

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 776 202**

51 Int. Cl.:

C12Q 1/6806 (2008.01)

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **11.04.2014** **PCT/SE2014/050451**

87 Fecha y número de publicación internacional: **16.10.2014** **WO14168575**

96 Fecha de presentación y número de la solicitud europea: **11.04.2014** **E 14782776 (0)**

97 Fecha y número de publicación de la concesión europea: **11.12.2019** **EP 2984182**

54 Título: **Captura de conformación cromosómica dirigida**

30 Prioridad:

11.04.2013 SE 1300264

13.06.2013 SE 1350724

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
29.07.2020

73 Titular/es:

SAHLÉN, PELIN (100.0%)

Armborstvägen 6

192 59 Sollentuna, SE

72 Inventor/es:

SANDBERG, RICKARD y

SAHLÉN, PELIN

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 776 202 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

Captura de conformación cromosómica dirigida

5 CAMPO DE LA INVENCION

La invención se refiere a un procedimiento que permite capturar la conformación cromosómica, así como a un kit que contiene componentes útiles para ser utilizados en este procedimiento para detectar relaciones de promotor y potenciador.

10 ANTECEDENTES DE LA INVENCION

En general, se cree que los cromosomas y los genomas están organizados en tres dimensiones, de modo que los elementos genómicos relacionados funcionalmente, p. ej., silenciadores y potenciadores, y sus genes diana, interactúan directamente o se encuentran lejos unos de otros.

15 Se cree que los genomas son complejos y están compuestos de ácidos nucleicos y proteínas, así como de otros componentes biológicos.

20 La actividad de los genes está estrechamente regulada para lograr las funciones biológicas en el momento y lugar adecuados. Cada gen porta una región llamada promotor, que es una secuencia corta de ADN responsable de interpretar las señales en el entorno celular para decidir si el gen debe activarse o no. Las proteínas específicas (factores de transcripción) se unen a la secuencia promotora para iniciar el ensamblaje o desensamblaje de la maquinaria de las proteínas y activar o inactivar su gen. Tanto las estructuras conformacionales secundarias como las terciarias de los genomas, así como los elementos reguladores, constituyen la arquitectura que inicia y dirige los eventos que ocurren dentro de una célula, eventos que podrían dar lugar a diferentes enfermedades o trastornos, o simplemente ser actividades normales dentro de la célula.

30 La tecnología desarrollada por Lieberman-Aiden y col., 2009 (Metodología de captura de conformación cromosómica (Hi-C)) se ha utilizado para mapear las interacciones de largo alcance y que sondea la arquitectura tridimensional de genomas completos.

Van Berkum L y col., 2010, J. Vis. Exp. Vol. 39, el1869doi:10.3791/1869 describe la tecnología convencional Hi-C, que podría combinarse con otras técnicas. Sin embargo, hasta ahora todas las técnicas han tenido algunos inconvenientes y han requerido mucho tiempo, aparte de que la mayoría de las técnicas tienen una resolución muy baja.

35 La nueva técnica inventada da lugar a un mapeo significativamente mayor de los promotores y potenciadores en comparación con otras técnicas (véase el ejemplo 2 de la solicitud).

40 Sin embargo, existen algunas limitaciones en la tecnología Hi-C, entre las que se incluye la necesidad de realizar una secuencia muy profunda para poder resolver las interacciones reguladoras entre los elementos promotores y potenciadores y, por lo tanto, la tecnología Hi-C requiere mucho tiempo y es cara, y es necesario desarrollar nuevas técnicas que puedan resolver esos problemas y permitan evaluar y detectar interacciones directas intra e intercromosómicas entre elementos reguladores remotos, y utilizar la información para diagnosticar las afecciones médicas y/o biológicas específicas.

45 RESUMEN DE LA INVENCION

50 El objeto de la presente invención es proporcionar un nuevo procedimiento mejorado que podría usarse junto con, por ejemplo, la tecnología Hi-C mencionada anteriormente para revelar más datos de una manera rápida y eficiente en relación con los genomas y la regulación de las actividades dentro de una célula. Por ejemplo, los potenciadores constituyen una gran fracción del genoma y son esenciales para la regulación de los genes. El alcance de la invención se encuentra definido en las reivindicaciones.

55 La invención se refiere a un procedimiento, que combina la captura de conformación cromosómica con una secuenciación basada en objetivos para aumentar la resolución (hasta ~600 pb) y la sensibilidad de las interacciones regulatorias. Usando este método, que aquí se llama HiCap, se encontraron 144751 interacciones promotor-potenciador que implican a 15 042 promotores y 101 856 potenciadores en células madre embrionarias de ratón, muchos de los cuales son potenciadores novedosos. A diferencia de los procedimientos anteriores de captura de conformación de cromatina en todo el genoma, los datos incluyen promotores de miles de genes no expresados. Mediante el procedimiento inventado también se pudo detectar una gran cantidad de interacciones promotor-promotor y potenciador-potenciador, organizadas de una manera que sugiere grupos de socios de interacción que permiten el procesamiento complejo de información entre potenciadores. El procedimiento tiene la resolución de la que carecían los procedimientos anteriores de captura de

conformación de cromatina, proporcionando finalmente una forma de mapear las interacciones reguladoras a nivel de potenciador/promotor único.

Mediante el uso del nuevo procedimiento inventado, ahora es posible utilizar la información revelada para diagnosticar afecciones médicas y/o biológicas específicas. En el entorno de la genética clínica, la importancia de la variación del ADN en las regiones reguladoras (potenciadores) al inicio y durante la progresión de la enfermedad compleja se ha descuidado durante mucho tiempo debido a la falta de una metodología para generar un conjunto funcional reducido. El enfoque HiCap permite anotar socios reguladores no codificadores de promotores a nivel de potenciador único. Esto extenderá la búsqueda de las variantes causantes de enfermedades a elementos reguladores además de las regiones codificantes de genes, proporcionando una forma de estudiar el papel de la variación no codificante en la patología de la enfermedad. Se podrían descubrir factores de riesgo genéticos adicionales que de otro modo podrían haberse pasado por alto. Con una mejor comprensión de la contribución de la variación no codificante a la progresión de la enfermedad, la elección de la terapia puede ser más precisa. Además, esto creará una demanda de nuevas terapias clínicas para mejorar las tasas de supervivencia de los pacientes.

En un primer aspecto, la invención se refiere a un procedimiento que comprende: i) proporcionar un ADN genómico reticulado, donde el ADN se conserva de modo que esté intacto, donde el ADN comprende un primer y un segundo conjunto de regiones ii) fragmentar el genoma reticulado creando una pluralidad de fragmentos con uniones, iii) agregar un marcador de unión marcado y ligar los fragmentos con uniones y marcador en condiciones tales que el marcador esté ligado a las uniones; iv) purificar los fragmentos que contienen un marcador ligado en la unión; v) agregar sondas de captura marcadas y seleccionar los fragmentos que se hibridan con las sondas de captura marcadas y vi) analizar los fragmentos que contienen un marcador ligado a la unión y los que se hibridan con la sonda de captura marcada para determinar la identidad de los fragmentos, donde las sondas de captura son las sondas que se hibridan con regiones reguladoras, como las secuencias promotoras.

Mediante el procedimiento inventado, la información que se obtiene de las interacciones genómicas y las regulaciones dentro de una célula se ha incrementado y se ha reducido la carga de trabajo y, por lo tanto, se podrá utilizar dicha tecnología en el diagnóstico de enfermedades médicas y condiciones biológicas específicas.

La invención permite investigar las interacciones, tales como las interacciones reguladoras del promotor de una manera más barata y menos intensiva en cuanto a los datos. En un segundo aspecto, la invención se refiere a un kit que comprende: i) una enzima que eliminará los nucleótidos biotinilados de los extremos de fragmentos no ligados, ii) un conjunto de enzimas para preparar las bibliotecas de secuenciación para una secuenciación de alto rendimiento, que comprende una enzima que repara los extremos de ADN, una enzima y adaptadores de secuenciación para ligar los adaptadores a los fragmentos, iii) perlas de estreptavidina para seleccionar fragmentos biotinilados, iv) un conjunto de sonda de captura de secuencia para capturar solo fragmentos complementarios a las regiones de interés, donde el conjunto de sonda de captura se hibrida con regiones reguladoras, tales como las secuencias promotoras, y v) productos químicos necesarios para realizar la reacción de captura de secuencia usando el conjunto de sonda de captura. El kit podría usarse en el procedimiento descrito anteriormente.

Se describirán con más detalle otras ventajas y objetos con la presente invención, entre otras cosas con referencia a los dibujos adjuntos.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La figura 1 muestra una presentación esquemática de secuencias quiméricas que informan de la interacción genómica de largo alcance, como entre promotores y potenciadores.

La figura 2 muestra una descripción general de la metodología HiCap. Se aplica un protocolo estándar Hi-C con modificaciones menores (la primera subfigura se adoptó de Lieberman-Aiden y col., 2009). Se prepara una biblioteca de secuenciación ligando adaptadores de secuenciación a los fragmentos (no mostrados aquí) y los fragmentos se amplifican por PCR. Se eliminan los materiales biotinilados originales (no mostrados aquí). Luego, estos fragmentos se hibridan con las sondas de captura de secuencia según las instrucciones del fabricante (la tercera subfigura se adopta de Roche Nimblegen, Inc.). El material no capturado es arrastrado y el material capturado se amplifica. El material capturado amplificado se secuenciará luego en formato pair-end.

La figura 3 muestra una representación esquemática de sondas de captura de secuencia por promotor. Los cuatro sitios de enzimas de restricción más cercanos (sitio RE) al sitio de inicio de la transcripción (TSS) se eligen como regiones objetivo para las sondas. Cada sonda tiene alrededor de 90 bases de largo y se encuentran en mosaico alrededor de 150-300 bases de regiones, cada una a aproximadamente 5 bases de distancia de otra.

La figura 4 muestra la tecnología inventada.

La figura 5 muestra la ubicación de unión de todo el genoma de 15 factores de transcripción diferentes y el análisis usando la tecnología inventada.

- 5 La figura 6 muestra la investigación de si las interacciones globales de HiCap podrían informar sobre la organización general de las interacciones regulatorias.

La figura 7 muestra el tamaño medio teórico del fragmento para el cortador de 4 (*Mbol*)

- 10 La figura 8 muestra las interacciones ancladas en las regiones de control negativo (las regiones intergénicas y exónicas objetivo) que se utilizaron para calcular el fondo en función de la distancia entre los pares de lectura.

La figura 9 muestra las interacciones filtradas basadas en el soporte de pares en réplicas.

- 15 La figura 10 muestra los resultados de interacción HiCap de alta confianza.

La figura 11 muestra que el cálculo de las fracciones esperadas utilizando un modelo de fondo completamente aleatorio produjo superposiciones más altas con marcas potenciadoras, pero también otras marcas de cromatina.

- 20 La figura 12 muestra la comparación de HiCap con otras técnicas.

La figura 13 muestra la comparación entre diferentes técnicas.

La figura 14 muestra una prueba de validación.

25

DESCRIPCIÓN DETALLADA DE LA INVENCION

Definiciones

- 30 En el contexto de la presente solicitud e invención, se aplican las siguientes definiciones: El término "unión" pretende referirse al sitio de ligación entre dos fragmentos de ADN que no residen uno al lado del otro en la secuencia genómica, sino que se unen debido a que residen uno cerca del otro en el espacio tridimensional del núcleo (figura 1)

- El término "marcador de unión" pretende referirse a un compuesto que puede incorporarse dentro de un ácido nucleico y
35 puede proporcionar una base para la purificación selectiva en uno o más pasos. Los ejemplos de marcadores de unión pueden incluir un enlazador de nucleótidos marcado, un nucleótido marcado y/o modificado, traslación de nick, enlazadores de cebadores o enlazadores etiquetados. La mayoría de estos son conocidos por un experto en la materia.

- El término "marcador de unión etiquetado" pretende referirse a un marcador que comprende cualquier secuencia de ácido
40 nucleico que comprenda una etiqueta que y puede incorporarse (es decir, ligarse) a otra secuencia de ácido nucleico. Por ejemplo, la etiqueta puede servir para purificar selectivamente la secuencia de ácido nucleico (es decir, por ejemplo, por cromatografía de afinidad). Tal etiqueta puede incluir, pero no está limitada a, una etiqueta de biotina o una etiqueta de digoxigenina.

- 45 El término "sonda de captura etiquetada" pretende referirse a una secuencia corta de nucleótidos que comprende una etiqueta y que es capaz de hibridarse con otra secuencia de nucleótidos. Por ejemplo, la etiqueta puede servir para purificar selectivamente secuencias específicas de ácido nucleico de interés. Tal etiqueta puede incluir, pero no está limitada a, una etiqueta de biotina o de digoxigenina.

- 50 El término "conservado e intacto" pretende referirse a que el ADN genómico y las proteínas están en la misma configuración que en la célula y, por lo tanto, a elementos reguladores que se ubican a distancia en la secuencia primaria del genoma entre sí, pero que están espacialmente cerca en el espacio nuclear del material conservado que también podría ser detectado e identificado por la técnica inventada, como el contacto promotor y potenciador.

- 55 El término "primer y segundo conjunto de regiones" pretende referirse a secuencias de nucleótidos que se encuentran en diferentes posiciones dentro del genoma pero que, en condiciones específicas, entran en contacto entre sí y por eso pueden cooperar y dirigir los eventos que ocurren dentro de la célula, como la expresión o el silenciamiento de genes específicos.

- 60 El término "fragmentos" pretende referirse a cualquier secuencia de ácido nucleico que sea más corta que la secuencia de la que deriva. Los fragmentos pueden ser de cualquier tamaño, desde varias megabases y/o kilobases a solo unos pocos nucleótidos de longitud. Las condiciones experimentales pueden determinar un tamaño de fragmento esperado,

que incluye, entre otros, digestión con enzimas de restricción, sonicación, incubación de ácidos, incubación de bases, microfluidización, etc.

El término "cromosoma" pretende referirse a una secuencia de ácido nucleico natural.

