios fuera del objetivo solo identificados en la muestra que comprende ADN genómico de células no tratadas con la una o más nucleasas programables específicas de objetivo, pero no en la muestra que comprende ADN genómico de células tratadas con la una o más nucleasas programables específicas de objetivo como sitios fuera del objetivo *in vivo* de la una o más nucleasas programables específicas de objetivo.

Figura 1

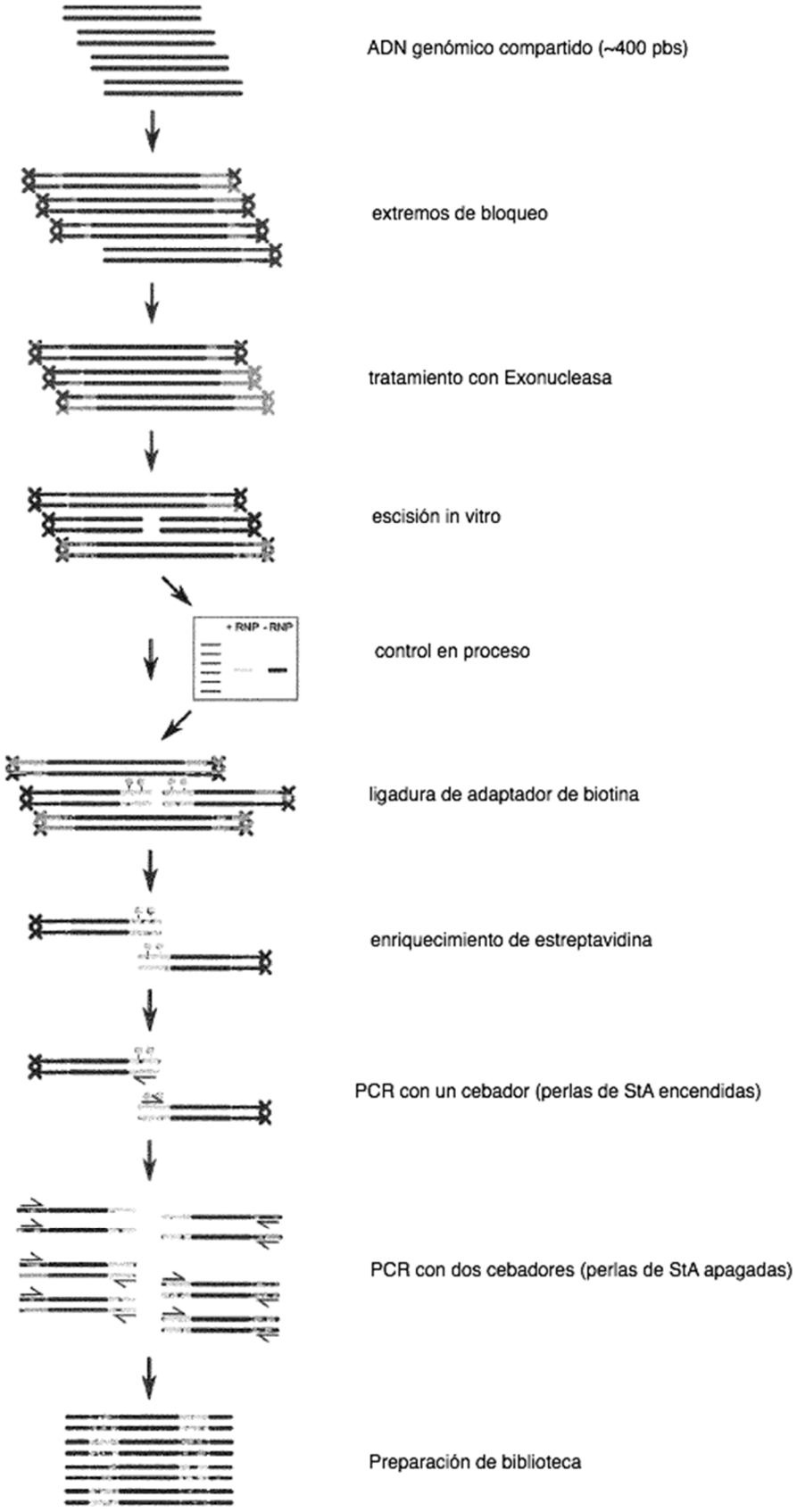
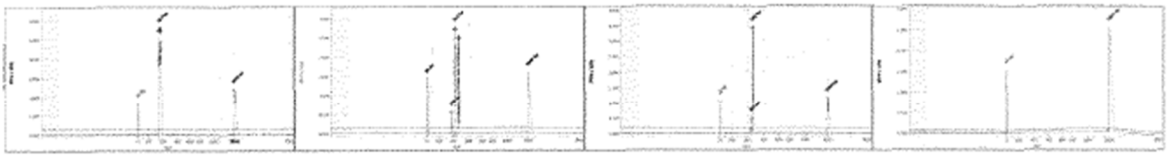

