

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 937 986**

51 Int. Cl.:

**C12Q 1/6848** (2008.01)

**C12Q 1/6869** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **19.09.2019 PCT/EP2019/075101**

87 Fecha y número de publicación internacional: **02.04.2020 WO20064478**

96 Fecha de presentación y número de la solicitud europea: **19.09.2019 E 19769176 (9)**

97 Fecha y número de publicación de la concesión europea: **09.11.2022 EP 3856928**

54 Título: **Método de caracterización de las modificaciones causadas por el uso de nucleasas de diseño**

30 Prioridad:

**25.09.2018 EP 18196438**

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

**03.04.2023**

73 Titular/es:

**ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG  
(50.0%)**

**Fahnenbergplatz**

**79098 Freiburg, DE y**

**DEUTSCHES KREBSFORSCHUNGSZENTRUM**

**STIFTUNG DES ÖFFENTLICHEN RECHTS (50.0%)**

72 Inventor/es:

**CATHOMEN, TONI;**

**TURCHIANO, GIANDOMENICO;**

**BLATTNER, GEORGES;**

**MONACO, GIANNI;**

**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**

ES 2 937 986 T3

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

## DESCRIPCIÓN

Método de caracterización de las modificaciones causadas por el uso de nucleasas de diseño

La edición del genoma describe la modificación dirigida de los genomas de cualquier tipo de célula de interés con las llamadas "nucleasas de diseño".

5 Se conocen varias nucleasas de diseño, que también se denominan "nucleasas programables" o "nucleasas modificadas". Ciertos ejemplos son las nucleasas de dedos de zinc ("zinc-finger nucleases", ZFN), las nucleasas efectoras similares a activadores transcripcionales ("transcriptional activator-like effector nucleases", TALEN) y las nucleasas modificadas dirigidas por ARN ("RNA-guided engineered nucleases", RGEN) que pueden derivarse del sistema de inmunidad adaptativa procariota de grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR/Cas). Estos medios son importantes y se utilizan ampliamente para la edición del genoma no solo en células cultivadas, sino también en organismos enteros. Las nucleasas de diseño tienen su origen en la naturaleza, pero se modifican artificialmente para que actúen de la forma prevista.

10 La edición del genoma tiene amplias aplicaciones y se ha empleado con éxito para modificar genéticamente microorganismos procariotas y eucariotas, cultivos, ganado, organismos modelo para la investigación, líneas celulares para la selección de fármacos y diversos tipos celulares u órganos para aplicaciones terapéuticas. Para la mayoría de estas aplicaciones, la especificidad de la nucleasa de diseño empleada es un parámetro clave para garantizar el mantenimiento de la integridad del genoma del tipo celular editado.

15 En el contexto de las aplicaciones clínicas en seres humanos, los tipos celulares pertinentes que se han editado con nucleasas de diseño incluyen células madre hematopoyéticas, linfocitos B y T, células madre epidérmicas, células madre pluripotentes, células hepáticas, células musculares y células retinianas. Las enfermedades diana pertinentes incluyen, entre otras, trastornos hereditarios, en concreto, trastornos hereditarios con herencia dominante o enfermedades causadas por mutaciones en genes estrechamente regulados, enfermedades infecciosas o cáncer.

20 Antes de emplear la edición genómica en tipos celulares trasplantables *ex vivo* o antes de aplicar herramientas de edición génica *in vivo* directamente en el paciente, las nucleasas de diseño deben evaluarse cuidadosamente con respecto a su actividad y especificidad. La especificidad de las endonucleasas modificadas es la clave para cualquier aplicación clínica de la edición de genes con el fin de mantener la integridad del genoma y reducir el riesgo de inducir mutaciones oncogénicas. La consecuencia de la mutagénesis inducida por nucleasas de diseño en los denominados sitios diana inespecíficos y/o las aberraciones cromosómicas resultantes suelen denominarse genotoxicidad que, en última instancia, podría provocar cáncer.

25 La expresión "sitio diana" ("on-target site") se utiliza en la presente solicitud para denominar un sitio en el que se pretende introducir una ruptura bicatenaria del ADN mediante el uso de "nucleasas de diseño". Este sitio de acción previsto suele denominarse "sitio diana".

30 Las nucleasas de diseño tienen una cierta especificidad de secuencia y, por tanto, actúan en dicho "sitio diana". Sin embargo, las nucleasas de diseño también pueden actuar en los denominados "sitios diana inespecíficos" ("off-target sites"), que muestran cierto grado de homología de secuencia con el "sitio diana". La expresión "sitio diana inespecífico", tal como se utiliza en el presente documento, se refiere a un sitio en el que las nucleasas de diseño tienen actividad y que suele tener una secuencia que no es idéntica a la secuencia diana de la secuencia de diseño. Un "sitio diana inespecífico" se refiere a una secuencia distinta de un "sitio diana" que cortan las nucleasas de diseño. El hecho de que las nucleasas de diseño tengan actividad incluso en sitios distintos del sitio diana puede deberse a fenómenos que pueden deberse a diversas razones. La desventaja de las nucleasas de diseño que cortan en sitios diana inespecíficos es que esto puede dar lugar a efectos secundarios no deseados como mutaciones, deleciones, inversiones de secuencia y otras alteraciones en el genoma que deben evitarse.

35 En general, la actividad en dianas inespecíficas inducida por nucleasas de diseño puede dar lugar a mutaciones cortas de inserción/delección (indel), grandes deleciones cromosómicas, inversiones cromosómicas, así como translocaciones cromosómicas. A nivel molecular, la actividad en dianas inespecíficas se produce cuando los restos de unión al ADN de las nucleasas de diseño se unen a secuencias del genoma que comparten homología con el sitio diana real. En la última década se ha invertido mucho esfuerzo en aumentar la seguridad de las herramientas de edición genómica, lo que ha permitido mejorar las nucleasas de diseño con una especificidad mucho mayor.

40 No obstante, una evaluación preclínica exhaustiva de la especificidad de las nucleasas de diseño es un requisito claramente establecido por los organismos reguladores, como el Instituto Paul Ehrlich de Alemania o la Administración de Alimentos y Medicamentos de los Estados Unidos (FDA). Existe la necesidad de métodos de diagnóstico aplicados que sean muy sensibles y que permitan a los expertos no solo medir la mutagénesis en dianas inespecíficas, sino también las aberraciones cromosómicas y/o cualquier otra modificación genómica inesperada con alta sensibilidad.

45 Para evaluar el riesgo genotóxico asociado a la aplicación de nucleasas de diseño, tales como las nucleasas CRISPR-Cas, se han desarrollado varios métodos para determinar la actividad en dianas inespecíficas de las nucleasas de diseño o las aberraciones cromosómicas inducidas por nucleasas de diseño. En principio, estos métodos pueden

subdividirse en algoritmos informáticos de predicción (métodos informáticos), métodos de ensayo *in vitro* y métodos basados en células. Todos estos métodos se basan en la secuenciación de nueva generación ("next generation sequencing", NGS) y suelen emplearse en un proceso de dos etapas: En primer lugar, se utiliza un "ensayo de selección" para identificar todos los posibles sitios diana inespecíficos en el genoma de interés. Se utiliza un "ensayo de confirmación" posterior para secuenciar, en el genoma de las células editadas genéticamente, los posibles sitios diana inespecíficos definidos en el ensayo de selección.

Los algoritmos de predicción informáticos se basan en parámetros bien definidos, incluida la similitud con la secuencia diana (Lee *et al.* (2016), *Mol. Ther.*, 24, 475-487). Representan un "ensayo de selección" rápido y relativamente barato, pero, la mayoría de las veces, esos algoritmos pasan por alto sitios diana inespecíficos críticos. A diferencia del análisis informático, los métodos experimentales permiten la identificación de dianas inespecíficas independientemente de los parámetros predeterminados y, por consiguiente, están menos sesgados. Sin embargo, los métodos experimentales son más laboriosos y caros. Además, están sujetos a limitaciones técnicas y algunos carecen de sensibilidad.

En la actualidad, se emplean varios "ensayos de selección" experimentales para determinar los sitios diana inespecíficos y es probable que sean lo suficientemente sensibles como para ser considerados para la evaluación preclínica de la especificidad de las nucleasas de diseño, como, por ejemplo:

- a) documento EP 3 219 810 (secuenciación del genoma completo)
- b) Guide-Seq (Tsai *et al.* (2015), *Nat. Biotechnol.*, 33, 187-197),
- c) BLISS (Yan *et al.* (2017), *Nat. Commun.*, 8, 15058),
- d) Digenome-Seq (Kim *et al.* (2015), *Nat. Methods*, 12, 237-243), y
- e) Circle-Seq (Tsai *et al.* (2017), *Nat. Methods*, 14, 607-614).

Guide-Seq es un método basado en células que introduce oligodesoxinucleótidos cortos bicatenarios (ODNbc) en la célula, además de la nucleasa de diseño. Una vez que la nucleasa de diseño corta el genoma, el ODNbc corto es integrado por la maquinaria celular de reparación del ADN en las rupturas bicatenarias del ADN resultantes, y puede actuar entonces como punto de partida para la secuenciación de alto rendimiento. Este método funciona bien, pero solo en determinadas líneas celulares humanas cuyo genoma puede diferir considerablemente del genoma del paciente.

BLISS detecta rupturas bicatenarias del ADN reales en células mediante un ligamiento de oligos *in vitro* a los extremos de ADN disponibles. El ADN ligado se transcribe *in vitro* y el banco se secuencia mediante secuenciación de alto rendimiento. Digenome-Seq y Circle-Seq son métodos *in vitro*, que se basan en el corte del genoma completo o de fragmentos circularizados del genoma con CRISPR-Cas.

Para Digenome-Seq, la secuenciación del genoma completo se realiza en los genomas cortados *in vitro*, lo que producirá lecturas de secuencias con los mismos extremos 5' en los sitios de corte que luego pueden ser identificados de forma computacional. Para alcanzar la cobertura necesaria y, por tanto, la sensibilidad suficiente, Digenome-Seq debe realizarse en máquinas de secuenciación de alto rendimiento, tales como la línea HiSeq de Illumina. En consecuencia, la aplicación de Digenome-Seq es bastante cara.

En Circle-Seq, los adaptadores de secuenciación se ligan a los extremos 5' cortados, que luego se pueden utilizar para NGS para identificar los sitios diana inespecíficos. Sin embargo, Circle-Seq puede adolecer de posibles sesgos derivados de la necesidad de circularizar el ADN genómico y requiere grandes cantidades de ADN de entrada. Por lo tanto, Circle-Seq no puede realizarse si solo se dispone de una cantidad limitada de muestra, por ejemplo, una biopsia. En todos los casos, estos sitios diana inespecíficos determinados de modo experimental deben validarse en las células del paciente mediante "ensayos de confirmación" que utilizan NGS, tales como la secuenciación multiplexada de amplicones diana, para establecer un perfil de especificidad real de las nucleasas en las células diana clínicamente pertinentes.

Es importante destacar que los métodos descritos anteriormente permiten a los investigadores predecir los sitios diana inespecíficos que son cortados por la nucleasa de diseño elegida, pero ninguno de ellos permite una evaluación de las aberraciones cromosómicas graves inducidas por nucleasas programables, tales como las descritas recientemente (Kosicki *et al.* (2018), *Nat. Biotechnol.*, 36, 765-771).

Otros dos métodos descritos, la secuenciación masiva de traslocaciones en el genoma completo ("high-throughput genome-wide translocation sequencing", HTGTS) y la metodología de secuenciación unidireccional dirigida ("unidirectional targeted sequencing methodology", UDiTaS) pueden identificar translocaciones u otras aberraciones cromosómicas inducidas por nucleasas de diseño. HTGTS (documento WO 2016/081798) y UDiTaS (documento WO 2018/129368) divulgan métodos relacionados con la detección de rupturas bicatenarias del ADN inespecíficas en el genoma. Estos dos métodos también permitieron la identificación de acontecimientos de translocación, pero los análisis bioinformáticos descritos, así como la fragmentación genómica sesgada (uso de tagmentación Tn5 para UDiTaS, enzimas de restricción para HTGTS) limitan considerablemente la sensibilidad de estos enfoques. La HTGTS

no establece un límite inferior de detección ("lower limit of detection", LLoD) ni una sensibilidad. El LLoD de UDiTas se indica como del 0,1 %.

5 Todos los métodos conocidos no pueden identificar reordenamientos cromosómicos que no estén relacionados con la actividad en dianas inespecíficas de una nucleasa de diseño. En concreto, estos métodos no pueden identificar reordenamientos cromosómicos mediados por homología que se desencadenan por la actividad en la diana de una nucleasa de diseño.

HTGTS y UDiTas no son cuantitativos con respecto a acontecimientos de translocación desconocidos.

HTGTS y UDiTas no demostraron funcionar en ADN genómico obtenido de un tipo celular editado genéticamente y que fuera clínicamente pertinente, es decir, células madre hematopoyéticas.

10 El documento WO 2016/081798 divulga la "secuenciación masiva de traslocaciones en el genoma completo" (HTGTS), que es un método para detectar translocaciones cromosómicas inducidas por la expresión de nucleasas de diseño.

15 En concreto, (i) no se conoce la sensibilidad de HTGTS, (ii) HTGTS no es cuantitativa, sino que solo indica frecuencias relativas de unión (es decir, indica las frecuencias relativas de acontecimientos de translocación en una población celular), (iii) HTGTS no identifica ni cuantifica deleciones grandes (>1 kb) en el sitio diana y no permite la identificación de reordenamientos cromosómicos mediados por RH. Por último, (iv) el HTGTS necesita entre 20 y 100 microgramos de ADN genómico (en contraste con los 0,5 microgramos que necesita el método reivindicado) y, por tanto, no es aplicable para analizar muestras clínicamente pertinentes.

20 El documento WO 2018/129368 divulga la "secuenciación dirigida unidireccional" (UDiTas) que detecta intencionadamente reordenamientos cromosómicos inducidos, es decir, reordenamientos cromosómicos entre dos rupturas bicatenarias del ADN inducidas intencionadamente. Este método no muestra ningún dato que demuestre que es capaz de identificar reordenamientos cromosómicos no intencionados, es decir, no puede identificar ni cuantificar reordenamientos cromosómicos desconocidos. Además, no detecta deleciones genómicas no intencionadas en el sitio diana y no permite la identificación de reordenamientos cromosómicos mediados por RH. Es importante señalar que no se han presentado pruebas que demuestren que el método puede detectar aberraciones cromosómicas inducidas por una única ruptura bicatenaria del ADN.

25 Es un objetivo de la presente invención proporcionar un método que pueda identificar modificaciones genómicas, incluyendo aberraciones cromosómicas graves, y que sea (i) muy sensible, (ii) muy específico, (iii) cuantitativo, (iv) capaz de detectar tipos de reordenamientos cromosómicos no descritos previamente, y (v) que se realice directamente en el ADN genómico aislado del tipo celular clínicamente pertinente. El método se denomina en el presente documento CAST-Seq (análisis de aberraciones cromosómicas mediante PCR mediada por conector dirigido único). La invención se define además en las reivindicaciones.

30 CAST-Seq se basa en la PCR mediada por conector dirigido único (LM-PCR) y utiliza cebadores señuelo para mejorar la relación señal-ruido. Este método permite identificar sitios diana inespecíficos y detectar modificaciones genómicas derivadas de la actividad tanto dentro en la diana como en dianas inespecíficas de nucleasas de diseño, incluidas grandes deleciones, inversiones y translocaciones con una sensibilidad inigualable. Es importante destacar que, gracias a la alta sensibilidad de CAST-Seq, el ensayo puede realizarse con menos de 1 µg de ADN genómico como entrada. Por lo tanto, CAST-Seq puede aplicarse directamente a cualquier tipo de célula humana clínicamente pertinente elegida, incluidas las células editadas genéticamente *ex vivo* antes del trasplante o las células derivadas de una biopsia de órganos editados genéticamente. Esta configuración única y el hecho de que CAST-Seq se realiza directamente en el tipo de célula editada o tejido de interés, hace que un ensayo de confirmación basado en NGS sea redundante al unir directamente la "prueba de selección" con la "prueba de confirmación". Por lo tanto, CAST-Seq puede mejorar sustancialmente el proceso al detectar aberraciones cromosómicas en "sitios diana" y "sitios diana inespecíficos".

#### Realizaciones preferidas de la presente invención

45 La presente invención proporciona un método novedoso que se utiliza para caracterizar las modificaciones genómicas causadas por el uso de endonucleasas de diseño en cualquier tipo de célula eucariota, incluidas, entre otras, células humanas, células de primates no humanos, tipos de células de mamífero, tipos de células de vertebrado, levaduras, células vegetales.

50 CAST-Seq puede caracterizar la aberración cromosómica causada tanto por la actividad en dianas inespecíficas como por la actividad en la diana de las nucleasas de diseño. Como tal, también proporciona un nuevo método de diagnóstico para clasificar translocaciones acéntricas/dicéntricas poco frecuentes derivadas de la fusión de dos cromosomas hermanos en sitios diana, o grandes deleciones cromosómicas originadas en el sitio de corte diana. Es importante destacar que CAST-Seq también puede detectar aberraciones cromosómicas inducidas por nucleasas de diseño que se inician en sitios comunes de ruptura ("common breaking sites", CBS) o en sitios naturales de ruptura ("naturally occurring breaking sites", NBS) del genoma.

- En el contexto de las aplicaciones clínicas de la edición genómica en seres humanos, CAST-Seq puede aplicarse con eficacia durante la fase preclínica para caracterizar la especificidad de cualquier endonucleasa (por ejemplo, entre otras, nucleasas de diseño de los tipos CRISPR-Cas, TALEN, ZFN, MegaTAL) con el fin de elegir, por ejemplo, una endonucleasa que combine alta actividad con alta especificidad. En este contexto, CAST-Seq también puede utilizarse para caracterizar el impacto de las modificaciones introducidas en una endonucleasa programable, tales como las modificaciones que afectan a la afinidad, la especificidad y/o la estabilidad de la endonucleasa.
- Además, debido a la alta sensibilidad, son suficientes cantidades mínimas de ADN genómico para realizar un análisis completo. Por lo tanto, CAST-Seq también puede emplearse para caracterizar un producto de edición génica fabricado antes de su aplicación al paciente como parte de un análisis de control de calidad.
- CAST-Seq también puede emplearse en la fase de seguimiento del paciente. Por ejemplo, CAST-Seq puede utilizarse para evaluar la integridad genómica de diversos tipos de células de sangre periférica tras el trasplante de células madre hematopoyéticas editadas genéticamente.
- Además, dado que CAST-Seq es un método semicuantitativo, la alteración en las frecuencias de modificaciones específicas puede seguirse a lo largo del tiempo, por ejemplo, para evaluar la expansión clonal de ciertas modificaciones en células precancerosas tempranas. Una vez se disponga de datos suficientes, esto también permitirá utilizar CAST-Seq para predecir el resultado y/o el riesgo de mutaciones genotóxicas en el desarrollo del cáncer.
- Las aplicaciones de CAST-Seq incluyen, entre otras, trastornos en los que se aplica la edición genómica *ex vivo*, tales como defectos del sistema inmunitario, hemofilia, hemoglobinopatías, trastornos metabólicos, enfermedades infecciosas y mejoras de las inmunoterapias basadas en linfocitos T para combatir el cáncer.
- CAST-Seq también puede aplicarse para evaluar el resultado de la edición génica realizada *in vivo*, es decir, aplicando directamente las herramientas de edición genómica al paciente a través, por ejemplo, del transporte viral o el transporte mediante nanopartículas o cualquier otro medio. En este contexto, una pequeña biopsia tomada del órgano diana (por ejemplo, el hígado) será suficiente para evaluar el impacto de las modificaciones genéticas o cromosómicas inducidas por nucleasas de diseño. Este enfoque no solo puede aplicarse al órgano diana, sino que también podría emplearse para evaluar el impacto de los enfoques de edición genética en órganos diana inespecíficos. También se pueden emplear estudios longitudinales para rastrear el destino de las células editadas genéticamente.
- El método de la presente invención puede utilizarse preferentemente en trastornos en los que se aplica la edición del genoma *in vivo*, tales como la hemofilia, los trastornos metabólicos, los trastornos oculares genéticos, los trastornos auditivos hereditarios, los trastornos musculares hereditarios, las enfermedades neuromusculares y los trastornos que afectan al sistema nervioso central.
- La presente invención proporciona una herramienta de diagnóstico novedosa para el estudio del genoma del cáncer. Dada una mutación o un estímulo concretos, CAST-Seq puede cartografiar el sitio común de ruptura (CBS) y retratar la firma de mutación de un modelo cancerígeno dado. Mediante el uso de CAST-Seq en un enfoque de este tipo, será posible definir nuevos enfoques convencionales para predecir y diagnosticar los resultados del cáncer.
- El método de la presente invención se refiere a la detección de modificaciones no deseadas en un ácido nucleico, preferentemente un ácido genómico, causadas por la actividad de una nucleasa de diseño. Dichas modificaciones se producen preferentemente en los denominados "sitios diana inespecíficos", pero también pueden producirse en los denominados "sitios diana". Para detectar tales modificaciones no deseadas, el método según la presente invención realiza una etapa de amplificación del ácido nucleico que es preferentemente una PCR (reacción en cadena de la polimerasa). También pueden utilizarse otros métodos adecuados para amplificar ácidos nucleicos, tales como los métodos de amplificación isotérmica, las reacciones en cadena de la ligasa, la amplificación isotérmica mediada por bucles, la amplificación por desplazamiento múltiple o la amplificación basada en secuencias de ácidos nucleicos ("nucleic acid sequence based amplification", NASBA).
- En la primera etapa, se prepara un banco a partir de células eucariotas expuestas a una nucleasa de diseño en condiciones que permiten a la nucleasa de diseño introducir al menos una ruptura bicatenaria del ADN (etapa a). Las denominadas nucleasas de diseño adecuadas son, preferentemente, las nucleasas CRISPR-Cas; TALEN; ZFN; MegaTAL, por nombrar solo algunas.
- A continuación, el ácido nucleico del banco se convierte en "fragmentos aleatorios" (etapa b). En una realización preferida, los fragmentos obtenidos tienen una longitud de aproximadamente 350 pares de bases. Esto significa que la mayoría de los fragmentos oscila entre aproximadamente 200 y aproximadamente 500 pares de bases, siendo la mediana del tamaño de los fragmentos de aproximadamente 350 pares de bases. La fragmentación puede obtenerse mediante medidas físicas, tales como la aplicación de fuerzas de cizallamiento o la sonicación o, como alternativa, la fragmentación también puede obtenerse mediante digestión con enzimas adecuadas que corten el ácido nucleico bicatenario en sitios aleatorios. Esta etapa no implica la acción de enzimas de restricción definidas ni de transposones.
- Para tener extremos uniformes en cada fragmento, se realiza una reparación para obtener extremos que preferentemente se modifican para tener una A prominente en el extremo 3' (etapa c). Esos "fragmentos aleatorios"

que tienen una A prominente se acoplan entonces con un conector adecuado que tiene también una T 3' prominente que es complementaria con la A de los fragmentos reparados. Esto mejora la tasa de ligamiento del conector a los extremos reparados de los "fragmentos aleatorios".