5

El término "análisis" pretende referirse a cualquier proceso o procedimiento por el que se utiliza una recopilación de información para llegar a una conclusión basada en la información de la secuencia. Los ejemplos incluyen la búsqueda de secuencias codificantes de proteínas dentro de un genoma utilizando información de secuencia obtenida en experimentos que perfilan la transcripción de genes.

10

El término "fragmentación" como se usa en esta invención pretende referirse a un método mediante el cual una secuencia de nucleótidos se fragmenta/separa en fragmentos de unidades más pequeñas. Las técnicas que se utilizarán para la fragmentación incluyen la escisión enzimática (es decir, por ejemplo, enzimas de restricción que actúan sobre ácidos nucleicos o enzimas proteasas que actúan sobre proteínas), hidrólisis de bases, hidrólisis ácida, sonicación o

15 desestabilización térmica inducida por calor.

El término "sonda de captura marcada" pretende referirse a una secuencia corta de nucleótidos que comprende una etiqueta y que es capaz de hibridarse con otra secuencia de nucleótidos. Por ejemplo, la etiqueta puede servir para purificar selectivamente secuencias específicas de ácido nucleico de interés. Tal etiqueta puede incluir, pero no está

20 limitada a, una etiqueta de biotina o de digoxigenina.

El término "retícula", "reticulación" o "crosslinking" pretende referirse a una asociación química estable entre dos compuestos, de modo que puedan procesarse adicionalmente como una unidad. Dicha estabilidad puede basarse en enlaces covalentes y/o no covalentes. Por ejemplo, los ácidos nucleicos y/o las proteínas pueden estar reticulados por

25

agentes químicos (es decir, un fijador) de modo que mantengan sus relaciones espaciales durante los procedimientos de laboratorio de rutina (es decir, por ejemplo, extracción, lavado, centrifugación, etc.) Muchos productos químicos son capaces de proporcionar reticulación, incluidos, entre otros, el formaldehído, el adipimidato de dimetilo (DMA) o el glutaraldehído.

30 El término "ligado o ligación" pretende referirse al enlace de dos secuencias de ácido nucleico que generalmente comprenden un enlace fosfodiéster. El enlace se ve facilitado normalmente por la presencia de una enzima catalítica (es decir, una ligasa) en presencia de reactivos cofactores y una fuente de energía (es decir, por ejemplo, trifosfato de adenosina (ATP)).

35 El término "marco de lectura abierto" pretende referirse a cualquier secuencia de ácido nucleico que codifique una proteína.

El término "elemento regulador" pretende referirse a una secuencia de ácido nucleico que afecta la expresión de otra secuencia genómica. Ejemplos son los potenciadores, represores, aislantes y regiones de control de locus.

40 El término "hibridar o hibridación" pretende referirse al emparejamiento de ácidos nucleicos complementarios usando cualquier proceso mediante el cual una cadena de ácido nucleico se une a una cadena complementaria a través del emparejamiento de bases para formar un complejo de hibridación. Muchos factores influyen sobre la hibridación y la fuerza de la hibridación (es decir, la fuerza de la asociación entre los ácidos nucleicos) tales como el grado de complementariedad entre los ácidos nucleicos, la rigurosidad de las condiciones implicada, la T_m del híbrido formado y la relación G:C dentro

45

Los términos "endonucleasas de restricción" y "enzimas de restricción" pretenden referirse a enzimas, que cortan ADN bicatenario en o cerca de una secuencia de nucleótidos específica. Se dice que las moléculas de ADN tienen "extremos 5'" y "extremos 3'" porque los mononucleótidos se hacen reaccionar para formar oligonucleótidos de manera tal que el

50

fosfato 5' de un anillo de pentosa mononucleotídico se une al oxígeno 3' de su vecino en una dirección mediante un enlace fosfodiéster. Por lo tanto, un extremo de un oligonucleótido se denomina "extremo 5'" si su fosfato 5' no está unido al oxígeno 3' de un anillo de pentosa mononucleotídico. Un extremo de un oligonucleótido se denomina "extremo 3'" si su oxígeno 3' no está unido a un fosfato 5' de otro anillo de pentosa mononucleotídico. Como se usa en la presente invención, también se puede decir que una secuencia de ácido nucleico, incluso si es interna con respecto a un oligonucleótido más

55

grande, tiene extremos 5' y 3'. En una molécula de ADN lineal o circular, los elementos discretos se denominan elementos "cadena arriba" o 5' o elementos "cadena abajo" o 3'. Esta terminología refleja el hecho de que la transcripción se realiza de forma 5' a 3' a lo largo de la cadena de ADN. Los elementos promotores y potenciadores que dirigen la transcripción de un gen enlazado generalmente se encuentran ubicados 5' o cadena arriba de la región codificante. Sin embargo, los elementos potenciadores pueden ejercer su efecto incluso cuando se encuentran cadena abajo de la región codificante.

60

La terminación de la transcripción y las señales de poliadenilación se encuentran en el extremo 3' de la región codificante.

El término "elemento regulador" pretende referirse a un elemento genético, que controla algún aspecto de la expresión de

secuencias de ácido nucleico. Por ejemplo, un promotor es un elemento regulador, que facilita el inicio de la transcripción de una región codificante operativamente unida. Otros elementos reguladores son las señales de empalme, señales de poliadenilación, señales de terminación, etc. Las señales de control de transcripción en eucariotas comprenden elementos "promotores" y "potenciadores". Los promotores y potenciadores consisten en series cortas de secuencias de ADN que interactúan específicamente con proteínas celulares implicadas en la transcripción (Maniatis, T. y col., Science 236:1237 (1987)). Los elementos promotores y potenciadores se han aislado de una variedad de fuentes eucariotas que incluyen genes en células y virus de plantas, levaduras, insectos y mamíferos (elementos de control análogos, es decir, promotores, también se encuentran en procariotas). La selección de un promotor y potenciador particular depende del tipo de célula que se utilizará para expresar la proteína de interés.

El término "etiquetado" pretende referirse a composiciones detectables por medios espectroscópicos, fotoquímicos, bioquímicos, inmunoquímicos, eléctricos, ópticos o químicos. Dichas etiquetas incluyen biotina para la tinción con conjugado de estreptavidina etiquetado o etiqueta de fluoróforo Cy3 para la detección fluorescente en microarrays.

15 *El procedimiento*

La presente invención se refiere a las interacciones entre elementos genéticos que se producen dentro de la célula viva. En una célula viva, la actividad de los genes está estrechamente regulada para lograr las funciones biológicas en el momento y lugar adecuados. Cada gen porta una región llamada promotor, que es una secuencia corta de ADN responsable de interpretar las señales en el entorno celular para decidir si el gen debe activarse o no. Las proteínas específicas (factores de transcripción) se unen a la secuencia promotora para iniciar el ensamblaje o desensamblaje de la maquinaria de las proteínas y activar o inactivar su gen. A menudo hay regiones de ADN ubicadas distalmente en el pliegue del genoma sobre las secuencias promotoras.

La conformación tridimensional de los cromosomas puede estar implicada en la compartimentación del núcleo y en la proximidad espacial de elementos funcionales ampliamente separados. Comprender cómo se pliegan los cromosomas puede proporcionar información sobre las complejas relaciones entre la estructura de la cromatina, la actividad de los genes y el estado funcional de la célula. Sin embargo, más allá de la escala de nucleosomas, actualmente se sabe poco sobre la organización de la cromatina. Debido a que el ácido desoxirribonucleico (ADN) es una molécula lineal, el genoma se considera a menudo lineal. Sin embargo, los cromosomas no son rígidos, por lo que la distancia espacial entre dos loci genómicos no necesita corresponder a su distancia a lo largo del genoma. Las regiones separadas por muchas megabases pueden ser inmediatamente adyacentes en un espacio tridimensional. Desde el punto de vista de la regulación, puede ser útil comprender las interacciones de largo alcance entre los loci genómicos. Por ejemplo, los silenciadores, potenciadores de genes y elementos aislantes podrían funcionar a través de grandes distancias genómicas.

Estas secuencias de ADN distales se denominan potenciadores y también pueden unirse a proteínas específicas. Las interacciones entre las proteínas unidas al potenciador y unidas al promotor contribuyen a la decisión de si el gen se activará o no. Este proceso se llama regulación distal de los genes. Los promotores de los genes siempre se encuentran proximales a los genes; sin embargo, las regiones reguladoras distales pueden estar muy lejos en la secuencia primaria del genoma y no es posible saber qué elementos reguladores distales se pliegan y actúan sobre qué promotor de la secuencia primaria de ADN en sí. Mediante el procedimiento/invención propuesta, ahora es posible saber qué elementos reguladores regulan qué promotor.

He aquí un ejemplo de una descripción paso a paso de la metodología HiCap en la presente invención (figura 2).

1. Las sondas de secuencia dirigidas a promotores de organismos relevantes están diseñadas alrededor de los sitios de enzimas de restricción más cercanos al sitio de inicio de la transcripción (figura 3). Si la secuencia alrededor de la enzima de restricción más cercana no es adecuada para el diseño de la sonda debido al alto contenido de GC y la presencia de elementos repetidos, las sondas se diseñan para la siguiente enzima de restricción más cercana. Las sondas deben diseñarse de modo que estén lo suficientemente cerca del promotor. Las sondas están diseñadas para al menos dos enzimas de restricción en cada lado del promotor (en total al menos 4 sitios de enzimas de restricción) para aumentar la probabilidad de capturar las interacciones del promotor. Además, las sondas están diseñadas para alrededor de 400-1000 regiones que no están cerca de ningún promotor anotado o elemento regulador. Estos constituirán regiones de control negativo para determinar los niveles de interacción de fondo.
2. Se aplica un protocolo estándar HiC a la línea celular de interés. Es importante utilizar una enzima de restricción frecuente como DpnII para alcanzar una mejor resolución de secuencia por elemento regulador.
3. Los materiales biotinilados originales se eliminan uniéndolo el material amplificado por PCR a perlas recubiertas con estreptavidina y se conserva el sobrenadante.
4. El sobrenadante se hibrida con sondas de captura de secuencia diseñadas a medida según las instrucciones del fabricante.
5. Las sondas de captura hibridadas se lavan según las instrucciones del fabricante y el material se amplifica por PCR y está listo para la secuenciación.

6. El material capturado y amplificado por PCR se secuenciará usando la plataforma Illumina HiSeq.
7. Los pares de lectura están alineados con el genoma apropiado.
8. Las frecuencias de interacción de fondo se calculan utilizando pares de lectura que se asignan a sondas de destino que se dirigen a regiones no reguladoras.
- 5 9. Se seleccionan pares de lectura para los que solo se selecciona una lectura del mapeo de pares a una región promotora y la coordenada genómica del otro par se mantiene como interactor putativo.
10. Para cada interactor putativo, se cuenta el número de pares de apoyo y los que están significativamente por encima de los niveles de fondo se asignan como potenciadores putativos de los promotores a los que estaban vinculados o como interacciones promotor-promotor.
- 10 En un aspecto, la invención se refiere a un procedimiento que comprende: i) proporcionar un ADN genómico reticulado, donde el ADN se conserva de modo que esté intacto, donde el ADN comprende un primer y un segundo conjunto de regiones ii) fragmentar el genoma reticulado creando una pluralidad de fragmentos con uniones, iii) agregar un marcador de unión marcado y ligar los fragmentos con uniones y marcador en condiciones tales que el marcador esté ligado a las uniones; iv) purificar los fragmentos que contienen un marcador ligado en la unión; v) agregar sondas de captura marcadas y purificar selectivamente los fragmentos que se hibridan con las sondas de captura marcadas y vi) analizar los fragmentos que contienen un marcador ligado a la unión y que se hibridan con la sonda de captura marcada para determinar la identidad de los fragmentos, donde las sondas de captura son las sondas que se hibridan con regiones reguladoras, como las secuencias de promotores.
- 15
- 20 El ADN genómico puede ser reticulado/inmovilizado por fijación, como por formaldehído y haciendo que el ADN esté conservado e intacto. Este paso asegura que todo el ADN genómico y las proteínas unidas a él se unan covalentemente entre sí para que la conformación del ADN genómico permanezca intacta durante el resto del procedimiento.
- 25 Luego, el ADN se fragmenta, lo que podría hacerse mediante una o más enzimas de restricción que dejan varias bases sin emparejar en el sitio de digestión creando una mezcla de fragmentos que incluyen un primer y un segundo conjunto de fragmentos que están cerca uno del otro e interactúan entre sí. El ADN comprende extremos adhesivos después del tratamiento con la enzima de restricción. Los ejemplos de enzimas de restricción incluyen HindIII, DpnII y BglII. El tratamiento del ADN con la enzima de restricción se puede realizar durante varias horas hasta toda la noche, dependiendo de la enzima utilizada.
- 30 Los fragmentos, cuyas bases se llenan con un marcador de unión etiquetado, como los nucleótidos etiquetados con biotina. El ADN se trata con calor para desactivar las enzimas presentes en la muestra.
- 35 Luego, el material se diluye y se liga para que los extremos libres del ADN se puedan unir entre sí, de modo que la primera y la segunda de las regiones se unan entre sí. Este es el paso donde se captura el plegamiento del genoma: dado que la estructura tridimensional del genoma se preserva mediante la reticulación, las regiones que estaban cercanas entre sí (es decir, que interactúan) en el momento de la reticulación pueden ligarse entre sí incluso aunque en realidad están muy lejos en la secuencia primaria de ADN. Luego se invierte la reticulación y se extrae el ADN. El material está ahora compuesto de ADN que contiene secuencias que estaban cerca unas de otras en el espacio tridimensional (figura 2). Estas secuencias quiméricas se denominan secuencias de unión y el sitio en el que dos secuencias distales son adyacentes se denomina unión (figura 2). Los primeros fragmentos biotinilados no ligados se eliminan usando ADN polimerasa T4. Luego, el ADN se fragmenta por sonicación, y los fragmentos biotinilados se extraen usando perlas de estreptavidina. Este paso garantiza que solo se investigarán en los pasos posteriores los fragmentos que contienen una unión. Esos primeros pasos son ya conocidos y se describen en van Berkum, N. L. y col., J. Vis. Exp. volumen 39, (2010) (Tecnología Hi-C).
- 40 Actualmente en el campo, los fragmentos de unión resultantes se secuencian utilizando tecnologías de secuenciación de próxima generación para desentrañar las interacciones genómicas. Las interacciones que ocurren entre los promotores y sus elementos reguladores (potenciadores) se denominan interacciones reguladoras. Sin embargo, la metodología Hi-C captura cualquier par de región genómica que estuviera cerca de otro en el momento de la reticulación como una interacción. Muchos de estos pares representan regiones en proximidad espacial debido a la estructura del genoma. Estos eventos se llaman interacciones estructurales. Dado que muchas células comparten una estructura genómica similar, las interacciones estructurales son más abundantes que las interacciones reguladoras, que a menudo están estrictamente reguladas y pueden estar presentes solo en un pequeño número de células en el momento de la reticulación. Por lo tanto, cuando uno secuenciar un experimento Hi-C, las interacciones estructurales se resuelven con una mejor resolución que las interacciones reguladoras. A pesar de que Hi-C proporciona grandes conocimientos sobre la estructura tridimensional del genoma, los dominios de cromatina, etc., se necesita una secuencia muy profunda para poder resolver las interacciones reguladoras que ocurren entre los promotores y sus secuencias reguladoras (potenciadores).
- 45
- 50
- 55
- 60 En una realización, la invención se refiere a un procedimiento, que puede usarse para seleccionar/encontrar secuencias de unión que, por ejemplo, informan sobre las interacciones potenciador/promotor. Hoy en día, se secuenciarán las

secuencias de unión en Hi-C que contienen interacciones con promotores que realmente no representan regulación/potenciadores distales