ADN genómico compartido (~400 pbs)
extremos de bloqueo
tratamiento con Exonucleasa
escisión in vitro
+ RNP - RNP
control en proceso
ligadura de adaptador de biotina
enriquecimiento de estreptavidina
PCR con un cebador (perlas de StA encendidas)
PCR con dos cebadores (perlas de StA apagadas)
Preparación de biblioteca

Figura 2

A

sin adaptador de bloqueo	sin exonucleasa	1ra Exonucleasa III+I 2da exonucleasa lambda	sin adaptador de bloqueo Exonucleasa
			
sin adaptador	9 %	0 %	-
1 adaptador	48 %	0 %	-
2 adaptadores	**43 %**	**89 %**	-
Intermedio	-	11 %	-

B

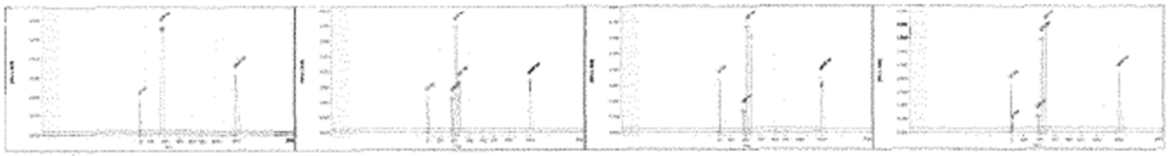
sin adaptador de bloqueo	37,5 pmol adaptador de bloqueo	75,0 pmol adaptador de bloqueo	112,5 pmol adaptador de bloqueo
			
sin adaptador	17 %	9 %	7 %
1 adaptador	57 %	48 %	43 %
2 adaptadores	**26 %**	**43 %**	**50 %**

