

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 949 801**

51 Int. Cl.:

C12N 5/10 (2006.01)

C12N 15/11 (2006.01)

C12N 15/67 (2006.01)

C07K 14/435 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **09.01.2018 PCT/US2018/013003**

87 Fecha y número de publicación internacional: **12.07.2018 WO18129544**

96 Fecha de presentación y número de la solicitud europea: **09.01.2018 E 18736585 (3)**

97 Fecha y número de publicación de la concesión europea: **12.04.2023 EP 3565891**

54 Título: **Métodos para alterar la expresión génica mediante la perturbación de multímeros de factores de transcripción que estructuran bucles reguladores**

30 Prioridad:

09.01.2017 US 201762444341 P

07.12.2017 US 201762596093 P

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

03.10.2023

73 Titular/es:

**WHITEHEAD INSTITUTE FOR BIOMEDICAL
RESEARCH (100.0%)**

**455 Main Street
Cambridge, MA 02142, US**

72 Inventor/es:

**YOUNG, RICHARD, A.;
WEINTRAUB, ABRAHAM, S.;
LI, CHARLES, H. y
SIGOVA, ALLA, A.**

74 Agente/Representante:

TORO GORDILLO, Ignacio

ES 2 949 801 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

Métodos para alterar la expresión génica mediante la perturbación de multímeros de factores de transcripción que estructuran bucles reguladores

5

Antecedentes de la invención

Los programas de expresión de genes específicos de tipo celular en seres humanos generalmente están controlados por elementos reguladores de genes llamados potenciadores (Buecker y Wysocka, 2012; Bulger y Groudine, 2011; Levine *et al.*, 2014; Ong y Corces, 2011; Ren y Yue, 2016). Los factores de transcripción (FT) se unen a estos elementos potenciadores y regulan la transcripción de los promotores de genes cercanos o distantes a través de contactos físicos que implican bucles de ADN entre potenciadores y promotores (Bonev y Cavalli, 2016; Fraser *et al.*, 2015; Heard y Bickmore, 2007; de Laat y Duboule, 2013; Pombo y Dillon, 2015; Spitz, 2016). A pesar de la importancia fundamental del control génico adecuado para la identidad y el desarrollo celular, las proteínas que contribuyen a las interacciones estructurales entre potenciadores y promotores son poco conocidas.

Existe evidencia considerable de que las interacciones potenciador-promotor pueden ser facilitadas por cofactores transcripcionales como mediador, complejos de proteínas de mantenimiento estructural de los cromosomas (SMC) tales como cohesina y proteínas de unión al ADN tales como CTCF. El mediador puede unir físicamente los factores de transcripción (FT) unidos al potenciador y el aparato de transcripción unido al promotor (Allen y Taatjes, 2015; Jeronimo *et al.*, 2016; Kagey *et al.*, 2010; Malik y Roeder, 2010; Petrenko *et al.*, 2016). La cohesina se carga en potenciadores y promotores activos por la proteína NIPBL asociada al mediador, y puede estabilizar transitoriamente las interacciones potenciador-promotor (Kagey *et al.*, 2010; Schmidt *et al.*, 2010). Las proteínas CTCF unidas a potenciadores y promotores pueden interactuar entre sí y, por lo tanto, pueden facilitar las interacciones potenciador-promotor (Guo *et al.*, 2015; Splinter *et al.*, 2006), pero CTCF generalmente no ocupa estos elementos que interactúan (Cuddapah *et al.*, 2009; Kim *et al.*, 2007; Phillips-Cremins *et al.*, 2013; Wendt *et al.*, 2008).

Las interacciones potenciador-promotor generalmente ocurren dentro de estructuras de bucle cromosómico más grandes formadas por la interacción de proteínas CTCF unidas a cada uno de los anclajes de bucle (Gibcus y Dekker, 2013; Gorkin *et al.*, 2014; Hnisz *et al.*, 2016a; Merkenschlager y Nora, 2016). Estas estructuras de bucle, llamadas TAD de diversas formas, dominios de bucle, dominios de contacto CTCF y vecindades aisladas, tienden a aislar potenciadores y genes dentro de los bucles CTCF-CTCF de elementos fuera de esos bucles (Dixon *et al.*, 2012; Downen *et al.*, 2014; Hnisz *et al.*, 2016b; Ji *et al.*, 2016; Lupiáñez *et al.*, 2015; Narendra *et al.*, 2015; Nora *et al.*, 2012; Phillips-Cremins *et al.*, 2013; Rao *et al.*, 2014; Tang *et al.*, 2015). La restricción de las interacciones de ADN dentro de las estructuras de bucle CTCF-CTCF de esta manera puede facilitar los contactos adecuados entre el potenciador y el promotor.

La evidencia de que las interacciones CTCF-CTCF desempeñan un papel global importante en las estructuras de los bucles cromosómicos, pero solo ocasionalmente están directamente involucradas en los contactos potenciador-promotor (Phillips y Corces, 2009), nos llevó a considerar la posibilidad de que una proteína puente análoga a CTCF podría participar en general en interacciones potenciador-promotor.

McClellan M.J. et al: "Modulation of Enhancer Looping and Differential Gene Targeting by Epstein-Barr Virus Transcription Factors Directs Cellular Reprogramming", PLOS PATHOGENS, vol. 9, n.º 9, (2013), página e1003636, muestra que las proteínas EBNA del virus de Epstein-Barr inhiben la expresión génica mediante la prevención del bucle potenciador-promotor y que las proteínas EBNA por lo tanto compiten con la unión a los sitios potenciadores de locus *CTBP2*, *ITGAL* y *WEE1*.

El documento WO 2014/071247 divulga el control de la expresión génica mediante la inhibición de la dimerización de c-Myc con Max y métodos de selección para la identificación de inhibidores de la dimerización. También se divulga el uso de compuestos que inhiben la actividad de Myc en el tratamiento de enfermedades oncogénicas y proliferativas.

El documento US 2009/082470 describe la inhibición de la dimerización del factor de transcripción STAT3, inhibiendo así la expresión génica regulada por STAT3. Los compuestos que interfieren con la dimerización de las moléculas STAT3 se describen como adecuados para el tratamiento de enfermedades vasculares tales como la degeneración macular relacionada con la edad.

El documento WO 2015/033293 divulga métodos para silenciar la expresión génica al interferir con el bucle promotor-potenciador del ADN, por ejemplo, induciendo una rotura de doble cadena en un sitio de unión de CTCF.

Benjamin Bonavida: "Therapeutic YY1 Inhibitors in Cancer; ALL in ONE", Critical reviews in oncogenesis, vol. 22, n.º 1-2, (2017), páginas 37-47, revisa los inhibidores de YY1 y su uso en la terapia del cáncer.

Sumario de la invención

En el presente documento se demuestra que el factor de transcripción YY1 actúa para estructurar interacciones de bucle entre potenciadores y promotores. YY1 es un factor de transcripción de dedo de zinc esencial y ampliamente

expresado que ocupa la mayoría de los potenciadores y promotores. YY1 estructura las interacciones de bucles potenciador-promotor, y la perturbación de la unión de YY1 interrumpe los bucles potenciador-promotor. YY1 puede estructurar bucles potenciadores-promotores mediante la multimerización (por ejemplo, dimerización) de moléculas YY1 unidas a dos elementos de ADN distantes. Dada la capacidad de otros factores de transcripción para formar

En el presente se divulga un método *in vitro* para modular la expresión de uno o más genes en una célula, que comprende modular la multimerización (por ejemplo, dimerización) de YY1 y, por lo tanto, modular la expresión de uno o más genes. En algunos aspectos, la multimerización (por ejemplo, dimerización) se modula con una composición que comprende un ácido nucleico y/o una molécula pequeña.

También se revela en el presente documento un método *in vitro* para modular la expresión de uno o más genes en una célula, que comprende modular la formación de un bucle de ADN potenciador-promotor en el genoma de la célula, en donde la formación depende de YY1 y en donde la formación es modulada por (a) la modulación de la unión de YY1 a la región promotora y/o potenciadora del bucle de ADN promotor-potenciador modificando la región promotora y/o potenciadora, o (b) modulando la multimerización de YY1 en la célula. En algunos aspectos, la multimerización (por ejemplo, dimerización) se modula poniendo en contacto la célula con un ácido nucleico, polipéptido y/o una molécula pequeña.

También se divulga en el presente documento una composición que modula la formación de bucles de ADN potenciadores-promotores, en donde la formación del bucle de ADN potenciador-promotor depende de YY1, para su uso en el tratamiento de una enfermedad o condición asociada con la actividad aberrante de una expresión génica en un sujeto, en donde la formación es modulada por (a) la modulación de la unión de YY1 a la región promotora y/o potenciadora del bucle de ADN promotor-potenciador modificando la región promotora y/o potenciadora, o (b) modulando la multimerización de YY1 en la célula. En algunos aspectos, la enfermedad o afección asociada con la expresión génica aberrante es cáncer.

También se divulga en el presente documento un método *in vitro* de detección de un compuesto que modula la expresión de uno o más genes en una célula, que comprende poner en contacto la célula con un agente de prueba, y medir la formación del bucle de ADN del promotor potenciador dependiente de YY1 en la célula, en donde el agente de prueba se identifica como un modulador de la expresión génica si el nivel de formación de bucles de ADN potenciador-promotor en la célula en contacto con el agente de prueba es diferente del nivel de formación de bucles de ADN potenciador-promotor en una célula de control no contactada con el agente de prueba.