5 Esa realización particular se usa para seleccionar solo secuencias de interés usando secuencias cortas de ADN específicamente diseñadas llamadas sondas de captura, que se etiquetan y luego se denominan sondas de captura etiquetadas. Las sondas de captura pueden diseñarse contra secuencias promotoras ya que se conoce su secuencia. Estas secuencias cortas llamadas sondas de captura se etiquetarán como biotiniladas. Por ejemplo, el primer conjunto de regiones son regiones promotoras y el segundo conjunto de regiones comprende una secuencia reguladora, como potenciadores, silenciadores, aislantes, que se ubican cerca o distantes entre sí en el ADN o en el mismo o diferentes cromosomas y en otro ejemplo, el segundo conjunto de regiones comprende secuencias potenciadoras.

15 Las sondas de captura marcadas se hibridan con sus secuencias complementarias y luego se arrastran hacia abajo mediante, por ejemplo, perlas magnéticas recubiertas con estreptavidina. Aquí, en nuestro procedimiento, las sondas de captura están diseñadas para secuencias promotoras, y luego el material Hi-C se hibrida con esas sondas para seleccionar solo las interacciones de los promotores. La secuencia capturada contendrá la secuencia promotora pero también la secuencia que era proximal en el espacio en el momento de la reticulación. El material capturado se secuenciará entonces, solo secuenciando las interacciones de las regiones objetivo de las sondas de captura. Este procedimiento permite resolver las interacciones del promotor con regiones reguladoras distales con menos secuenciación.

20 Primero, las sondas de captura de secuencia están diseñadas para regiones promotoras. Para este fin, se selecciona una enzima de restricción adecuada para dar una resolución apropiada de las regiones reguladoras. La enzima debe dar un tamaño de fragmento medio entre 500-1000 pb. HiCap se ha aplicado a células madre embrionarias de ratón (mESC) para mapear los potenciadores del desarrollo a alta resolución. Un conjunto de sondas de captura de secuencia (Roche Nimblegen Inc) que consta de 53 735 sondas de captura de secuencia dirigidas a 31 000 promotores de ratones y 400 regiones que están al menos a 100 kb de distancia de los promotores para servir como controles negativos. La captura de secuencia se realiza entonces en material Hi-C de células madre embrionarias de ratón usando estas sondas y se secuenciará en la plataforma Illumina HiSeq. Un estudio dio como resultado 144 751 interacciones ancladas por promotores que implicaron 19 000 interacciones de promotores y 46 873 interacciones promotor-promotor adicionales. En total, 101 856 potenciadores putativos (PE) podrían mapearse con una resolución promedio de 609 pb. Hasta ahora, ChIA-PET ha sido el mejor procedimiento de rendimiento medio/alto disponible para estudiar elementos reguladores (Li, G, et. al, Cell, 2012). Comparando HiCap con ChIA-PET (en células K562, pull-down de ARN polimerasa II), a profundidades de secuencia similares, encontramos 3,3 veces más potenciadores con una resolución 6 veces mayor (tabla 1). Al comparar la superposición con los picos de ChIP-seq, también encontramos que HiCap tenía una mayor especificidad para los potenciadores (en promedio, un enriquecimiento 15 veces mayor para la superposición con los picos de ChIP-seq).

35 Tabla 1. Comparación de ChIA-PET con la metodología HiCap.

Procedimiento	Total de pares de lectura utilizados	Número de potenciadores	Longitud promedio del potenciador (pb)
ChIA-PET (K562 Pol2)	30M	33.682	3.789
HiCap (mES)	37M	109.864	609

40 Para investigar el potencial regulador de los potenciadores encontrados por HiCap, se realizó una comparación con datos públicos. Con este fin, se descargaron 12 conjuntos de datos de hipersensibilidad ChIP-Seq y DNase (conjuntos de datos ChIP-Seq para H3K4me1, H3K27Ac, H3K9Ac, CTCF, Cohesin, complejo Mediator, una colección de factores de transcripción específicos de mESC, p300, conjunto de potenciador específico de mESC ENCODE, regiones de baja metilación y sitios de hipersensibilidad a ADNasa específicos de mESC, disponibles a través del proyecto ENCODE en ratones, PLoS Biol, e1001046, 2011)). Se encontró un 67 % de los potenciadores de HiCap intracromosómicos superpuestos con al menos un pico de ChIP-seq.

50 Se realizó RNA-Seq se realizó para generar perfiles de expresión génica de mESC. De los 20 558 promotores anotados de genes codificantes proteínas, detectamos al menos una interacción distal para el 91 % de genes altamente expresados (428 de 469), el 85 % (10 423 de 12 210) de genes expresados en niveles medios y el 69 % (4043 de 7879) de genes poco expresados. Con la ayuda de los datos de expresión, también fue posible comprobar si la regulación por el mismo potenciador da como resultado niveles de expresión similares. Mientras que la expresión de genes conectados a los mismos potenciadores tenía un coeficiente de variación de 0,66 en promedio, para muestras estrechamente relacionadas (líneas celulares para trofoectodermo, endodermo primitivo y epiblasto) fueron mayores, a 0,68 ($P=2 \times 10^{-2}$, prueba de suma de rangos de Wilcoxon), 0,67 ($P=0,005$) y 0,67 ($P=0,01$) respectivamente. Por lo tanto, hubo una coregulación utilizando las interacciones reguladoras que se detectaron, mientras se controlaba el efecto de los genes expresados de manera similar que residían cerca uno del otro.

En otro aspecto, la invención se refiere a un kit que comprende: i) una enzima que eliminará los nucleótidos biotinilados de los extremos de fragmentos no ligados, ii) un conjunto de enzimas para preparar las bibliotecas de secuenciación para una secuenciación de alto rendimiento, que comprende una enzima que repara los extremos de ADN, una enzima y adaptadores de secuenciación para ligar los adaptadores a los fragmentos, iii) perlas de estreptavidina para seleccionar fragmentos biotinilados, iv) un conjunto de sonda de captura de secuencia para capturar solo fragmentos complementarios a las regiones de interés, donde el conjunto de sonda de captura se hibrida con regiones reguladoras, tales como las secuencias promotoras, y v) productos químicos necesarios para realizar la reacción de captura de secuencia usando el conjunto de sonda de captura.

Los siguientes ejemplos pretenden ilustrar, pero no limitar, la invención de cualquier manera, forma o forma, ya sea explícita o implícitamente.

EJEMPLOS

EJEMPLO 1

Las sondas de secuencia dirigidas a promotores de organismos relevantes se diseñaron alrededor de los sitios de enzimas de restricción más cercanos a los sitios de inicio de la transcripción (figura 3). Esto se repitió para todas las regiones promotoras, luego, estas regiones objetivo se enviaron a la empresa (Roche Nimblegen, Inc.) para el diseño y la fabricación de las sondas. Además, las sondas también se diseñaron para alrededor de 500-1000 regiones que no están cerca de ningún promotor anotado o elemento regulador para calcular la frecuencia de las interacciones que ocurren al azar debido a la proximidad en lugar de a la función (frecuencia de interacción de fondo).

Tenga en cuenta que no es necesario conocer de antemano ninguna secuencia potenciadora, ya que se capturarán junto con las secuencias promotoras mediante sondas de captura.

Las sondas pueden diseñarse no solo contra promotores, sino también para cualquier conjunto de regiones de interés, como regiones genómicas particulares o regiones genómicas no anotadas asociadas a la enfermedad.

Estos constituirán regiones de control negativo para determinar los niveles de fondo.

Se aplicó un protocolo estándar HiC a la línea celular de interés. En este experimento, se usaron células madre embrionarias de ratón. Es imperativo usar una enzima de restricción frecuente para alcanzar una mejor resolución de secuencia por elemento regulador y usamos DpnII; fragmenta el ADN genómico humano en promedio cada 800 bases ($\pm$ bases).

Los materiales biotinilados originales se eliminaron uniendo el material amplificado por PCR a perlas recubiertas con estreptavidina y se conservó el sobrenadante. Este fue un paso necesario ya que el material biotinilado original representa todo el genoma y no deben secuenciarse. Se capturarán durante el paso de captura de secuencia junto con las sondas de captura de secuencia biotiniladas. Por lo tanto, los eliminamos y solo hibridamos el material amplificado del material biotinilado original con las sondas de secuencia. La hibridación se realiza según las instrucciones del fabricante (Roche Nimblegen, Inc.).

Las sondas de captura hibridadas se lavaron según las instrucciones del fabricante (Roche Nimblegen, Inc.) y el material se amplifica por PCR y está listo para la secuenciación. El material capturado y amplificado por PCR se secuencia usando la plataforma HiSeq (Illumina Inc).

Los pares de lectura se alinearon con el genoma del ratón. Los pares de lectura que abarcan menos de 1 kilobase de distancia entre sí se descartaron ya que no informan sobre las interacciones distales. Entonces, las frecuencias de interacción de fondo se calculan utilizando pares de lectura que se asignan a sondas de destino que se dirigen a regiones no reguladoras.

Se seleccionaron pares de lectura para los que solo se selecciona una lectura del mapeo de pares a una región promotora y la coordenada genómica del otro par se mantuvo como interactor putativo. Para cada interactor putativo, se cuenta el número de pares de apoyo y los que estaban significativamente por encima de los niveles de fondo se asignan como potenciadores putativos de los promotores a los que estaban vinculados o como interacciones promotor-promotor.