Figura 3

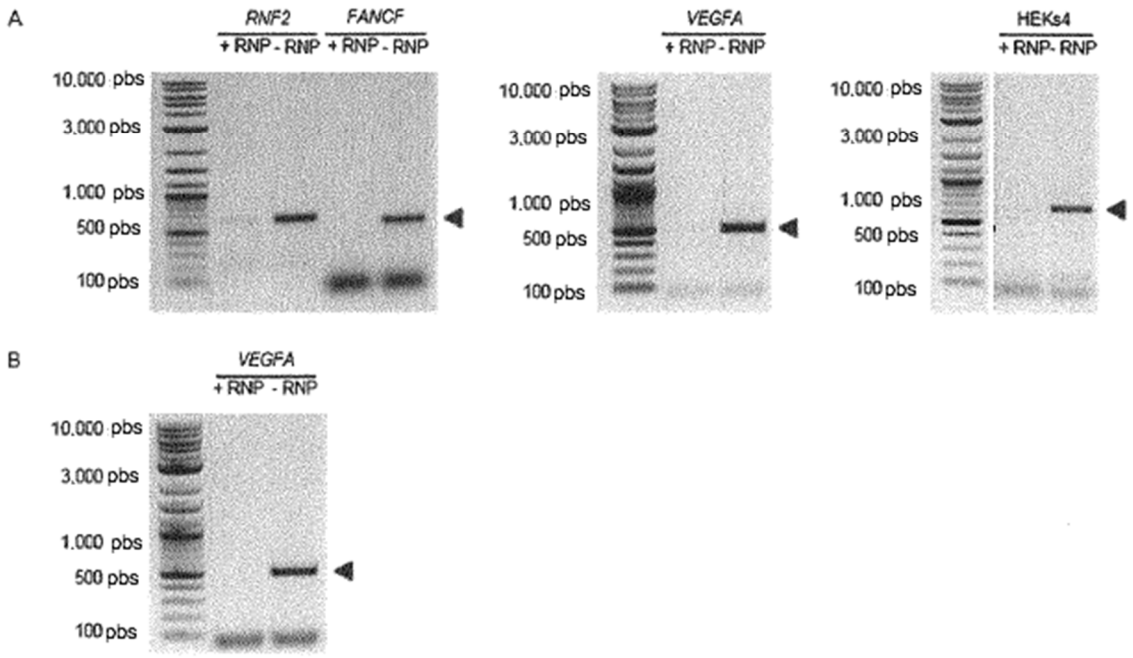

A
RNF2
+ RNP - RNP
FANCF
+ RNP - RNP
VEGFA
+ RNP - RNP
HEKs4
+ RNP- RNP
10.000 pbs
3.000 pbs
1.000 pbs
500 pbs
100 pbs
B
VEGFA
+ RNP - RNP