También se divulga en el presente documento un método *in vitro* para identificar uno o más genes con expresión dependiente de un potenciador en una célula, que comprende identificar uno o más bucles de ADN promotor-potenciador dependientes de YY1 que comprenden el potenciador en la célula, e identificar uno o más genes expresados en el bucle de ADN promotor-potenciador, en donde el uno o más genes expresados en el bucle de ADN potenciador-promotor se identifican como genes cuya expresión depende del potenciador; opcionalmente, en donde la etapa de identificar uno o más bucles de ADN promotor-potenciador que comprenden el potenciador comprende realizar un ensayo ChIP-MS; y/o en donde el potenciador es un potenciador asociado a una enfermedad.

Breve descripción de los dibujos

El archivo de patente o solicitud contiene al menos un dibujo realizado a color. La Oficina proporcionará copias de esta publicación patente o de solicitud de patente con dibujos a color, previa solicitud y pago de la tasa correspondiente.

FIG. 1A-1H. YY1 es un modelo de factor estructurante del potenciador-promotor candidato (FIG. 1A) que representa un bucle de potenciador-promotor contenido dentro de un bucle de vecindad aislado más grande. Los factores de transcripción estructurantes potenciadores-promotores candidatos se identificaron mediante ChIP-MS de histonas con modificaciones características de la cromatina potenciadora y promotora. (FIG. 1B) Puntuaciones CRISPR (CS) de todos los genes en células KBM7 de Wang *et al.* (2015). Los factores de estructuración de potenciadores-promotores candidatos identificados por ChIP-MS se indican como puntos y los identificados como esenciales (CS < -1) se muestran en rojo. (FIG. 1C) Histograma que muestra el número de tejidos en los que cada factor de estructuración potenciador-promotor candidato se expresa en 53 tejidos examinados por GTEx. Los candidatos que se expresan ampliamente (expresados en más del 90 % de los tejidos examinados) y son esenciales para las células se muestran en rojo. (FIG. 1D) Análisis Metagene que muestra la ocupación de YY1 y CTCF en elementos potenciadores, promotores y aisladores en CME de ratón. (FIG. 1E) Resumen de las clases de interacciones de alta confianza identificadas por YY1 y CTCF ChIA-PET en células MER. (FIG. 1F) Ejemplo de una interacción potenciador-promotor YY1-YY1 en el locus Raf1 en células MER. (FIG. 1G) Modelo que representa el ensayo de coimmunoprecipitación para detectar la dimerización de YY1 y evaluar la dependencia del ARN para la dimerización de YY1. (FIG. 1H) Resultados de transferencia Western que muestran la co-inmunoprecipitación de la proteína YY1 marcada con FLAG y YY1 marcada con HA a partir de lisados nucleares preparados a partir de células transfectadas. La cuantificación de la señal restante normalizada a la entrada después del tratamiento con RNasa A para el YY1 marcado co-inmunoprecipitado se muestra debajo de las bandas relevantes. Véase también

la Tabla 11, FIG. 18. Véanse los métodos STAR para obtener una descripción detallada de los análisis genómicos. Los conjuntos de datos utilizados en esta figura se enumeran en la Tabla S4.

FIG. 2A-2D - El agotamiento de YY1 provoca la pérdida de interacciones potenciador-promotor. (FIG. 2A) YY1 ChIA-PET detecta interacciones entre el superpotenciador Sox2 (barras rojas) y el promotor Sox2. (FIG. 2B) Histograma que muestra los transcritos de Sox2 por célula determinados por FISH de molécula única para células infectadas con una pequeña horquilla dirigida a GFP (shGFP, n = 195 células) o YY1 (shYY1, n = 224 células) (FIG. 2C) YY1 ChIA-PET detecta interacciones entre el superpotenciador Oct4 (barras rojas) y el promotor Oct4. (FIG. 2D) Histograma que muestra las transcripciones de Oct4 por célula determinadas por FISH de molécula única para células infectadas con una pequeña horquilla dirigida a GFP (shGFP, n = 195 células) o YY1 (shYY1, n = 224 células).

FIG. 3A-3F. YY1 puede mejorar las interacciones del ADN *in vitro*. (FIG. 3A y FIG. 3D) Modelos que representan los ensayos de circularización de ADN *in vitro* utilizados para detectar la capacidad de YY1 para mejorar las interacciones de bucle de ADN. (FIG. 3B y FIG. 3E) Resultados del ensayo *in vitro* de circularización de ADN visualizado por electroforesis en gel. La banda inferior dominante refleja la plantilla de ADN lineal inicial, mientras que la banda superior corresponde al producto de ligadura de ADN circularizado. (FIG. 3C y FIG. 3F) Cuantificaciones de la circularización de la plantilla de ADN en función del tiempo de incubación con T4 ADN ligasa. Los valores corresponden al porcentaje de molde de ADN que se circulariza y representa la media y la desviación estándar de cuatro experimentos. Véase también la Figura 17.

FIG. 4A-4E -La pérdida de YY1 provoca la pérdida de las interacciones potenciador-promotor. (FIG. 4A) Pista génica para el gen Zfp518a que muestra datos de ChIA-PET, ChIP-seq y GRO-seq. El esquema representa el promotor y el potenciador. La secuencia del potenciador Zfp518a que fue objetivo de CRISPR se muestra con la secuencia de ARN guía resaltada en azul y la secuencia PAM resaltada en rojo [GCGTCGGCCATGACAGTTACATCCGGGTATGATGCCTAGC (SEQ ID NO: 2)]. En la parte inferior está la secuencia del mutante homocigoto obtenido después del direccionamiento de CRISPR [GCGTCG-GCCTGTGACATCCGGGTATGATGCCTAGC (SEQ ID NO: 3)] y analizado en la FIG. 4C a la FIG. 4E. También se muestra la secuencia de ARN guía [ACUGUCAAUGUAGGCCCAUA (SEQ ID NO: 1)] (FIG. 4B) El análisis 4C-seq detecta una frecuencia de interacción disminuida entre el potenciador Zfp518a y el promotor Zfp518a en la línea celular mutante. La línea gruesa indica la frecuencia media de interacción de dos experimentos biológicos repetidos. (FIG. 4C) ChIP-qPCR muestra una unión de YY1 disminuida en el potenciador Zfp518a en la línea celular mutante. (FIG. 4D) La RT-qPCR muestra una disminución de la expresión de Zfp518a en la línea celular mutante. (FIG. 4E) Cuantificación del cambio en la frecuencia de interacción (señal 4C-seq) entre el potenciador mutado y el promotor que se muestra en b (región en recuadro).

FIG. 5- Modelo de YY1 como factor estructurante potenciador-promotor. Modelo que representa YY1 (glóbulos rojos) que estructura el bucle de potenciador-promotor. El bucle potenciador-promotor está contenido dentro de una vecindad aislada que está estructurada por CTCF (glóbulos morados). Tanto YY1 como CTCF estructuran bucles de ADN mediante homodimerización.

FIG. 6A-6C. La delección de los sitios de unión de YY1 provoca la pérdida de interacciones potenciador-promotor: (FIG. 6A) Modelo que representa la delección mediada por CRISPR/Cas9 de un motivo de unión a YY1 en la región reguladora de un gen. (FIG. 6B y FIG. 6C) Delección mediada por CRISPR/Cas9 de motivos de unión a YY1 en las regiones reguladoras de dos genes, *Raf1* (FIG. 6B) y *Etv4* (FIG. 6C), y se midieron los efectos sobre la ocupación YY1, bucle potenciador-promotor y los niveles de ARNm. Las posiciones de los motivos de unión YY1 objetivo, se indica el genotipo de las líneas de tipo salvaje y mutante, y el punto de vista de 4C-seq. La señal media de 4C-seq se representa como una línea (las réplicas individuales se muestran en la FIG. 14) y el área sombreada representa el intervalo de confianza del 95 %. Se analizaron tres duplicados biológicos para experimentos 4C-seq y ChIP-qPCR, y seis duplicados biológicos para experimentos RT-qPCR. La SEQ ID NO: 4 en la FIG. 6B es una parte del gen *Raf1* de tipo salvaje. La SEQ ID NO: 5 en la FIG. 6B es una parte del gen *Raf1* mutado. La SEQ ID NO: 6 en la FIG. 6C es una parte del gen *Etv4* de tipo salvaje. La SEQ ID NO: 7 en la FIG. 6C es una parte del gen *Etv4* mutado. Las barras de error representan la desviación estándar. Todos los valores de p se determinaron utilizando la prueba t de Student. Véase también la Figura 13. Véanse los métodos STAR para obtener una descripción detallada de los análisis genómicos. Los conjuntos de datos utilizados en esta figura se enumeran en la Tabla S4.