5 En una realización preferida, el conector comprende también una secuencia que es complementaria al cebador directo o al cebador inverso, respectivamente. Esta construcción permite una fácil amplificación del fragmento que tiene el conector.

10 A continuación, se lleva a cabo una primera reacción de amplificación del ácido nucleico con un "cebador para diana" adecuado y un "cebador para conector" adecuado que son complementarios con una secuencia muy próxima a la secuencia diana o con una posición de unión introducida preferentemente por el conector. En una realización preferida, los sitios de unión de los cebadores para diana están situados a una distancia de al menos 25 nucleótidos, preferentemente de al menos 35 nucleótidos y más, preferentemente de al menos 50 nucleótidos cadena arriba del sitio diana. Los cebadores señuelo aumentan la sensibilidad y la especificidad del método según la invención.

15 Además de los cebadores directos e inversos, se añaden al menos uno, preferentemente al menos dos cebadores señuelo. El propósito de los cebadores señuelo es suprimir o al menos reducir sustancialmente la amplificación de dichos fragmentos que solo contienen secuencias diana en un "fragmento aleatorio", es decir, que no contienen un acontecimiento de aberración cromosómica. Cuando se reduce la amplificación de los fragmentos que contienen la secuencia diana, hay más posibilidades de identificar los sitios diana inespecíficos, ya que el número de fragmentos que contienen dichos sitios diana inespecíficos aumenta en comparación con los fragmentos que solo contienen secuencias diana. Un "cebador para diana" es un cebador que se une específicamente al sitio diana. Tiene una alta identidad y una longitud suficiente para proporcionar una alta especificidad de unión.

20 En una realización preferida de la presente invención, se utilizan al menos dos cebadores señuelo diferentes, en los que ambos cebadores señuelo son complementarios con secuencias muy cercanas cadena abajo del sitio diana. En función de las secuencias específicas que rodean el sitio diana deseado, se deben seleccionar las secuencias con las que el cebador señuelo es complementario.

25 En una realización preferida, los sitios de unión de los cebadores señuelo no se solapan. En una realización preferida, un cebador es complementario con la cadena superior de la secuencia de ADN, mientras que el otro cebador es complementario con la cadena inferior de la secuencia de ADN. Preferentemente, las secuencias están situadas a una distancia de al menos 10 nucleótidos, preferentemente de al menos 15 nucleótidos y más, preferentemente de al menos 30 nucleótidos cadena abajo del sitio diana. Las condiciones óptimas de ubicación de las secuencias adecuadas para unirse a los cebadores señuelo tienen que evaluarse para cada sitio diana. El efecto que se obtiene utilizando los cebadores señuelo es que se reduce la aparición de secuencias amplificadas que contienen el sitio diana y que, por lo tanto, aumenta sustancialmente la probabilidad de detectar sitios diana inespecíficos. Dado que se conocen las secuencias cadena arriba y cadena abajo del sitio diana, se puede seleccionar fácilmente una secuencia adecuada para el cebador señuelo. Preferentemente, los cebadores señuelo no están bloqueados en ninguno de sus extremos para permitir que la polimerasa extienda los cebadores señuelo.

35 Las secuencias obtenidas por el método según la presente invención se someten entonces a secuenciación de alto rendimiento y la información de las secuencias obtenidas se analiza con las medidas bioinformáticas que son bien conocidas por los expertos en la materia.

40 Aunque se han descrito previamente métodos para detectar la actividad en dianas inespecíficas o las aberraciones cromosómicas inducidas por nucleasas, el método según la presente invención, abreviado como CAST-Seq, es una nueva herramienta fundamental para la evaluación del riesgo clínico en la edición terapéutica del genoma mediante la inclusión de algunas características críticas novedosas.

Las ventajas que pueden obtenerse por el método de la presente invención son, en concreto:

- (i) es muy sensible y muy específico,
- 45 (ii) es cuantitativo,
- (iii) es capaz de detectar tipos de aberraciones cromosómicas no descritas anteriormente, y
- (iv) puede realizarse directamente en el tipo de célula clínicamente pertinente.

Las propiedades ventajosas se describen en los ejemplos expuestos en el presente documento:

- (i) Mayor sensibilidad y especificidad
- 50 Los datos actuales indican que CAST-Seq es capaz de detectar 10 acontecimientos de translocación (= 1 acierto) en 150 000 genomas haploides (500 ng de ADN genómico de entrada), lo que corresponde a un límite inferior de detección (LLOD) de aproximadamente el 0,007 %. Esta alta sensibilidad, junto con una mayor especificidad, podría alcanzarse mediante el uso de cebadores señuelo, que se describen por primera vez en el presente enfoque.

(ii) Cuantitativa

5 Los puntos de ruptura cromosómica, en combinación con el sitio de ligamiento del adaptador, crean identificadores moleculares distintivos, lo que permite determinar un número de translocaciones individuales, agruparlas en acontecimientos provocados por un desencadenante concreto, y cuantificar las frecuencias de acontecimientos muy poco frecuentes basándose en la cantidad conocida de genomas de entrada. La correlación lineal entre el número de aciertos de CAST-Seq y el número real de reordenamientos cromosómicos, determinado mediante ddPCR cuantitativa, confirma la naturaleza cuantitativa del método y su alta sensibilidad.

(iii) Aberraciones cromosómicas no descritas anteriormente

10 CAST-Seq identificó por primera vez reordenamientos cromosómicos no relacionados con la actividad en dianas inespecíficas de una nucleasa de diseño. En concreto, se descubrió que las rupturas bicatenarias del ADN inducidas por nucleasas son solo uno de los factores que impulsan las translocaciones. CAST-Seq demuestra, por primera vez, que las regiones que comparten una homología sustancial con el gen diana, incluso si no contienen un sitio diana inespecífico, probablemente estén sujetas a reordenamientos cromosómicos.

(iv) Realización en un tipo de célula clínicamente pertinente

15 A diferencia de HTGTS/UDiTaS, se ha demostrado que CAST-Seq puede realizarse en ADN genómico obtenido de células madre hematopoyéticas editadas genéticamente, es decir, un tipo celular clínicamente pertinente.

20 El método según la presente invención se ilustra y describe con más detalle en las figuras, las tablas y los experimentos. El experto en la materia es consciente de que los resultados expuestos representan una realización preferida, en la que las características individuales de los experimentos o las figuras pueden combinarse con facilidad con otras características expuestas en otros experimentos del presente documento. Normalmente, no es necesario que todas las características de un ejemplo solo puedan utilizarse juntas.

25 En otra realización, la presente invención se refiere además a kits para realizar un método de la presente invención. Dicho kit comprende los componentes necesarios para llevar a cabo el método específico descrito en el presente documento. En concreto, el kit contiene los cebadores, los conectores específicos y los cebadores señuelo, así como las enzimas necesarias para llevar a cabo la reacción. Todos los componentes descritos en los métodos divulgados en el presente documento pueden estar contenidos por sí solos o juntos en dicho kit.

Las realizaciones preferidas de la presente invención se describen e ilustran con más detalle en las figuras y los ejemplos de la presente solicitud.

Se utilizaron las siguientes abreviaturas en las figuras y tablas, así como en los experimentos:

Abreviatura	Explicación
CBS	Sitio común de ruptura
CCR2	Receptor de quimiocinas C-C de tipo 2
CCR5	Receptor de quimiocinas C-C de tipo 5
CD34	Agrupamiento de diferenciación 34
ddPCR	PCR digital en nanogotas
DSB	Ruptura bicatenaria del ADN
FASTQ	Formato de archivo de calidad de la herramienta de búsqueda rápida de alineamientos
FANCF	Grupo de complementación F de la anemia de Fanconi
FLASH	Ajuste rápido de la longitud de las lecturas cortas
HTGTS	Secuenciación masiva de traslocaciones en el genoma completo

Abreviatura	Explicación
RH	Recombinación homóloga
IGV	Visor integrado del genoma
LM-PCR	PCR mediada por conector
NBS	Sitio natural de ruptura
NGS	Secuenciación de nueva generación
DI	Sitio diana inespecífico
PAM	Motivo adyacente al protoespaciador
PCR	Reacción en cadena de la polimerasa
SAMtools	Herramientas de cartografiado/alineamiento de secuencias
UDiTaS	Metodología de secuenciación dirigida unidireccional
VEGFA	Factor A de crecimiento endotelial vascular

En concreto, las figuras muestran los resultados de los experimentos como sigue:

Figura 1. Esquema del proceso CAST-Seq

5 (a) Preparación del banco. La actividad simultánea en la diana (cromosoma gris oscuro) y en una diana inespecífica (cromosoma gris claro) de las nucleasas de diseño (ilustradas con una tijera) en las células puede inducir una translocación entre las dos rupturas bicatenarias ("double strand breaks", DSB) del ADN, dando lugar, por ejemplo, a una translocación recíproca. De este modo, el cromosoma diana se separa en una parte centromérica (c) y otra telomérica (t). En la mayoría de los casos, no se producirá ninguna translocación (lado derecho). El ADN genómico derivado de células no tratadas y de células editadas genéticamente se fragmenta aleatoriamente y se repara en sus extremos para permitir la adición de una proyección 3'-A. Esta proyección corta se utiliza para el posterior ligamiento de un conector corto (negro). Para simplificar, solo se representa la reacción con el extremo centromérico. Una segunda reacción con el extremo telomérico (muy a la izquierda) se realiza de forma análoga. La primera PCR se realiza con cebadores (flechas blancas) que se unen al sitio diana y a las secuencias conectoras. A la reacción de PCR se añaden los denominados cebadores "señuelo" (flechas negras), diseñados para unirse muy cerca del sitio de corte en la diana. No pueden unirse a acontecimientos de translocación (izquierda), pero impiden la amplificación de sitios diana no modificados (derecha) generando amplicones cortos que no pueden seguir amplificándose en las siguientes etapas de la PCR. La segunda PCR se realiza con cebadores internos que portan proyecciones 5' que se utilizan en la tercera PCR para añadir los códigos de barras para NGS. (b) Proceso bioinformático. Los archivos FASTQ derivados de NGS se procesan según el esquema general. Los recuadros agrupan las principales etapas del flujo bioinformático: apareamiento emparejamiento y filtrado, recorte, alineamiento, definición de agrupamientos y análisis de agrupamientos.

Figura 2. Definición bioinformática de categorías.

25 (a) Distancia base de lectura. Para calcular la probabilidad de que una lectura esté dentro de un agrupamiento por azar, en lugar de ser un acontecimiento provocado por una nucleasa de diseño, la muestra CAST-Seq de células editadas genéticamente se comparó con un banco de lecturas aleatorias creado de modo informático que contiene el mismo número de lecturas. La distribución de la distancia de lecturas consecutivas se muestra en una escala logarítmica. En este ejemplo, la línea umbral de 2500 pb describe un área <5 % en el banco aleatorio, lo que significa que la probabilidad de que una lectura esté dentro de un agrupamiento por azar es inferior al 5 % ( $p < 0,05$ ). El análisis CAST-Seq de células no tratadas se muestra como control. (b) Puntuación del alineamiento de la secuencia diana. Se comparó una región genómica de 500 pb que rodeaba estos sitios de translocación con 10 000 secuencias aleatorias de 500 pb. Cada sitio se alineó con la secuencia diana de la nucleasa de diseño utilizando una tabla de puntuación (tabla 12). Si la puntuación de

alineamiento de la secuencia diana de la secuencia era superior al 5 % de la mejor puntuación en las secuencias aleatorias, el acontecimiento se clasificaba como translocación derivada de la actividad en una diana inespecífica (DI). (c) Tramo máximo de la región de homología. En el caso de los sitios que no son dianas inespecíficas, se buscó la subcadena homóloga común más larga entre la región diana y la región de translocación dentro de una ventana de 5 kb alrededor del sitio de translocación. Si la longitud de la subcadena homóloga era superior al 5 % de la subcadena más larga de las secuencias aleatorias, el acontecimiento se clasificaba como translocación mediada por recombinación homóloga (RH). Todos los demás se clasificaron como translocación derivada del sitio común de ruptura (CBS).

Figura 3. Efecto de los cebadores oligonucleótidos señuelo

(a) Esquema del sistema de ensayo de señuelos. La eficacia de los cebadores señuelo (flechas negras) se ensayó en el locus *CCR5* utilizando dos cebadores específicos del locus (flechas blancas) que amplifican un fragmento de 412 pb. La presencia de los cebadores señuelo debería reducir o impedir la amplificación del fragmento de 412 pb. F, cebador directo; R, cebador inverso. (b) Utilización de cebadores señuelo bloqueados. La PCR se realizó con cebadores de *CCR5* en combinación con cebadores señuelo bloqueados por una fosforilación 3' (barras negras). Se muestran las siguientes amplificaciones: Controles: F, reacción solo con el cebador de *CCR5* directo; 1D, solo se utilizó uno de los dos cebadores señuelo; H<sub>2</sub>O, sin plantilla en la reacción. Las proporciones 1:1; 1:5 y 1:10 reflejan la proporción de cebadores de diana *CCR5* a los cebadores señuelo. (c) Cebadores señuelo no bloqueados. La PCR se realizó con cebadores específicos de *CCR5* en combinación con cebadores señuelo no bloqueados. Se muestran las siguientes amplificaciones: Control H<sub>2</sub>O, sin plantilla en la reacción. Las proporciones 1:1; 1:5 y 1:10 reflejan la proporción de cebadores de diana *CCR5* a los cebadores señuelo. (d) Cebador señuelo no bloqueado individual. La PCR se realizó con cebadores de *CCR5* en combinación con un solo cebador señuelo de orientación inversa. Se muestran las siguientes amplificaciones: Control F, cebador directo de *CCR5* en combinación con cebador señuelo inverso. (b-d) Los tamaños de los amplicones se indican a la izquierda, las diferentes proporciones de *CCR5* a cebadores señuelo que se ensayaron se indican en la parte inferior como 1:1, 1:5, 1:10. Todas las secuencias de cebadores se indican en la tabla 2.

Como puede observarse en la figura 3, el cebador señuelo no bloqueado podía reducir o impedir eficazmente la amplificación del fragmento de 412 pb (c) y (d). Esto sugiere que el uso de cebadores señuelo puede cambiar la proporción de la amplificación del sitio diana a la amplificación de las plantillas de PCR que contienen un acontecimiento de translocación (véase la figura 1a). Por lo tanto, en la primera ronda de amplificación (véase la figura 1a), los cebadores señuelo no bloqueados impiden o reducen sustancialmente la amplificación de los acontecimientos que no son de translocación, mientras que el uso de cebadores señuelo bloqueados no tiene tal efecto (b).

Figura 4. Modificaciones genómicas cartografiadas mediante CAST-seq.

El ADN genómico aislado de células madre y progenitoras hematopoyéticas positivas a CD34, que fueron editadas con complejos de ribonucleoproteínas CRISPR-Cas9 dirigidos al exón 3 del locus *CCR5* (sitio diana: 5'-GTGAGTAGAGCGGAGGCAGGAGG (SEQ ID NO:1), PAM subrayado), se sometió a CAST-Seq. (a) Cartografiado de las modificaciones genómicas. Todos los sitios pertinentes de modificación genómica identificados por CAST-Seq se muestran en un ideograma cromosómico. Los sitios cartografiados pueden subdividirse en tres categorías principales: aberraciones cromosómicas mediadas por sitios diana inespecíficos (DI), por recombinación mediada por homología (RH) o por sitios comunes de ruptura (CBS). (b) Análisis DI. Los gráficos circulares indican las fracciones de desapareamientos y protuberancias encontradas en los sitios cartografiados. Se indican los números de desapareamientos/protuberancias de 0 a 5 y más.

Figura 5. Alineación.

El ADN genómico aislado de células madre y progenitoras hematopoyéticas positivas a CD34, que fueron editadas con complejos de ribonucleoproteínas CRISPR-Cas9 dirigidos al exón 3 del locus *CCR5* (sitio diana: 5'-GTGAGTAGAGCGGAGGCAGGAGG (SEQ ID NO:1), PAM subrayado), se sometió a CAST-Seq. (a) Alineación. El sitio diana de referencia (fila superior: 5'-GTGAGTAGAGCGGAGGCAGGNRG (SEQ ID NO:2); PAM subrayado; N, cualquier nucleótido; R, purina) y los primeros 25 sitios diana inespecíficos (DI) identificados por CAST-Seq. Se resaltan los nucleótidos desapareados y las protuberancias, es decir, las inserciones/delecciones de nucleótidos dentro de los sitios diana inespecíficos con respecto al sitio diana de referencia. "1" significa una inserción de 1 nucleótido, "-" una delección de 1 nucleótido. La posición de inicio del agrupamiento se indica a la izquierda. (b) Diversidad de secuencias diana inespecíficas. Se creó un logotipo de secuencia a partir de la colección de sitios diana inespecíficos alineados, que representa la secuencia consenso y la diversidad de las secuencias diana inespecíficas. (c) Tolerancia a desapareamiento y protuberancias. Los sitios diana inespecíficos identificados se alinearon con la secuencia diana de 23 nucleótidos de longitud y luego se agruparon en 4 regiones de nucleótidos de longitud reconocidas por el ARNg (1-4, 5-8, 9-12, 13-16, 17-20) y el tramo de 3 nucleótidos de longitud unido por la proteína Cas9 (PAM, 21-23). Se indica la fracción de desapareamientos y protuberancias que se encuentran en cada uno de estos grupos.

Figura 6. Representación esquemática de la lectura de CAST-Seq.