Figura 4

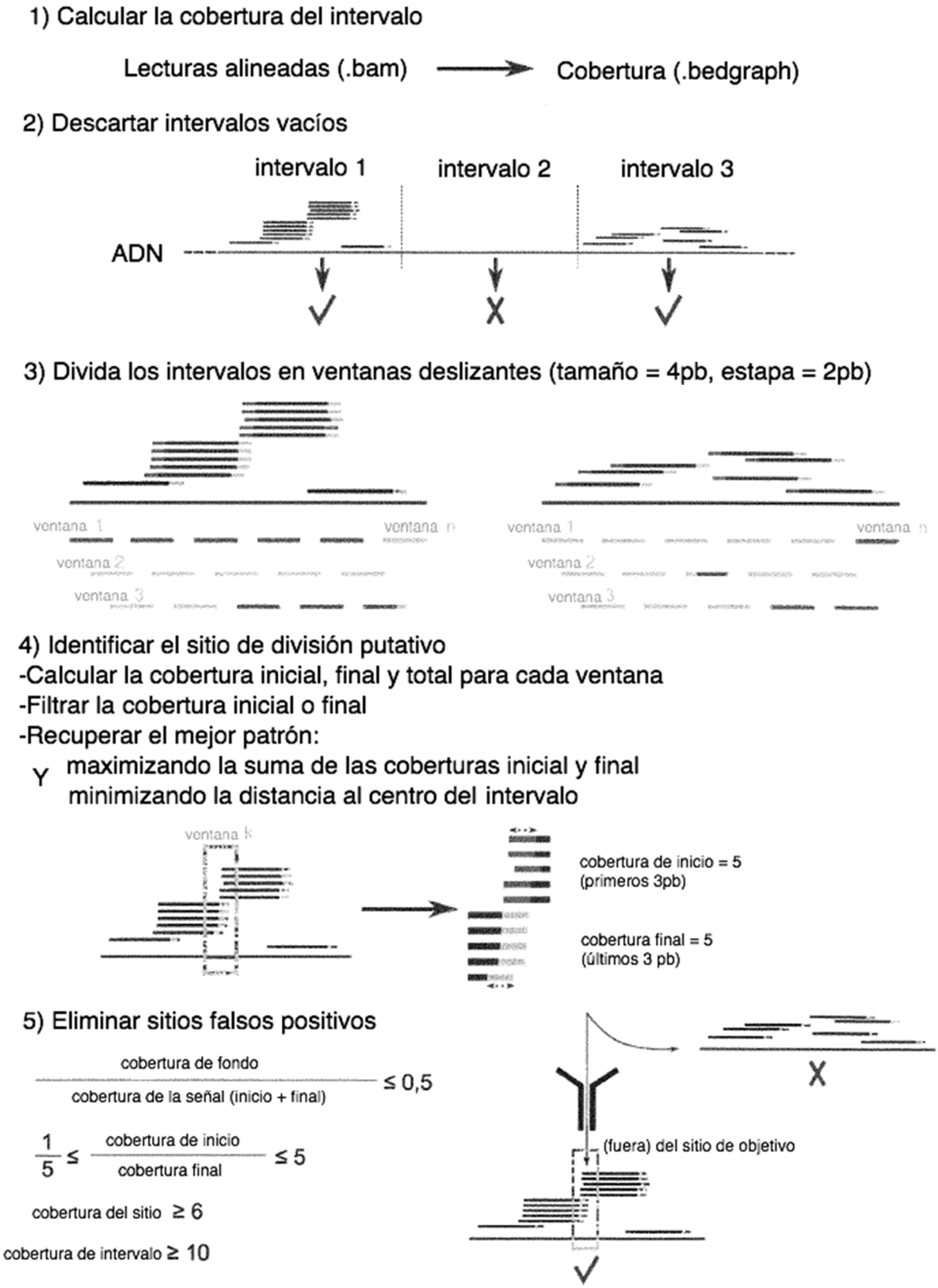

1) Calcular la cobertura del intervalo
Lecturas alineadas (.bam)
Cobertura (.bedgraph)
2) Descartar intervalos vacíos
intervalo 1
intervalo 2
intervalo 3
ADN
3) Divida los intervalos en ventanas deslizantes (tamaño = 4pb, estapa = 2pb)
ventana 1
ventana n
ventana 2
ventana 3
4) Identificar el sitio de división putativo
-Calcular la cobertura inicial, final y total para cada ventana
-Filtrar la cobertura inicial o final
-Recuperar el mejor patrón:
Y maximizando la suma de las coberturas inicial y final
minimizando la distancia al centro del intervalo
ventana k
cobertura de inicio = 5
(primeros 3pb)
cobertura final = 5
(últimos 3 pb)
5) Eliminar sitios falsos positivos
cobertura de fondo / cobertura de la señal (inicio + final) ≤ 0,5
1/5 ≤ cobertura de inicio / cobertura final ≤ 5
cobertura del sitio ≥ 6
cobertura de intervalo ≥ 10
(fuera) del sitio de objetivo