FIG. 7A-7H. El agotamiento de YY1 interrumpe la expresión génica. (FIG. 7A) Modelo que representa el sistema dTAG utilizado para agotar rápidamente la proteína YY1. (FIG. 7B) Validación de transferencia Western de la activación de la etiqueta de degron FKBP y la capacidad para degradar induciblemente la proteína YY1. (FIG. 7C) Cambio en la expresión génica ($\log_2$ del factor de cambio) tras la degradación de YY1 para todos los genes representados frente a la expresión en células no tratadas. Los genes que mostraron cambios significativos en la expresión (valor de p ajustado por FDR < 0,05) están coloreados con los genes regulados al alza representados en rojo y los genes regulados a la baja representados en azul. (FIG. 7D) Mapas de calor que muestran el cambio en la expresión de cada gen tras la degradación de la señal ChIP-seq YY1 y YY1 de tipo salvaje en una región de ± 2 kb centrada en el TSS de cada gen. Cada fila representa un solo gen y los genes se clasifican por su valor de p ajustado para el cambio en la expresión tras la degradación de YY1. (FIG. 7E) Modelo que representa el esquema experimental para probar el efecto de la degradación de YY1 en la diferenciación de células madre embrionarias en las tres capas germinales a través de la formación de cuerpos embrioides a partir de células no tratadas (YY1) y células tratadas con compuesto dTAG para degradar YY1 (YY1⁻). (FIG. 7F) Imágenes microscópicas de cuerpos embrioides formados a partir de células YY1⁺ e YY1⁻. (FIG. 7G) Imágenes de inmunohistoquímica de cuerpos embrioides formados a partir de células YY1⁺ e YY1⁻. GATA4 se muestra en verde y el ADN teñido con DAPI se muestra en azul. La barra de escala representa 50 μ m. (FIG. 7H) Cuantificación de resultados de RNA-seq de una

sola célula para cuerpos embrioides formados a partir de células YY1⁺ e YY1⁻. El porcentaje de células que expresan varios genes específicos de diferenciación se muestra para cuerpos embrioides YY1⁺ e YY1⁻. Véase también la Tabla S3 y la FIG. 15. Véanse los métodos STAR para obtener una descripción detallada de los análisis genómicos. Los conjuntos de datos utilizados en esta figura se enumeran en la Tabla S4.

FIG. 8A-8E. El agotamiento de YY1 interrumpe el bucle potenciador-promotor. (FIG. 8A) Gráfica de dispersión que muestra para todas las interacciones potenciador-promotor YY1-YY1 el cambio en la frecuencia de interacción normalizada ($\log_2$ del factor de cambio) tras la degradación de YY1, medido por H3K27ac HiChIP y representado frente a la frecuencia de interacción normalizada en células no tratadas. (FIG. 8B) Cambio en la frecuencia de interacción normalizada ($\log_2$ del factor de cambio) tras la degradación de YY1 para tres clases diferentes de interacciones: todas las interacciones, interacciones no asociadas con los picos de YY1 ChIP-seq y las interacciones potenciador-promotor YY1-YY1. (FIG. 8C) Gráfico de dispersión que muestra para cada gen asociado con una interacción potenciador-promotor YY1-YY1 el cambio en la expresión génica ($\log_2$ del factor de cambio) tras la degradación de YY1 representada frente a la expresión en células no tratadas. Los genes que mostraron cambios significativos en la expresión (valor de p ajustado por FDR < 0,05) están coloreados con los genes regulados al alza representados en rojo y los genes regulados a la baja representados en azul. (FIG. 8D y FIG. 8E) Efecto de la degradación de YY1 en el locus *Slc7a5* (FIG. 8D) y el locus *Klf9* (FIG. 8E) en interacciones potenciador-promotor y la expresión génica. La parte superior de cada panel muestra un arco que representa una interacción potenciador-promotor detectada en los datos de HiChIP. La señal en los píxeles delineados se usó para cuantificar el cambio en la frecuencia de interacción normalizada tras la degradación de YY1. Se analizaron tres duplicados biológicos por afección para H3K27ac HiChIP y dos duplicados biológicos se analizaron para RNA-seq. Las barras de error representan la desviación estándar. Los valores de p para HiChIP se determinaron utilizando la prueba t de Student. Los valores P para RNA-seq se determinaron mediante una prueba de Wald. Véase también la Figura 14. Véanse los métodos STAR para obtener una descripción detallada de los análisis genómicos. Los conjuntos de datos utilizados en esta figura se enumeran en la Tabla S4.

FIG. 9A-9F. Rescate de interacciones potenciador-promotor en células. (FIG. 9A) Modelo que representa el uso de dCas9-YY1 para unir artificialmente YY1 a un sitio adyacente a la mutación del sitio de unión de YY1 en la región proximal del promotor de *Etv4* para determinar si YY1 unido artificialmente puede rescatar interacciones potenciador-promotor. (FIG. 9B) Modelo que representa los experimentos de rescate de dCas9-YY1. Las células mutantes del motivo de unión YY1 proximal al promotor de *Etv4* se transdujeron con lentivirus para expresar de manera estable dCas9 o dCas9-YY1, y dos ARNGs para dirigir su localización a las secuencias adyacentes al motivo de unión YY1 eliminado en la región proximal del promotor de *Etv4*. La capacidad para rescatar el bucle promotor potenciador se ensayó mediante 4C-seq. (FIG. 9C) Resultados de transferencia Western que muestran que las células mutantes del motivo de unión YY1 proximal al promotor de *Etv4* transducidas con lentivirus para expresar de forma estable dCas9 o dCas9-YY1 expresan con éxito dCas9 o dCas9-YY1. (FIG. 9D) El anclaje artificial de YY1 usando dCas9-YY1 se realizó en sitios adyacentes a la mutación del sitio de unión de YY1 en la región proximal del promotor de *Etv4*. Los efectos de anclar YY1 usando dCas9-YY1 en el bucle potenciador-promotor y la expresión del *Etv4* se midió el gen y se comparó con dCas9 solo. Se muestran el genotipo de las células mutantes del motivo de unión YY1 proximal al promotor de *Etv4* y el punto de vista 4C-seq (VP). La señal 4C-seq se muestra como el promedio suavizado de lecturas por millón por par de bases. La señal media de 4C-seq se representa como una línea y el área sombreada representa el intervalo de confianza del 95 %. Se analizaron tres duplicados biológicos para experimentos 4C-seq y CAS9 ChIP-qPCR, y seis duplicados biológicos para experimentos RT-qPCR. Las barras de error representan la desviación estándar. Todos los valores de p se determinaron utilizando la prueba t de Student. (FIG. 9E) Modelo que representa la pérdida de interacciones de bucle después de la degradación inducible de los factores de estructuración CTCF y YY1 seguido de la restauración del bucle tras el lavado de los compuestos de degradación. (FIG. 9F) Cambio en la frecuencia de interacción normalizada ($\log_2$ factor de cambio) después de la degradación de YY1 y CTCF (tratadas) y la recuperación (lavado) en relación con las células no tratadas. Para la degradación de YY1, el cambio en la frecuencia de interacción normalizada se representa gráficamente para las interacciones potenciador-promotor YY1-YY1. Para la degradación de CTCF, el cambio en la frecuencia de interacción normalizada se representa gráficamente para las interacciones CTCF-CTCF. Véase también la Figura 14. Véanse los métodos STAR para obtener una descripción detallada de los análisis genómicos. Los conjuntos de datos utilizados en esta figura se enumeran en la Tabla S4.

FIG. 10A-10K. YY1 generalmente ocupa potenciadores y promotores en células de mamíferos. (FIG. 10A-FIG. 10B) Mapas de calor que muestran la ocupación YY1 en potenciadores (FIG. 10A) y promotores activos (FIG. 10B) en seis tipos de células humanas. (FIG. 10C-FIG. 10E) Resúmenes de las principales clases de interacciones de alta confianza identificadas con YY1 HiChIP en tres tipos de células humanas. (FIG. 10F-FIG. 10K) Ejemplos de interacciones potenciador-promotor YY1-YY1 en tres tipos de células humanas: cáncer colorrectal (FIG. 10F y FIG. 10I), leucemia linfoblástica aguda de linfocitos T (FIG. 10G y FIG. 10J) y leucemia mieloide crónica (FIG. 10H y FIG. 10K). Los ejemplos mostrados muestran interacciones potenciador-promotor YY1-YY1 que implican potenciadores típicos (FIG. 10F-FIG. 10H) y superpotenciadores (FIG. 10I-FIG. 10K). Véase también la Figura 12. Véanse los métodos STAR para obtener una descripción detallada de los análisis genómicos. Los conjuntos de datos utilizados en esta figura se enumeran en la Tabla S4.

FIG. 11A-11C - YY1 y CTCF se expresan de forma ubicua y son esenciales. (FIG. 11A) Transcripción de lecturas por millón (RPKM) para YY1 en una variedad de tipos de células y tejidos humanos primarios. (FIG. 11B) Transcripción de lecturas por millón (RPKM) para CTCF en una gama de tipos de células y tejidos humanos primarios. (FIG. 11C) Puntuaciones CRISPR de una pantalla CRISPR de todo el genoma en células KBM7 para

YY1 y CTCF.

FIG. 12 -YY1 se multimeriza *in vivo*. Coinmunoprecipitación (Co-IP) de construcciones YY1 marcadas con FLAG y HA muestra que YY1 se dimeriza *in vivo*.

FIG. 13A-13B. La pérdida de la unión de YY1 provoca la pérdida de las interacciones potenciador-promotor, relacionado con la figura 6. (FIG. 13A y FIG. 13B) Deleción mediada por CRISPR/Cas9 de motivos de unión a YY1 en las regiones reguladoras de dos genes, *Raf1* (FIG. 13A) y *Etv4* (FIG. 13B). La parte superior de cada panel muestra una interacción potenciador-promotor YY1-YY1 de alta confianza y perfiles de unión de ChIP-seq para YY1 y H3K27ac que se muestran como lecturas por millón por par de bases. Se muestran la posición del motivo de unión al ADN YY1 objetivo y el genotipo de las líneas de tipo salvaje y mutante. La parte inferior de cada panel muestra los perfiles de interacción de cromatina en células de tipo salvaje y mutantes ancladas en el punto de vista indicado (VP) para tres duplicados biológicos. La señal 4C-seq se muestra como lecturas suavizadas por millón por par de bases. La SEQ ID NO: 4 en la FIG. 13A es una parte del gen *Raf1* de tipo salvaje. La SEQ ID NO: 5 en la FIG. 13A es una parte del gen *Raf1* mutado. La SEQ ID NO: 6 en la FIG. 13B es una parte del gen *Etv4* de tipo salvaje. La SEQ ID NO: 7 en la FIG. 13B es una parte del gen *Etv4* mutado. Las fuentes de los conjuntos de datos utilizados en esta figura se enumeran en la Tabla S4.