Las células madre y progenitoras hematopoyéticas positivas a CD34 se editaron con complejos de ribonucleoproteínas CRISPR-Cas9 dirigidos al locus *CCR5* en el exón 3 (sitio diana: 5'-GTGAGTAGAGCGGAGGCAGGAGG (SEQ ID NO:1), PAM subrayado). Se extrajo el ADN genómico al cabo de 7 días y se sometió a CAST-Seq. (a) Visualización de los resultados de CAST-seq. Se utilizó IGV para visualizar los resultados de CAST-Seq en la vecindad del locus diana *CCR5*. Cada lectura de CAST-Seq cartografiadas está representada por una barra. Las barras de color gris claro indican la orientación inversa y las barras de color gris oscuro la orientación directa, respectivamente. La cobertura, es decir, el número de lecturas cartografiadas, se indica en la parte superior, y la ubicación de los loci *CCR5* y *CCR2* en la parte inferior. (b) Ejemplos de aberraciones cromosómicas. Se indican dos ejemplos sobre cómo interpretar los resultados: (1) una translocación dicéntrica y una inversión de secuencia inducidas por la actividad simultánea en la diana en el locus *CCR5* y la actividad del sitio diana inespecífico muy cerca de *CCR2*; (2) una gran deleción provocada por la actividad en la diana inespecífica en el locus *CCR5* que provocó un acontecimiento de recombinación homóloga (RH) con un sitio en el locus *CCR2* que comparte una alta homología de secuencia con el sitio diana en *CCR5*. (c) Resumen esquemático de todas las aberraciones cromosómicas graves identificadas por CAST-Seq tras el corte en el sitio diana.

15 Figura 7. Proceso bioinformático más restrictivo.

(a) Descripción general. Los archivos FASTQ derivados de NGS se procesaron según con la descripción general. Los recuadros agrupan las principales etapas del flujo bioinformático: apareamiento y filtrado, recorte, alineamiento, definición de agrupamientos, análisis de agrupamientos, filtrado. (b) Distancia base de lectura. Para calcular la probabilidad de que una lectura esté dentro de un agrupamiento por azar y no por un acontecimiento provocado por una nucleasa de diseño, la muestra de CAST-Seq procedente de células editadas genéticamente se comparó con un banco de lecturas aleatorias creado de modo informático que contiene el mismo número de lecturas. La distribución de la distancia de lecturas consecutivas se muestra en una escala logarítmica. En este ejemplo, la línea umbral de 2500 pb describe un área <5 % en el banco aleatorio, lo que significa que la probabilidad de que una lectura esté dentro de un agrupamiento por azar es inferior al 5 % ( $p < 0,05$ ). El análisis CAST-Seq de células no tratadas se muestra como control. (c) Puntuación del alineamiento de la secuencia diana. Se comparó una región genómica de 500 pb que rodeaba estos sitios de translocación con 10 000 secuencias aleatorias de 500 pb. Cada sitio se alineó con la secuencia diana de la nucleasa de diseño. Si la puntuación de alineamiento de la secuencia diana del sitio era superior a la mejor puntuación del 5 % en las secuencias aleatorias, el acontecimiento se clasificaba como translocación derivada de la actividad fuera del objetivo (DI). (d) Tramos máximos de la región de la región de homología. Se buscó la subsecuencia homóloga común más larga entre la región diana y la región de translocación dentro de una ventana de 5 kb alrededor del sitio de translocación. Si la longitud de la subcadena homóloga era superior a 24 pb, el acontecimiento se clasificaba como translocación mediada por recombinación homóloga (RH). Todos los demás se clasificaron como translocación derivada de un sitio natural de ruptura (NBS).

Figura 8. Análisis CAST-Seq de la nucleasa CRISPR-Cas9 dirigida a *CCR5*<sup>#1</sup> con un algoritmo bioinformático más restrictivo.

(a) Esquema de la estrategia señuelo. Los cebadores "presa" y "cebo" se unen al conector y al sitio diana, respectivamente, para amplificar las aberraciones cromosómicas. Los cebadores señuelo se unen muy cerca del sitio diana, pero opuestos al cebador cebo, para evitar la formación de amplicones de longitud completa en los sitios diana no modificados (izquierda). (b) Análisis cualitativo CAST-Seq. Los gráficos del visor integrado del genoma ("Integrative Genomics Viewer", IGV) ilustran las lecturas de CAST-Seq que rodean a la región diana dentro de una ventana de 33 kb. Cada lectura de CAST-Seq cartografiadas está representada por una barra (solo se muestran las 7 primeras líneas). Las barras de color gris oscuro indican las secuencias que se alinean con la cadena negativa y las barras de color gris claro con la cadena positiva. La cobertura, es decir, el número de lecturas cartografiadas, se indica en el centro, y la ubicación de los genes en la parte inferior. Las posiciones del sitio diana y del agrupamiento de *CCR2* RH se resaltan con líneas discontinuas. (c) Alineación del sitio diana. El sitio diana de referencia *CCR5*<sup>#1</sup> se muestra en la parte superior (N, cualquier nucleótido; R, purina). Se resaltan los nucleótidos desapareados y las deleciones/inserciones (-1/1). El número de aciertos aparece a la izquierda, y las categorías a la derecha. (d) Análisis de indel. Se llevó a cabo una secuenciación profunda de amplicones dirigida en los sitios RH y/o DI identificados del ADN genómico recogido 4 días después de la edición génica con Cas9 o HiFi-Cas9. Las diferencias estadísticamente significativas se indican con "\*\*\*" ( $p < 0,05$ ; prueba Z corregida por la desviación estándar calculada en células no tratadas (NT)). (e) Representación gráfica de los reordenamientos complejos seleccionados descubiertos en el sitio diana. Por ejemplo, una combinación de secuencias derivadas de *CCR5* invertida (gris claro) y *CCR5* (gris oscuro) (arriba) o un tramo largo de una secuencia *CCR5* invertida/duplicada (gris oscuro, abajo). (f) Cartografiado de reordenamientos cromosómicos. Todos los sitios de aberración cromosómica pertinentes identificados por CAST-Seq se muestran en un ideograma cromosómico. Los sitios cartografiados se subdividen en el agrupamiento de sitio diana (ON), así como en reordenamientos cromosómicos mediados por corte en dianas inespecíficas (DI), por recombinación mediada por homología (RH) o por sitios naturales de ruptura (NBS). Las barras amarillas indican una clasificación ambigua (RH/DI). (g) Cuantificación. El número de reordenamientos cromosómicos

cuantificados mediante CAST-Seq o ddPCR se representa en el gráfico de dispersión. Se indican la línea de regresión lineal (línea de puntos) y el coeficiente de correlación al cuadrado ( $R^2$ ).

Figura 9. Análisis CAST-Seq de sitios genómicos diana de CRISPR-Cas9 o TALEN con un algoritmo bioinformático más restrictivo.

- 5 (a-d) Cartografiado de aberraciones cromosómicas. Ideogramas cromosómicos que presentan el análisis CAST-Seq de un par TALEN dirigido a *HBB* (a), así como CRISPR-Cas9 dirigida a *CCR5<sup>#2</sup>* (b), *FANCF* (c) y *VEGFA* (d). Se destacan todos los sitios de aberración cromosómica pertinentes identificados por CAST-Seq. (e-f) Comparación con GUIDE-Seq y CIRCLE-Seq. Los datos obtenidos del análisis CAST-Seq de nucleasas CRISPR-Cas9 dirigidas a *FANCF* (e) y *VEGFA* (f) se compararon con los datos publicados de GUIDE-Seq (PMC4320685) y CIRCLE-Seq (PMC5924695) y se visualizaron en diagramas de Venn.

Figura 10. Dinámica de las aberraciones cromosómicas.

- 15 (a-c) Visualización cualitativa. Los gráficos del visor integrado del genoma (IGV) muestran la región diana, *CCRS<sup>#1</sup>* (a), *I<sup>#2</sup>* (b) y *HBB* (c), dentro de una ventana de 33 kb. Solo se muestran las filas superiores. Las flechas blancas indican la orientación del cebo y las líneas verticales discontinuas el sitio diana. A la izquierda se indica el momento de recolección en días tras la electroporación (D1, D4, D14). (d-f) Análisis cuantitativo. Los gráficos muestran el número de lecturas de CAST-Seq agrupadas (aciertos) para las muestras D1 a D14 de CRISPR-Cas dirigidas a *CCR5<sup>#1</sup>* (d) y *CCR5<sup>#2</sup>* (e) o TALEN dirigidas a *HBB* (f). Se indica la categoría del agrupamiento (RH y/o DI).

Figura 11. Cinética de reparación del ADN y cuantificación de aberraciones cromosómicas.

- 20 (a) Estrategia de ddPCR. El "amplicón de borde" (-200 pb) abarca el sitio de escisión y está flanqueado por amplicones 5' o 3' a ambos lados del sitio diana. Se prevé que las translocaciones reduzcan la cantidad de productos de amplicones de borde, mientras que las grandes deleciones también reducirán la cantidad de amplicones flanqueantes. Se utilizaron amplicones situados en el lado telomérico (telo.) y en el brazo cromosómico opuesto (brazo q) en relación con el sitio diana, así como dos amplicones de control (cto.) en otro cromosoma, para establecer el cambio relativo de copias amplificables de la diana. (b-d) Variación del número de copias del sitio diana. Los gráficos muestran la variación relativa del número de copias ("copy number variation", CNV) de los sitios diana amplificables en células CD34+ editadas con CRISPR-Cas dirigido a *CCR5<sup>#1</sup>* (b) o *CCR5<sup>#2</sup>* (c), o con un TALEN dirigido a *HBB* (d), en diferentes puntos temporales (del día 1 al día 14) tras la transfección. (e-g) Resumen de datos. Los resultados de la ddPCR se utilizaron para normalizar (Norm.) las frecuencias de indel determinadas por el ensayo T7E1 para los puntos temporales D4. "Gran deleción" indica la disminución relativa del número promedio de amplicones flanqueantes, mientras que "otras aberraciones" se especifica como la diferencia relativa entre el número de amplicones de borde y el número promedio de amplicones flanqueantes.

Tabla 1: Secuencias diana

Se enumeran los sitios diana de nucleasas de diseño en *CCR5*, *VEGFA* y *FANCF*.

35 Tabla 2: Diseño de cebadores y conectores

Se enumeran los desoxiligonucleótidos utilizados para realizar CAST-Seq para evaluar aberraciones cromosómicas en células editadas con nucleasas CRISPR-Cas9 dirigidas a *CCR5*, *VEGFA* y *FANCF*. Las secuencias ilustradas en la tabla 2 pueden servir como ejemplo de cómo se pueden diseñar los cebadores apropiados también para otros sitios diana.

40 Tabla 3: Efecto de los cebadores señuelo

Para evaluar el impacto de los cebadores señuelo en la relación señal-ruido de CAST-Seq, se realizaron análisis paralelos en presencia o ausencia de cebadores señuelo. Los datos se basan en todas las lecturas en agrupamientos identificados por CAST-Seq realizadas en ADN genómico aislado de células madre y progenitoras hematopoyéticas CD34+ que se editaron con nucleasas CRISPR-Cas9 dirigidas al locus *VEGFA* o al locus *FANCF*.

45 Tabla 4: Análisis CAST-Seq para la nucleasa CRISPR-Cas9 dirigida a *CCR5*

- Se enumeran todos los sitios identificados por CAST-Seq (análisis completo, es decir, directo e inverso) en células madre y progenitoras hematopoyéticas CD34+ editadas con una nucleasa CRISPR-Cas9 dirigida al locus *CCR5* (sitio diana: 5'-GTGAGTAGAGCGGAGGCAGGAGG (SEQ ID NO:1, PAM subrayado). La tabla presenta la ubicación cromosómica de la aberración cromosómica, el número de lecturas desduplicadas (aciertos), el número de lecturas y la categoría asignada al acontecimiento de translocación.

Tabla 5: Sensibilidad de CAST-Seq

Se utilizó la PCR digital en nanogotas (ddPCR) para cuantificar el número de acontecimientos de delección grande que se producen entre los loci *CCR5* y *CCR2* en células no tratadas y en células madre hematopoyéticas editadas con nucleasas CRISPR-Cas9 dirigidas al locus *CCR5*. Quinientos ng de ADN genómico contienen aproximadamente 152 000 genomas haploides.

5 Tabla 6: Análisis CAST-Seq para la nucleasa CRISPR-Cas9 dirigida a *VEGFA*

Se enumeran todos los sitios pertinentes identificados por CAST-Seq (mostrados a modo de ejemplo para el análisis en sentido directo) en células madre y progenitoras hematopoyéticas CD34+ editadas con una nucleasa CRISPR-Cas9 dirigida al locus *VEGFA* (sitio diana: 5'-GGTGAGTGAGTGTGTGCGTGTGG (SEQ ID NO:3), PAM subrayado). La tabla presenta la ubicación cromosómica de la aberración cromosómica, el número de lecturas desduplicadas (aciertos), el número de lecturas y la categoría asignada al acontecimiento de translocación.

10

Tabla 7: Análisis CAST-Seq para la nucleasa CRISPR-Cas9 dirigida a *FANCF*

Se enumeran todos los sitios pertinentes identificados por CAST-Seq (mostrados a modo de ejemplo para el análisis en sentido directo) en células madre y progenitoras hematopoyéticas CD34+ nucleofectadas con una nucleasa CRISPR-Cas9 dirigida al locus *FANCF* (sitio diana: 5'-GGAATCCCTTCTGGCAGCACCTGG (SEQ ID NO:4), PAM subrayado). La tabla presenta la ubicación cromosómica de la aberración cromosómica, el número de lecturas desduplicadas (aciertos), el número de lecturas y la categoría asignada al acontecimiento de translocación.

15

Tabla 8: Programas informáticos utilizados para CAST-Seq

Se enumeran todos los programas informáticos utilizados para CAST-Seq. Las versiones indicadas estaban disponibles en la fecha de prioridad en las direcciones facilitadas.

20 Tabla 9: Paquetes R utilizados para CAST-Seq

A continuación, se muestra el paquete R utilizado para CAST-Seq. Las versiones indicadas estaban disponibles en la fecha de prioridad en las direcciones facilitadas.

Tabla 10: Matriz de puntuación

Matriz de puntuación de la sustitución de nucleótidos utilizada para el alineamiento de los sitios de translocación con la secuencia del sitio diana, incluidos los pesos para los desapareamientos y las protuberancias (inserciones/delecciones). Se utiliza el código IUPAC. A, adenina; C, citosina; G, guanina; T (o U), timina (o uracilo); R, A o G; Y, C o T; S, G o C; W, A o T; K, G o T; M, A o C; B, C o G o T; D, A o G o T; H, A o C o T; V, A o C o G; N, cualquier base.

25

Tabla 11: Diseño de cebadores para el sitio diana HBB (diana de TALEN)

30 Se muestra la secuencia diana, así como las secuencias pertinentes necesarias para la amplificación.

Tabla 12: Diseño de cebadores para el sitio diana 2 *CCR5*

Se muestran la secuencia diana pertinente y las secuencias de los cebadores.

Tabla 13: Diseño de cebadores para ddPCR

Para varios sitios diana se proporcionan las secuencias de cebadores directos e inversos.

35 **Realización preferida de la invención**

Los resultados de los experimentos utilizando el método de la presente invención mostrados en las figuras y las tablas pueden interpretarse como sigue:

La identificación de acontecimientos mutagénicos muy poco frecuentes inducidos por nucleasas de diseño, tales como mutagénesis en dianas inespecíficas, translocaciones, grandes delecciones o grandes inversiones, mediante secuenciación de alto rendimiento plantea diversos retos. Para que sea rentable, el método debe basarse en unos requisitos mínimos de secuenciación sin comprometer la sensibilidad. Para que tenga pertinencia clínica, el método debe ser aplicable a células derivadas de pacientes en lugar de realizarse en una línea celular sustituta con un trasfondo genético y epigenético diferente. Además, la prueba debe poder ejecutarse con un aporte mínimo de ADN genómico, de modo que pueda realizarse con material celular valioso derivado del paciente. Por último, los sesgos técnicos y de análisis, tales como los sesgos de amplificación de la PCR y los fallos en el proceso bioinformático, deben reducirse al mínimo para evitar resultados falsos positivos o falsos negativos.

40

45

CAST-Seq se desarrolló para cumplir estos requisitos e identificar acontecimientos muy poco frecuentes de aberración cromosómica con una sensibilidad sin precedentes. Para ello, CAST-Seq emplea una estrategia de PCR de 3 etapas que incluye el uso de cebadores internos y señuelo, respectivamente. En la figura 1 se muestra un esquema general

de CAST-Seq. Tras el aislamiento del ADN genómico a partir de células expuestas a nucleasas de diseño, el ADN genómico se fragmenta mediante ultrasonidos enfocados o digestión enzimática para producir fragmentos con un tamaño medio de 350 pb. Tras la reparación de los extremos y el ligamiento de los conectores a ambos extremos, se realiza una primera etapa de la PCR que incluye un cebador específico del sitio diana (cebador para diana, tabla 2), un cebador que se une al conector (cebador para conector, tabla 2) y uno o dos cebadores señuelo (tabla 2). Los cebadores señuelo están diseñados para unirse muy cerca sitio diana, pero en el sitio opuesto con respecto al cebador para diana. Se añaden a la reacción para evitar la generación de productos de amplificación de longitud completa a partir de plantillas derivadas de acontecimientos de no translocación (figura 1a derecha, figura 3). Los cebadores señuelo no pueden unirse a las plantillas derivadas de acontecimientos de translocación (u otras aberraciones cromosómicas) (figura 1a izquierda) y, por lo tanto, no impiden su amplificación. Para la segunda etapa de la PCR se utilizan dos cebadores internos (cebador interno para diana y cebador interno para conector, tabla 2) que contienen adaptadores para la tercera PCR. Los productos derivados del cebador señuelo (figura 1a derecha) no se amplificarán en esta etapa. Por último, la tercera PCR se utiliza para añadir los adaptadores de Illumina y los códigos de barras para NGS.

El proceso bioinformático para identificar y anotar acontecimientos de aberración cromosómica se muestra esquemáticamente en la figura 1b y se describe en detalle en el ejemplo 2. CAST-Seq se diseñó no solo para detectar acontecimientos de translocación, sino también otras aberraciones cromosómicas, incluidas grandes deleciones e inversiones de secuencia, de forma semicuantitativa. Es probable que los acontecimientos anotados a una región cromosómica específica se deriven de un único modo de acción, directa o indirectamente relacionado con la actividad en la diana o en dianas inespecíficas de la nucleasa de diseño. Tales acontecimientos se definen como agrupamientos si se produjeron al menos 2 lecturas desduplicadas dentro de una distancia de 2500 pb. Para calcular la probabilidad de que una lectura esté dentro de un agrupamiento por azar y no por un determinado modo de acción, la muestra de CAST-Seq analizada se comparó con un banco de lecturas aleatorias creado de modo informático que contiene el mismo número de lecturas (figura 2). La distribución de la distancia de lecturas consecutivas se muestra a modo de ejemplo para un análisis CAST-Seq realizado en células madre hematopoyéticas editadas con nucleasas CRISPR-Cas9 dirigidas a *CCR5*, en comparación con una muestra sin tratar y el banco de control aleatorio (figura 2a). En este ejemplo, la línea de umbral de 2500 pb describe un área <5 % en el banco aleatorio, lo que significa que la probabilidad de que una lectura esté dentro de un agrupamiento por azar es inferior al 5 %.

Al evaluar los resultados de CAST-Seq, los inventores se dieron cuenta de que no todos los acontecimientos de aberración cromosómica identificados podían vincularse directamente a la actividad en la diana de la nucleasa de diseño. Una vía de reparación del ADN activa en la mayoría de las células es la reparación dirigida por homología. La reparación dirigida por homología se basa en la recombinación homóloga (RH) entre regiones homólogas de ADN para reparar el sitio dañado. A menudo, la cromátida hermana se utiliza en este proceso para una reparación perfecta, pero otras secuencias homólogas también pueden ser reclutadas. Por lo tanto, un acontecimiento de corte en la diana puede desencadenar una translocación intercromosómica o intracromosómica (incluidas grandes deleciones e inversiones de secuencia) con regiones homólogas en el genoma. Por último, una ruptura bicatenaria del ADN inducida por una nucleasa de diseño en el sitio diana también puede desencadenar la recombinación con rupturas naturales del ADN en el genoma. Los acontecimientos identificados se clasificaron como translocaciones mediadas por dianas inespecíficas (DI), recombinaciones mediadas por homología (RH) o translocaciones mediadas por un sitio común de ruptura (CBS), en función de si se podía identificar un sitio diana inespecífico y de la extensión de las secuencias homólogas que aparecen en el sitio de translocación. La bioinformática subyacente se describe en el ejemplo 3. En resumen, la secuencia diana de la nucleasa de diseño se alinea con las regiones de agrupamientos de CAST-Seq y se selecciona la mejor puntuación de alineamiento. Un banco aleatorio informático se examina de la misma manera para comprobar si la puntuación del alineamiento supera el umbral del 5 % (figura 2b). Todos los agrupamientos que superaron este umbral se clasificaron como translocación desencadenada por DI de la nucleasa de diseño. Para todos los agrupamientos que no cumplían este requisito, se examinó una ventana de 5 kb alrededor del acontecimiento de translocación en busca de secuencias homólogas con el sitio diana y se comparó con el banco aleatorio para definir el umbral de longitud común. Un acontecimiento de translocación se clasificó como RH si la longitud de las regiones flanqueantes era superior al 5 % de subcadenas más largas en las secuencias aleatorias (figura 2c). Todos los demás agrupamientos se marcaron como translocación desencadenada por CBS.