Figura 5

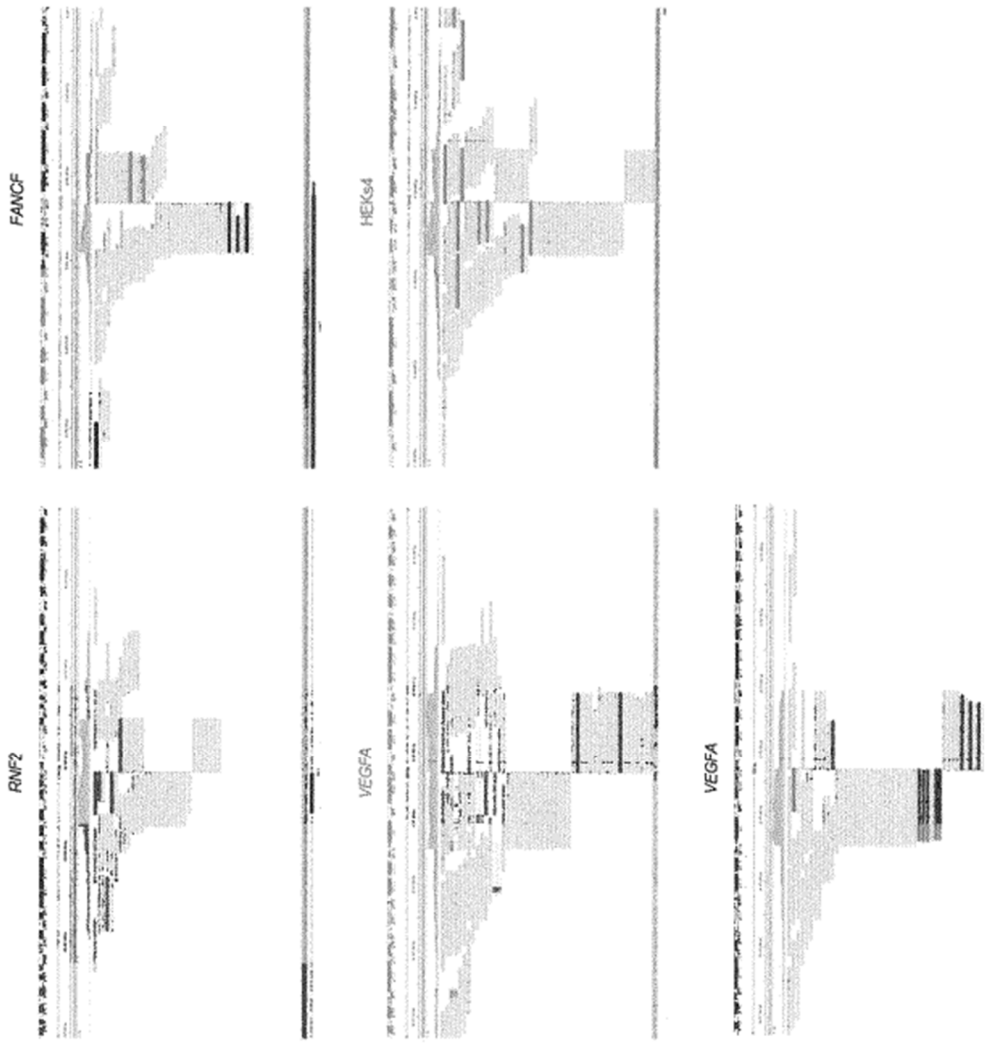


Figura 6

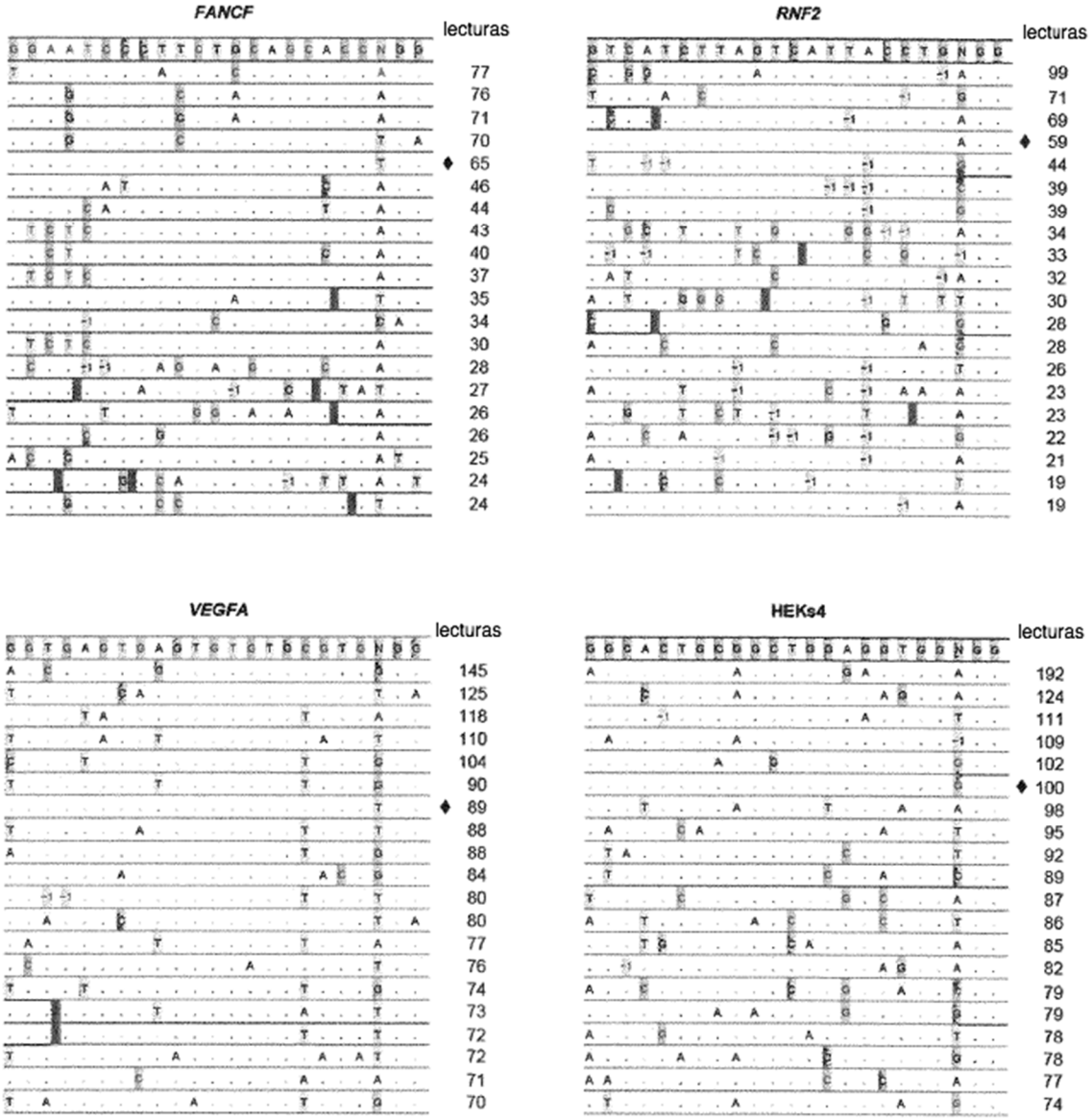

FANCF
lecturas
77
76
71
70
65
46
44
43
40
37
35
34
30
28
27
26
26
25
24
24
RNF2
lecturas
99
71
69
59
44
39
39
34
33
32
30
28
28
26
23
23
22
21
19
19
VEGFA
lecturas
145
125
118
110
104
90
89
88
88
84
80
80
77
76
74
73
72
72
71
70
HEKs4