FIG. 14A-14E. El agotamiento de YY1 interrumpe el bucle potenciador-promotor, relacionado con la figura 8. (FIG. 14A y FIG. 14B) Resúmenes de las principales clases de interacciones de alta confianza identificadas por YY1 ChIA-PET (FIG. 14A) y H3K27ac HiChIP (FIG. 14B). Las interacciones se clasifican en función de la presencia de elementos potenciadores, promotores y aisladores en los anclajes de cada interacción. Las interacciones se muestran como arcos entre estos elementos y el grosor de los arcos refleja aproximadamente el porcentaje de interacciones de esa clase en relación con el número total de interacciones que se clasificaron. (FIG. 14C) Porcentaje de picos YY1 ChIP-seq en células ME que están asociados con interacciones potenciador-promotor, asociado con interacciones no potenciador-promotor, y no asociado con una interacción detectada para interacciones de alta confianza identificadas por YY1 ChIA-PET y H3K27ac HiChIP. (FIG. 14D) Porcentaje de genes que aumentan significativamente en expresión, disminuyen significativamente en la expresión, o no se expresan diferencialmente en respuesta a la degradación de YY1 para tres clases de genes: todos los genes, genes implicados en interacciones potenciador-promotor que no tienen picos YY1 en ambos extremos, y genes implicados en interacciones potenciador-promotor YY1-YY1. (FIG. 14E) Expresión de los genes *Raf1* y *Etv4* antes (0 h) y después de la degradación de YY1 (24 h) medida por RNA-seq. Las fuentes de los conjuntos de datos utilizados en esta figura se enumeran en la Tabla S4.

FIG. 15A-15D. El agotamiento de YY1 afecta a la diferenciación de células ME, relacionado con la figura 7. (FIG. 15A) Modelo que representa la diferenciación de células ME pluripotenciales en células de las tres capas germinales. Se indican los marcadores específicos de pluripotencia y diferenciación que se examinaron. (FIG. 15B) Imágenes inmunohistoquímicas de cuerpos embrioides formados a partir de células no tratadas (YY1+) y células tratadas con compuesto dTAG para degradar YY1 (YY1-). GFAP y TUBB3, que se expresan en células pertenecientes al linaje del ectodermo se muestran en verde y rojo, respectivamente. El ADN teñido con DAPI se muestra en azul. (FIG. 15C) Representación basada en análisis de componentes principales (PCA) de datos de secuencia de ARN de una sola célula para cuerpos embrioides formados a partir de células no tratadas (YY1+) y células tratadas con compuesto dTAG para degradar YY1 (YY1-). Cada punto representa una sola celda y los puntos se organizan en función de PCA. Las células de los cuerpos embrioides YY1+ se muestran en beige y las células de los cuerpos embrioides YY1 se muestran en azul. (FIG. 15D) Expresión de genes específicos de diferenciación y pluripotencia (FIG. 15A) medidos por ARN-seq de una sola célula de cuerpos embrioides formados a partir de células no tratadas (YY1+) y células tratadas con compuesto dTAG para degradar YY1 (YY1-). Cada punto representa una sola celda y los puntos están sombreados según su valor de expresión normalizado. Las fuentes de los conjuntos de datos utilizados en esta figura se enumeran en la Tabla S4.

FIG. 16A-16C. Rescate de interacciones potenciador-promotor en células, relacionado con la figura 9. (FIG. 16A) Modelo que representa el sistema dTAG utilizado para degradar rápidamente la proteína YY1. La etiqueta FKBP degenon se introdujo en ambos alelos del locus del gen *Yy1* endógeno. La adición del compuesto dTAG da como resultado el reclutamiento de la ligasa cereblon E3 a la proteína YY1 etiquetada con degenon FKBP, dando como resultado una rápida degradación mediada por proteasomas. Los efectos de la degradación de YY1 se examinaron 24 horas después del tratamiento con el compuesto dTAG. El lavado del compuesto dTAG durante 5 días permitió la recuperación de la proteína YY1. (FIG. 16B) Validación de transferencia Western de la degradación de YY1 después de 24 horas de tratamiento con compuesto dTAG y recuperación de YY1 después de 5 días de lavado del compuesto dTAG. (FIG. 16C) Modelo que representa el sistema de degradación AID utilizado para degradar rápidamente la proteína CTCF en Nora *et al.* (2017). El marcador AID se introdujo en el locus del gen *Ctcf* endógeno. La adición de auxina da como resultado el reclutamiento de la ligasa TIR1 E3 a la proteína CTCF marcada con AID, dando como resultado una degradación mediada por proteasomas. Los efectos de la degradación de CTCF se examinaron 48 horas después del tratamiento con el compuesto dTAG. El lavado de auxina durante 2 días permitió la recuperación de la proteína CTCF.

FIG. 17A-17B. YY1 puede mejorar las interacciones del ADN *in vitro*, relacionada con la Figura 3A-3F. (FIG. 17A) La pureza de la proteína His6-YY1 recombinante se validó mediante electroforesis en gel del material purificado seguido de tinción con azul de Coomassie y análisis de transferencia Western con anticuerpo anti-YY1. (FIG. 17B) La actividad de la proteína YY1 recombinante purificada fue validada por EMSA. El YY1 purificado se incubó con una sonda de ADN biotinilado en presencia o ausencia de un ADN competidor no biotinilado. La actividad de la proteína recombinante se evaluó por la capacidad de unirse al ADN y se determinó por resolución en un gel nativo. La sonda biotinilada "libre" no unida se encuentra en la parte inferior del gel, mientras que la sonda unida por YY1

migra más lentamente y aparece como una banda más alta. La adición de ADN competidor anula este efecto, lo que indica que la actividad es específica.

FIG. 18A-18J. Las interacciones asociadas con YY1 conectan potenciadores y promotores, relacionado con la figura 1. (FIG. 18A) Mapa de calor que muestra YY1, H3K27ac, y señal CTCF ChIP-seq y señal GRO-seq en los promotores, potenciadores y aisladores en células madre embrionarias de ratón (células MEr). La señal de ChIP-seq y GRO-seq se representa como lecturas por millón por par de bases en una región de ± 2 kb centrada en cada promotor, potenciador y aislador. (FIG. 18B) Análisis metagénico ampliado que muestra la ocupación de YY1 y CTCF en los elementos potenciadores, promotores y elementos aisladores en células MEr. Adicionalmente, la ocupación de YY1 se trazó en los picos de YY1 que no se clasificaron como potenciadores, promotor o aislador, y la ocupación de CTCF se representó en los picos de CTCF que no se clasificaron como potenciadores, promotor o aislador. Los perfiles de ChIP-seq se muestran como lecturas medias por millón por par de bases para elementos de cada clase en una región de ± 2 kb centrada en cada región. El número de potenciadores, promotores y aisladores examinados se indican entre paréntesis. Para facilitar las comparaciones del mismo factor entre diferentes regiones, se cuantificó la señal total de ChIP-seq en la región y se muestra en la esquina superior derecha de la gráfica para cada análisis de metagen. (FIG. 18C) Análisis Metagene que muestra la señal GRO-seq y la señal H3K27ac ChIP-seq en los picos YY1 y CTCF en células MEr que no se clasificaron como parte de un potenciador, promotor o aislador. Los perfiles de ChIP-seq se muestran como lecturas medias por millón por par de bases para elementos de cada clase en una región de ± 2 kb centrada en cada región. El número de picos de YY1 y CTCF examinados se indica entre paréntesis. Para facilitar las comparaciones del mismo factor entre diferentes regiones, se cuantificó la señal total de ChIP-seq en la región y se muestra en la esquina superior derecha de la gráfica para cada análisis de metagen. (FIG. 18D) Resumen ampliado de las principales clases de interacciones de alta confianza identificadas en los conjuntos de datos YY1 y CTCF ChIA-PET presentados en la FIG. 1E. Las interacciones se clasifican en función de la presencia de elementos potenciadores, promotores y aisladores en los anclajes de cada interacción. Las interacciones se muestran como arcos entre estos elementos y el grosor de los arcos refleja aproximadamente el porcentaje de interacciones de esa clase en relación con el número total de interacciones que se clasificaron. (FIG. 18E) Un ejemplo de amplias interacciones potenciador-promotor asociadas con YY1. Las interacciones YY1 de alta confianza se representan como arcos rojos, mientras que las interacciones CTCF de alta confianza se representan como arcos azules. Perfiles de unión de ChIP-seq para YY1, CTCF, H3K27ac y la señal GRO-seq trenzada se muestran como lecturas por millón por par de bases en el locus *Klf9* en células MEr. El gen *Klf9* se indica en el modelo de gen y los superpotenciadores que interactúan se marcan bajo la pista H3K27ac ChIP-seq. (FIG. 18F) Análisis Metagene que muestra la ocupación de YY1 en constituyentes potenciadores típicos y constituyentes superpotenciadores. Los perfiles de ChIP-seq se muestran como lecturas medias por millón por par de bases para elementos de cada clase en una región de ± 2 kb centrada en cada región. Para facilitar las comparaciones del mismo factor entre diferentes regiones, se cuantificó la señal total de ChIP-seq en la región y se muestra en la esquina superior derecha de la gráfica para cada análisis de metagen. El número de elementos examinados se enumera en la parte superior de la gráfica. Ambos gráficos se basan en la cantidad mínima de señal constituyente del potenciador típico. (FIG. 18G) Mapas de calor que muestran para cada interacción YY1 de alta confianza la cantidad de PET que respaldan la interacción, para interacciones que tienen al menos un ancla superpuesta a un superpotenciador (izquierda) y para interacciones que no tienen extremos superpuestos a un superpotenciador (derecha). Cada fila representa una interacción y la intensidad del color de cada fila representa el recuento de PET para esa interacción. (FIG. 18H) Diagrama de caja que muestra los recuentos PET de interacciones YY1 ChIA-PET de alta confianza que no están asociadas con superpotenciadores o asociadas con superpotenciadores. (FIG. 18I) Modelo que representa el ensayo de coimmunoprecipitación para detectar la dimerización de YY1. (FIG. 18J) Resultados de transferencia Western que muestran la co-inmunoprecipitación de la proteína YY1 marcada con FLAG e YY1 marcada con HA a partir de lisados nucleares preparados a partir de células transfectadas. Se observa interacción entre la proteína YY1 etiquetada con FLAG y YY1 marcada con HA, mientras que no se observa interacción con la proteína OCT4. Las fuentes de los conjuntos de datos utilizados en esta figura se enumeran en la Tabla S4.