La estrategia de señuelo de CAST-Seq fue diseñada para reducir las lecturas de fondo derivadas de sitios diana no modificados y puede aplicarse con facilidad para cada sitio diana elegido. Su eficacia se muestra a modo de ejemplo para el sitio diana *CCR5* (figura 3a). Se utilizaron dos cebadores señuelo en orientación directa (F) e inversa (R). Si los cebadores señuelo se modificaban por fosforilación 3' para bloquear la extensión 3', eran capaces de reducir, pero no bloquear completamente la formación de los amplicones de longitud completa de 412 pb (figura 3b). Sin embargo, cuando se añadieron a la reacción cebadores señuelo no modificados, se impidió eficazmente la generación del amplicón de longitud completa (figura 3c). En su lugar, se generaron dos productos con los tamaños esperados de 264 pb y 140 pb. Por último, incluso la presencia de un único cebador señuelo fue suficiente para suprimir con eficacia la generación de amplicones de longitud completa (figura 3d), lo que demuestra la eficacia de la estrategia señuelo. Para calcular la eficacia global de esta estrategia, se realizaron análisis CAST-Seq paralelos para nucleasas CRISPR-Cas dirigidas a *VEGFA* o *FANCF* en presencia o ausencia de cebadores señuelo. Al considerar todas las secuencias translocadas, con la excepción de las lecturas derivadas del agrupamiento de la diana, se observó un aumento de aproximadamente 5 veces en la relación señal-ruido (tabla 3).

Para ilustrar el potencial de este nuevo método, se realizó CAST-Seq en ADN genómico aislado de células madre hematopoyéticas que fueron editadas con CRISPR-Cas9 dirigida a *CCR5* (tabla 4, figura 4). Todos los acontecimientos de translocación identificados, estratificados en DI, RH y CBS, se muestran en un gráfico cromosómico (figura 4a). Cabe destacar que la mayoría de los sitios DI contienen más de 5 desapareamientos en combinación con protuberancias (figura 4b).

Un examen más detallado de los primeros 30 agrupamientos alineados permitió realizar análisis adicionales, tales como la ubicación de desapareamientos y protuberancias, así como la caracterización de los PAM identificados en sitios diana inespecíficos (figura 5a). Cabe destacar que, mientras que todos los PAM de sitios diana inespecíficos tenían una G en la posición 3, dos sitios DI no contenían una purina (R: A o G) en la posición 2. Además, 2 de 30 centros toleraron una protuberancia de -1 en la posición 1. Como se indicó anteriormente, la tolerancia a aceptar desapareamientos y protuberancias es más pronunciada en la región distal a PAM del sitio diana. Un análisis del logotipo (figura 5b) muestra la secuencia consenso de los sitios diana, confirmando una menor promiscuidad en la región proximal a PAM. Esta observación fue corroborada por un análisis cuantitativo realizado en subgrupos de sitios diana (figura 5c). Por lo tanto, el alineamiento de la secuencia diana mediada por CAST-Seq duplica los hallazgos previamente notificados de que las regiones más conservadas en los sitios diana de CRISPR-Cas9 son los nucleótidos en el PAM y en la región proximal a PAM. Por otra parte, CAST-Seq identificó nuevas características relativas a la tolerancia a los desapareamientos y las protuberancias en la secuencia PAM.

Tal como se mencionó anteriormente, las aberraciones cromosómicas también pueden ser desencadenadas por la actividad en la diana de la nucleasa de diseño. El locus *CCR2* está situado junto al locus diana *CCR5* y comparte una elevada homología de secuencia con *CCR5*. Una inspección más detallada de esta región reveló un gran número de aberraciones cromosómicas alrededor del sitio de corte de la diana *CCR5* (figura 6a). Cabe destacar un punto caliente en el gen *CCR2* que fue anotado por CAST-Seq como una gran delección inducida por RH (figura 6b). El genotipado mediante el ensayo T7E1 confirmó que este sitio no fue cortado por CRISPR-Cas9, lo que confirma que la actividad en la diana puede desencadenar aberraciones cromosómicas al potenciar la recombinación entre secuencias homólogas.

Dado que CAST-Seq tiene una orientación de secuenciación forzada, es posible definir la orientación de los acontecimientos de translocación. Por lo tanto, CAST-Seq permitió a los inventores identificar aberraciones cromosómicas adicionales desencadenadas por la actividad en la diana, tales como translocaciones intercromosómicas que condujeron a la formación de cromosomas acéntricos y dicéntricos (figura 6c). En resumen, CAST-Seq puede identificar diversas aberraciones cromosómicas, incluidos acontecimientos no descritos previamente por otros métodos.

Esto comprende aberraciones cromosómicas desencadenadas por la actividad en la diana, tales como translocaciones recíprocas a cromosomas homólogos que conducen a la formación de cromosomas acéntricos/dicéntricos, grandes inserciones/delecciones e inversiones.

Para evaluar la sensibilidad de CAST-Seq, los inventores evaluaron los acontecimientos poco frecuentes de recombinación entre *CCR5* y *CCR2* en ADN genómico editado por nucleasas CRISPR-Cas9 dirigidas a *CCR5* o en el control no tratado. Utilizando 500 ng de ADN genómico, el acontecimiento de recombinación se detectó en ambas muestras con 60 o 63 011 lecturas, respectivamente (tabla 5). Para determinar el número de copias de los loci *CCR2-CCR5* fusionados en esas muestras, se realizó una cuantificación con PCR digital en nanogotas (ddPCR) colocando un cebador en *CCR5* y el otro en *CCR2*. El ensayo produjo 9,8 copias en la muestra no tratada y 1280 copias en la muestra editada con CRISPR-Cas9. Esto significa que CAST-Seq fue capaz de detectar 9,8 acontecimientos de aberración cromosómica en aproximadamente 152 000 genomas haploides, lo que corresponde a una sensibilidad del -0,006 %. Dado que la sensibilidad de CAST-Seq es directamente proporcional al número de células analizadas, pueden alcanzarse sensibilidades aún mayores (es decir, <1:10 000) aumentando el número de células utilizadas para aislar el ADN genómico y, en consecuencia, poder procesar más ADN genómico mediante CAST-Seq.

Para verificar que CAST-Seq puede aplicarse a otros loci diana, también se editaron células madre hematopoyéticas con nucleasas CRISPR-Cas9 dirigidas a los loci *VEGFA* y *FANCF*. Los resultados de estos experimentos se muestran en la tabla 6 y la tabla 7, respectivamente.

Ejemplo 1: Preparación de bancos de CAST-seq.

El ADN genómico de células hematopoyéticas no tratadas y editadas genómicamente se extrae con kit QIAmp DNA Blood Mini Kit (Qiagen) y posteriormente se fragmenta por sonicación (Covaris) o por una reacción enzimática (kit NEBNext® Ultra™ II FS DNA Library Prep Kit, NEB) para obtener fragmentos de ADN con un tamaño promedio de aproximadamente 350 pb (véase la figura 1a). A continuación, el ADN genómico fragmentado se repara en sus extremos y se le aplica una cola de A (kit NEBNext® Ultra II End Repair/dA-Tailing Module, NEB). El ADN conector, generado por hibridación de dos desoxioligonucleótidos asimétricos con una proyección T-3' en la cadena positiva o un grupo fosfo-5' en combinación con un grupo amino-C7-3' en la cadena negativa, se liga al ADN cizallado y posteriormente se purifica con el kit QIAquick PCR Purification Kit (Qiagen). En la primera ronda de PCR, se mezclan 500 ng de ADN con oligos señuelo, así como con un cebador específico de conector y otro específico de sitio diana. La PCR se realiza utilizando las siguientes condiciones: 20 ciclos de 95 °C durante 15 s, 63 °C durante 20 s y 72 °C

durante 20 s. En la segunda ronda de PCR, se realiza una PCR con cebadores internos con un cebador específico de conector y un cebador específico de sitio diana con secuencias adaptadoras de Illumina en sus extremos 5' en las siguientes condiciones: 20 ciclos de 95 °C durante 15 s, 68 °C durante 20 s, 72 °C durante 20 s. Tras la amplificación, la reacción de PCR se purifica con el kit QIAquick PCR Purification Kit (Qiagen) y se cuantifica con NanoDrop (ThermoFisher). Para ambas rondas se utilizó la polimerasa Hot-start Q5 (Q5® High-Fidelity DNA Polymerase, NEB) en reacciones de 50 µl. Los adaptadores con código de barras de Illumina se añaden mediante una reacción PCR de 7 ciclos utilizando oligos NEBNext® Multiplex Oligos for Illumina (NEB) según el protocolo convencional. A continuación, los amplicones (aproximadamente 0,1-1 µg) se purifican con esferas magnéticas AMPure XP (Beckman Coulter) diluidas hasta 4 nM y se cuantifican mediante ddPCR utilizando el "ddPCR Library Quantification Kit for Illumina TruSeq" (BIORAD) siguiendo las instrucciones del fabricante. Por último, se carga una concentración final de 8-10 pM de ADN desnaturalizado en 600 µl en un cartucho MiSeq Reagent Kit v2 (Illumina) y se secuencian.

#### Ejemplo 2: Análisis bioinformático.

Las lecturas por extremos apareados de la secuenciación de alto rendimiento de Illumina se fusionaron utilizando FLASH (v1.2.11) (<https://ccb.jhu.edu/software/FLASH/>) con 10 y 250 como solapamiento mínimo y máximo, respectivamente. A continuación, se utilizó BBmap (versión 38.22) (<https://sourceforge.net/projects/bbmap/>) para aplicar primero una selección positiva de las lecturas que contienen el sitio diana de la nucleasa de diseño con el fin de eliminar los productos de PCR mal cebados. Después, las lecturas se recortaron para eliminar las secuencias conectoras, las secuencias adaptadoras de Illumina y la secuencia de alargamiento dirigido (figura 1b). Al final del proceso de filtrado se comprobó la inversión de secuencias diana cortas mediante BBmap. Los parámetros se establecieron de la siguiente manera: la longitud kmero para encontrar contaminantes se definió en función del sitio diana de la nucleasa, el conector o la longitud del cebador. Se permitieron hasta 2 desapareamientos y/o protuberancias en los kmeros de referencia. Los parámetros no mencionados en el presente documento se establecieron por defecto. A continuación, las lecturas seleccionadas se alinearon con el genoma humano de referencia (GRCh38/hg38) con Bowtie2 (versión 2.3.4.2) (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) utilizando el preajuste muy sensible de parámetros para garantizar el alineamiento más fiable. Se eliminaron los alineamientos de baja calidad para reducir las posibilidades de identificar sitios de translocación falsos positivos (MAPQ < 15). Los programas informáticos utilizados para el procesamiento FASTQ se detallan en la tabla 8. Los análisis posteriores se realizaron con scripts R internos. La lista de paquetes R dedicados a la anotación genómica puede consultarse en la tabla 9. A continuación, las lecturas alineadas se desduplicaron según el número de cromosoma y los valores de inicio, fin y hebra. Las lecturas que se encuentran dentro de 2 pb para el inicio y el fin, respectivamente, se consideraron lecturas "duplicadas" y, por lo tanto, se desduplicaron. Posteriormente, se determinó el punto de translocación y se utilizó para el análisis de agrupamientos mediante la comparación de la distribución de distancias de lecturas consecutivas con un banco generado de modo informático que contiene un número equivalente de sitios aleatorios. Las tasas de error de agrupamiento se cuantificaron con la curva de distribución de distancias de un banco aleatorio (figura 2). Se aplicó una distancia umbral de 2500 pb para alcanzar un valor de  $p$  significativo ( $p < 0,05$ ) en todas las muestras de los inventores analizadas. A continuación, se compararon los agrupamientos pertinentes derivados de la muestra tratada con las de la muestra no tratada para restar el fondo. Se realizó una prueba hipergeométrica para evaluar la importancia de la diferencia entre las muestras editadas genéticamente y las muestras de control no tratadas mediante la comparación del número de lecturas con el tamaño de la población (es decir, el número total de lecturas en los archivos FASTQ sin procesar). Los conglomerados significativos se seleccionaron según el valor de  $p$  ajustado ( $p < 0,05$ ). Por último, se rechazaron los agrupamientos que contenían solo 1 lectura desduplicada.

#### Ejemplo 3: Clasificación de los acontecimientos de translocación.

Los sitios de translocación se dividieron en tres categorías: translocaciones derivadas de la actividad en dianas inespecíficas (DI), translocaciones mediadas por recombinación homóloga (RH) y translocaciones derivadas del sitio común de ruptura (CBS) (figura 2). Para asignar cada sitio a una de estas categorías, los inventores compararon una región genómica de 500 pb que rodeaba los sitios de translocación con 10 000 secuencias aleatorias de 500 pb. A continuación, cada uno de los sitios derivados de secuencias reales o aleatorias se alineó con la secuencia diana de la nucleasa de diseño, con una puntuación entre 1 para el apareamiento y -1 para el desapareamiento, la apertura de huecos y la extensión de huecos (tabla 10), y se seleccionó el mejor apareamiento de las secuencias complementarias directa e inversa. A continuación, se buscó la subcadena común más larga entre cada secuencia, incluidas las regiones flanqueantes izquierda y derecha, en las secuencias complementarias directa e inversa dentro de una ventana de 5 kb. Por último, cada acontecimiento se clasificó de la siguiente manera: DI, si la puntuación de alineamiento de la secuencia diana de la secuencia era superior al 5 % de la mejor puntuación en las secuencias aleatorias. RH, si el acontecimiento no era DI, pero la longitud de las regiones flanqueantes era superior al 5 % de la subcadena más larga en las secuencias aleatorias. Todos los demás acontecimientos que no cumplían estos criterios se clasificaron como ECB.

#### Ejemplo 4: Análisis bioinformático más restrictivo (figura 7).

*Alineamiento:* Las lecturas apareadas de la secuenciación miSeq de Illumina se fusionaron utilizando el programa informático FLASH (Bioinformatics, 27 (2011), 2957-2963). Se utilizó BBmap (<https://sourceforge.net/projects/bbmap/>)

para filtrar y recortar de la siguiente manera: las lecturas fusionadas que contenían el sitio diana de la nucleasa de diseño se filtraron y se mantuvieron, mientras que las lecturas de productos de PCR mal cebados se filtraron y se desecharon. Se recortaron las secuencias conectoras, las secuencias adaptadoras de Illumina, la secuencia de alargamiento dirigido y las lecturas de mala calidad. Las lecturas seleccionadas se alinearon con el genoma humano GRCh38 (hg38) mediante Bowtie2 (Nat. Methods, 9 (2012), 357-359) y los parámetros preestablecidos muy sensibles para maximizar la precisión del alineamiento. Para reducir la probabilidad de encontrar falsos positivos, se seleccionaron lecturas alineadas con buena calidad de cartografiado (MAPQ > 15). El archivo BAM alineado se convirtió en archivo bed utilizando BEDTools (Bioinformatics. 26 (2010), 841-842).

*Desduplicación/definición del agrupamiento:* Las lecturas situadas en las mismas coordenadas se consideraron duplicados derivados de la PCR y, por lo tanto, se desduplicaron. Para hacer frente a los sesgos de secuenciación/alineamiento del punto de translocación o del ligamiento del conector, se añadió una tolerancia de +/-3 pb. Por lo tanto, todas las lecturas dentro de esta ventana de +/-3 pb se desduplicaron y la cantidad total de lecturas se almacenó para cuantificar el acontecimiento de translocación. Las regiones de alta densidad de lecturas se determinaron utilizando un conjunto aleatorio de regiones del genoma humano para calcular la distribución de distancias entre dos lecturas consecutivas. Una distancia umbral de 2500 pb alcanzó un valor de  $p$  significativo ( $p < 0,05$ ) en todas las muestras analizadas. Posteriormente, las lecturas consecutivas separadas por menos de 2500 pb se fusionaron en agrupamientos, que representaban todos los sitios de translocación putativos. Al comparar más de una réplica de una muestra, durante el proceso bioinformático se fusionaron dos agrupamientos proximales (CCR5/CCR2 y HBB/HBD), y los agrupamientos individuales se recuperaron manualmente reajustando los bordes. Por último, se evaluó la importancia de los agrupamientos identificados en comparación con una muestra de control no tratada mediante una prueba exacta de Fisher. El umbral de significación se fijó para un valor de  $p$  ajustado (Benjamini-Hochberg) inferior a 0,05.

*Clasificación del acontecimiento de translocación:* Los sitios de translocación se clasificaron en tres grupos: translocaciones mediadas por dianas inespecíficas (DI) y mediadas por recombinación homóloga (RH), y translocaciones derivadas de sitios naturales de ruptura (NBS). Para evaluar la significación estadística de los grupos, se eligió al azar un conjunto de 10 000 secuencias del genoma humano de 500 pb de longitud. Para DI, los sitios de translocación se alinearon con la secuencia diana. Se construyó una matriz de sustitución de nucleótidos utilizando +1 y -1 como pesos para el apareamiento y el desapareamiento, respectivamente. Se permitieron los huecos con el mismo peso de penalización que el desapareamiento. Se utilizó un alineamiento por pares del paquete Biostrings R (<https://rdrr.io/bioc/Biostrings/>) con un tipo de alineamiento "local-global". Se calcularon las puntuaciones de alineamiento DI para los sitios de translocación identificados y las secuencias aleatorias. Para RH, se eligió la subcadena común más larga ("longest common substring", LCS) entre las regiones flanqueantes izquierda y derecha, definiendo una ventana circundante de 5 kb alrededor del sitio de translocación, y la ventana conocida de 5 kb alrededor de la diana esperada. Se utilizaron secuencias aleatorias para calcular la longitud de LCS entre las regiones diana y aleatorias. Por último, cada sitio de translocación se clasificó de la siguiente manera: DI si la puntuación de alineamiento DI era superior a las mejores puntuaciones del 5 % en secuencias aleatorias; RH si LCS era más largo que las mejores puntuaciones del 5 % de LCS en secuencias aleatorias; NBS en caso contrario.

*Anotación.* Los sitios de translocación seleccionados se anotaron con el gen o región génica más cercana (por ejemplo, promotor, exón, intrón, etc.), basándose en la distancia al sitio de inicio de la transcripción ("transcriptional start site", TSS) indicado en el Bioconductor Annotation Package TxDb.Hsapiens.UCSC.hg38.knownGene (<http://bioconductor.org/packages/TxDb.Hsapiens.UCSC.hg38.knownGene/>). Se notifica el conjunto completo de genes que se encuentra dentro de una ventana de 100 kb alrededor del sitio de translocación, destacando específicamente los genes relacionados con el cáncer basados en la base de datos OncoKB (JCO Precis Oncol., 2017, 1-16).

Ejemplo 5: Análisis moleculares.