lecturas
192
124
111
109
102
100
98
95
92
89
87
86
85
82
79
79
78
78
77
74

Figura 7

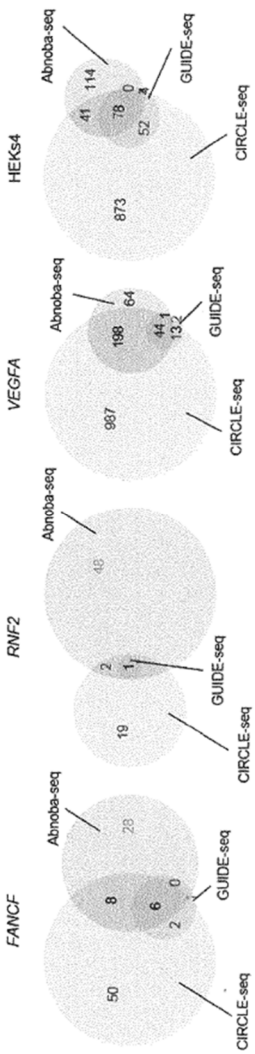

HEKs4
Abnoba-seq
GUIDE-seq
CIRCLE-seq
41
114
78
0
4
52
873
VEGFA
Abnoba-seq
GUIDE-seq
CIRCLE-seq
198
64
44
1
13
2
987
RNF2
Abnoba-seq
GUIDE-seq
CIRCLE-seq
48
2
1
19
FANCF
Abnoba-seq
GUIDE-seq
CIRCLE-seq
28
8
6
0
2
50