EJEMPLO 2

Configuración experimental

- Con el fin de generar un mapa de alta resolución de interacciones regulatorias de largo alcance, se desarrolló HiCap que amplía Hi-C mediante una captura de secuencia adicional de fragmentos que contienen promotores para mapear simultáneamente interacciones de largo alcance para genes transcritos y no transcritos (**fig. 4a**). Para mejorar la resolución, el Hi-C inicial con un cortador de 4 (*MboI*) con un tamaño de fragmento medio teórico de solo 422 pb (**fig. 7**).
- 5 Las sondas de captura se diseñaron para fragmentos de restricción que contienen promotores de ratón (31 127 promotores en 16 696 genes únicos) y regiones de control adicionales en regiones intergénicas y exones ($n = 184$). Se generaron dos bibliotecas biológicas HiCap replicadas de células madre embrionarias de ratón (mESC) y se secuenciaron las bibliotecas de ambos extremos (2x100 pb) a una profundidad de 200-300 M de pares de lectura. Las lecturas se mapearon de forma independiente y los pares de lectura se descartaron si se mapearon dentro de 1 kb unos de otros
- 10 (para eliminar fragmentos autoligados). La eficiencia estimada de la enzima de restricción fue del 71 %. Las interacciones ancladas en las regiones de control negativo (las regiones intergénicas y exónicas objetivo) se utilizaron para calcular el fondo como una función de la distancia entre los pares de lectura (**fig. 8**). Luego, llamamos las interacciones ancladas al promotor, que requerían ser respaldadas por significativamente más pares (2 desviaciones estándar por encima de la media) que el fondo en ambas réplicas biológicas. Posteriormente filtramos las interacciones basadas en el soporte de
- 15 pares en ambas réplicas (**fig. 9**) para obtener 151 740 interacciones (mínimo 4 pares) y 13 527 interacciones de alta confianza (mínimo 14 pares), implicando 17 255 genes (7808 para interacciones de alta confianza). Estas interacciones incluyeron tanto interacciones de promotores a regiones distales (68 % de todas las interacciones) como interacciones entre dos regiones promotoras. Las interacciones intracromosómicas dominaron las interacciones promotor-distal (94 %) y promotor-promotor (99 %).
- 20 Con el fin de validar nuestras interacciones HiCap de los promotores a las regiones distales, investigamos en qué medida se enriquecieron para los potenciadores de mESC conocidos. Evaluamos el enriquecimiento como la proporción de superposición observada respecto a la esperada. Para incorporar las ubicaciones no aleatorias de promotores y potenciadores en nuestro modelo de fondo, calculamos las fracciones esperadas a través del muestreo aleatorio de
- 25 fragmentos de los promotores anotados utilizando las distribuciones de distancia observadas de las interacciones HiCap. Encontramos que las interacciones HiCap de alta confianza se enriquecieron significativamente ($P = 3,4 \times 10^{-34}$ a $P < 10^{-30}$ para todos los datos de potenciadores, prueba de Chi-cuadrado) para potenciadores previamente mapeados (**fig. 4b** y **fig. 10**), pero no para la cromatina transcripcionalmente silenciosa con marcas H3K27me3 ($P = 0,6$, prueba de Chi-cuadrado). Además, las regiones de heterocromatina marcadas con H3K9me3 se agotaron entre las interacciones de
- 30 HiCap (**fig. 4b**). El cálculo de las fracciones esperadas utilizando un modelo de fondo completamente aleatorio produjo superposiciones más altas con marcas potenciadoras, pero también otras marcas de cromatina (**fig. 11**). Estos resultados demostraron que nuestras interacciones ancladas a los promotores estaban altamente enriquecidas para los potenciadores conocidos.
- 35 Para determinar cómo HiCap se comparó con los procedimientos anteriores, a continuación, comparamos nuestras interacciones con los datos de Hi-C publicados en mESC y los datos de ChIA-PET generados con la inmunoprecipitación de ARN polimerasa II (ARN pol II) en células K562. La comparación de las frecuencias de mapeo sin procesar con los promotores y potenciadores demostró que HiCap tenía un porcentaje significativamente mayor de lecturas alineadas con ambos promotores conocidos ($P < 10^{-86}$, prueba de Chi-cuadrado) y potenciadores ($P < 10^{-13}$ para H3K27Ac; $P < 10^{-3}$ para
- 40 Cohesin, prueba de Chi-cuadrado) que tanto Hi-C como ChIA-PET (**fig. 12**). Los datos de ChIA-PET tuvieron una longitud de fragmento promedio de 3789 nts, pero las interacciones de HiCap tuvieron tamaños de fragmentos ($P = 2,2 \times 10^{-16}$, prueba de Chi-cuadrado) significativamente más cortos (media = 704 nts) (**fig. 4c**), una resolución adecuada para mapear potenciadores individuales. Además, las interacciones de HiCap fueron enriquecidas más significativamente ($P < 1,4 \times 10^{-2}$, prueba exacta de Fisher) para los potenciadores publicados que las interacciones ChIA-PET, al comparar su
- 45 superposición con varios datos de localización de potenciadores en células K562 (**fig. 4 d**), y en comparaciones usando interacciones recientes de ChIA-PET publicadas de mESC (**fig. 13**). Luego volvimos a analizar los datos de RNA-seq para investigar la expresión del ARN potenciador (eRNA), y las interacciones HiCap tuvieron significativamente más expresión de eRNA ($P = 3 \times 10^{-46}$, prueba de Chi-cuadrado) que las interacciones ChIA-PET (**fig. 4e**), proporcionando soporte por separado para la mayor pureza de las interacciones de buena fe capturadas con HiCap. Otra posible fortaleza con HiCap
- 50 es el potencial de mapear interacciones para genes transcritos activamente y silenciosos, mientras que las interacciones ChIA-PET capturan solo genes unidos por la proteína objetivo (a menudo ARN pol II). Encontramos que el 34 % de los genes no transcritos ($< 0,1$ RPKM) tenían una interacción mapeada, en comparación con solo el 18 % en ChIA-PET (**fig. 4f**). Volvimos a analizar las ubicaciones de unión de todo el genoma de 15 factores de transcripción (TF) diferentes en mESC junto con los análisis de expresión diferencial de todo el genoma después de la sobreexpresión de TF para
- 55 determinar si los genes con interacciones HiCap que los unían a los potenciadores mapeados se encontraban regulados positivamente con mayor frecuencia. Primero nos centramos en los genes más cercanos de TF mapeados (**fig. 5a**) y descubrimos que los genes con soporte de interacción HiCap se regulaban positivamente más a menudo que aquellos sin soporte de interacción HiCap (**fig. 5b**). El mayor enriquecimiento fue significativo ($P < 0,001$, prueba exacta de Fisher) para interacciones HiCap en varios umbrales de lectura (**fig. 5b**), así como interacciones Hi-C (solo en lecturas ≥ 2) (**fig. 5b**).
- 60 Para investigar la relevancia funcional de las interacciones entre potenciadores y genes no más cercanos, evaluamos su enriquecimiento para genes regulados positivamente. Las interacciones HiCap de alta confianza con genes más distantes tuvieron un enriquecimiento similar y, a veces, incluso significativamente mayor para los genes regulados

positivamente que el conjunto de genes más cercanos (**fig. 5c-d**). En particular, identificamos enriquecimientos significativamente mejores que el conjunto de genes más cercano para experimentos con Tcfcp2l1 y Myc (**fig. 13**), mientras que las interacciones con sitios de unión conocidos de otros factores como Klf4 tuvieron enriquecimientos similares a los conjuntos de genes más cercanos. Sin embargo, el Hi-C publicado en mESC fue consistentemente peor que el conjunto de genes más cercano. Esta validación funcional da la confianza de que las interacciones HiCap pueden predecir cambios en la expresión génica después de las perturbaciones de los TF.

La regulación génica en el núcleo tiene un componente espacial e investigamos si las interacciones globales de HiCap podrían informar sobre la organización general de las interacciones reguladoras. Observamos un aparente enriquecimiento para grupos interconectados (camarillas) de solo promotores (**fig. 6a**), solo potenciadores (**fig. 6j**) en comparación con los motivos que implican tanto a potenciadores como a promotores (**fig. 6b-i**). Las interacciones que involucran a dos promotores probablemente estaban sobrerepresentadas debido a la captura de secuencia, pero sorprendentemente detectamos un alto soporte de lectura para interacciones que implican dos potenciadores (a pesar de la captura de secuencia). Además, el soporte de lectura para las interacciones que implican dos potenciadores fue mayor que las interacciones entre un promotor y un potenciador (**fig. 6k**), proporcionando soporte adicional para las interacciones potenciador-potenciador e indicando que son prevalentes, incluso más prevalentes que las interacciones promotor-promotor. Los análisis adicionales de las interacciones potenciador-potenciador revelaron que estaban más a menudo unidas por el mismo factor TF de lo que se esperaría por casualidad, con enriquecimientos significativos para E2f1 y una tendencia hacia el enriquecimiento de otros factores (**fig. 6l**). Curiosamente, se anotó con más frecuencia que los genes conectados a través de interacciones de promotores o interacciones mutuas de promotor-potenciador se pertenecían a las mismas categorías de ontología génica (**fig. 6m**) respaldando el hallazgo anterior de que tales interacciones podrían estar implicadas en la coordinación transcripcional. Finalmente, observamos que dos genes con promotores interactuantes tenían más soporte de pares si interactuaban adicionalmente con uno, dos o más potenciadores (**fig. 6n**).

25 Procedimientos

Células ES de ratón. Se obtuvieron células madre embrionarias de ratón (línea R1) del laboratorio de Janet Rossant (Toronto, Canadá). Las células se mantuvieron en placas recubiertas con gelatina al 0,1 % en medio de Eagle modificado (DMEM) de Dulbecco suplementado con suero fetal de ternero (FCS) al 10 %, aminoácidos no esenciales 0,1 mM, L-glutamina 0,3 mg/ml, piruvato 1 mM (Invitrogen), 1000 U/ml de LIF murina (Chemicon International ESGRO), y se mantuvieron en una atmósfera de CO₂ al 5 % a 37 °C. El medio de células indiferenciadas se cambió diariamente.

HiCap. Se realizó Hi-C en células madre embrionarias de ratón como se ha descrito anteriormente⁹, excepto por las siguientes modificaciones. Generamos experimentos de réplica a partir de 20 millones de células madre embrionarias de ratón (mESC) que se reticularon con formaldehído al 1 % durante 10 minutos. Se lisaron las células y se aislaron los núcleos. Los núcleos aislados se digirieron con FastDigest MboI de 4 cortadores (Thermo Scientific, 1 µl/µg de ADN) durante 4 horas a 37 °C. Los extremos del material digerido se rellenaron con dATP, dGTP, dCTP y dTTP biotinilados utilizando fragmento de Klenow (Fermentas, 0,1 U por 1 µg de ADN). El Klenow se desactivó usando 0,01 M de EDTA en incubación a 75 °C durante 15 minutos. Luego, el material se diluyó a 3,5 ng/µl y se ligó usando T4 DNA Ligase (Promega). La reticulación se revirtió agregando proteinasa K e incubando durante la noche a 65 °C. Se eliminaron las proteínas y se purificó el ADN usando fenol-cloroformo seguido de precipitación con etanol. Los extremos biotinilados pero no ligados se eliminaron usando T4 DNA Polymerase incubando a 12 °C durante 15 minutos. El material se fragmentó a 300-600 bases por sonicación. Los extremos del fragmento se repararon y se les añadió una cola A. Luego, los fragmentos biotinilados se unieron a perlas de estreptavidina y los fragmentos no unidos se eliminaron por lavado. Los adaptadores de secuenciación se ligaron a continuación a los fragmentos unidos a las perlas. El material se amplificó 6-9 ciclos mientras se unió a perlas para obtener una cantidad suficiente para la captura de secuencia. El material biotinilado original se eliminó, el sobrenadante se hibridó al conjunto de sonda de captura de secuencia según las instrucciones del fabricante (Roche Nimblegen Inc.). El material hibridado se lavó según las instrucciones del fabricante y se amplificó con PCR durante 3-6 ciclos. Las siguientes bibliotecas de ADN se secuenciaron a 100 pb desde ambos extremos (secuenciación por pares) en un HiSeq 2000 (Illumina Inc.).

Mapeo de datos de secuencia. Las secuencias por pares se alinearon con el genoma del ratón (versión mm9) usando Bowtie versión 0.12.7 con la opción -m 1, en modo de extremo único para los dos extremos por separado, y con recorte iterativo desde el extremo 3' para lecturas no alineadas, 5 bases a la vez, hasta que se alineen. Las lecturas de mapeo múltiple se descartaron. El mapeo por pares no es adecuado para las bibliotecas HiCap ya que los 100 pares de bases en cada extremo contienen a menudo el punto de ligación para que un mapeador de pares recorte suavemente ese extremo de secuencia y así eliminar efectivamente la información de emparejamiento. Por lo tanto, utilizamos scripts personalizados para emparejar los extremos de secuencia asignados independientemente e indexamos cada par de secuencias a su correspondiente fragmento de restricción MboI

Sondas de captura de secuencia. Diseñamos sondas de captura de secuencia contra promotores de ratón que se compilaban de múltiples fuentes. Las anotaciones RefSeq y Ensembl se utilizaron junto con los sitios de inicio de

transcripción de DBTSS (desde el 25 de mayo de 2010) y MPromDb (desde el 28 de mayo de 2010). DBTSS se basa en ARNm de longitud completa, y corresponde principalmente a RefSeq y Ensembl. MPromDb se basa en datos de ARN PolII y H3K4me3 ChIP-seq para diferentes tipos de células, incluidas las células ES. A partir de los sitios de inicio de transcripción anotados, buscamos los sitios de corte de restricción más cercanos (GATC) en cada lado, y elegimos como regiones capturadas los 150 últimos pb antes del sitio de corte. Cuando los sitios de restricción estaban separados por <300 pb, elegimos la región completa entre ellos. A partir de estas regiones, Nimblegen diseñó las secuencias de sonda reales. También seleccionamos las regiones de control exónicas e intergénicas de entre las que se incluyeron en la misma canalización de selección de sonda.

- 10 **Llamada de interacciones.** Llamamos las interacciones significativas para todos los promotores que contenían fragmentos de restricción. Con este fin, se seleccionaron pares alineados de los cuales se seleccionó al menos un mapeo de parejas en un promotor. Las regiones promotoras se definieron como 1000 bases cadena abajo y 3000 bases cadena arriba del sitio de inicio de la transcripción. La extensión permitió que los mapeos que implicaban al fragmento directamente circundante al fragmento anclado al promotor se incorporaran a los análisis, ya que la eficiencia del corte de restricción era solo del 71 %. Recogimos todas las secuencias por pares con un extremo procedente de una región promotora y el otro extremo a una distancia de al menos 1000 pb de la región promotora. Luego contamos la frecuencia de interacciones con todos los fragmentos de restricción *MboI* en el genoma de cada región promotora capturada. Se descartaron los pares de lectura con exactamente las mismas posiciones de mapeo (para eliminar cualquier efecto potencial de los duplicados de PCR). Se aplicó el mismo procedimiento a las regiones de control negativo para obtener pares de lectura de las interacciones que luego se usaron como probabilidades de interacción de fondo. Agrupamos todas las distancias de interacciones de control negativo (tamaño de grupo de 1 kb) y calculamos el promedio y la desviación estándar del número de interacciones encontradas por fragmento para las regiones de control negativo (descartando de los cálculos los fragmentos con cero interacciones). Estas probabilidades de fondo de las interacciones se utilizaron para evaluar si cada interacción anclada al promotor era significativa en cada réplica biológica de forma independiente, utilizando una prueba Z. Ajustamos los valores *P*, para tener en cuenta las múltiples pruebas realizadas, utilizando el procedimiento de Benjamini-Hochberg y necesitamos una interacción significativa para tener unos valores *P* ajustados por debajo de 0,2 en ambas réplicas biológicas, lo que tuvo como resultado un umbral de valor *P* ajustado efectivo de 0,04 ya que se requería que las interacciones estuvieran presentes en ambas réplicas biológicas. Además, requerimos al menos 4 pares de lectura de soporte en cada réplica biológica y para las interacciones de alta confianza requerimos 14 pares de lectura por réplica. Las interacciones promotor-promotor fueron llamadas de manera similar, pero requerían que ambos extremos de las lecturas por pares se alinearan dentro de las regiones promotoras anotadas. También extrajimos los pares de lectura sin procesar para aquellas interacciones que solo implican a los potenciadores. Para este propósito, recogimos todas las regiones distales de interacciones promotor-elementos distales y realizamos análisis similares para los pares de lectura con ambos extremos procedentes de una región distal mapeada de HiCap. Las regiones potenciadoras no se extendieron, por lo que su resolución correspondía a fragmentos de restricción.