Para el análisis mediante PCR digital en nanogotas (ddPCR), se digirieron 150-550 ng de ADN genómico con 5 U de HindIII HF o AvrII (NEB) a 37 °C durante 30 min para reducir la viscosidad de la muestra. Después de la digestión, se añadieron 100 ng (translocación) o 20 ng (delección grande) de ADN genómico digerido a la mezcla de reacción de ddPCR que contenía QX200TM EvaGreen ddPCR Supermix TM (Bio-Rad, n.º de catálogo 1864034). Cada reacción se complejó con 100 nM de cebadores y se cargó en el generador de nanogotas QX200 (Bio-Rad). Las nanogotas generadas se transfirieron a una placa de PCR de 96 pocillos (Bio-Rad, n.º de catálogo 12001925) y la placa se selló con un sellador de placas PCR PX1 (Bio-Rad). Para todos los ensayos, la PCR final se realizó del siguiente modo: precalentamiento de la tapa a 95 °C durante 5 min, 50 ciclos de 95 °C durante 30 s, 62 °C durante 60 s, 72 °C durante 2 min, seguido de 5 min a 4 °C y 5 min a 90 °C (tasa de incremento ajustada a 2 °C/s). Tras la PCR, los datos se adquirieron en un lector de nanogotas QX200 Droplet Reader y los resultados se analizaron con QuantaSoftTM Analysis Pro (Bio-Rad). Los resultados se consideraron significativos si se generaban al menos 10 000 nanogotas/20 µl de reacción. Para calcular las frecuencias de "grandes deleciones" y "otras aberraciones" en las muestras editadas, el valor promedio de ddPCR de un duplicado técnico se normalizó primero con respecto a la muestra de control apareada no tratada para minimizar la variación entre ensayos, y después se normalizó con respecto a la cantidad de ADN genómico de entrada dividiendo el número por el promedio de los dos valores obtenidos para los genes de control (RAD1, STAT3). El valor promedio de los ensayos 5' y 3' se utilizó para determinar la fracción de grandes deleciones.

La fracción de translocaciones se calculó restando la fracción de grandes deleciones del valor "Edge". El porcentaje de indel del ensayo T7E1 se recalculó basándose en la fórmula:  $(100 - (\text{gran deleción} \times 100) - (\text{translocación} \times 100)) \times \%$  de indel.

Tabla 1: Secuencias diana

<b>Seq ID</b>	<b>Secuencia (5'→3')</b>	<b>Comentario</b>
1	GTGAGTAGAGCGGAGGCAGGAGG	CCR5 sitio diana (PAM subrayado)
2	GTGAGTAGAGCGGAGGCAGGNRG	CCR5 sitio diana consenso (PAM subrayado)
3	GGTGAGTGAGTGTGTGCGTGTGG	VEGFA sitio diana (PAM subrayado)
4	GGAATCCCTTCTGCAGCACCTGG	FANCF sitio diana (PAM subrayado)

5

Tabla 2: Diseño de cebadores y conectores

<b>Objetivo/diana</b>	<b>Función</b>		<b>n.º ID</b>	<b>Secuencia 5'→3'</b>	<b>SEQ ID NO:</b>
Conectores	Ligamiento del conector	cadena positiva	4038	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACT	5
		cadena negativa	4039	P-GTCCCTTAAGCGGAGC-NH3	6
Primer conector	Primera PCR	conector	4032	GTAATACGACTCACTATAGGGC	7
	Segunda PCR	conector, interno	4033	ACACTCTACACTCTTTCCCTACAGGACGCTCTTCCGATCTAGGGCTCCGCTTAAGGGAC	8
Cebador inverso de CCR5 (lado centromérico)	15' PCR	Diana	4034	AGGTAGATGTCAGTCATGCT	9
		señuelo directo	4036	ATCAATGTGAAGCAAATCGCA	10
		señuelo inverso	4037	AGGGCTCCGATGTATAATAATTG	11
	Segunda PCR	Diana, interno	4035	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCTTCAGCCTTTGCAGTTTATCAG	12
Cebador directo de CCR5 (lado telomérico)	Primera PCR	Diana	4272	GGATTATCAAGTGCAAGTCC	13
		señuelo directo	3779	CTGGTCATCCTCATCCTG	14
		señuelo inverso	4261	AAAACCAAAGATGAACACCAGT	15
	Segunda PCR	Diana, interno	4262	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTATAACATCGGAGCCCTGCCA	16
Cebador VEGFA	Primera PCR	Diana	4382	GAGAGGGACACAGATC	17
		señuelo directo	4380	CGTCTTCGAGTGAGGAC	18
		señuelo inverso	4381	CTGCTCGCTCCATTAC	19

Objetivo/diana	Función		n.º ID	Secuencia 5'→3'	SEQ ID NO:
	Segunda PCR	Diana, interno	4383	GACTGGAGTTCAGACGTGTCTCTTCCGATCTACACAGATCTGG AATCCTGGAGTG	20
Cebador <i>FANCF</i>	Primera PCR	Diana	4362	GTTCCAATCAGTACGCAG	21
		señuelo directo	4360	CTTGAGACCGCCAGAAG	22
		señuelo inverso	4361	CACTACCTACGTCAGCAC	23
	Segunda PCR	Diana, interno	4363	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCCGTCTCCAA GGTGAAAGC	24

Tabla 3: Efecto de los cebadores señuelo.

Objetivo	LECTURAS EN DIANAS INESPECÍFICAS (CAMBIO EN NÚMERO DE VECES PROMEDIO)	STDEV
<i>VEGFA</i>	5,1	±0,4
<i>FANCF</i>	5,0	±0,2

5 Para evaluar el impacto de los cebadores señuelo en la relación señal-ruido, se realizaron análisis CAST-Seq paralelos en presencia o ausencia de cebadores señuelo. El cambio en número de veces se calculó mediante la fórmula siguiente:

$$\frac{(\text{total de lecturas en agrupamientos con señuelo} - \text{lecturas en el agrupamiento de la diana con señuelo}) / \text{total de lecturas con señuelo}}{(\text{total de lecturas en agrupamientos sin señuelo} - \text{lecturas en el agrupamiento de la diana sin señuelo}) / \text{total de lecturas sin señuelo}}$$

Tabla 4: Análisis CAST-Seq para la nucleasa CRISPR-Cas9 dirigida a *CCR5*

Cromosoma	Inicio	Fin	Aciertos	Lecturas	Categoría
chr3	46367384	46381002	13684	2564351	DIANA
chr13	24886065	24888172	69	5930	DIANA INESPECÍFICA
chr19	35352634	35353338	17	7538	DIANA INESPECÍFICA
chr11	133747455	133747970	8	6045	DIANA INESPECÍFICA
chr10	11712420	11712929	8	518	DIANA INESPECÍFICA
chr3	33147914	33148416	7	211	DIANA INESPECÍFICA
chr22	29073798	29074309	7	208	DIANA INESPECÍFICA
chr2	27731141	27731654	6	2916	DIANA INESPECÍFICA
chr11	396300	396818	6	849	DIANA INESPECÍFICA
chr3	192973003	192973517	5	69	DIANA INESPECÍFICA

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chr12	52697472	52697979	4	417	DIANA INESPECÍFICA
chr6	1490592	1491094	4	302	DIANA INESPECÍFICA
chr10	91608416	91608926	4	138	DIANA INESPECÍFICA
chr16	9384008	9384516	4	47	DIANA INESPECÍFICA
chr2	7907009	7907511	4	13	DIANA INESPECÍFICA
chr2	203870062	203870562	3	3621	DIANA INESPECÍFICA
chr16	3054688	3055207	3	66	DIANA INESPECÍFICA
chr3	46356995	46358671	694	63011	RH
chr3	46360849	46364874	19	4431	RH
chr3	37046851	37047366	12	217	RH
chr1	235196336	235196844	6	364	RH
chr1	220970449	220970959	5	90	RH
chr2	108473896	108474401	5	13	RH
chr19	11128451	11128957	4	155	RH
chr3	46351798	46352301	4	92	RH
chr1	233407233	233407736	4	53	RH
chr13	20707652	20708155	2	396	RH
chr11	87001121	87001628	2	320	RH
chr2	70200237	70200737	2	16	RH
chr4	126697877	126698400	22	1344	CBS
chr1	15404594	15405115	13	2861	CBS
chr3	45888052	45888571	13	418	CBS
chr7	148990482	148991011	13	168	CBS
chr7	5397004	5397524	12	269	CBS
chr8	105275605	105276115	11	890	CBS
chr19	44844225	44844734	11	618	CBS
chr7	36875934	36876451	11	193	CBS

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chr3	173891336	173891854	10	502	CBS
chr4	56315538	56316061	10	79	CBS
chr7	45528063	45528572	10	32	CBS
chr18	31539684	31540191	9	3630	CBS
chr2	47793539	47794039	9	3174	CBS
chr1	1535558	1536071	9	1052	CBS
chr13	58455707	58456225	9	461	CBS
chr2	172156140	172156661	9	109	CBS
chr2	12715962	12716463	8	1461	CBS
chr10	132462758	132463271	8	927	CBS
chr16	85033587	85034097	8	514	CBS
chr11	55807139	55807651	8	468	CBS
chr2	183611509	183612020	8	368	CBS
chr18	50028215	50028720	8	339	CBS
chrX	96709289	96709797	8	286	CBS
chr1	168791914	168792420	8	109	CBS
chr3	64787649	64788160	8	48	CBS
chr5	126888190	126888709	8	30	CBS
chr15	27783853	27784356	7	1023	CBS
chr19	32703187	32703693	7	610	CBS
chr11	69750753	69751263	7	580	CBS
chr9	131914502	131915008	7	345	CBS
chr2	65853977	65854487	7	332	CBS
chr6	27247206	27247715	7	280	CBS
chr15	33693648	33694159	7	274	CBS
chr4	173907279	173907789	7	234	CBS
chr7	7769275	7769801	7	120	CBS

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chr19	2121817	2122322	7	117	CBS
chr8	94762064	94762573	6	2427	CBS
chr16	59245045	59245555	6	496	CBS
chr2	3579370	3579875	6	341	CBS
chr1	186338264	186338773	6	308	CBS
chr6	40451542	40452052	6	245	CBS
chr7	131308704	131309211	6	220	CBS
chr11	49963855	49964372	6	208	CBS
chr11	31515788	31516297	6	186	CBS
chr6	143628747	143629252	6	184	CBS
chr7	91394256	91394759	6	177	CBS
chr9	102600198	102600701	6	140	CBS
chr5	177371842	177372363	6	109	CBS
chr6	130485685	130486199	6	101	CBS
chr17	79978041	79978546	6	85	CBS
chr6	143636867	143637377	6	71	CBS
chrX	72272282	72272791	6	69	CBS
chr13	114101805	114102315	5	820	CBS
chr4	76774790	76775295	5	809	CBS
chr3	46411418	46411925	5	766	CBS
chr10	55973993	55974499	5	745	CBS
chr9	38254525	38255028	5	488	CBS
chr9	18557072	18557577	5	410	CBS
chr3	46445017	46445531	5	380	CBS
chr18	63835995	63836497	5	339	CBS
chr5	127232193	127232697	5	317	CBS
chr1	112168534	112169040	5	294	CBS

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chr20	32614709	32615211	5	292	CBS
chr2	80914457	80914959	5	284	CBS
chr10	2406025	2406532	5	272	CBS
chr10	28696454	28696961	5	272	CBS
chr14	88256869	88257375	5	251	CBS
chr19	19505880	19506387	5	240	CBS
chr9	70388652	70389158	5	235	CBS
chr9	33409401	33409908	5	230	CBS
chr17	38415258	38415770	5	228	CBS
chr1	40698677	40699183	5	219	CBS
chr8	136204368	136204872	5	211	CBS
chr1	216502468	216502974	5	209	CBS
chr7	26267708	26268214	5	202	CBS
chr17	49401535	49402043	5	193	CBS
chr12	12787084	12787589	5	171	CBS
chr4	1693018	1693534	5	162	CBS
chr8	38858536	38859046	5	150	CBS
chr4	2620355	2620855	5	127	CBS
chr4	109557413	109557919	5	126	CBS
chr19	28522506	28523013	5	119	CBS
chr11	86139833	86140336	5	107	CBS
chr9	123514549	123515054	5	103	CBS
chr10	86576951	86577457	5	78	CBS
chrX	29244651	29245157	5	63	CBS
chr11	36240842	36241345	5	49	CBS
chr15	96641850	96642361	5	39	CBS
chr9	89213638	89214142	5	38	CBS

## ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chr1	200635695	200636199	4	503	CBS
chr14	38655983	38656486	4	451	CBS
chr15	89371851	89372356	4	358	CBS
chr4	2869887	2870392	4	354	CBS
chr6	137829231	137829737	4	342	CBS
chr11	77252689	77253194	4	326	CBS
chr15	51628367	51628874	4	248	CBS
chr10	53066645	53067158	4	224	CBS
chr3	15452244	15452751	4	213	CBS
chr3	65002198	65002704	4	138	CBS
chr1	60997000	60997510	4	130	CBS
chr12	54236355	54237067	4	117	CBS
chr2	173548865	173549369	4	113	CBS
chr3	46393593	46394096	4	112	CBS
chr6	156725581	156726092	4	101	CBS
chr11	66915449	66915954	4	100	CBS
chr4	1055772	1056284	4	95	CBS
chr1	184825832	184826338	4	87	CBS
chr2	69215725	69216234	4	83	CBS
chr10	129225376	129225883	4	73	CBS
chr8	69492701	69493206	4	73	CBS
chr8	68587575	68588079	4	71	CBS
chr5	139124212	139124716	4	63	CBS
chr6	15884860	15885550	4	61	CBS
chr21	26072559	26073063	4	58	CBS
chr3	153091026	153091529	4	55	CBS
chr3	50861382	50861886	4	46	CBS

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chr7	108610938	108611440	4	43	CBS
chr19	17219771	17220274	4	26	CBS
chr15	78960842	78961356	4	18	CBS
chr3	165425162	165425675	4	16	CBS
chr7	157319162	157319668	4	11	CBS
chr11	113417981	113418481	3	2848	CBS
chr8	141479410	141479920	3	1404	CBS
chr5	11169204	11169704	3	724	CBS
chr10	112920191	112920700	3	676	CBS
chr11	112588300	112588810	3	621	CBS
chr4	18234048	18234552	3	458	CBS
chr17	4380511	4381020	3	411	CBS
chr4	11274428	11274929	3	381	CBS
chr15	29948282	29948787	3	352	CBS
chr18	36530300	36530800	3	295	CBS
chrX	153791109	153791615	3	279	CBS
chr12	18509306	18509816	3	266	CBS
chr17	45218569	45219074	3	185	CBS
chr4	42518552	42519056	3	176	CBS
chr6	155540040	155540544	3	147	CBS
chrX	115645818	115646318	3	146	CBS
chr10	44544941	44545450	3	143	CBS
chr17	48063737	48064241	3	115	CBS
chr5	97535476	97535976	3	109	CBS
chr3	46348421	46348930	3	106	CBS
chr2	239988540	239989042	3	95	CBS
chr4	100650839	100651349	3	95	CBS

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chr13	97026065	97026568	3	88	CBS
chr7	37672389	37672898	3	85	CBS
chr7	23678576	23679085	3	83	CBS
chr8	129635414	129635916	3	75	CBS
chr6	84500191	84500694	3	64	CBS
chr6	27090092	27090598	3	63	CBS
chr11	98499993	98500495	3	55	CBS
chr19	9555661	9556162	3	44	CBS
chr4	105211836	105212339	3	39	CBS
chr13	101369378	101369883	3	36	CBS
chr13	56755987	56756487	3	32	CBS
chr2	108338695	108339216	3	27	CBS
chr1	75920748	75921253	3	21	CBS
chr1	185205447	185205953	3	17	CBS
chr13	67889173	67889687	3	16	CBS
chr16	69102160	69102664	3	12	CBS
chr17	82038731	82039235	2	477	CBS
chr14	20631760	20632273	2	431	CBS
chr4	43014401	43014902	2	316	CBS
chrX	118259173	118259681	2	256	CBS
chr7	17621451	17621951	2	238	CBS
chr3	74033894	74034397	2	232	CBS
chr7	98990484	98990986	2	212	CBS
chr4	26631521	26632024	2	203	CBS
chr10	24406690	24407193	2	188	CBS
chr1	151017119	151017621	2	179	CBS
chr19	33134450	33134950	2	165	CBS

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chrM	7851	8360	2	159	CBS
chr5	166609122	166609625	2	157	CBS
chr1	34453057	34453560	2	155	CBS
chr18	45147024	45147531	2	146	CBS
chr7	152294192	152294697	2	138	CBS
chr20	11786155	11786665	2	130	CBS
chr3	126350843	126351346	2	122	CBS
chr9	80433013	80433513	2	118	CBS
chr15	94746835	94747338	2	111	CBS
chr5	87461581	87462083	2	101	CBS
chr6	115217939	115218443	2	91	CBS
chr1	157064187	157064689	2	80	CBS
chr2	180392626	180393127	2	76	CBS
chr9	42951103	42951605	2	70	CBS
chr5	153617451	153617951	2	68	CBS
chr2	180755987	180756487	2	62	CBS
chr7	15108412	15108914	2	43	CBS
chrX	33590983	33591485	2	35	CBS
chr7	50689655	50690158	2	32	CBS
chr16	57658360	57658863	2	31	CBS
chr7	13091707	13092209	2	29	CBS
chr21	41109456	41109957	2	28	CBS
chr7	15987330	15987832	2	28	CBS
chr18	50291373	50291873	2	22	CBS
chr6	138450255	138450757	2	20	CBS
chr6	74442580	74443080	2	19	CBS
chr8	472578	473082	2	19	CBS

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chr3	95094312	95094816	2	16	CBS
chr6	119200268	119200768	2	16	CBS
chr20	23519480	23519982	2	12	CBS

Tabla 5: Sensibilidad de CAST-Seq

<b>Muestra</b>	<b>Método</b>	<b>ng/reacción</b>	<b>Lecturas positivas</b>	<b>Copias/500 ng</b>
<b>CCR5/CCR2</b> células sin tratar	CAST-Seq	500	60	-
	ddPCR	0,003	4/163 185	9,8
<b>CCR5/CCR2</b> células editadas con CRISPR-Cas9	CAST-Seq	500	63 011	-
	ddPCR	0,003	174/33 325	1280

5 Se utilizó la PCR digital en nanogotas (ddPCR) para determinar el número de acontecimientos de delección grande en CCR5/CCR2 entre los loci CCR5 y CCR2.