Figura 8

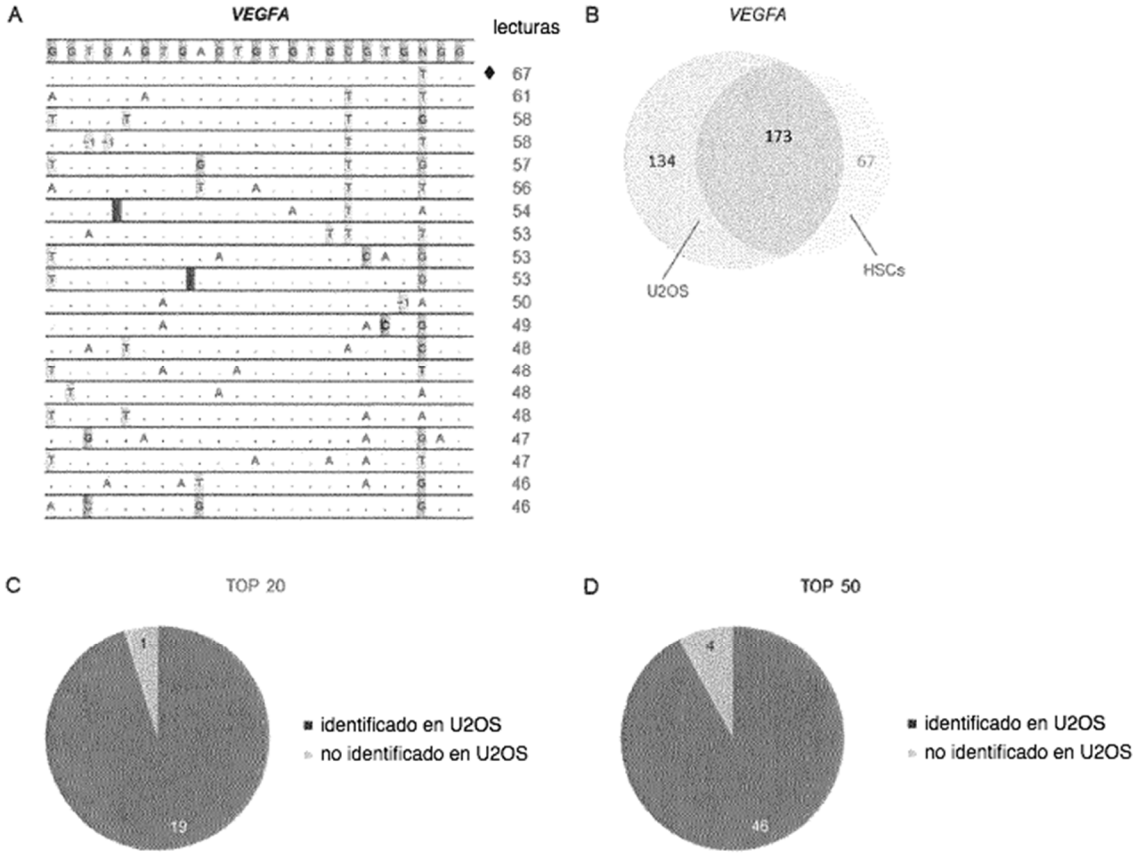
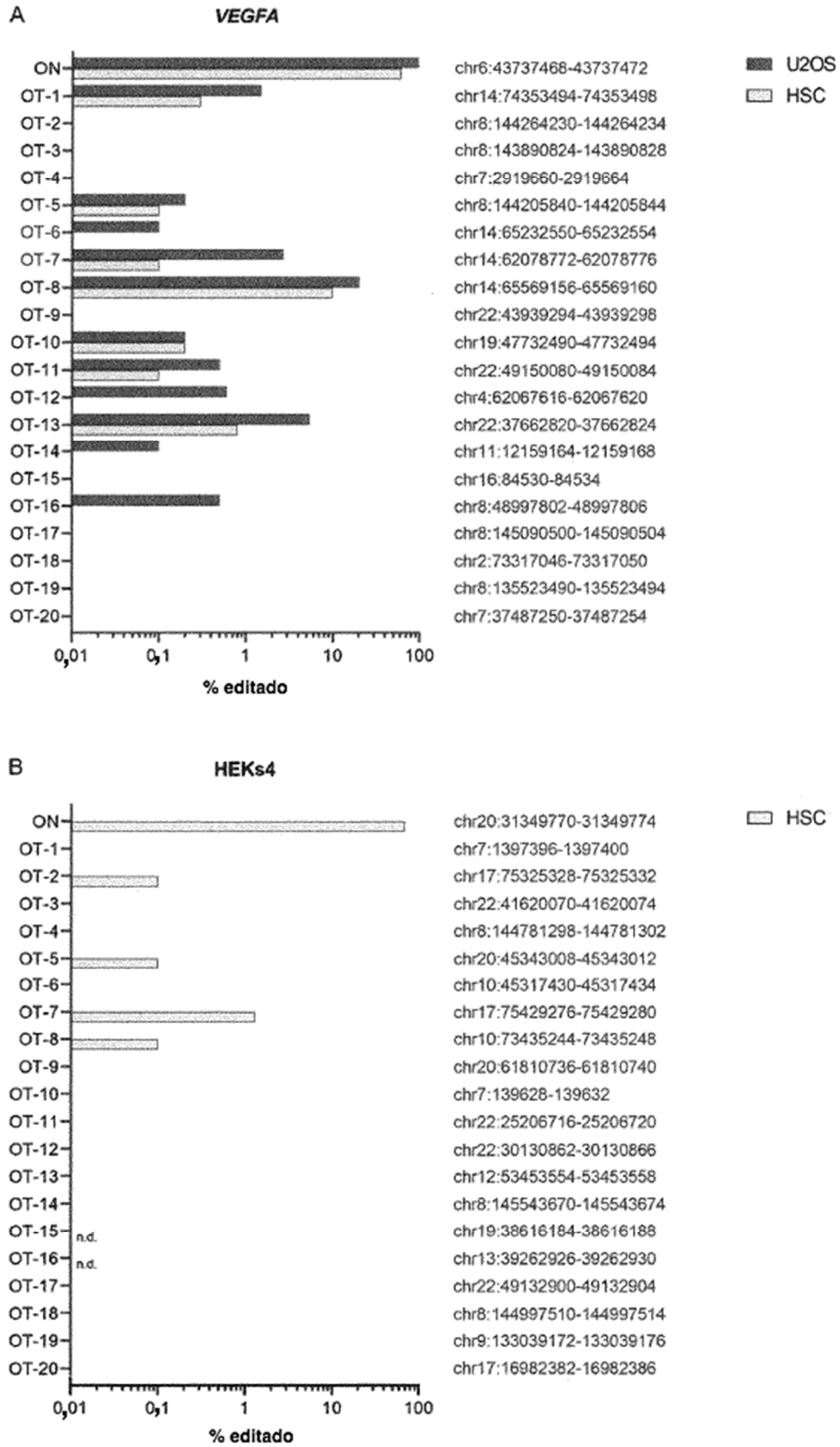