Descripción detallada de la invención

La práctica de la presente invención normalmente empleará, a menos que se indique otra cosa, técnicas convencionales de biología celular, cultivo celular, biología molecular, biología transgénica, microbiología, tecnología de ácidos nucleicos recombinantes (por ejemplo, ADN), inmunología y ARN de interferencia (ARNi) que están dentro de la experiencia en la técnica. En las siguientes publicaciones se encuentran descripciones no limitantes de algunas de estas técnicas: Ausubel, F., *et al.*, (eds.), *Current Protocols in Molecular Biology*, *Current Protocols in Immunology*, *Current Protocols in Protein Science* y *Current Protocols in Cell Biology*, todos John Wiley & Sons, N.Y., edición a diciembre de 2008; Sambrook, Russell y Sambrook, *Molecular Cloning: A Laboratory Manual*, 3.^a ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2001; Harlow, E. y Lane, D., *Antibodies - A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1988; Freshney, R.I., "Culture of Animal Cells, A Manual of Basic Technique", 5.^a ed., John Wiley & Sons, Hoboken, NJ, 2005. Se encuentra información no limitante con respecto a agentes terapéuticos y enfermedades humanas en Goodman and Gilman's *The Pharmacological Basis of Therapeutics*, 11.^a Ed., McGraw Hill, 2005, Katzung, B. (ed.) *Basic and Clinical Pharmacology*, McGraw-Hill/Appleton & Lange; 10.^a ed. (2006) u 11.^a edición (julio de 2009). Información no limitante sobre genes y trastornos genéticos se encuentra en McKusick, V.A.: *Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders*. Baltimore: Johns Hopkins University Press, 1998 (12.^a edición) o la base de datos en línea más reciente: Online

Mendelian Inheritance in Man, OMIM™. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) y National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), a partir del 1 de mayo de 2010, ncbi.nlm.nih.gov/omim/ y en Online Mendelian Inheritance in Animals (OMIA), una base de datos de genes, trastornos y rasgos hereditarios en especies animales (que no sean humanos ni ratones), en omia.angis.org.au/contact.shtml, los significados estándar de los términos aceptados en el arte se utilizan en el presente documento a menos que se indique lo contrario. En el presente documento se utilizan abreviaturas estándar para varios términos.

La divulgación en el presente documento demuestra que la multimerización (por ejemplo, dimerización) de factores de transcripción unidos a potenciadores y promotores puede estructurar interacciones en bucle entre potenciadores y promotores que son funcionalmente importantes en el control génico. Los potenciadores están frecuentemente mal regulados en enfermedades, incluida la adquisición de elementos potenciadores específicos de la enfermedad a través de la expresión aberrante de factores de transcripción o la adquisición de variantes de ADN que nuclea la formación de potenciadores. El descubrimiento de que los factores de transcripción median su actividad a través de la multimerización (por ejemplo, dimerización) para estructurar interacciones en bucle entre dos elementos de ADN distintos implica que la perturbación de las interfases de multimerización (por ejemplo, dimerización) de la proteína del factor de transcripción o la interacción con el ADN (por ejemplo, mediante la metilación) ADN se puede utilizar para interrumpir los bucles potenciadores-promotores específicos de la enfermedad. Con múltiples factores de transcripción que se unen a diferentes potenciadores, también puede implicar un mecanismo para determinar la especificidad del potenciador-promotor que se puede usar para identificar los genes diana de los elementos potenciadores asociados a la enfermedad.

Multimerización moduladora del factor de transcripción

En un aspecto, la invención se refiere a un método *in vitro* para modular la expresión de uno o más genes en una célula, que comprende modular la multimerización (por ejemplo, dimerización) de YY1 y, por lo tanto, modular la expresión de uno o más genes.

"Modular" o "modificar" se usa consistentemente con su uso en el arte, es decir, que significa causar o facilitar un cambio cualitativo o cuantitativo, alteración o modificación en un proceso, vía o fenómeno de interés. Sin limitación, tal cambio puede ser un aumento, disminución o cambio en la fuerza o actividad relativa de diferentes componentes o ramas del proceso, vía o fenómeno. Un "modulador" o "modificador" es un agente que provoca o facilita un cambio cualitativo o cuantitativo, alteración o modificación en un proceso, vía o fenómeno de interés. En ciertas realizaciones, modular se refiere a reducir, ralentizar o eliminar de otro modo la expresión de uno o más genes. La modulación de la expresión de un gen puede lograrse o facilitarse, por ejemplo, por cualquier agente (por ejemplo, una molécula o compuesto de ácido nucleico) que cause o facilite un cambio cualitativo o cuantitativo, alteración o modificación en la expresión del gen en un sujeto.

Los factores de transcripción (FT) contienen dominios de unión al ADN que reconocen y se unen a sitios o secuencias de reconocimiento en los promotores de genes transcripcionalmente activos, y también contienen dominios de activación o represión que activan o suprimen la transcripción génica cuando el FT se une al sitio o secuencia de reconocimiento. Los motivos de unión a FT son conocidos en la técnica. Véase, por ejemplo, el documento PCT/US16/59399, presentado el 28 de octubre de 2016, los métodos, enseñanzas y realizaciones de esta solicitud se pueden combinar libremente con las descritas en el presente documento. Las secuencias de motivos de unión a FT también se pueden encontrar en bases de datos disponibles públicamente. En algunas realizaciones, el motivo de unión de FT es un motivo de unión YY1. En algunas realizaciones, el motivo de unión YY1 es GCGCCATnTT (SEQ ID NO: 44), CCGCCATnTT, CGCCATnTT, GCCGCCATTTTG (SEQ ID NO: 45), GCCAT o CCAT.

Según la invención, el factor de transcripción de los métodos y composiciones divulgados en el presente documento es YY1. YY1 (ID del gen: 7528 (ser humano); ID del gen: 22632 (ratón)) es un factor de transcripción distribuido de forma amplia o ubicua que pertenece a la clase GLI-Kruppel de proteínas con dedos de cinc y está implicado en la represión y activación de un número diverso de promotores.

En algunas realizaciones, el factor de transcripción se une a un potenciador ya una región promotora del genoma de la célula. En algunas realizaciones, las regiones potenciadora y promotora están ubicadas ambas en la misma vecindad aislada del genoma de la célula.

El término "unión" pretende significar a lo largo de la descripción una asociación física entre una molécula diana (por ejemplo, una secuencia de ADN, un sitio de unión del factor de transcripción en una región potenciadora o promotora de un genoma, sitio de unión al ADN genómico en un factor de transcripción) o complejo y un agente de unión (por ejemplo, factor de transcripción, ácido nucleico de interferencia, molécula pequeña, anticuerpo). La asociación generalmente depende de la presencia de una característica estructural particular del objetivo (por ejemplo, sitio de unión del factor de transcripción, sitio de unión al ADN en el factor de transcripción). Debe entenderse que no es necesario que la especificidad de unión sea absoluta, sino que en general se refiere al contexto en el que se produce la unión. Como se usa en el presente documento, un "sitio de unión del factor de transcripción" se refiere a una región de ADN genómico que se asocia con un factor de transcripción. Se entiende que cada nucleótido del ADN genómico

puede no interactuar con el factor de transcripción; en cambio, solo partes del sitio de unión pueden interactuar. Como se usa en el presente documento, un compuesto que se une a un sitio de unión del factor de transcripción y modula la unión del factor de transcripción puede unirse o no a los nucleótidos que interactúan con el factor de transcripción.

Como se usa en el presente documento, una "vecindad aislada" es una región de un cromosoma limitada por uno o más marcadores. En algunos aspectos, una "vecindad aislada" es una estructura de bucle cromosómico formada por la interacción de dos sitios de ADN unidos por la proteína CTCF y ocupados por el complejo de cohesina. Véase Hnisz, *et al.*, "Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control," *Cell*. 2016 Nov 17;167(5):1188-1200. doi: 10.1016/j.cell.2016.10.024.