Tabla 6: Análisis CAST-Seq para la nucleasa CRISPR-Cas9 dirigida a *VEGFA*

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chr6	43748797	43786167	6248	996445	DIANA
chr14	65102179	65102710	49	1646	DIANA INESPECÍFICA
chr22	37266517	37267041	34	1417	DIANA INESPECÍFICA
chr5	90144882	90145413	27	1572	DIANA INESPECÍFICA
chr5	116098720	116099230	9	243	DIANA INESPECÍFICA
chr20	59400339	59400861	9	148	DIANA INESPECÍFICA
chr6	43737908	43739102	8	493	DIANA INESPECÍFICA
chr11	18369652	18370170	8	156	DIANA INESPECÍFICA
chr19	40601908	40602418	8	24	DIANA INESPECÍFICA
chr3	179462091	179462601	7	233	DIANA INESPECÍFICA
chr16	12170504	12171007	7	84	DIANA INESPECÍFICA
chr12	58444469	58444973	6	31	DIANA INESPECÍFICA
chr6	39060603	39061111	5	126	DIANA INESPECÍFICA
chr6	43720476	43720979	4	97	DIANA INESPECÍFICA

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chr11	69083417	69083928	3	129	DIANA INESPECÍFICA
chr10	113678795	113679299	3	11	DIANA INESPECÍFICA
chr18	55588355	55588859	2	74	DIANA INESPECÍFICA
chr10	97000584	97001084	2	54	DIANA INESPECÍFICA
chr11	122681489	122681989	2	39	DIANA INESPECÍFICA
chr14	61611810	61612312	2	21	DIANA INESPECÍFICA
chr5	11938880	11939383	2	12	DIANA INESPECÍFICA
chr10	128219323	128219823	2	11	DIANA INESPECÍFICA
chr16	25475787	25476289	2	11	DIANA INESPECÍFICA
chr6	42465570	42466081	4	315	RH
chr6	43741170	43743375	11	599	CBS
chr8	133122122	133122635	8	433	CBS
chr3	43438528	43439033	8	211	CBS
chr6	43187467	43187977	7	193	CBS
chr6	43733353	43733861	7	191	CBS
chr11	132132426	132132932	7	98	CBS
chr17	8611719	8612229	6	337	CBS
chr6	43715825	43716338	6	206	CBS
chr12	91984719	91985230	6	149	CBS
chr9	80241133	80241648	6	141	CBS
chr12	120678790	120679309	6	123	CBS
chrX	126695885	126696392	5	505	CBS
chr2	144496514	144497032	5	285	CBS
chr3	51431062	51431579	5	223	CBS
chr5	93772512	93773014	5	189	CBS
chrUn_KI270442v1	98918	99420	5	96	CBS
chr20	16983569	16984078	5	73	CBS

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chr11	65289784	65290286	5	71	CBS
chr11	62744864	62745368	4	332	CBS
chr10	62979116	62979625	4	246	CBS
chr8	103827952	103828465	4	149	CBS
chrX	37912288	37912790	4	120	CBS
chr8	51018324	51018841	4	99	CBS
chr6	81829314	81829818	4	88	CBS
chr11	46891576	46892079	4	62	CBS
chr4	46405134	46405638	4	45	CBS
chr6	31492043	31492545	3	521	CBS
chr2	86112989	86113496	3	227	CBS
chr14	24347314	24347817	3	215	CBS
chr3	129911170	129911671	3	142	CBS
chr3	158845758	158846260	3	128	CBS
chr15	70227078	70227584	3	108	CBS
chr12	124923985	124924489	3	72	CBS
chr9	112200162	112200664	3	71	CBS
chr7	51813287	51813796	3	65	CBS
chr1	109067369	109067871	3	61	CBS
chr22	27094649	27095151	3	43	CBS
chr6	108630850	108631354	3	36	CBS
chr1	110417674	110418189	3	28	CBS
chr21	45699101	45699603	3	23	CBS
chr16	75303846	75304352	3	21	CBS
chr5	18617756	18618262	2	292	CBS
chr11	69569487	69569990	2	268	CBS
chr19	12369785	12370286	2	86	CBS

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lecturas</b>	<b>Categoría</b>
chr22	38055609	38056112	2	78	CBS
chr11	3468924	3469431	2	50	CBS
chr19	33384998	33385498	2	50	CBS
chr5	62393455	62393960	2	48	CBS
chr3	171773292	171773802	2	29	CBS
chr17	12195155	12195655	2	27	CBS
chr20	57997602	57998104	2	20	CBS
chr6	47302001	47302503	2	16	CBS
chr6	43729146	43729649	2	14	CBS
chr17	21170726	21171226	2	13	CBS

Tabla 7: Análisis CAST-Seq para la nucleasa CRISPR-Cas9 dirigida a *FANCF*

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lectura</b>	<b>Categoría</b>
chr11	22606881	22632976	16940	1194160	DIANA
chr8	107462104	107462619	3	202	DIANA INESPECÍFICA
chr17	75942773	75943281	8	427	RH
chr11	547497	548003	8	13	RH
chr10	3343323	3343830	6	334	RH
chr6	54512069	54512584	5	602	RH
chr22	31700499	31700999	4	508	RH
chr16	65201096	65201596	2	29	RH
chr11	22638379	22640795	16	756	CBS
chr12	118006840	118007350	14	327	CBS
chr11	22635293	22635993	13	1591	CBS
chr17	45061400	45061910	11	869	CBS
chr11	22603294	22604583	11	306	CBS
chr2	217749840	217750351	11	200	CBS

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lectura</b>	<b>Categoría</b>
chr10	92944396	92944911	10	628	CBS
chr11	22596928	22597439	10	513	CBS
chr4	170721395	170721913	10	245	CBS
chr5	223850	224355	10	112	CBS
chr2	156410822	156411330	9	564	CBS
chr11	22553782	22554292	9	399	CBS
chr3	131301701	131302212	9	184	CBS
chr11	62226588	62227098	9	172	CBS
chr15	74789553	74790064	8	1597	CBS
chr2	137754238	137754752	8	537	CBS
chr11	22493017	22493531	8	437	CBS
chr14	37848648	37849160	8	361	CBS
chr5	117503633	117504145	8	236	CBS
chr3	141960302	141960813	8	183	CBS
chr9	134110671	134111182	7	853	CBS
chr6	64592782	64593296	7	833	CBS
chr15	41979815	41980319	7	634	CBS
chr1	234873070	234873586	7	631	CBS
chr11	22497191	22497702	7	516	CBS
chr1	240683448	240683973	7	346	CBS
chr10	16143699	16144212	7	221	CBS
chr1	78612968	78613479	7	202	CBS
chrX	96728443	96728954	6	530	CBS
chr11	22546650	22547161	6	310	CBS
chr2	129780736	129781243	6	273	CBS
chr11	22170550	22171062	6	260	CBS
chr11	62851806	62852311	6	117	CBS

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lectura</b>	<b>Categoría</b>
chr16	173343	173857	5	1377	CBS
chr8	142938854	142939365	5	959	CBS
chr20	21331851	21332361	5	679	CBS
chr11	22580420	22580932	5	424	CBS
chr7	102837794	102838306	5	411	CBS
chr11	22566342	22566853	5	233	CBS
chr1	247258418	247258927	5	226	CBS
chr12	45103309	45103828	5	223	CBS
chr14	105323378	105323887	5	139	CBS
chr1	178305158	178305668	5	47	CBS
chr10	90062787	90063289	4	568	CBS
chr16	2350790	2351306	4	310	CBS
chr7	75678856	75679361	4	167	CBS
chr3	172715492	172716001	4	143	CBS
chr1	235016437	235016944	4	74	CBS
chr6	18644615	18645122	4	69	CBS
chr18	46622512	46623014	3	1106	CBS
chr17	40424371	40424885	3	583	CBS
chr5	149868431	149868932	3	382	CBS
chr11	40812298	40812806	3	336	CBS
chr3	50181773	50182276	3	319	CBS
chr22	45412128	45412628	3	270	CBS
chr1	53577913	53578425	3	252	CBS
chr4	134558964	134559473	3	251	CBS
chr18	71341064	71341570	3	247	CBS
chr18	44811004	44811516	3	212	CBS
chr11	30496682	30497192	3	206	CBS

ES 2 937 986 T3

<b>Cromosoma</b>	<b>Inicio</b>	<b>Fin</b>	<b>Aciertos</b>	<b>Lectura</b>	<b>Categoría</b>
chr10	110052643	110053143	3	120	CBS
chr10	2906798	2907498	3	92	CBS
chr3	122911294	122911794	3	66	CBS
chr13	24842172	24842680	3	52	CBS
chr11	20802893	20803396	3	41	CBS
chr2	105413692	105414194	3	26	CBS
chr1	206110190	206110694	3	13	CBS
chr2	121678953	121679461	3	9	CBS
chr17	38879882	38880386	2	660	CBS
chr4	138331185	138331693	2	555	CBS
chr8	127002239	127002739	2	303	CBS
chr13	82512617	82513120	2	258	CBS
chr11	22526098	22526600	2	247	CBS
chr21	32488958	32489461	2	220	CBS
chr4	124168649	124169153	2	149	CBS
chr11	22334210	22334711	2	147	CBS
chr11	30548665	30549165	2	75	CBS
chr11	66802270	66802770	2	41	CBS
chr18	30656154	30656661	2	31	CBS
chr16	87951936	87952438	2	24	CBS
chr21	14495238	14495740	2	22	CBS
chr5	159229053	159229555	2	22	CBS
chr10	8393633	8394133	2	17	CBS
chr13	64752705	64753213	2	16	CBS
chr19	39397176	39397676	2	13	CBS
chr5	153083290	153083792	2	13	CBS
chr11	22666156	22666659	2	11	CBS

**Tabla 8: Programas informáticos usados para CAST-Seq**

<i>Programa informático</i>	<i>Versión</i>	<i>Uso</i>	<i>Alternativa</i>
FLASH ( <a href="https://ccb.jhu.edu/software/FLASH/">https://ccb.jhu.edu/software/FLASH/</a> )	1.2.11	lecturas de apareamiento	Bbmerge ( <a href="https://jgi.doe.gov/data-and-tools/bbtools/">https://jgi.doe.gov/data-and-tools/bbtools/</a> )
Bbmap ( <a href="https://jgi.doe.gov/data-and-tools/bbtools/">https://jgi.doe.gov/data-and-tools/bbtools/</a> )	38.22	selección de sitios diana de nucleasas de diseño, recorte de conector y adaptador	Trimomatic ( <a href="http://www.usadellab.org/cms/?page=trimomatic">http://www.usadellab.org/cms/?page=trimomatic</a> )
Bowtie2 ( <a href="http://bowtie2.bio.sourceforge.net/bowtie2/index.shtml">http://bowtie2.bio.sourceforge.net/bowtie2/index.shtml</a> )	2.3.4.2	alineamiento con el genoma hg38	BWA ( <a href="http://bio-bwa.sourceforge.net/index.shtml">http://bio-bwa.sourceforge.net/index.shtml</a> )
samtools ( <a href="http://samtools.sourceforge.net">http://samtools.sourceforge.net</a> )	1.9	conversión de SAM a BAM	Picard ( <a href="https://broadinstitute.github.io/picard/">https://broadinstitute.github.io/picard/</a> )
bedtools ( <a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a> )	2.27.1	conversión de SAM a Bed, generación de secuencias aleatorias	BEDOPS ( <a href="https://bedops.readthedocs.io/en/latest/">https://bedops.readthedocs.io/en/latest/</a> )

Tabla 9: Paquete R usado para CAST-Seq

Programa informático	Versión	URL	Uso
BScgenome.Hsapiens.UC SC.hg38	1.4.1	<a href="http://bioconductor.org/packages/release/data/annotation/html/BSgenome.Hsapiens.UCSC.hg38.html">http://bioconductor.org/packages/release/data/annotation/html/BSgenome.Hsapiens.UCSC.hg38.html</a>	obtener la secuencia a partir de coordenadas genómicas
Biostrings	2.46.0	<a href="https://bioconductor.org/packages/release/bioc/html/Biostrings.html">https://bioconductor.org/packages/release/bioc/html/Biostrings.html</a>	alineamiento de la secuencia con ARN guía
ChIPseeker	1.14.2	<a href="https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html">https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html</a>	anotación de genes de sitios de translocación
TxDb.Hsapiens.UCSC.hg38.knownGene	3.2.2	<a href="https://bioconductor.org/packages/release/data/annotation/html/TxDb.Hsapiens.UCSC.hg38.knownGene.html">https://bioconductor.org/packages/release/data/annotation/html/TxDb.Hsapiens.UCSC.hg38.knownGene.html</a>	coordenadas génicas y regiones génicas conocidas
org.Hs.eg.db	3.5.0	<a href="https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html">https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html</a>	equiparar el símbolo del gen con la ID de entrada
biomaRt	2.34.2	<a href="https://bioconductor.org/packages/release/bioc/html/biomaRt.html">https://bioconductor.org/packages/release/bioc/html/biomaRt.html</a>	extraer el TSS de oncogenes

Tabla 10: Matriz de puntuación

	A	C	G	T	M	R	W	S	Y	K	V	H	D	B	N	PRO TUB.	
A	1	-1	-1	-1	0	0	0	-1	-1	-1	-	-	-	-1	-	-1	
											0,33 333	0,33 333	0,33 333		0, 5		
C	-1	1	-1	-1	0	-1	-1	0	0	-1	-	-	-1	-	-	-1	
											0,33 333	0,33 333		0,33 333	0, 5		
G	-1	-1	1	-1	-1	0	-1	0	-1	0	-	-1	-	-	-	-1	
											0,33 333		0,33 333	0,33 333	0, 5		
T	-1	-1	-1	1	-1	-1	0	-1	0	0	-1	-	-	-	-	-1	
												0,33 333	0,33 333	0,33 333	0, 5		
M	0	0	-1	-1	0	-0,5	-0,5	-0,5	-0,5	-1	-	-	-	-	-	-1	
											0,33 333	0,33 333	0,66 667	0,66 667	0, 5		
R	0	-1	0	-1	-0,5	0	-0,5	-0,5	-1	-0,5	-	-	-	-	-	-1	
											0,33 333	0,66 667	0,33 333	0,66 667	0, 5		
W	0	-1	-1	0	-0,5	-0,5	0	-1	-0,5	-0,5	-	-	-	-	-	-1	
											0,66 667	0,33 333	0,33 333	0,66 667	0, 5		
S	-1	0	0	-1	-0,5	-0,5	-1	0	-0,5	-0,5	-	-	-	-	-	-1	
											0,33 333	0,66 667	0,66 667	0,33 333	0, 5		
Y	-1	0	-1	0	-0,5	-1	-0,5	-0,5	0	-0,5	-	-	-	-	-	-1	
											0,66 667	0,33 333	0,66 667	0,33 333	0, 5		
K	-1	-1	0	0	-1	-0,5	-0,5	-0,5	-0,5	0	-	-	-	-	-	-1	
											0,66 667	0,66 667	0,33 333	0,33 333	0, 5		
V	-	-	-	-1	-	-	-	-	-	-	-	-	-	-	-	-1	
	0,33 333	0,33 333	0,33 333		0,33 333	0,33 333	0,66 667	0,33 333	0,66 667	0,66 667	0,33 333	0,55 556	0,55 556	0,55 556	0, 5		
H	-	-	-1	-	-	-	-	-	-	-	-	-	-	-	-	-1	
	0,33 333	0,33 333		0,33 333	0,33 333	0,66 667	0,33 333	0,66 667	0,33 333	0,66 667	0,55 556	0,33 333	0,55 556	0,55 556	0, 5		
D	-	-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-1	
	0,33 333		0,33 333	0,33 333	0,66 667	0,33 333	0,33 333	0,66 667	0,66 667	0,33 333	0,55 556	0,55 556	0,33 333	0,55 556	0, 5		
B	-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-1	
		0,33 333	0,33 333	0,33 333	0,66 667	0,66 667	0,66 667	0,33 333	0,33 333	0,33 333	0,55 556	0,55 556	0,55 556	0,33 333	0, 5		
N	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-0,5	-1	
															0, 5		
PRO TUB.	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	n/a

Matriz de puntuación de la sustitución de nucleótidos utilizada para el alineamiento de los sitios de translocación con la secuencia del sitio diana, incluidos los pesos para los desapareamientos y las protuberancias (inserciones/delecciones).

5

Se utiliza el código IUPAC. A, adenina; C, citosina; G, guanina; T (o U), timina (o uracilo); R, A o G; Y, C o T; S, G o C; W, A o T; K, G o T; M, A o C; B, C o G o T; D, A o G o T; H, A o C o T; V, A o C o G; N, cualquier base; n/a, no aplicable.

Tabla 11: Diseño de cebadores para el sitio diana HBB (diana de TALEN)

ES 2 937 986 T3

FUNCIÓN		SECUENCIA 5'-3'
<b>SECUENCIA DIANA</b>		TGATAGGCACTGACTCTCT (subunidad TALEN izquierda) (SEQ ID NO:25)
		TAAGGGTGGGAAAATAGAC (subunidad TALEN derecha) (SEQ ID NO:26)
<b>CAST-SEQ, PRIMERA PCR</b>	señuelo para cebo interno	GTTGGTATCAAGGTTACAAGAC (SEQ ID NO:27)
		CTGCTGGTGGTCTACC (SEQ ID NO:28)
<b>CAST-SEQ, SEGUNDA PCR</b>		GACTGGAGTTTCAGACGTGTGCTCTTCCGATCTGACCAATAGAACTGGGCATGG (SEQ ID NO:29)

Tabla 12: Diseño de cebadores para el sitio diana 2 de CCR5 (CCR5<sup>#2</sup>, diana de CRISPR-Cas9)

FUNCIÓN		SECUENCIA 5'-3'
<b>SECUENCIA DIANA</b>		CAATGTGTCAACTCTTGACAGGG (SEQ ID NO:30)
<b>CAST-SEQ, PRIMERA PCR</b>	señuelo para cebo inverso de señuelo interno	AAACACAGCATGGACGAC (SEQ ID NO:31)
		CCAGTGGGACTTTGGAAATAC (SEQ ID NO:32)
		GCATAGTGAGCCCAGAAG (SEQ ID NO:33)
<b>CAST-SEQ, SEGUNDA PCR</b>		GACTGGAGTTTCAGACGTGTGCTCTTCCGATCTAGGAGGATGATGAAGAA GATTCCAGAG (SEQ ID NO:34)

Tabla 13: Diseño de cebadores para ddPCR

Diana	Función		Secuencia 5'-3'
CCR5#1	ddPCR-Edge	Directo	TTATTATACATCGGAGCCCTGCCAA (SEQ ID NO:35)
		Inverso	TGCTCTTCAGCCTTTTGCAGTTTATCAG (SEQ ID NO:36)
	ddPCR-5'	Directo	AGTTTGCATTCATGGAGGGCAAC (SEQ ID NO:37)
		Inverso	GGCAGGGCTCCGATGTATAATAATTG (SEQ ID NO:38)
	ddPCR-3'	Directo	CATGCTGGTCATCCTCATCCTG (SEQ ID NO:39)
		Inverso	CCCAGAAGGGGACAGTAAGAAGG (SEQ ID NO:40)
CCR5#2	ddPCR-Edge	Directo	TCCTTCTTACTGTCCCCTTCTGG (SEQ ID NO: 41)
		Inverso	AGCAAACACAGCATGGACGAC (SEQ ID NO:42)
	ddPCR-5'	Directo	CATGCTGGTCATCCTCATCCTG (SEQ ID NO:43)
		Inverso	CCCAGAAGGGGACAGTAAGAAGG (SEQ ID NO:44)

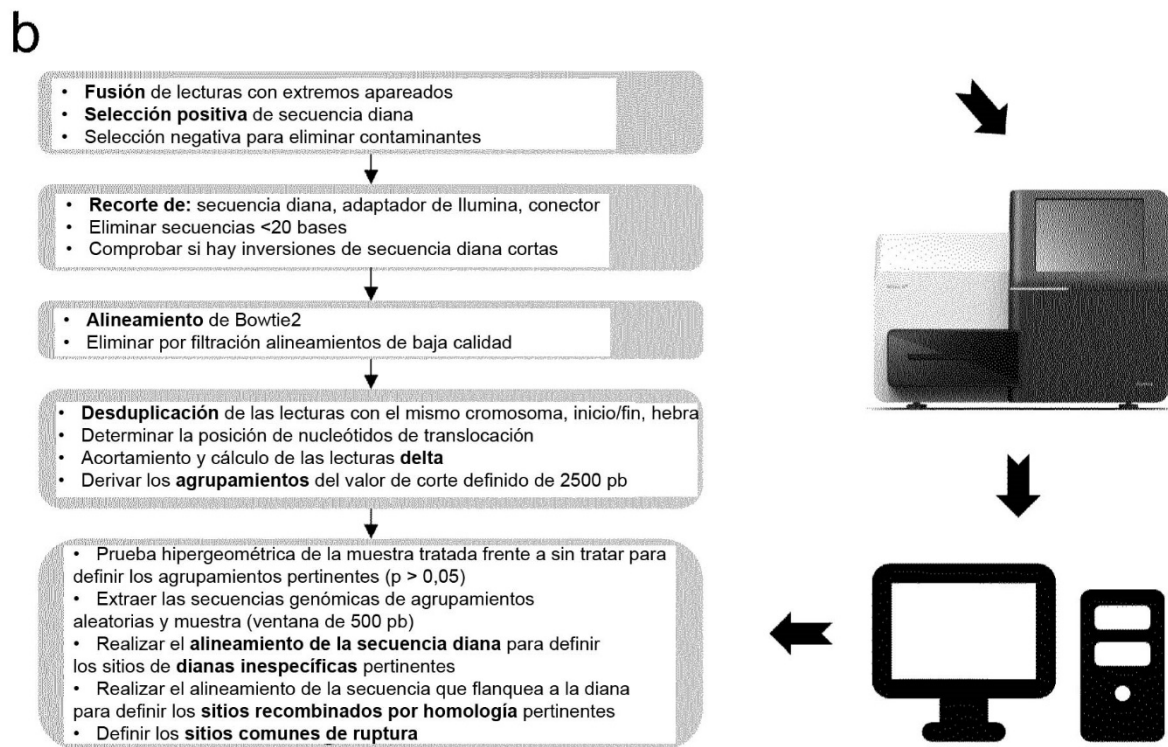
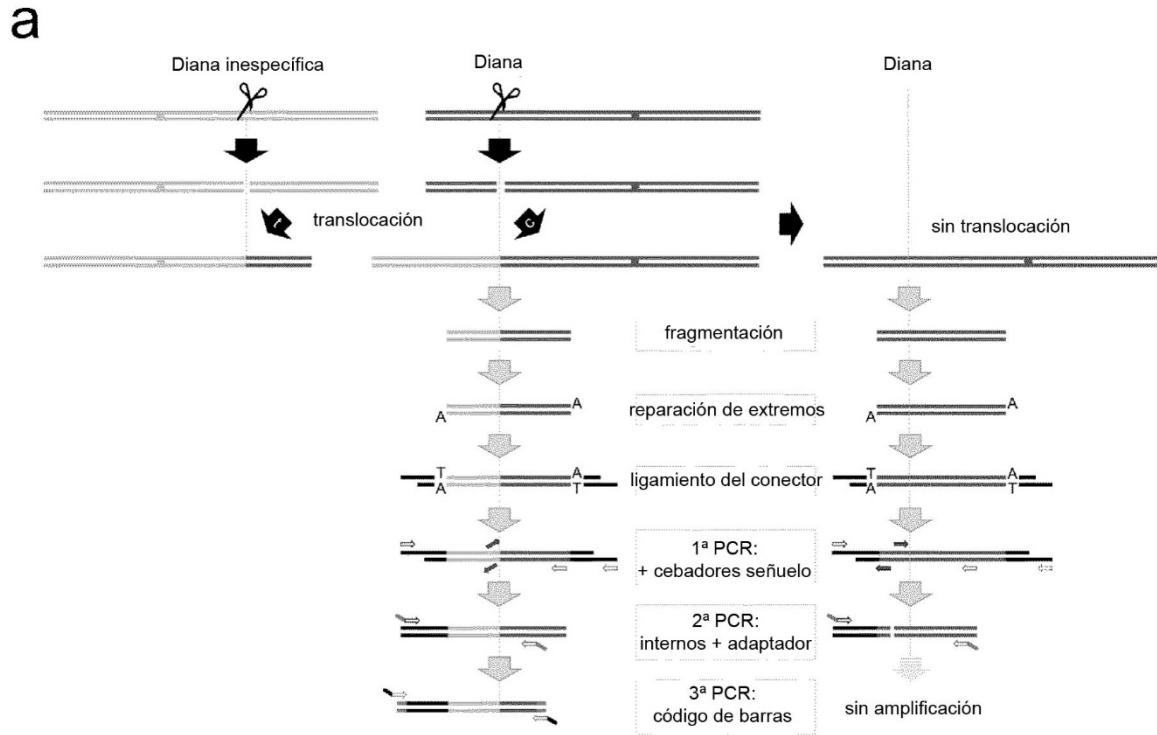
ES 2 937 986 T3