A
VEGFA
lecturas
67
61
58
58
57
56
54
53
53
53
50
49
48
48
48
48
47
47
46
46
B
VEGFA
173
134
67
U2OS
HSCs
C
TOP 20
1
19
identificado en U2OS
no identificado en U2OS
D
TOP 50
4
46
identificado en U2OS
no identificado en U2OS

Figura 9


A
VEGFA
ON
OT-1
OT-2
OT-3
OT-4
OT-5
OT-6
OT-7
OT-8
OT-9
OT-10
OT-11
OT-12
OT-13
OT-14
OT-15
OT-16
OT-17
OT-18
OT-19
OT-20
0,01
0,1
1
10
100
% editado
chr6:43737468-43737472
chr14:74353494-74353498
chr8:144264230-144264234
chr8:143890824-143890828
chr7:2919660-2919664
chr8:144205840-144205844
chr14:65232550-65232554
chr14:62078772-62078776
chr14:65569156-65569160
chr22:43939294-43939298
chr19:47732490-47732494
chr22:49150080-49150084
chr4:62067616-62067620
chr22:37662820-37662824
chr11:12159164-12159168
chr16:84530-84534
chr8:48997802-48997806
chr8:145090500-145090504
chr2:73317046-73317050
chr8:135523490-135523494
chr7:37487250-37487254
U2OS
HSC
B
HEKs4
ON
OT-1
OT-2
OT-3
OT-4
OT-5
OT-6
OT-7
OT-8
OT-9
OT-10
OT-11
OT-12
OT-13
OT-14
OT-15
OT-16
OT-17
OT-18
OT-19
OT-20
n.d.
n.d.
0,01
0,1
1
10
100
% editado
chr20:31349770-31349774
chr7:1397396-1397400
chr17:75325328-75325332
chr22:41620070-41620074
chr8:144781298-144781302
chr20:45343008-45343012
chr10:45317430-45317434
chr17:75429276-75429280
chr10:73435244-73435248
chr20:61810736-61810740
chr7:139628-139632
chr22:25206716-25206720
chr22:30130862-30130866
chr12:53453554-53453558
chr8:145543670-145543674
chr19:38616184-38616188
chr13:39262926-39262930
chr22:49132900-49132904
chr8:144997510-144997514
chr9:133039172-133039176
chr17:16982382-16982386
HSC