- Análisis de superposición con datos ChIP-seq de potenciador.** Descargamos las regiones potenciadoras inferidas en diferentes experimentos de ChIP-seq llevados a cabo con células mESC y K562. Clasificamos las regiones mapeadas en cada experimento para analizar solo las 5000 regiones mapeadas principales de cada experimento, a fin de controlar las diferentes señales y niveles de fondo en los diferentes experimentos. Para los datos de Mediator, descargamos las lecturas sin procesar para Med1 (SRX022694 y SRX022695) y Med12 (SRX022692 y SRX022693) y las alineamos con el genoma del ratón mm9. Realizamos llamadas pico utilizando SISSR versión 1.4, concatenamos y clasificamos los picos. Las regiones mapeadas de ChIP-seq se extendieron a 1000 pb si eran más cortas (relevante solo para las regiones unidas por Mediator). Para los análisis que comparaban la superposición de HiCap y ChIA-PET con potenciadores conocidos, calculamos la superposición observada con respecto a la esperada. La superposición observada se calculó simplemente como la fracción de interacciones HiCap o ChIA-PET que se superponían (con al menos 1 nt) con las regiones mapeadas de potenciadores. Para calcular la superposición esperada, muestreamos al azar regiones cercanas a sitios TSS anotados, utilizando la distribución de distancia real de las interacciones HiCap. Del mismo modo, las superposiciones esperadas de ChIA-PET utilizaron la distribución de distancia real en las interacciones ChIA-PET. Encontramos este procedimiento para controlar mejor las ubicaciones no aleatorias de genes y potenciadores en el genoma, mientras que el cálculo de la superposición esperada basado en un modelo completamente aleatorio (la fracción de fragmentos genómicos superpuestos con un potenciador conocido) hizo que todas las pruebas resultaran significativas.

- Comparaciones de superposición de potenciadores entre las interacciones HiCap y ChIA-PET.** Descargamos las interacciones promotor-potenciador mapeadas con ChIA-PET en células K562 y mESC. Analizamos 33 682 interacciones reportadas en K562 para las que la información de anclaje de la tabla de interacción publicada indicaba una interacción promotor-potenciador. Para las interacciones mESC que carecían de dicha información, derivamos las interacciones promotor-potenciador comparando los dos fragmentos emparejados con los sitios de inicio de la transcripción. Necesitábamos que solo uno de los dos fragmentos estuviera dentro de $\pm 2,5$ kb de cualquier sitio de inicio de la transcripción (anotaciones RefSeq, 18 de marzo de 2014), para lo cual se determinó que el otro fragmento fuera distal. Este procedimiento identificó 7738 interacciones de este tipo y se seleccionó la misma cantidad de interacciones HiCap de alta confianza para la comparación. En las comparaciones con los datos de ChIA-PET de K562 o mESC, clasificamos

las interacciones de HiCap según sus valores P y seleccionamos los 33 682 superiores (para la comparación de K562) o 7738 (para la comparación de mESC) para tener el mismo número de interacciones HiCap y ChIA-PET para los análisis. Los picos de ChIP-seq para H3K27Ac, H3K4me1, p300, smc3 (Cohesin) y CTCF para células K562 se recuperaron de la base de datos GEO con los siguientes ID de muestra: GSM733656, GSM733658, GSM733692, GSM1003583, 5 GSM935310 y GSM733719 respectivamente. Además, los picos de ChIP-seq correspondientes para mESC se recuperaron de la base de datos GEO con los siguientes ID de muestra: GSM1000099, GSM1000089, GSM1000121, GSM918750, GSM560343 y GSM918748 respectivamente. Todos los picos de ChIP-seq se ordenaron según su señal (signalValue, ENCODE broadPeak) y se utilizaron las 5000 regiones superiores para la comparación, a fin de controlar las diferentes señales y niveles de fondo en los diferentes experimentos. Superpusimos los potenciadores ChIA-PET y 10 HiCap a los picos relacionados con ChIP-seq y calculamos los valores observados para cada comparación. Calculamos la distribución de fondo mediante el muestreo aleatorio de regiones cercanas a sitios RefSeq TSS anotados, utilizando la distribución de distancia real promotor-potenciador del experimento HiCap o ChIA-PET. Las secuencias de control se compararon con los picos de ChIP-seq para calcular las superposiciones esperadas. Para permitir la superposición esperada variable (dado que las distribuciones de longitud de las interacciones difieren entre el experimento HiCap y 15 ChIA-PET), evaluamos el rendimiento de los procedimientos calculando la superposición observada menos la esperada para cada conjunto de datos ChIP-seq.

Comparación del mapeo de porcentajes de lectura sin procesar con los datos de promotores y potenciadores de los datos de HiCap, Hi-C y ChIA-PET. Utilizamos el mismo número de regiones promotoras RefSeq (25 267) del 20 ensamblaje de ratón mm9 para análisis mESC (para análisis Hi-C y HiCap) y del ensamblaje humano hg19 para análisis K562 de datos de ChIA-PET. Las regiones promotoras se definieron como la región cadena arriba de 1 kb de TSS anotado. Las ubicaciones de los potenciadores se basaron en dos experimentos representativos de ChIP-seq en Cohesin (Smc3) y H3K27Ac. Mapeamos 10 millones de lecturas sin procesar aleatorias de los experimentos HiCap, Hi-C y ChIA-PET e informamos de la alineación única de la fracción con los conjuntos de promotores y potenciadores. Los valores p se 25 calcularon usando la prueba de Chi-cuadrado en el número de lecturas alineadas frente al número total de lecturas para las comparaciones por pares entre procedimientos.

Análisis de nivel de expresión en células mESC y K562. Preparamos la biblioteca RNA-seq para mESC utilizando el protocolo Illumina mRNA-seq. La biblioteca fue secuenciada con un Illumina GAIIx con una longitud de lectura de 50 pb 30 en modo de extremo único (Fasteris). Las lecturas se alinearon con el genoma del ratón (ensamblaje mm9) y una colección completa de uniones de empalme utilizando Bowtie (versión 0.12.7). Descargamos los datos de RNA-seq de células K562 humanas del archivo de lectura de secuencia (SRX113647) y mapeamos las lecturas de secuencia hacia el genoma humano (hg19) usando STAR. Los niveles de expresión se estimaron como lecturas por kilobase de modelo de gen y millones de lecturas mapeadas de forma única (RPKM) utilizando Rpkmforgenes, donde solo se incluyeron posiciones de 35 mapeo únicas en la longitud del modelo de gen. La mapeabilidad se determinó usando MULTo y los modelos de genes se basaron en la anotación RefSeq descargada del navegador genómico UCSC el 31 de julio de 2011.

Prueba funcional en interacciones HiCap. Para evaluar la fuerza de las interacciones HiCap para predecir los genes expresados diferencialmente después de las perturbaciones de los factores de transcripción, creamos la siguiente prueba. 40 Los datos de unión del factor de transcripción se descargaron de Chen y col., y los datos de expresión después de la sobreexpresión del factor de transcripción. Para cada factor de transcripción presente en ambos conjuntos de datos, enumeramos el gen más cercano a cada punto medio de la región de unión. Identificamos las interacciones HiCap que conectaban a los promotores con el fragmento de restricción que contenía el punto medio del sitio de unión y enumeramos los genes de esos promotores. Para la figura 5b, comparamos el gen más cercano de picos sin interacciones HiCap con 45 los genes más cercanos que también tuvieron una interacción HiCap. Para la figura 5d, comparamos los genes no más cercanos con soporte HiCap con el conjunto de genes más cercanos (independientemente de las interacciones HiCap). Este procedimiento se realizó también en interacciones Hi-C (tanto las interacciones publicadas como en las inferidas utilizando nuestra canalización de llamadas de interacción). Comparamos los genes de la fracción regulada positivamente presentes dentro de los conjuntos de genes e informamos de las diferencias como fold changes. Para explicar la prueba 50 en detalle de la figura 5c-d y la figura 14, calculamos el número de genes más cercanos, Nc, y el número de genes conectados a HiCap, Nh. A partir de los datos de expresión, identificamos los genes expresados diferencialmente después de cada perturbación del factor de transcripción de forma independiente ($FDR \leq 0,05$ y fold change $> 1,5$). Luego, calculamos el número de genes únicos que se expresaron diferencialmente y que también estaban presentes en el conjunto de genes más cercanos (Uc) o en los genes inferidos (no más cercanos) de HiCap (Uh). Resumimos los 55 enriquecimientos como fold changes $[Uh/Nh]/[Uc/Nc]$ y calculamos un valor P utilizando la prueba de Chi-cuadrado con Uh y Uc como su relación observada y Nh/Nc como su relación esperada. Para la prueba compuesta que incluye todos los factores de transcripción, sumamos todos los valores de Nc Nh, Uc, Uh y realizamos las mismas pruebas.

Análisis de la expresión del ARN potenciador. Volvimos a analizar los datos mapeados de RNA-seq presentes en 60 Gene Expression Omnibus (GSM935897 y GSM935898) para determinar los niveles de expresión para potenciadores putativos inferidos por HiCap. Paralelamente, las lecturas de GSM854404 se mapearon en los potenciadores putativos de ChIA-PET usando Star. Se usaron alineamientos únicos para calcular los niveles de expresión de RPKM para cada

región inferida de HiCap o ChIA-PET. Los valores P se calcularon usando la prueba de Chi-cuadrado basada en la fracción de potenciadores putativos de HiCap y ChIA-PET con una expresión superior a 0,1 o 1,0 RPKM.

Motivos de las interacciones. Se extrajeron las interacciones de HiCap entre promotores y potenciadores para enumerar la frecuencia de los motivos (figura 6a-j). Para calcular los valores P, agrupamos las interacciones por distancia (1000-1999, 2000-3999, 4000-7999, etc., hasta 64000-127999) y por la suma de los grados de los nodos promotores (2, 3, 4, etc., hasta 20). Luego realizamos una prueba de suma de rangos de Wilcoxon de una cola para cada grupo (para los potenciadores 0 frente a 1 o 1 frente a 2+), y combinamos los valores P mediante el método de puntuación z de Stouffer, para calcular los valores P de dos colas. Los valores P también fueron significativos ($P < 1e^{-3^{**}}$) sin esta consideración de distancia y grado de red.

Análisis de ontología génica de pares de genes interconectados. Comprobamos si los pares de genes conectados a través de las interacciones promotor-promotor, promotor-potenciador-promotor comparten más a menudo la función anotada del gen. Para ello, utilizamos el servicio de ontología genética DAVID⁶. Primero calculamos para cada término de ontología génica cuántos pares de genes estaban conectados a través de una o más interacciones HiCap en los patrones descritos en la figura 6m para los genes dentro de ese término de ontología génica. Luego aleatorizamos (n=1000) todas las interacciones de HiCap entre todos los promotores y potenciadores y repetimos los mismos análisis anteriores. Calculamos los valores P como el número de aleatorizaciones con al menos tantos pares como los no aleatorios, o uno menos (para tener en cuenta la selección de términos con al menos un par real para ellos). Debido a las 1000 aleatorizaciones, el valor P mínimo posible fue de 0,001. Los valores P se ajustaron luego a las tasas de descubrimiento falso utilizando el método Benjamini-Hochberg.

EJEMPLO 3

- En otro caso ejemplar, se aplicó el método HiCap como sigue:
- Se diseñaron sondas de secuencia dirigidas a los promotores de organismos relevantes.
 - Alrededor de 5 millones de células (30 ug de ADN) se reticularon con formaldehído al 1 % durante 10 minutos a temperatura ambiente.
 - El núcleo del sedimento se resuspende en 240 ml de tampón de enzima de restricción 1,2X y 3,6 ul de SDS al 20 % (la concentración final es del 0,3 %). El sedimento se incubaba luego durante una hora a 37 °C agitándolo a 950 rpm.
 - Luego se añaden 27 ul de solución de Triton-X al 20 % (volumen final 2 %) al sedimento y se incubaba durante una hora a 37 °C agitándolo a 950 rpm.
 - Se agregan 30 ul de enzima FastDigest MboI (1 U/1 ug de ADN) al sedimento y se incubaba durante 4 horas a 37 °C agitándolo a 950 rpm.
 - La enzima se inactiva por calor mediante incubación a 65 °C durante 15 minutos.
 - Se añaden 156,8 ul de tampón de enzima de restricción 1X a la muestra digerida.
 - Se añaden 1,5 ul de dCTP 10 mM, dGTP y dTTP y 37,5 ul de 0,4 mM de biotina-dATP a la muestra digerida. Luego se añaden 1,2 ul de fragmento de Klenow (10 U/ul) y se incubaba durante 10 minutos a 37 °C.
 - Se añaden 10 ul de EDTA 0,5 M para inactivar la enzima y la muestra se incubaba a 75 °C durante 10 minutos.
 - Se preparan 9 ml de tampón de ligación de ADN ligasa T4 1X y se suplementan con 90 ul de ATP 100 mM. Se añaden 8090 ul de tampón de ligasa 1X suplementado a la muestra (la concentración final de ADN debe ser de alrededor de 3,5 ng/ul). Se añaden 50 unidades de ADN ligasa de T4 a la muestra y se incuban durante 4 horas a 16 °C y luego 1 hora a temperatura ambiente.
 - Se añaden 25 ul de proteinasa K (20 mg/ml) a la muestra y se incubaba a 65 °C durante 12 horas.
 - Se aplica una purificación estándar de fenol-cloroformo a la muestra. Se añade un volumen equivalente de fenol-cloroformo: alcohol isoamílico (25:24:1) a la muestra, se mezcla bien y se centrifuga a 3000 rpm durante 10 minutos. La capa acuosa se transfiere a un nuevo tubo y se añaden 2,5 volúmenes de etanol al 100 % y 0,1 volúmenes de acetato de sodio 3 M a pH 5,2, se mezcla bien y se incubaba durante una hora a -20 °C.
 - Centrifugar la muestra a 13,000 rpm durante 30 minutos y desechar cuidadosamente el sobrenadante sin alterar el sedimento.
 - Lavar el sedimento con 500 ul de etanol al 70 % mediante centrifugación a 13 000 rpm durante 5 minutos. Secar al aire el sedimento durante 10 minutos a temperatura ambiente y añadir 100 ul de agua destilada. La muestra está lista para la preparación de la biblioteca.
 - Eliminar la biotina de los fragmentos no ligados usando ADN polimerasa de T4 siguiendo el siguiente protocolo:

agua destilada	27,0
5X tampones (Fermentas)	20,0
10 mM dATP	1,0
10 mM dGTP	1,0
ADN (máx. 5ug por rxn)	50,0

ADN polimerasa T4 (3U/ul) (Fermentas)	1,0
Volumen total	100,0

Incubar la reacción a 12 °C durante 15 minutos en un termociclador. Detener las reacciones añadiendo 2 ul de 0,5 M EDTA y purificar el ADN usando fenol:cloroformo:alcohol isoamílico, seguido de precipitación con etanol y resuspender el sedimento con 100 ul de agua destilada.