La expresión "molécula pequeña" se refiere a una molécula orgánica que tiene menos de aproximadamente 2 kilodalton (kDa) en masa. En algunos ejemplos, la molécula pequeña tiene menos de aproximadamente 1,5 kDa, o menos de aproximadamente 1 kDa. En algunos ejemplos, la molécula pequeña es inferior a aproximadamente 800 dalton (D), 600 Da, 500 Da, 400 Da, 300 Da, 200 Da o 100 Da. A menudo, una molécula pequeña tiene una masa de al menos 50 Da. En algunos ejemplos, una molécula pequeña contiene múltiples enlaces carbono-carbono y puede comprender uno o más heteroátomos y/o uno o más grupos funcionales importantes para la interacción estructural con proteínas (por ejemplo, enlaces de hidrógeno), por ejemplo, un grupo amina, carbonilo, hidroxilo o carboxilo, y en algunas realizaciones al menos dos grupos funcionales. Las moléculas pequeñas a menudo comprenden una o más estructuras cíclicas de carbono o heterocíclicas y/o estructuras aromáticas o poliaromáticas, opcionalmente sustituidas con uno o más de los grupos funcionales anteriores. En algunos ejemplos, una molécula pequeña es una molécula artificial (que no se produce de forma natural). En algunos ejemplos, una molécula pequeña no es polimérica. En algunos ejemplos, una molécula pequeña no es un aminoácido. En algunos ejemplos, una molécula pequeña no es un nucleótido. En algunos ejemplos, una molécula pequeña no es un sacárido. En algunos ejemplos, la expresión "molécula pequeña" excluye moléculas que son ingredientes que se encuentran en un medio de cultivo tisular estándar.

El término "potenciador" se refiere a una región de ADN genómico a la que se unen proteínas (por ejemplo, factores de transcripción) para potenciar (aumentar) la transcripción de un gen. Los potenciadores pueden ubicarse a cierta distancia de los promotores y del sitio de inicio de la transcripción (TSS) de los genes cuya transcripción regulan y pueden ubicarse corriente arriba o corriente abajo del TSS. Los potenciadores se pueden identificar utilizando métodos conocidos por los expertos en la materia en función de una o más propiedades características. Por ejemplo, H3K27Ac es una modificación de histona asociada con potenciadores activos (Creyghton *et al.*, (2010) "Histone H3K27ac separates active from poised enhancers and predicts developmental state," *Proc Natl Acad Sci USA* 107, 21931-21936; Rada-Iglesias *et al.*, "A unique chromatin signature uncovers early developmental enhancers in humans," *Nature* 470, 279-283). En algunos ejemplos, los potenciadores se identifican como regiones de ADN genómico que, cuando están presentes en una célula, muestran un enriquecimiento de H3K27 acetilado (H3K27Ac), un enriquecimiento de H3K4 metilado (H3K4me1) o ambos. Los potenciadores pueden identificarse adicional o alternativamente como regiones de ADN genómico que, cuando están presentes en una célula, se enriquecen para ser ocupadas por factores de transcripción. Las modificaciones de histonas se pueden detectar usando inmunoprecipitación de cromatina (ChIP) seguida de hibridación de micromatriz (ChIP-Chip) o seguida de secuenciación (ChIP-Seq) u otros métodos conocidos en la técnica. Estos métodos pueden usarse también o alternativamente para detectar la ocupación de ADN genómico por factores de transcripción (u otras proteínas). Se puede usar un algoritmo de búsqueda de picos como el implementado en MACS versión 1.4.2 (análisis basado en modelos de ChIP-seq) o versiones posteriores del mismo para identificar regiones de enriquecimiento de ChIP-seq sobre el fondo (Zhang, Y., *et al.* (2008) "Model-based Analysis of ChIP-Seq (MACS)," *Genome Biol.* 9:R137). En algunos ejemplos puede usarse un umbral de valor de p de enriquecimiento de 10^{-9} . En algunos ejemplos, la región potenciadora es una región potenciadora distal. En algunos ejemplos, el potenciador es un superpotenciador. Véase, por ejemplo, el documento US20160237490 publicado el 18 de agosto de 2016.

En algunas realizaciones, la multimerización (por ejemplo, dimerización) del factor de transcripción YY1 se modula en una célula, modulando así la formación de bucles de ADN potenciadores-promotores en el genoma de la célula. En algunas realizaciones, se reduce la multimerización (por ejemplo, dimerización) del factor de transcripción YY1. La multimerización (por ejemplo, dimerización) del factor de transcripción en la célula se puede reducir en aproximadamente un 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más. En algunas realizaciones, aumenta la multimerización (por ejemplo, dimerización) del factor de transcripción YY1. La multimerización (por ejemplo, dimerización) del factor de transcripción en la célula se puede aumentar en aproximadamente un 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 110 %, 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, 190 %, 200 %, 210 %, 220 %, 230 %, 240 %, 250 %, 260 %, 270 %, 280 %, 290 %, 300 %, 310 %, 320 %, 330 %, 340 %, 350 %, 360 %, 370 %, 380 %, 390 %, 400 %, 500 %, 600 % o más. En algunos ejemplos, la multimerización (por ejemplo, dimerización) del factor de transcripción en la célula se puede aumentar o disminuir aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces o más.

En algunas realizaciones, la expresión de uno o más genes está disminuida. La expresión de uno o más genes se puede disminuir en aproximadamente un 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más. En algunas

realizaciones, aumenta la expresión de uno o más genes. La expresión de uno o más genes puede incrementarse en aproximadamente un 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 110 %, 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, 190 %, 200 %, 210 %, 220 %, 230 %, 240 %, 250 %, 260 %, 270 %, 280 %, 290 %, 300 %, 310 %, 320 %, 330 %, 340 %, 350 %, 360 %, 370 %, 380 %, 390 %, 400 %, 500 %, 600 % o más. En algunos ejemplos, la expresión de uno o más genes puede aumentar o disminuir aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces o más.

En algunas realizaciones, la multimerización (por ejemplo, dimerización) se modula con una composición que comprende una molécula pequeña, péptido, el polipéptido, ácido nucleico y/u oligonucleótido. En algunas realizaciones, se reduce la multimerización (por ejemplo, dimerización) del factor de transcripción YY1. La multimerización (por ejemplo, dimerización) del factor de transcripción en la célula se puede reducir en aproximadamente un 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más. En algunas realizaciones, aumenta la multimerización (por ejemplo, dimerización) del factor de transcripción YY1. La multimerización (por ejemplo, dimerización) del factor de transcripción en la célula se puede aumentar en aproximadamente un 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 110 %, 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, 190 %, 200 %, 210 %, 220 %, 230 %, 240 %, 250 %, 260 %, 270 %, 280 %, 290 %, 300 %, 310 %, 320 %, 330 %, 340 %, 350 %, 360 %, 370 %, 380 %, 390 %, 400 %, 500 %, 600 % o más. En algunos ejemplos, la multimerización (por ejemplo, dimerización) del factor de transcripción en la célula se puede aumentar o disminuir aproximadamente 1 vez, 2fold, 3 veces, 4 veces, 5 veces, 6 veces o más.

En algunos ejemplos, la composición comprende un polipéptido o un ácido nucleico que codifica un polipéptido que es una variante del factor de transcripción (por ejemplo, variante YY1) con aumento, disminución o ausencia de la afinidad (por ejemplo, afinidad de multimerización) por un factor de transcripción. En algunos ejemplos, la variante del factor de transcripción (por ejemplo, la variante YY1) tiene una afinidad de unión disminuida o nula por el sitio de unión del factor de transcripción. En algunos ejemplos, la composición comprende un polipéptido que se une a un factor de transcripción (por ejemplo, YY1) y disminuye o aumenta la multimerización del factor de transcripción (por ejemplo, dimerización).

En algunos ejemplos, la variante del factor de transcripción (por ejemplo, variante YY1) es una variante negativa dominante. La variante del factor de transcripción (por ejemplo, la variante YY1) puede (i) carecer de al menos una parte del dominio de unión al ADN y/o (ii) carecer de al menos una parte de la región que media la multimerización. El primer tipo podría inhibir la multimerización al unirse al factor de transcripción completamente funcional. El segundo tipo podría inhibir la formación de la estructura de bucle potenciador-promotor uniéndose a un factor de transcripción totalmente funcional (por ejemplo, un factor de transcripción totalmente funcional que se une a un promotor o potenciador). En algunas realizaciones, la composición comprende una molécula pequeña que se une al factor de transcripción YY1 y disminuye o aumenta la multimerización del factor de transcripción (por ejemplo, dimerización). En algunos ejemplos, la composición comprende un anticuerpo que se une a un factor de transcripción (por ejemplo, YY1) y disminuye la multimerización del factor de transcripción (por ejemplo, dimerización).

En algunas realizaciones, la célula diana es una célula madre (por ejemplo, una célula madre embrionaria, una célula madre embrionaria de mamífero, una célula madre embrionaria humana, una célula madre embrionaria murina). En algunas realizaciones, la célula es una célula madre embrionaria. En algunos ejemplos, la célula humana es una célula madre pluripotencial inducida.

En algunos ejemplos de los métodos y composiciones divulgados en el presente documento, las células incluyen células somáticas, células madre, células mitóticas o post-mitóticas, neuronas, fibroblastos o cigotos no humanos. Una célula, cigoto no humano, embrión no humano o mamífero posnatal no humano puede ser de origen vertebrado (por ejemplo, mamífero). En algunos aspectos, los vertebrados son mamíferos o aves. Los ejemplos particulares incluyen células primates no humanos, roedores (por ejemplo, ratón, rata), cánidos, felinos, bóvidos, équidos, caprinos, porcinos o aves (por ejemplo, pollos, patos, gansos, pavos) células, cigotos, embriones o mamíferos posnatales. En algunos ejemplos, se aísla la célula, cigoto no humano, embrión no humano o un mamífero posnatal no humano (por ejemplo, una célula aislada; un cigoto aislado; un embrión no humano aislado). En algunos ejemplos, se aísla una célula de ratón, cigoto de ratón, embrión de ratón o mamífero posnatal de ratón. En algunos ejemplos, se aísla una célula de rata, cigoto de rata, embrión de rata o de mamífero posnatal de rata. En algunos ejemplos, se utiliza una célula humana. Los métodos descritos en el presente documento se pueden utilizar en un mamífero (por ejemplo, un ratón *in vivo*).