Diana	Función		Secuencia 5'-3'	
	ddPCR-3'	Directo	ATCGATAGGTACCTGGCTGTCTG (SEQ ID NO:45)	
		Inverso	GT A TGGAAAA TGAGAGCTGCAGGTG (SEQ ID NO:46)	
CCR5#1 y CCR5#2	ddPCR-GADL1 (telómero)	Directo	TGCCAAGGCATCTTACCTCTTCC (SEQ ID NO:47)	
		Inverso	GCATCTGGTCTTCTGCTACACTGG (SEQ ID NO:48)	
	ddPCR-MYLK (brazo q)	Directo	CAGCCTTGTGATTCATGCTGTCC (SEQ ID NO:49)	
		Inverso	GGA CT CACCTTCTACTGTCAACTCC (SEQ ID NO:50)	
HBB	ddPCR-Edge	Directo	AGACCAATAGAAACTGGGCATGTGG (SEQ ID NO:51)	
		Inverso	ATCACTAAAGGCACCGAGCACT (SEQ ID NO:52)	
	ddPCR-5'	Directo	GGCTCATGGCAAGAAAGTGCTC (SEQ ID NO:53)	
		Inverso	CAGTGCAGCTCACTCAGTGTG (SEQ ID NO:54)	
	ddPCR-3'	Directo	CTGAGGAGAAGTCTGCCGTTAC (SEQ ID NO:55)	
		Inverso	CCACATGCCAGTTTCTATTGGT (SEQ ID NO:56)	
	ddPCR-CARS (telómero)	Directo	GGGCCAGGGAAGTGTATGATG (SEQ ID NO:57)	
		Inverso	ACAGACATCAGTGCCATTGCG (SEQ ID NO:58)	
	ddPCR-PODL1 (brazo q)	Directo	GCAGGTT CAGTCCCTCTTGG (SEQ ID NO:59)	
		Inverso	TGCTTGGCCTATGGACAGTTG (SEQ ID NO:60)	
	Diana común	ddPCR-RAD1 (ctl.)	Directo	CCTTCAGCTCTGTGGTGACG (SEQ ID NO:61)
			Inverso	CCCTTCTCAGCAAAGTCCCTG (SEQ ID NO:62)
ddPCR-STAT3 (ctl.)		Directo	ACTCTCACGGACGAGGAGC (SEQ ID NO:63)	
		Inverso	CAGTTTTCTAGCCGATCTAGGCAG (SEQ ID NO:64)	

**REIVINDICACIONES**

- 1.- Un método para la detección de alto rendimiento de modificaciones del genoma completo en el genoma de un ácido nucleico obtenido a partir de una célula o un tejido causadas por la actividad de una nucleasa de diseño, que comprende las siguientes etapas:
- 5 a) extracción del ADN genómico de células que fueron expuestas a una nucleasa de diseño en condiciones que permiten a la nucleasa de diseño introducir una ruptura bicatenaria del ADN (DSB) en el ADN genómico de la célula,
- b) fragmentación del ácido nucleico para obtener fragmentos aleatorios,
- 10 c) realización de una reparación de extremos para obtener extremos con una A prominente en el extremo 3' que se apareen con los conectores utilizados en la etapa d) que tengan una T complementaria,
- d) ligamiento con un conector que comprende una secuencia complementaria para producir un cebador denominado "cebador para conector",
- 15 e) realización de una primera reacción de amplificación de ácido nucleico que es una reacción en cadena de la polimerasa con un "cebador para conector" y un cebador denominado "cebador para diana", en el que un cebador se sitúa cadena arriba y otro cadena abajo del sitio diana, en el que está presente en la mezcla de reacción al menos un cebador señuelo diseñado para unirse muy cerca del sitio de corte de la diana en una zona flanqueada por el cebador para diana y el cebador para conector inverso,
- f) realización de una segunda reacción de amplificación del ácido nucleico en la que se añaden a la mezcla de reacción los denominados "cebadores internos", en los que un cebador es complementario con el locus diana y otro cebador es complementario con la secuencia conectora,
- 20 g) realización de otra reacción de amplificación de ácidos nucleicos en la que se añade a la mezcla de reacción al menos un código de barras que contenga cebadores,
- h) secuenciación del producto de amplificación con cebadores internos y con código de barras, y
- 25 i) alineación de los productos secuenciados con medios bioinformáticos adecuados con una secuencia de referencia para identificar una ubicación cromosómica que contenga una modificación genómica basada en al menos una ruptura bicatenaria del ADN.
- 2.- Un método según la reivindicación 1, caracterizado porque el cebador señuelo es complementario con una secuencia situada a una distancia de al menos 10 nucleótidos cadena abajo del sitio diana.
- 30 3.- Un método según cualquiera de las reivindicaciones 1-2, caracterizado porque el sitio de unión para un cebador señuelo está situado a una distancia de al menos 10 nucleótidos cadena abajo y el sitio de unión para el otro cebador señuelo está situado a una distancia de al menos 30 nucleótidos cadena abajo del sitio diana.
- 4.- Un método según la reivindicación 1, caracterizado porque el cebador para diana directo está situado a una distancia de al menos 25 nucleótidos cadena arriba del sitio diana.
- 35 5.- Un método según cualquiera de las reivindicaciones anteriores, caracterizado porque permite identificar aberraciones cromosómicas poco frecuentes.
- 6.- Un método según cualquiera de las reivindicaciones anteriores, caracterizado porque permite la detección de aberraciones cromosómicas en el sitio diana y/o en los sitios diana inespecíficos.
- 7.- Un método según cualquiera de las reivindicaciones 1 a 6, caracterizado porque la reacción de amplificación del ácido nucleico se realiza en la zona del sitio diana donde la nucleasa de diseño corta al ácido nucleico.
- 40 8.- Un método según cualquiera de las reivindicaciones 1 a 7, caracterizado porque la molécula que contiene el código proporciona un código de barras para la secuenciación de nueva generación.
- 9.- Un método según cualquiera de las reivindicaciones anteriores, caracterizado porque la información obtenida por la amplificación del ácido nucleico se analiza con la ayuda de la bioinformática que comprende el apareamiento de secuencias y la selección positiva para la secuencia diana.
- 45 10.- Un método según la reivindicación 9, en el que se eliminan las secuencias no relacionadas debidas a productos de cebado erróneo o productos de conector.
- 11.- Un método según cualquiera de las reivindicaciones 9 o 10, en el que las secuencias seleccionadas se alinean con el genoma de referencia y en el que se eliminan los sitios de translocación falsos positivos.

Figura 1. Esquema de CAST-Seq.



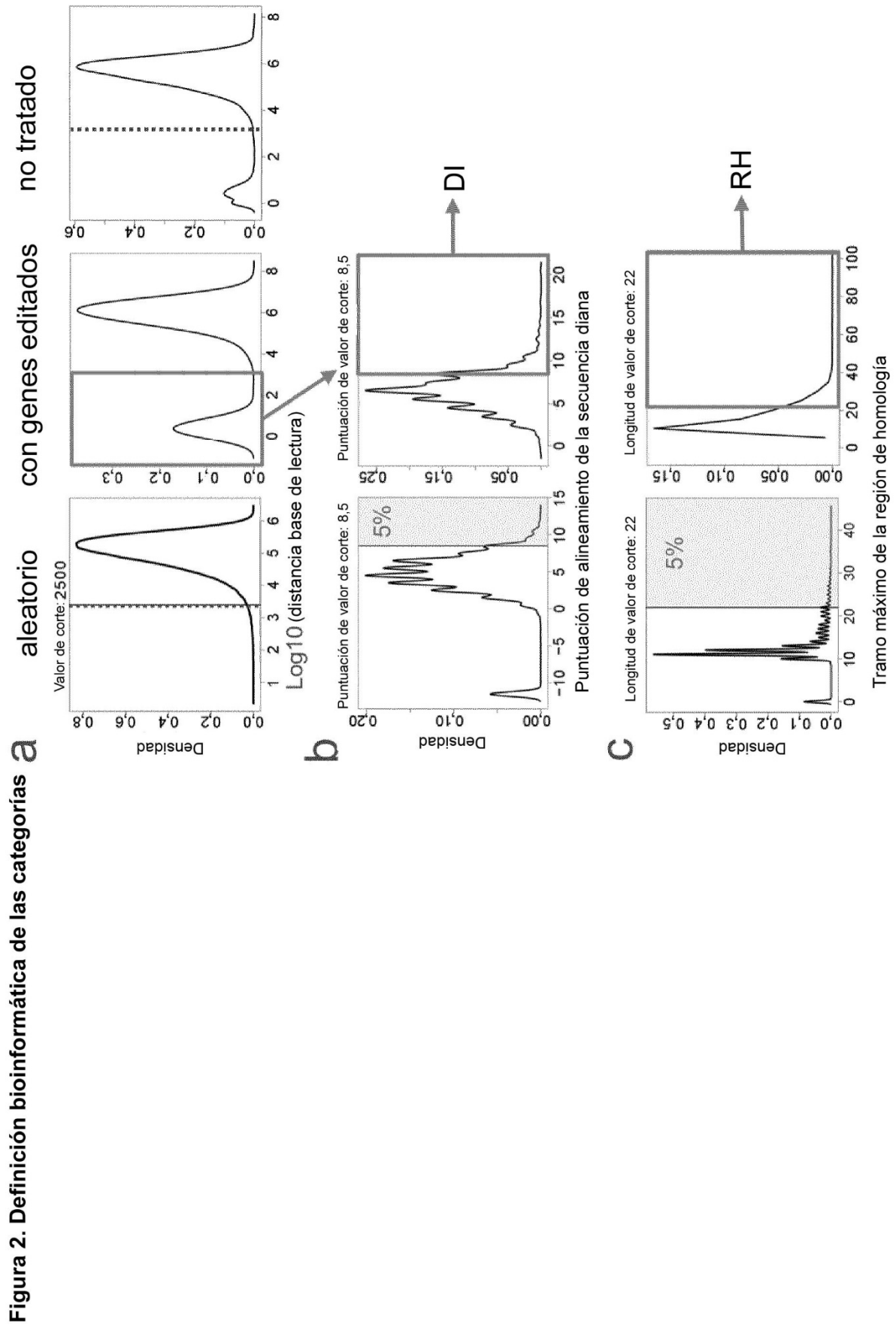


Figura 3. Efecto de los cebadores oligonucleotídicos señuelo

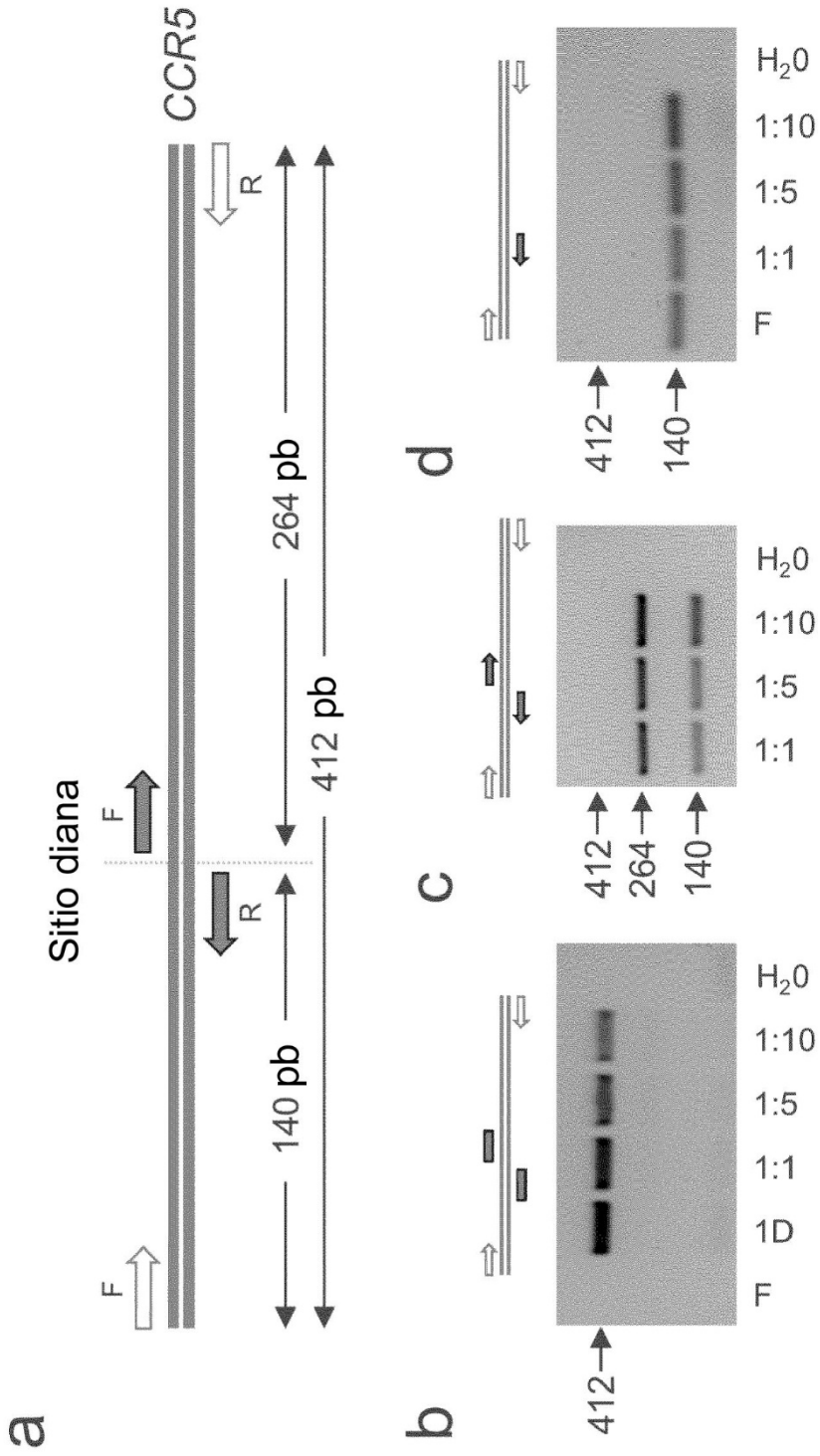


Figura 4. Modificaciones genómicas cartografiadas por CAST-Seq en células madre hemotopoyéticas humanas con genes editados

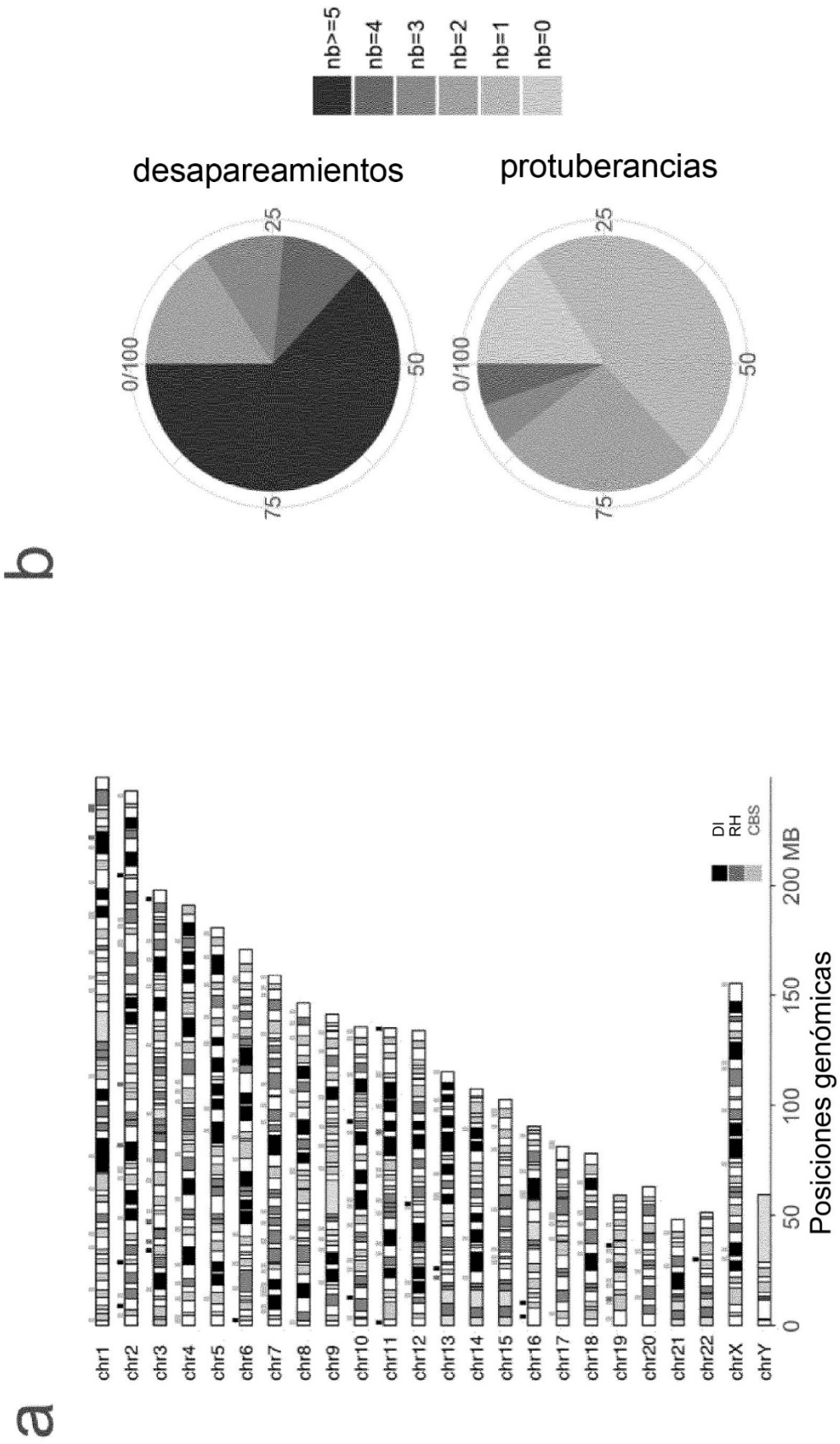


Figura 5. Alineamientos

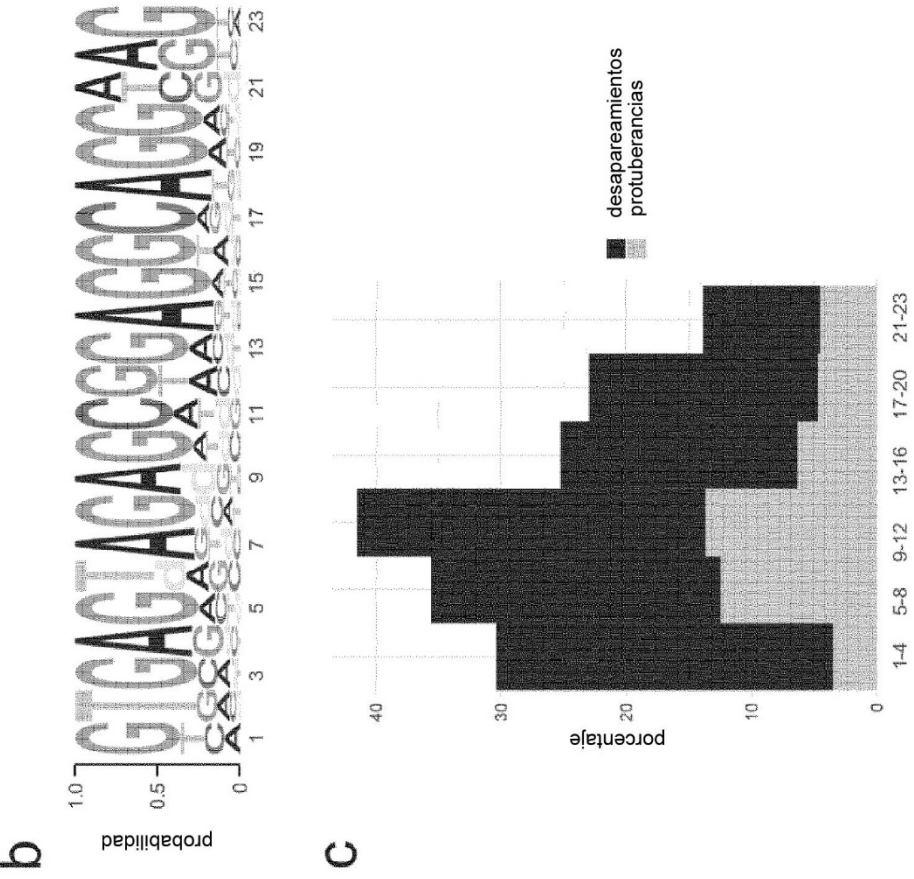


Figura 6. Representación esquemática de las lecturas de CAST-Seq en el locus CCR2-CCR5