14. Cortar el ADN usando el sonicador Covaris a 250-500 bases usando la siguiente configuración: Ciclo de trabajo: 10 %, Intensidad: 5, Ciclos por ráfaga: 200, Tiempo: 50x3 segundos, Tiempo total de 150 segundos. El volumen de la muestra es de 120 ul y cada tubo debe contener un máximo de 3 ug de ADN.
15. La preparación de la biblioteca de secuenciación de ADN por pares se realizará mediante el kit de preparación de la biblioteca Illumina TruSEQ según el protocolo del fabricante y se describirá en los siguientes pasos "Ocho reacciones de preparación de la biblioteca paralelas deberían producir suficiente material para la captura de secuencia".
16. Se mezclan 60 ul de ADN cizallado (la concentración total de ADN no debe exceder 1 ug) con 40 ul del kit de reparación final del kit de la biblioteca Illumina y se incuba a 30 °C durante 30 minutos.
17. Luego, el ADN se purifica usando una proporción perla:ADN de 1,6:1 usando perlas Ampure XP según las instrucciones del kit de la biblioteca y el ADN se eluye en 17,5 ul de tampón de resuspensión.
18. Se añaden 12,5 ul de tampón de cola A (20 mg/ml) al ADN y se incuba a 65 °C durante 30 horas.
19. En este paso, se extraerán los fragmentos biotinilados utilizando perlas magnéticas de estreptavidina. Se deben utilizar tubos de baja unión a partir de este paso. Primero, preparar el tampón no Tween (NTB) que contenga Tris-HCl 1 M 100 mM, pH 8,0, NaCl 2 M y EDTA 1 M. Para el lavado de las perlas, preparar el tampón de lavado Tween diluyendo el NTB dos veces y agregando 1 ul de Tween-20. A continuación, se preparan 100 ul de perlas de estreptavidina lavándolas dos veces con 400 ul de tampón de lavado Tween. El lavado de las perlas para los siguientes pasos del protocolo se realiza como sigue: agregar el tampón de lavado a las perlas, mezclar bien e incubar a temperatura ambiente girando durante 3 minutos. Colocar las perlas sobre un imán, esperar un minuto y desechar el sobrenadante.
20. Resuspender las perlas lavadas en 400 ul de NTB.
21. Combinar todo el ADN en un tubo y llevar el volumen a los 300 ul añadiendo agua destilada y combinarlo con 300 ul de perlas lavadas. Girar las perlas durante 15 minutos a temperatura ambiente.
22. Recuperar las perlas con un imán y desechar el sobrenadante. Lavar las perlas con 400 ul de NTB diluido dos veces y resuspender en 30 ul de tampón de resuspensión.
23. Añadir 2,5 ul de mezcla de ADN ligasa, 2,5 ul de tampón de resuspensión y 2,5 ul de índice de adaptador e incubar durante 10 minutos a 30 °C y añadir 5 ul de mezcla de parada de ligasa.
24. Limpiar el ADN usando perlas Ampure XP con una relación 1:1 de perlas/ADN según las instrucciones del kit de la biblioteca.
25. Realizar una PCR Illumina estándar de 9 ciclos según las instrucciones del kit de biblioteca y asegurarse de que haya al menos 1 ug de ADN ligado al adaptador.
26. Los fragmentos biotinilados originales se eliminan uniendo la muestra amplificada por PCR a 30 ul de perlas magnéticas de estreptavidina. Las perlas se incuban a temperatura ambiente durante 5 minutos y el sobrenadante se recoge en un tubo nuevo sin alterar las perlas.
27. 1 ug de ADN ligado al adaptador se hibrida con sondas de captura de secuencia según las instrucciones del fabricante (Roche Nimblegen Inc). Después de la hibridación, las sondas se lavan según las instrucciones del fabricante y se amplifican con 3-5 ciclos de PCR (Roche Nimblegen Inc). El material amplificado se secuenciará luego por pares usando la plataforma Illumina HiSeq. Es recomendable realizar una secuenciación superficial primero para verificar si la captura de secuencia funcionó antes de secuenciar la muestra en profundidad.
28. Se realiza un análisis adicional como se describió anteriormente.

REIVINDICACIONES

1. Un procedimiento que comprende: i) proporcionar un ADN genómico reticulado, donde el ADN se conserva de modo que esté intacto, donde el ADN comprende un primer y un segundo conjunto de regiones ii) fragmentar el genoma reticulado creando una pluralidad de fragmentos con uniones, iii) agregar un marcador de unión marcado y ligar los fragmentos con uniones y marcador en condiciones tales que el marcador esté ligado a las uniones; iv) purificar los fragmentos que contienen un marcador ligado en la unión; v) agregar sondas de captura marcadas y purificar selectivamente los fragmentos que se hibridan con las sondas de captura marcadas y vi) analizar los fragmentos que contienen un marcador ligado a la unión y que se hibridan con la sonda de captura marcada para determinar la identidad de los fragmentos, donde las sondas de captura marcadas son sondas que se hibridan con regiones reguladoras, como las secuencias de promotores.
2. El procedimiento según la reivindicación 1, donde la fragmentación se crea mediante enzimas de restricción.
3. El procedimiento según cualquiera de las reivindicaciones 1 y 2, donde el primer y el segundo conjunto de regiones están ligados entre sí.
4. El procedimiento según cualquiera de las reivindicaciones 1 a 3, donde el marcador de unión etiquetado está etiquetado con biotina.
5. El procedimiento según cualquiera de las reivindicaciones anteriores, donde el primer conjunto de regiones son regiones promotoras y el segundo conjunto de regiones comprende secuencias reguladoras, tales como potenciadores, silenciadores, aisladores, que se ubican cerca o distantes entre sí en el ADN o en los mismos o diferentes cromosomas.
6. El procedimiento según la reivindicación 5, donde dicho segundo conjunto de regiones comprende secuencias potenciadoras.
7. El procedimiento según cualquiera de las reivindicaciones anteriores, donde la primera región comprende un promotor o elemento regulador.
8. Un kit que comprende:
 - i) una enzima que eliminará los nucleótidos biotinilados de los extremos de fragmentos no ligados,
 - ii) un conjunto de enzimas para preparar bibliotecas de secuenciación para la secuenciación de alto rendimiento, que comprende una enzima que repara los extremos del ADN y una enzima y adaptadores de secuenciación para unir los adaptadores a los fragmentos,
 - iii) perlas de estreptavidina para seleccionar los fragmentos biotinilados,
 - iv) un conjunto de sondas de captura de secuencia para capturar solo los fragmentos complementarios a las regiones de interés, donde las sondas de captura se hibridan con regiones reguladoras, como las secuencias promotoras, y
 - v) productos químicos necesarios para realizar la reacción de captura de secuencia utilizando sondas de captura.
9. El kit según la reivindicación 8, donde el kit comprende un marcador de unión etiquetado y una sonda de captura etiquetada que están etiquetados con biotina.