Las células madre pueden incluir células madre totipotenciales no humanos, pluripotenciales, multipotenciales, oligopotenciales y unipotenciales. Los ejemplos específicos de células madre incluyen células madre embrionarias, células madre fetales, células madre adultas y células madre pluripotenciales inducidas (iPSC) (por ejemplo, véanse las solicitudes publicadas de EE.UU N.º 2010/0144031, 2011/0076678, 2011/0088107, 2012/0028821).

Las células somáticas pueden ser células primarias (células no inmortalizadas), tales como las recién aisladas de un animal, o pueden derivarse de una línea celular capaz de una proliferación prolongada en cultivo (por ejemplo, durante más de 3 meses) o una proliferación indefinida (células inmortalizadas). Las células somáticas adultas pueden obtenerse de individuos, por ejemplo, sujetos humanos, y cultivadas de acuerdo con los protocolos de cultivo celular

estándar disponibles para los expertos en la materia. Las células somáticas de uso en aspectos de la invención incluyen células de mamíferos, tales como, por ejemplo, células humanas, células de primates no humanos o roedores (por ejemplo, ratón, rata). Pueden obtenerse por métodos bien conocidos de diversos órganos, por ejemplo, piel, pulmón, páncreas, hígado, estómago, intestino, corazón, mama, órganos reproductores, músculo, sangre, vejiga, riñón, uretra y otros órganos urinarios, etc., generalmente de cualquier órgano o tejido que contenga células somáticas vivas. Las células somáticas de mamífero útiles en diversas realizaciones incluyen, por ejemplo, fibroblastos, células de Sertoli, células de la granulosa, neuronas, células pancreáticas, células epidérmicas, células epiteliales, células endoteliales, hepatocitos, células del folículo piloso, queratinocitos, células hematopoyéticas, melanocitos, condrocitos, linfocitos (linfocitos B y T), macrófagos, monocitos, células mononucleares, células de músculo cardíaco, células de músculo esquelético, etc.

En algunos ejemplos, el uno o más genes que se modulan comprenden genes reguladores de células. En algunas realizaciones, el uno o más genes comprenden Oct4, Nanog y/o Sox2. En algunos ejemplos, la célula es una célula cancerosa y el gen es un oncogén o un gen supresor de tumores. En algunos ejemplos, la célula (por ejemplo, célula cancerosa) puede albergar una mutación o variante polimórfica asociada con una actividad potenciadora aumentada o aberrante.

Modulación de la formación de bucles de ADN potenciador-promotor

Algunos aspectos de la invención están dirigidos a un *in vitro* método para modular la expresión de uno o más genes en una célula, que comprende modular la formación y/o la estabilidad de un bucle de ADN potenciador-promotor en el genoma de la célula, en donde la formación y/o la estabilidad depende del factor de transcripción, es decir, YY1. Como se usa en el presente documento, las indicaciones de que la formación del bucle de ADN potenciador-promotor es "dependiente del factor de transcripción" significa que el factor de transcripción es parcial o totalmente responsable de la formación y/o estabilidad del bucle de ADN potenciador-promotor. En algunos casos, el factor de transcripción es necesario pero no suficiente para la formación y/o estabilidad del bucle de ADN potenciador-promotor. En algunos casos, el factor de transcripción es necesario y suficiente para la formación y/o estabilidad del bucle de ADN potenciador-promotor. La expresión de uno o más genes en una célula puede disminuir en aproximadamente un 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más. La expresión de uno o más genes en una célula puede incrementarse en aproximadamente un 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 110 %, 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, 190 %, 200 %, 210 %, 220 %, 230 %, 240 %, 250 %, 260 %, 270 %, 280 %, 290 %, 300 %, 310 %, 320 %, 330 %, 340 %, 350 %, 360 %, 370 %, 380 %, 390 %, 400 %, 500 %, 600 % o más. En algunos ejemplos, la expresión de uno o más genes en una célula puede aumentar o disminuir aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces o más.

En algunas realizaciones, la formación del bucle de ADN potenciador-promotor se modula mediante la modulación de la unión del factor de transcripción YY1 a la región promotora y/o potenciadora (por ejemplo, el sitio de unión del factor de transcripción en la región promotora y/o potenciadora) del bucle de ADN potenciador-promotor. La unión del factor de transcripción se puede reducir en aproximadamente un 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más. La unión del factor de transcripción se puede aumentar en aproximadamente un 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 110 %, 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, 190 %, 200 %, 210 %, 220 %, 230 %, 240 %, 250 %, 260 %, 270 %, 280 %, 290 %, 300 %, 310 %, 320 %, 330 %, 340 %, 350 %, 360 %, 370 %, 380 %, 390 %, 400 %, 500 %, 600 % o más. En algunos ejemplos, la unión del factor de transcripción se puede aumentar o disminuir aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces o más.

En algunas realizaciones, la unión del factor de transcripción YY1 a la región promotora y/o potenciadora del bucle de ADN potenciador-promotor se modula modificando una región promotora y/o potenciadora (por ejemplo, el sitio de unión del factor de transcripción en una región promotora y/o potenciadora). En algunos ejemplos, la modificación comprende modificar el grado de metilación de la región promotora y/o potenciadora o una región dentro de aproximadamente 25 bases, 50 bases, 100 bases, 200 bases, 300 bases, 400 bases, 500 bases, 600 bases, 700 bases, 800 bases, 900 bases, 1000 bases, 1500 bases, 2000 bases, 5000 bases, 10000 bases, 20000 bases, 50000 bases o más corriente arriba o corriente abajo de la región promotora y/o potenciadora. En algunas realizaciones, la unión del factor de transcripción YY1 a la región promotora y/o potenciadora del bucle de ADN potenciador-promotor se modula modificando la metilación de uno o más sitios o motivos de unión a la transcripción YY1 en una región promotora y/o potenciadora. En algunos ejemplos, la unión del factor de transcripción (por ejemplo, YY1) al promotor y/o a la región potenciadora del bucle de ADN potenciador-promotor se modula modificando la metilación dentro de aproximadamente 25 bases, 50 bases, 100 bases, 200 bases, 300 bases, 400 bases, 500 bases, 600 bases, 700 bases, 800 bases, 900 bases, 1000 bases, 1500 bases, 2000 bases, 5000 bases, 10000 bases, 20000 bases, 50000 bases o más corriente arriba o corriente abajo de uno o más sitios/motivos de unión a la transcripción (por ejemplo, sitios/motivos de unión YY1) en una región promotora y/o potenciadora. En algunos ejemplos, el grado de metilación se puede aumentar o disminuir aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces o más. Los métodos para modular la metilación del ADN son conocidos en la técnica. Véase Liu *et al.*, "Editing DNA methylation in the mammalian genome," Cell, Vol. 167 (1):233-247.e17. En algunos casos, el grado de metilación puede modificarse mediante uno o más métodos descritos en la solicitud n.º: 62/377.520 (Rudolf Jaenisch, *et al.*, presentada

el 19 de agosto de 2016) y el documento PCT/US2017/047674 (Rudolf Jaenisch, *et al.*, presentada el 18 de agosto de 2017). En algunos ejemplos, el grado de metilación puede modificarse usando una nucleasa dirigida catalíticamente inactiva (por ejemplo, nucleasa específica de sitio catalíticamente inactiva). En algunos ejemplos, la unión de YY1 a un sitio o motivo de unión de YY1 se potencia al reducir la metilación del sitio o motivo de unión de YY1. En algunos ejemplos, la unión de YY1 a un sitio o motivo de unión de YY1 se potencia al reducir el nivel o grado de metilación del sitio o motivo de unión de YY1. En algunos ejemplos, la unión de YY1 a un motivo o sitio de unión de YY1 se reduce aumentando el nivel o grado de metilación del motivo o sitio de unión de YY1.

En algunas realizaciones, la modificación comprende modificar la secuencia de nucleótidos de una o más regiones promotoras y/o potenciadoras. En algunas realizaciones, la modificación comprende modificar la secuencia de nucleótidos del sitio de unión del factor de transcripción YY1 en una o más regiones promotoras y/o potenciadoras. En algunos ejemplos, la secuencia de nucleótidos de la región potenciadora o promotora (por ejemplo, el sitio de unión a la transcripción en una región promotora o potenciadora) se modifica con una nucleasa dirigida (por ejemplo, una nucleasa específica del sitio).

En algunos ejemplos, la modificación es una supresión de todo o parte de un motivo de unión, sustitución de uno o más nucleótidos en un motivo de unión donde la sustitución reduce la unión de FT, o alteración de un motivo de unión para aumentar la unión. En algunos ejemplos, una nucleasa específica del sitio catalíticamente inactiva dirigida hacia o cerca de un motivo de unión (por ejemplo, hasta aproximadamente 5, 10, 15, 20, 25, 50, 75, 100, 150, 200, 300 o 500 nucleótidos más allá de cualquiera de los extremos del motivo de unión) bloquea estéricamente la unión del FT al motivo de unión o bloquea la asociación del FT (por ejemplo, multimerización del FT) y la formación de estructuras de bucle de ADN. En algunos ejemplos, una nucleasa específica del sitio catalíticamente inactiva dirigida hacia o cerca de un motivo de unión (por ejemplo, hasta aproximadamente 5, 10, 15, 20, 25, 50, 75, 100, 150, 200, 300, o 500 nucleótidos del final del motivo de unión) modula (por ejemplo, aumenta o disminuye) la unión del FT al modular la metilación del ADN del motivo de unión o cerca del motivo de unión. En algunos ejemplos, la modificación es una modificación del ADN que inhibe o bloquea la unión de FT.