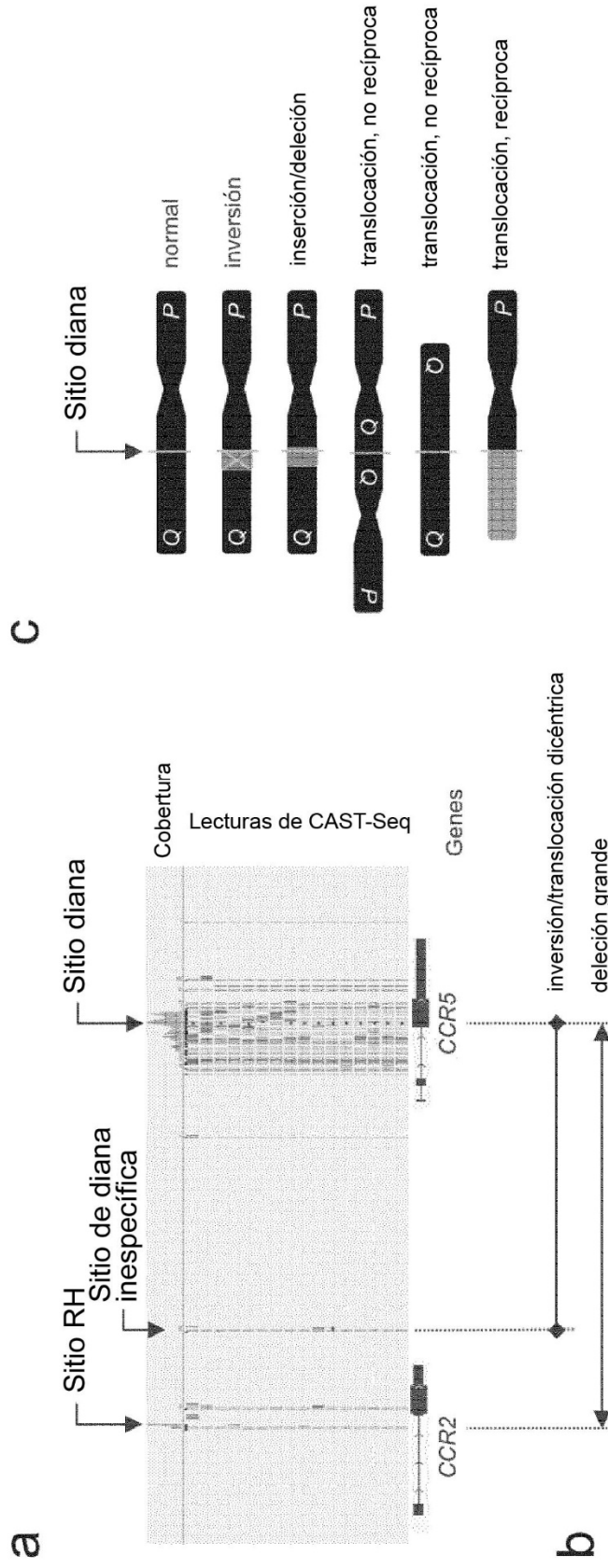


Figura 7

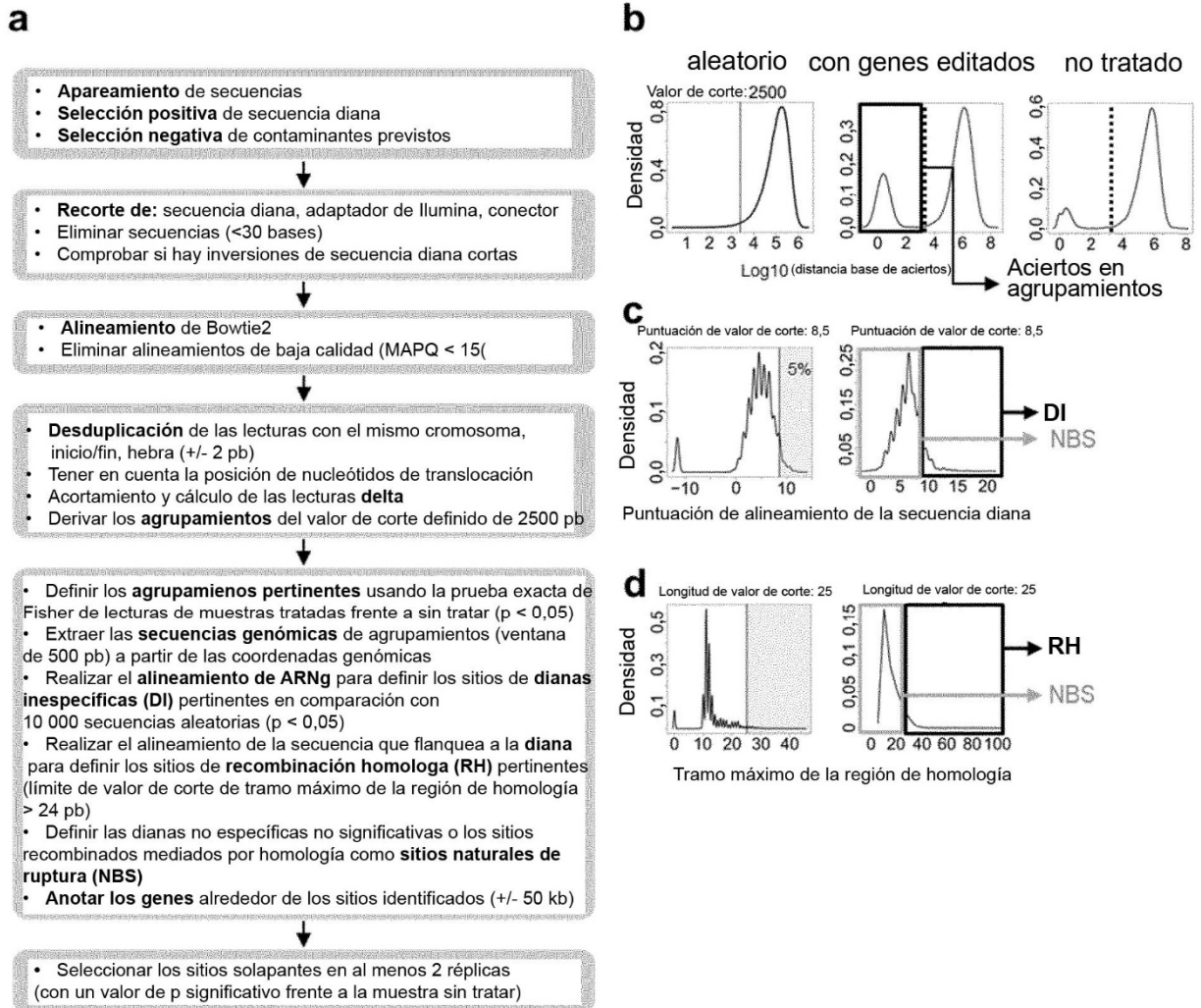


Figura 8

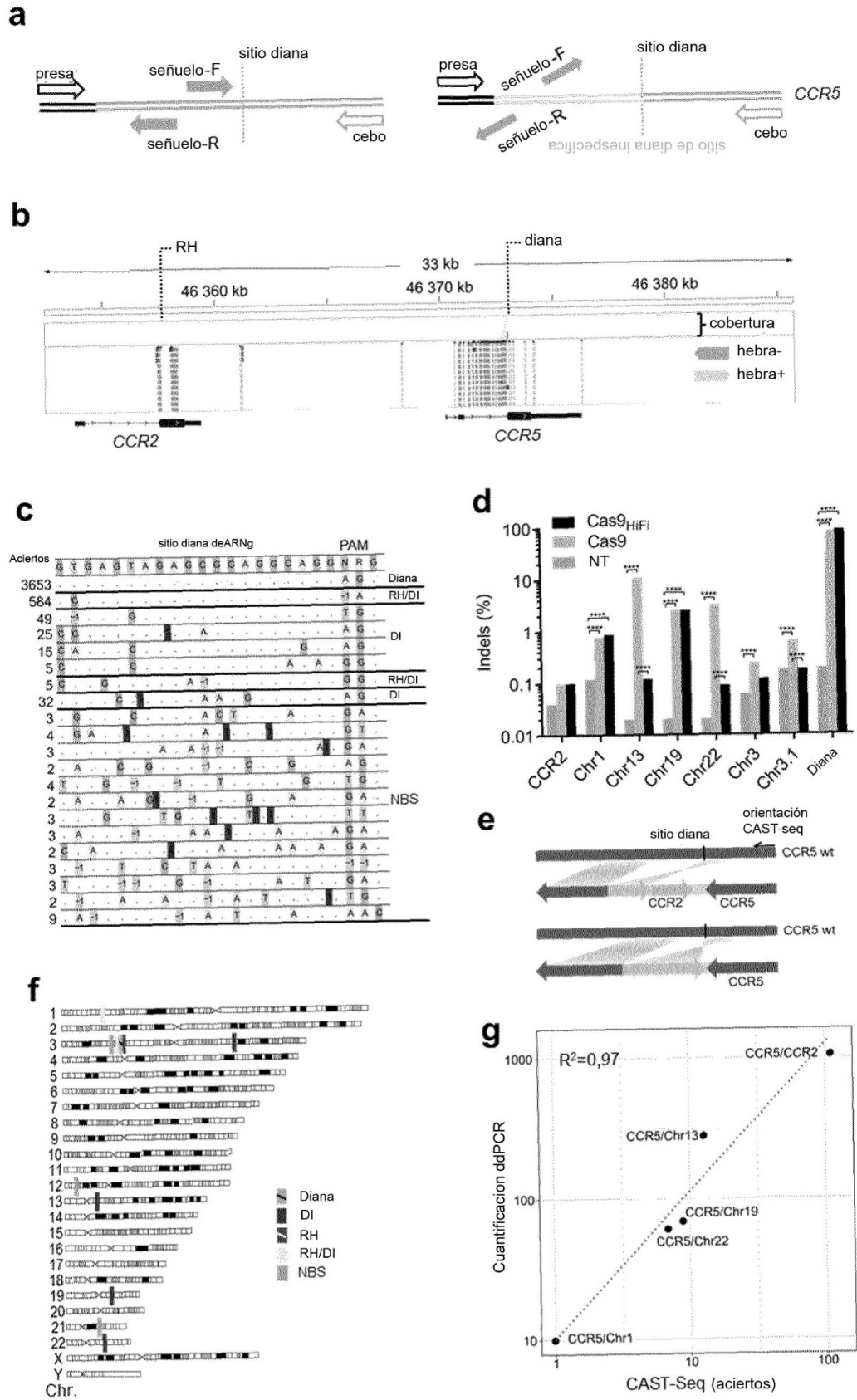


Figura 9

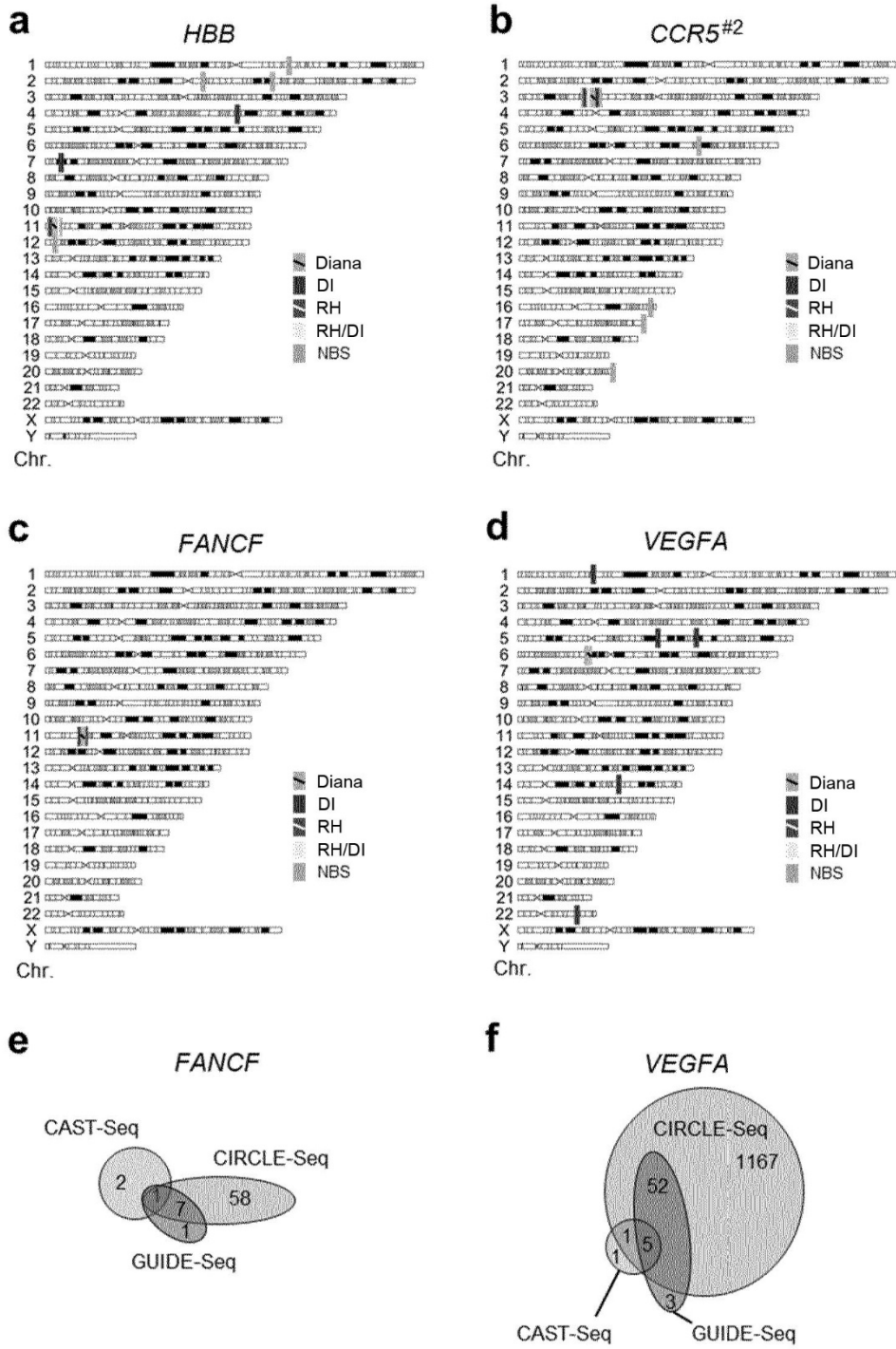


Figura 10

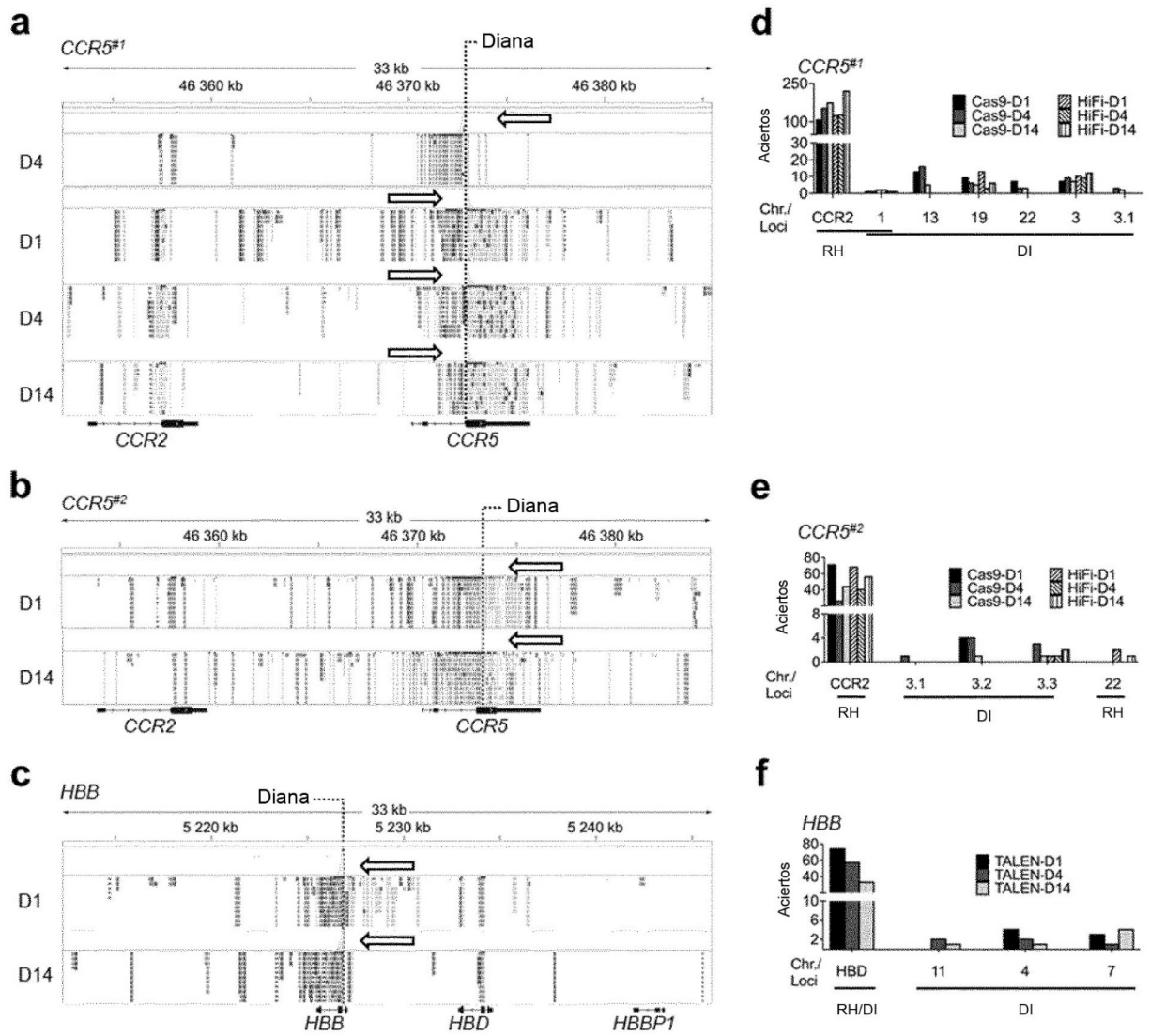


Figura 11