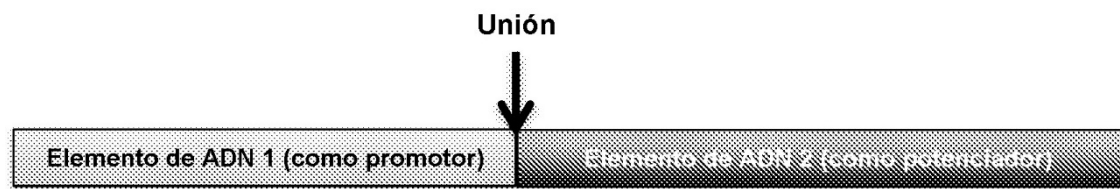


Fig 1

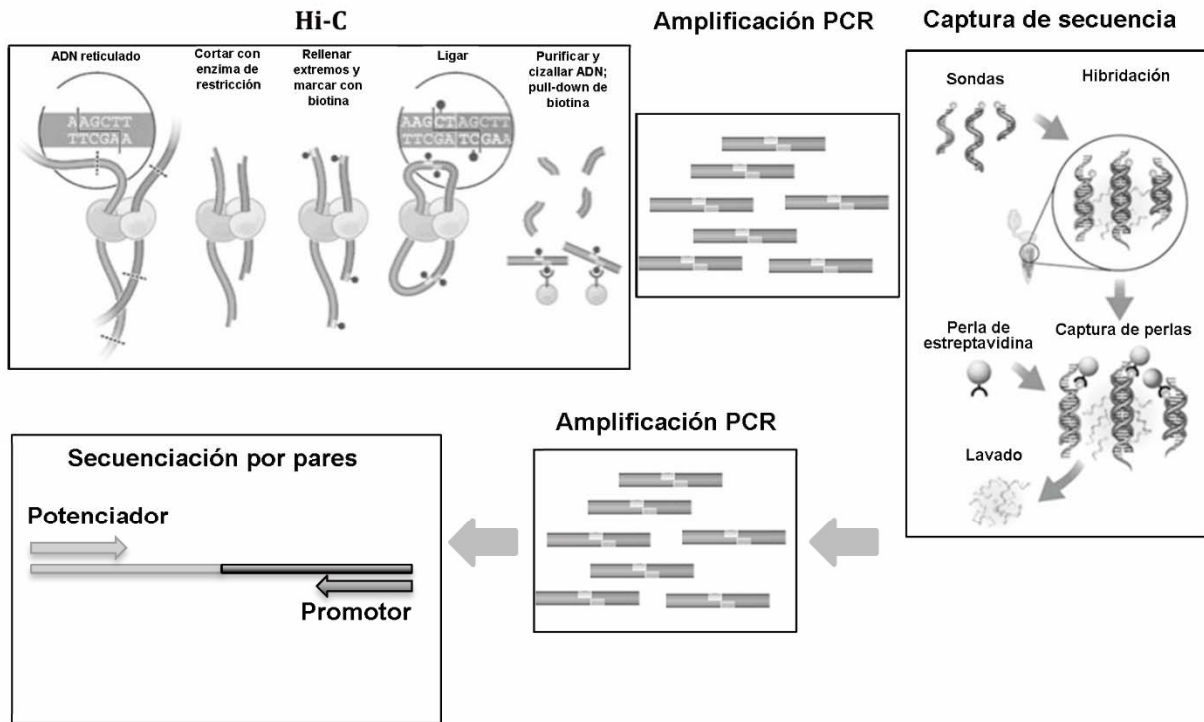


Fig 2

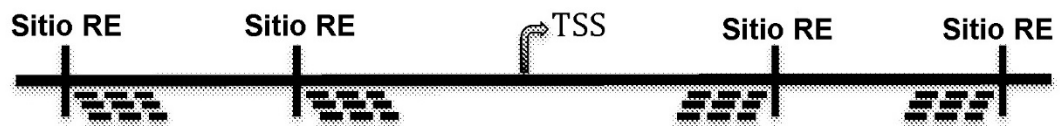


Fig 3

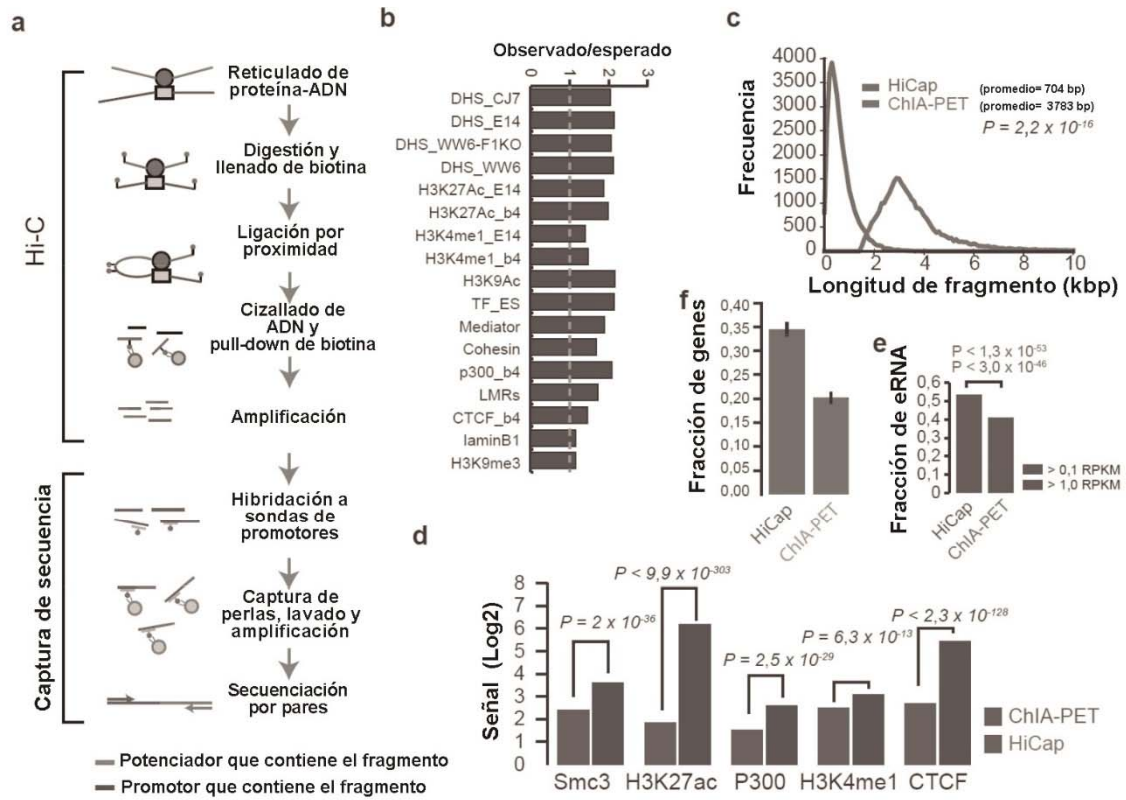


Fig 4

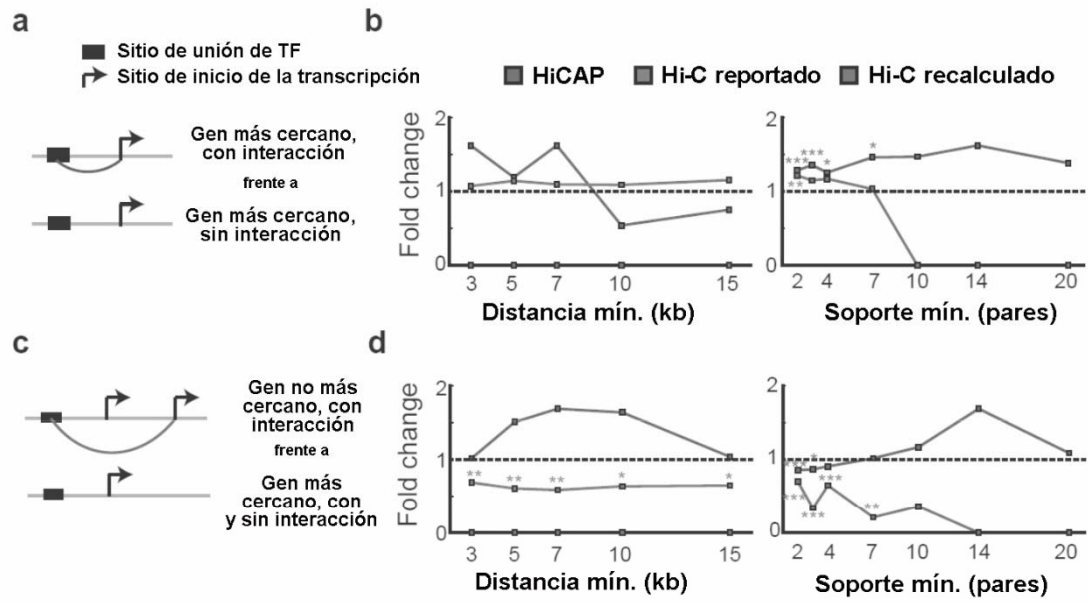


Fig 5

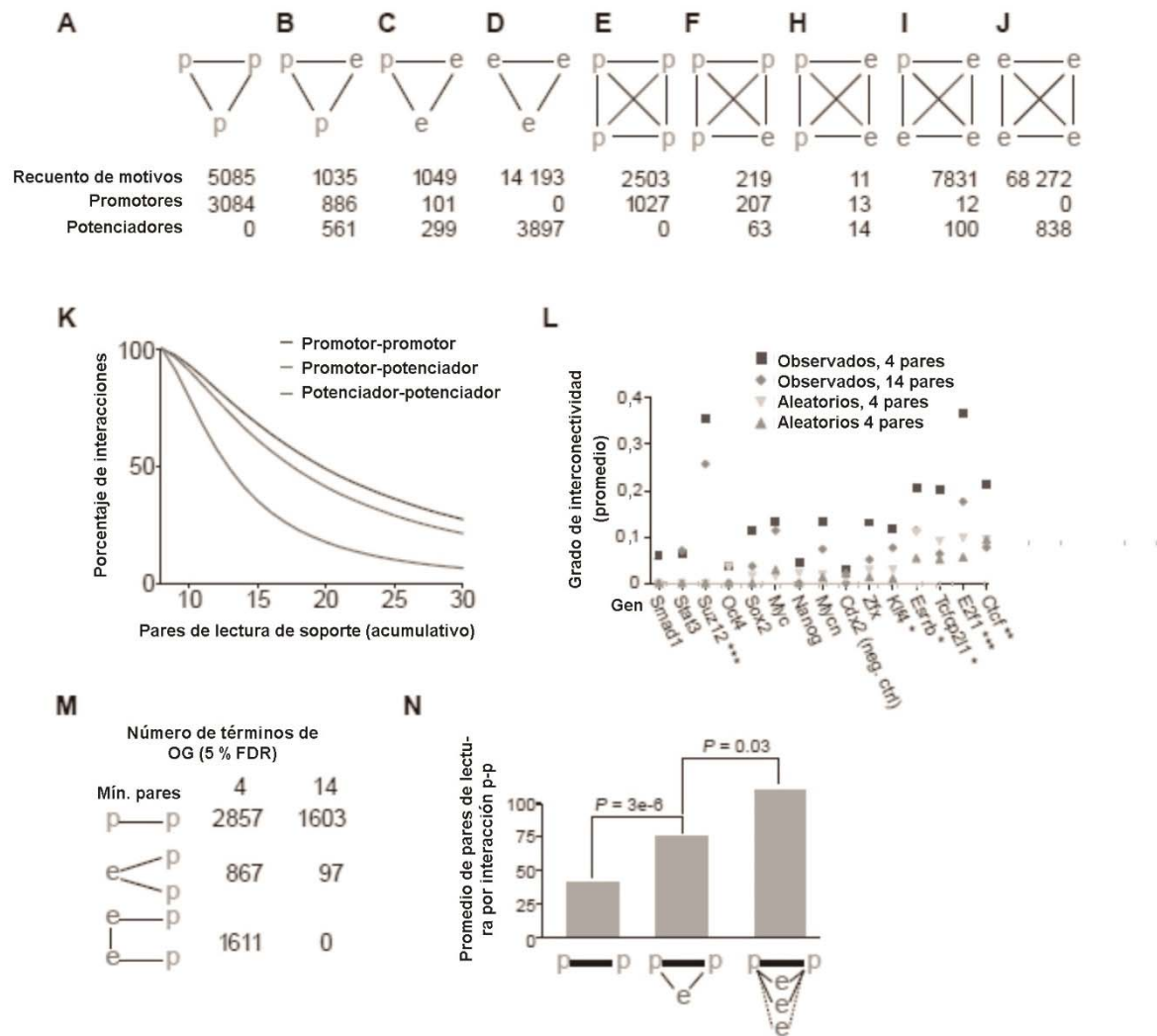


Fig 6

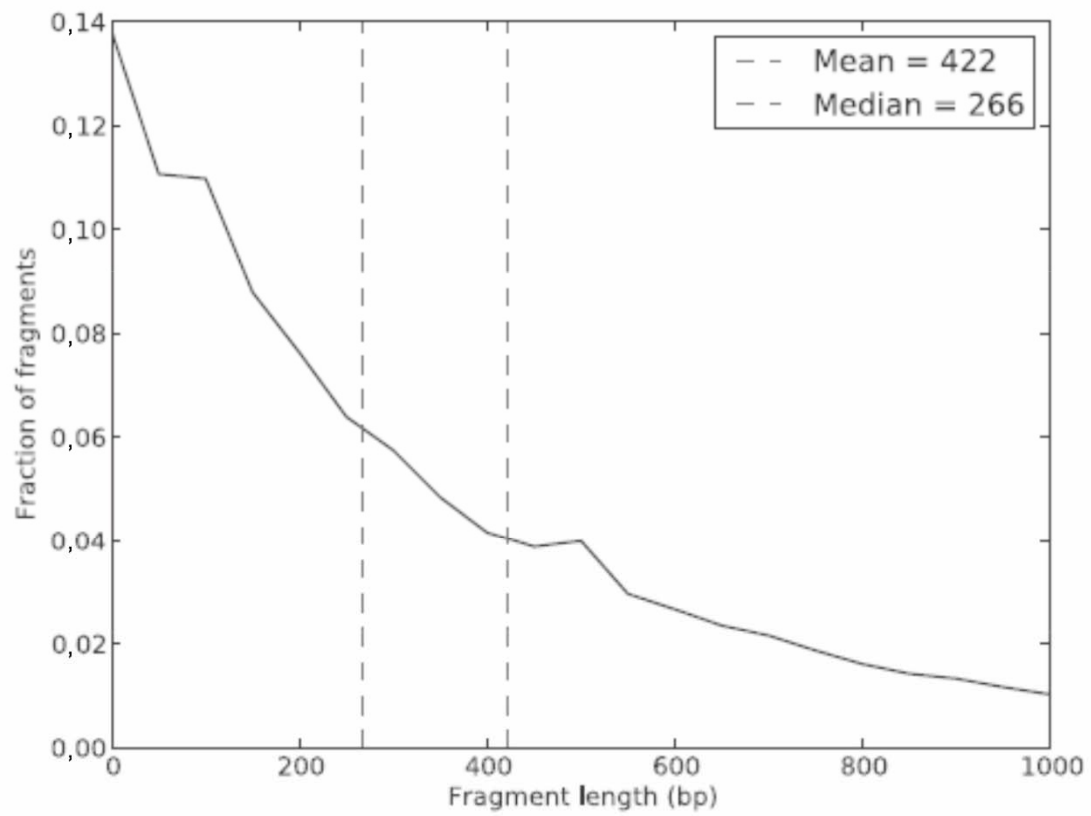


Fig 7

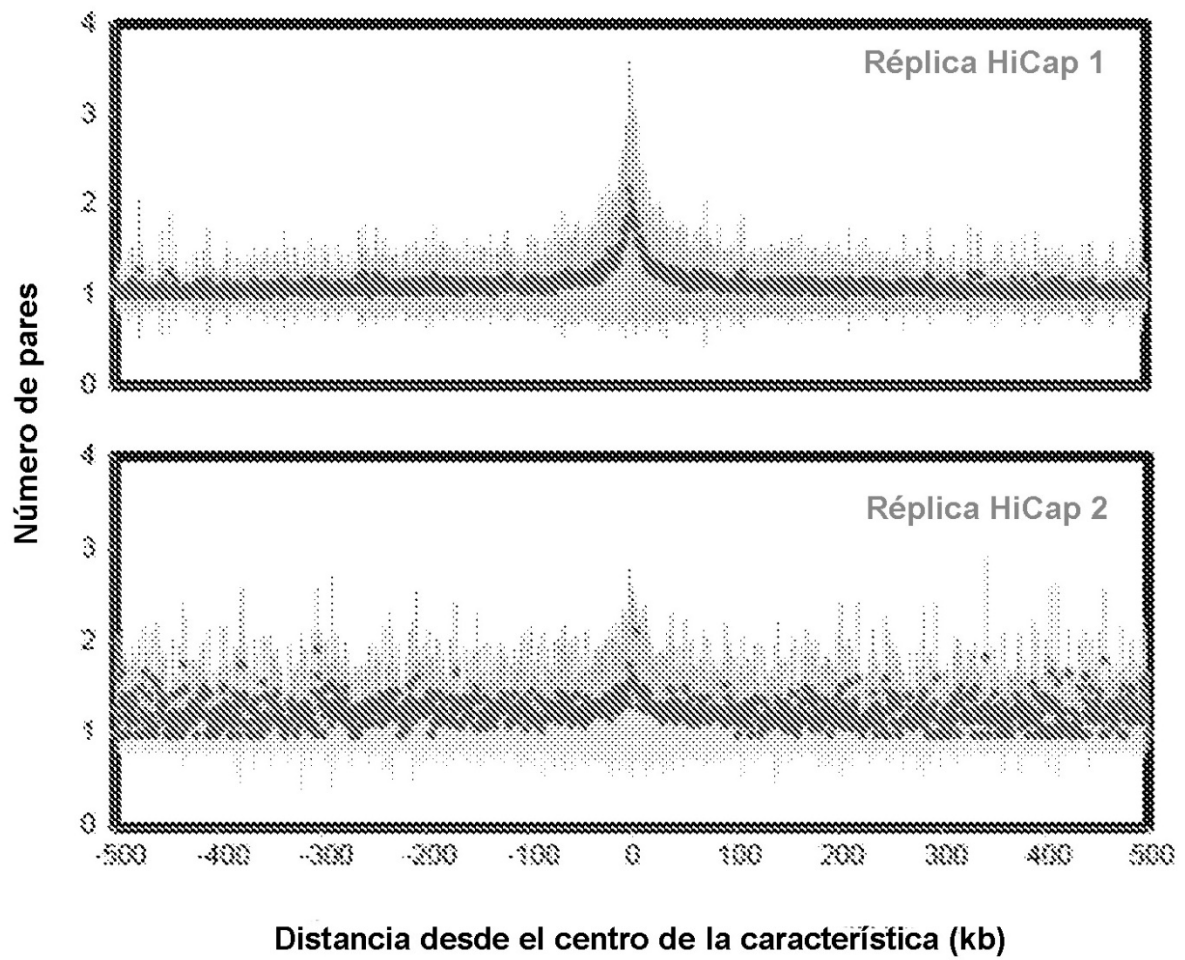


Fig 8

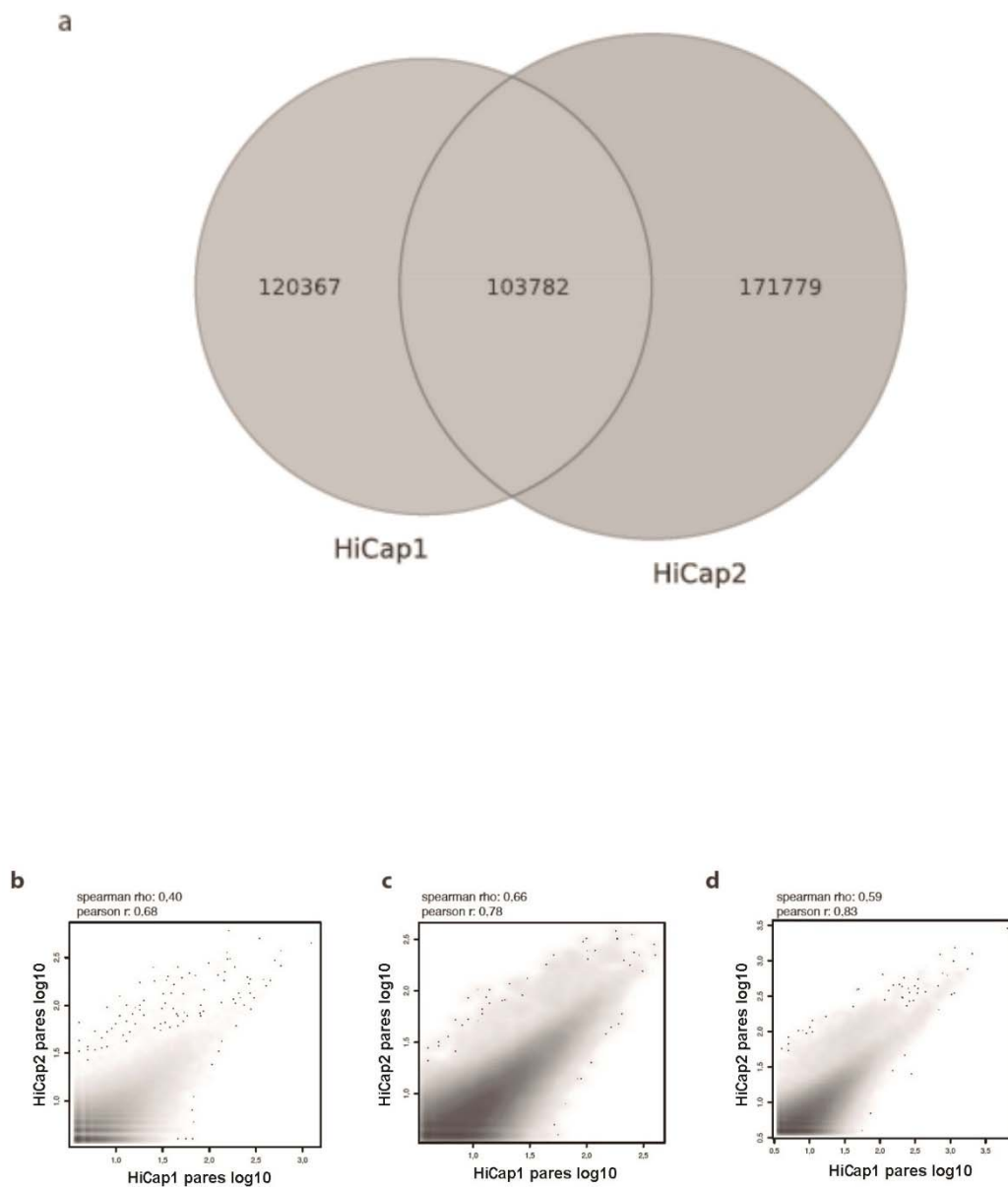


Fig 9

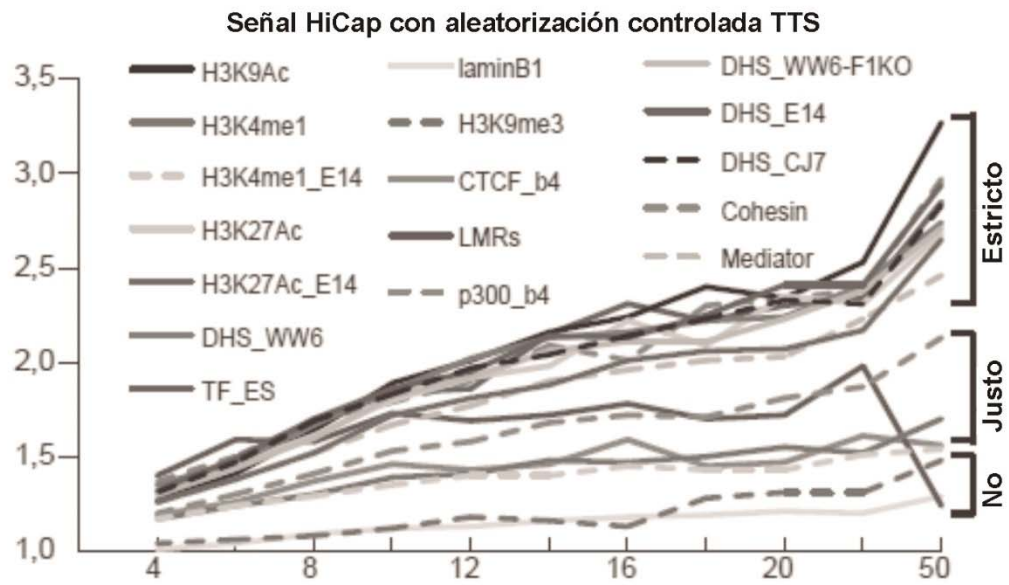


Fig 10

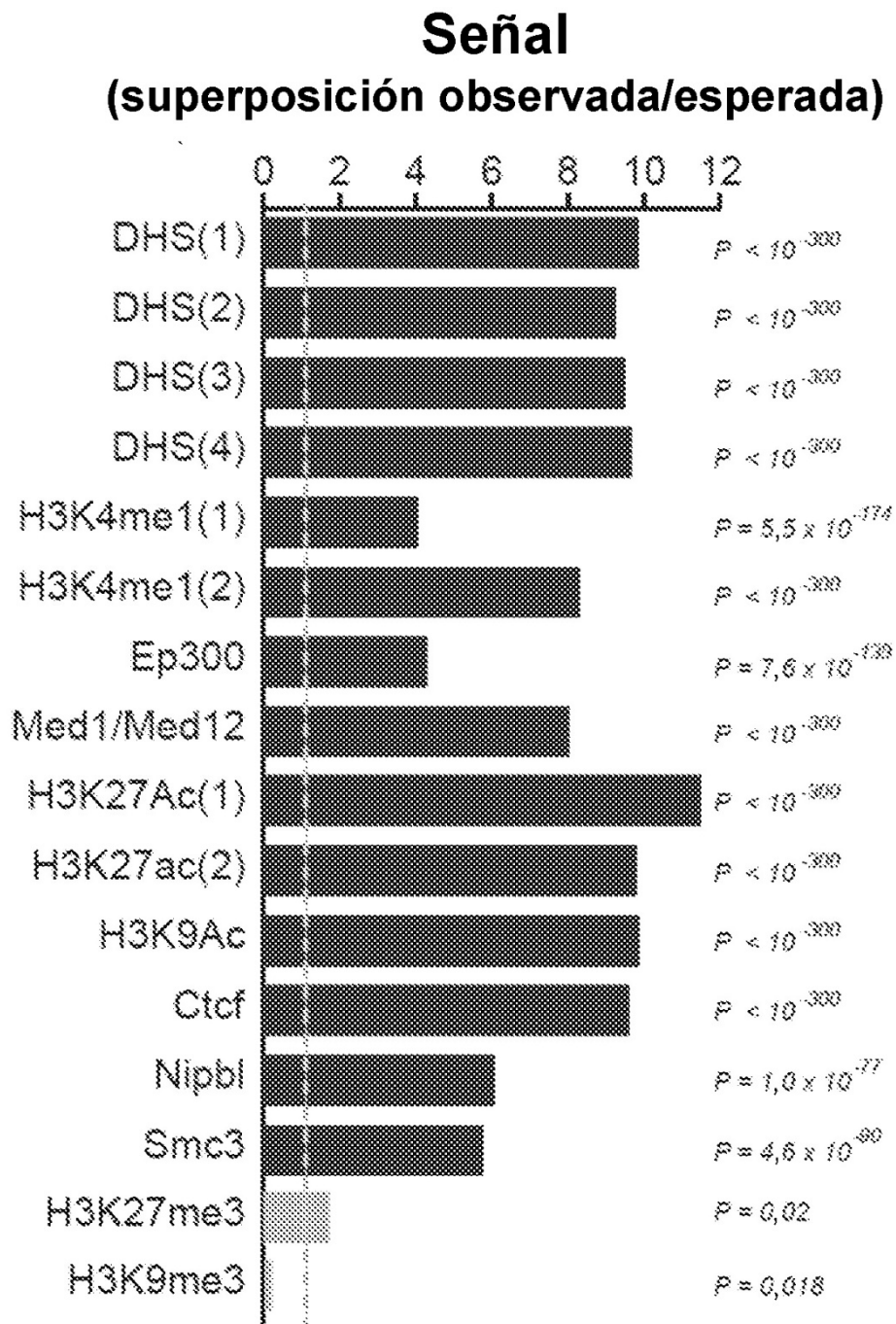


Fig 11

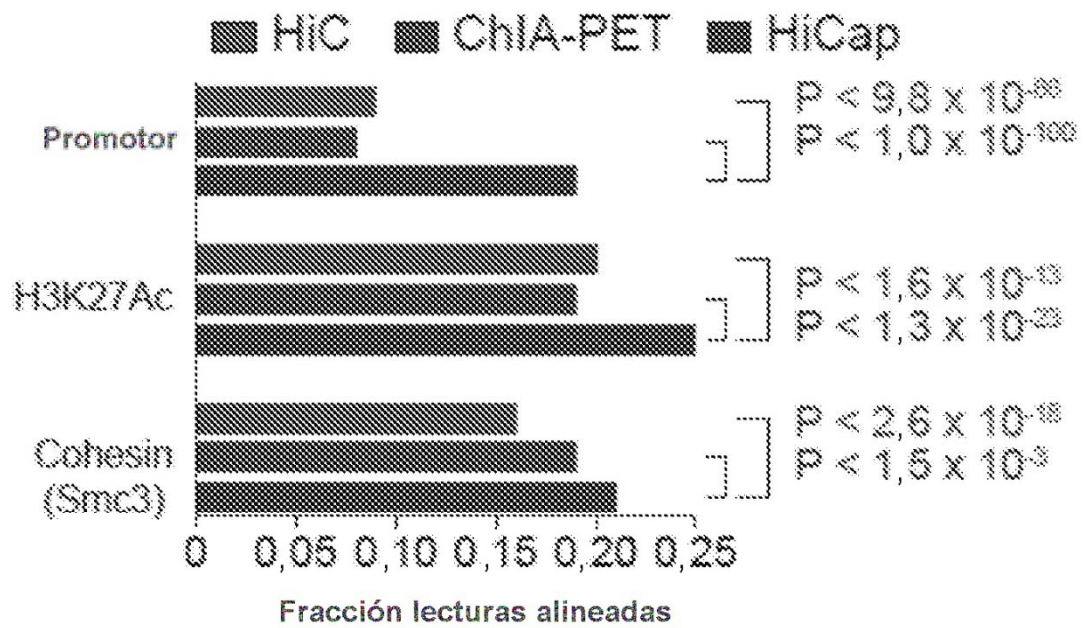


Fig 12

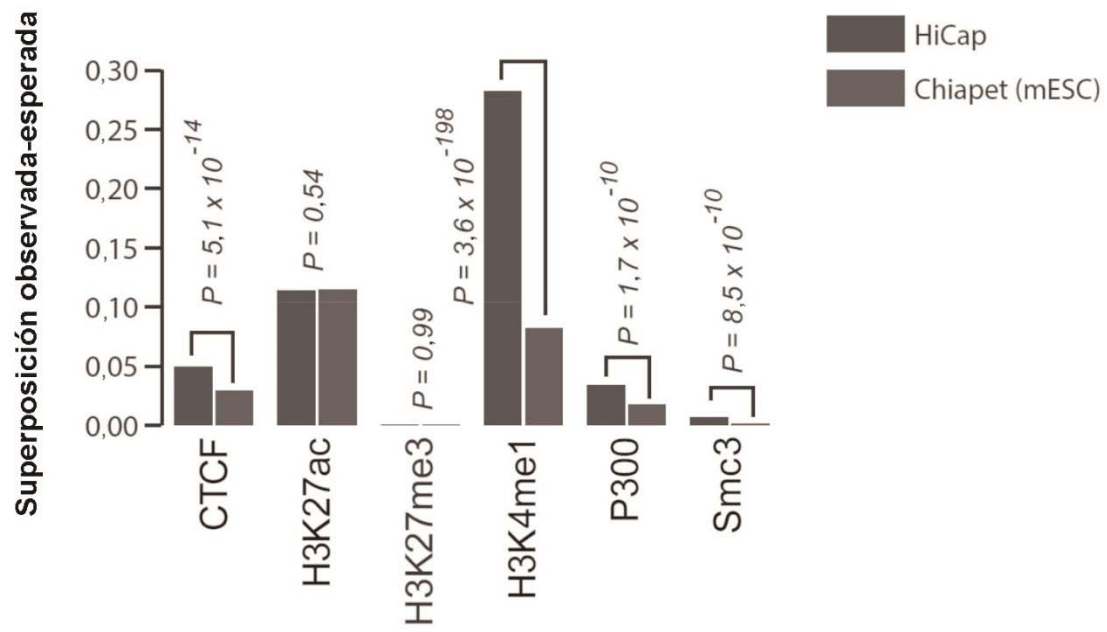


Fig 13

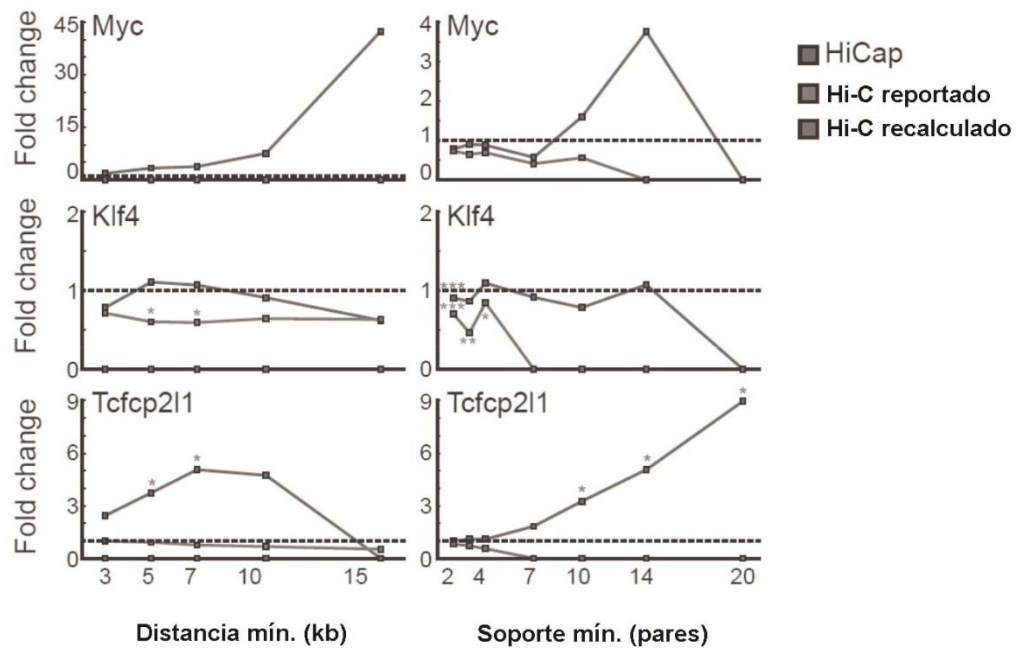


Fig 14