Actualmente hay cuatro tipos principales de nucleasas dirigidas (a veces también denominadas "nucleasas específicas del sitio") en uso: nucleasas de dedo de cinc (ZFN), nucleasas efectoras similares a activadores de transcripción (TALEN) y nucleasas guiadas por ARN (RGN) como las proteínas Cas del sistema CRISPR/Cas Tipo II y meganucleasas diseñadas. ZFN y TALEN comprenden el dominio de nucleasa de la enzima de restricción *FokI* (o una variante diseñada del mismo) fusionado con un dominio de unión a ADN (DBD) específico del sitio que está diseñado apropiadamente para dirigir la proteína a una secuencia de ADN seleccionada. En el caso de las ZFN, el dominio de unión al ADN comprende un dedo de cinc DBD. En el caso de las TALEN, el DBD específico del sitio está diseñado en base al código de reconocimiento de ADN empleado por los efectores similares al activador de la transcripción (TALE), una familia de proteínas de unión al ADN específicas del sitio que se encuentran en bacterias patógenas de plantas, tales como especies de *Xantomonas*. El sistema de repeticiones palindrómicas cortas agrupadas regularmente interesparadas (CRISPR) de tipo II es un sistema inmunitario adaptativo bacteriano que se ha modificado para su uso como tecnología de endonucleasa guiada por ARN para la ingeniería genómica. El sistema bacteriano comprende dos ARN bacterianos endógenos llamados crRNA y tracrRNA y una nucleasa asociada a CRISPR (Cas), por ejemplo, Cas9. El tracrRNA tiene una complementariedad parcial con el crRNA y forma un complejo con él. La proteína Cas es guiada a la secuencia diana por el complejo crRNA/tracrRNA, que forma un híbrido de ARN/ADN entre la secuencia de crRNA y la secuencia complementaria en el objetivo. Para su uso en la modificación del genoma, los componentes de crRNA y tracrRNA a menudo se combinan en un solo ARN guía quimérico (ARNgs o ARNg) en el que la especificidad de direccionamiento del crRNA y las propiedades del tracrRNA se combinan en una sola transcripción que localiza la proteína Cas en la secuencia objetivo para que la proteína Cas puede escindir el ADN. El ARNg a menudo comprende una secuencia guía de aproximadamente 20 nucleótidos complementaria u homóloga a la secuencia diana deseada seguida de aproximadamente 80 nt de crRNA-híbrido/tracrRNA. Un experto normal en la técnica apreciará que el ARN guía no necesita ser perfectamente complementario u homólogo a la secuencia diana. Por ejemplo, en algunos casos puede tener uno o dos apareamientos erróneos. La secuencia genómica con la que se hibrida el ARNg normalmente está flanqueada en un lado por una secuencia de motivo adyacente protoespaciador (PAM), aunque un experto en la técnica aprecia que ciertas proteínas Cas pueden tener un requisito relajado para una secuencia PAM. La secuencia PAM está presente en el ADN genómico pero no en la secuencia de ARNg. La proteína Cas se dirigirá a cualquier secuencia de ADN con la secuencia objetivo y la secuencia PAM correctas. La secuencia de PAM varía según la especie de bacteria de la que se derivó la proteína Cas. Los ejemplos específicos de proteínas Cas incluyen Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 y Cas10. En algunos ejemplos, la nucleasa específica del sitio comprende una proteína Cas9. Por ejemplo, se puede usar Cas9 de *Streptococcus pyogenes* (Esp), *Neisseria meningitidis*, *Staphylococcus aureus*, *Streptococcus thermophilus* o *Treponema denticola*. Las secuencias PAM para estas proteínas Cas9 son NGG, NNNNGATT, NNAGAA, NAAAAC, respectivamente. Se han desarrollado varias variantes diseñadas de las nucleasas específicas de sitio y se pueden usar en ciertas realizaciones. Por ejemplo, las variantes diseñadas de Cas9 y FokI son conocidas en la técnica. Por otra parte, se entenderá que se puede usar un fragmento o variante biológicamente activo. Otras variaciones incluyen el uso de nucleasas específicas de sitios híbridos. Por ejemplo, en las nucleasas FokI guiadas por ARN (RFN) de CRISPR, el dominio de la nucleasa FokI se fusiona con el extremo amino-terminal de una proteína Cas9 catalíticamente inactiva (dCas9). Los RFN actúan como dímeros y utilizan dos ARN guía (Tsai, QS, *et al.*, Nat Biotechnol. 2014; 32(6): 569-576). Las nucleasas específicas del sitio que producen una rotura de ADN monocatenario

también son útiles para la edición del genoma. Tales nucleasas, a veces denominadas "nickasas" se pueden generar mediante la introducción de una mutación (por ejemplo, una sustitución de alanina) en residuos catalíticos clave en uno de los dos dominios de nucleasa de una nucleasa específica del sitio que comprende dos dominios de nucleasa (tales como las proteínas ZFN, TALEN y Cas). Los ejemplos de dichas mutaciones incluyen D10A, N863A y H840A en SpCas9 o en posiciones homólogas en otras proteínas Cas9. Una mella puede estimular HDR con baja eficiencia en algunos tipos de células. Dos nickasas, dirigidas a un par de secuencias que están cerca una de la otra y en hebras opuestas puede crear una rotura de una sola hebra en cada hebra ("doble mella"), generando efectivamente un DSB, que opcionalmente puede ser reparado por HDR utilizando una plantilla de ADN donante (Ran, FA *et al.* Cell 154, 1380-1389 (2013). En algunos ejemplos, la proteína Cas es una proteína SpCas9. En algunos ejemplos, la variante SpCas9 es una variante triple R661A/Q695A/Q926A o una variante cuádruple N497A/R661A/Q695A/Q926A. Véase Kleinstiver *et al.*, "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects," Nature, Vol. 529, pp. 490-495 (y materiales complementarios) (2016). En algunos ejemplos, la proteína Cas es C2c1, una proteína CRISPR-Cas de clase 2 tipo V-B. Véase Yang *et al.*, "PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease," Cell, Vol. 167, pp. 1814-1828 (2016). En algunos ejemplos, la proteína Cas es una de las descritas en el documento US 20160319260 "Engineered CRISPR-Cas9 nucleases with Altered PAM Specificity".

En algunos ejemplos, la nucleasa objetivo (por ejemplo, nucleasa específica del sitio) tiene al menos un 90 %, 95 % o 99 % de identidad de la secuencia polipeptídica con una nucleasa dirigida de origen natural.

En algunos ejemplos, la secuencia de nucleótidos de la región potenciadora o promotora se modifica con una nucleasa específica del sitio (es decir, una nucleasa dirigida) y una o más secuencias guía. En algunos ejemplos, la nucleasa específica del sitio es una proteína Cas. Una variedad de genes o proteínas asociados a CRISPR (Cas) que se conocen en la técnica pueden usarse en los métodos de la invención y la elección de la proteína Cas dependerá de la situación particular (por ejemplo, www.ncbi.nlm.nih.gov/gen/?term=cas9). En un aspecto particular, el ácido nucleico Cas o la proteína utilizada en las composiciones es Cas9. En algunos ejemplos, una proteína Cas, por ejemplo, una proteína Cas9, puede ser de cualquiera de una variedad de especies procarióticas. En algunos ejemplos, una proteína Cas particular, por ejemplo, una proteína Cas9 particular, puede seleccionarse para reconocer una secuencia particular de motivo adyacente al protoespaciador (PAM). En ciertos ejemplos, una proteína Cas, por ejemplo, una proteína Cas9, puede obtenerse de una bacteria o arquea o sintetizarse usando métodos conocidos. En ciertas realizaciones, una proteína Cas puede ser de una bacteria grampositiva o de una bacteria gramnegativa. En determinados ejemplos, una proteína Cas puede ser de un estreptococo, (por ejemplo, *S. pyogenes*, un *S. thermophilus*) un criptococo, un *Corynebacterium*, un *Haemophilus*, una *Eubacteria*, una *Pasteurella*, una *Prevotella*, una *Veillonella* o una *Marinobacter*. En algunas realizaciones, los ácidos nucleicos que codifican dos o más proteínas Cas diferentes, o dos o más proteínas Cas, pueden estar presentes en la composición, por ejemplo, para permitir el reconocimiento y la modificación de los sitios que comprenden los motivos PAM similares o diferentes.

En algunos ejemplos, la proteína Cas es proteína Cpf1 o una porción funcional de la misma. En algunos ejemplos, la proteína Cas es Cpf1 de cualquier especie bacteriana o porción funcional de la misma. En determinados ejemplos, una proteína Cpf1 es una proteína U112 de *Francisella novicida* o una porción funcional de la misma, una proteína BV3L6 de *Acidaminococcus* sp. o una parte funcional de la misma o una proteína ND2006 de *Lachnospiraceae bacterium* o una parte funcional de la misma. La proteína Cpf1 es miembro de los sistemas CRISPR tipo V. La proteína Cpf1 es un polipéptido que comprende alrededor de 1300 aminoácidos. Cpf1 contiene un dominio de endonucleasa similar a RuvC. Véase Zetsche B, *et al.*, "Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system," Cell. 22 de octubre de 2015; 163(3):759-71. doi: 10.1016/j.cell.2015.09.038. Epub 2015 Sep 25.) y documento US20160208243. Un experto en la técnica aprecia que Cpf1 no utiliza tracrRNA y, por lo tanto, solo requiere un crRNA que contenga un solo tallo-bucle, que tolera cambios de secuencia que conservan la estructura secundaria.

En algunos ejemplos, se puede generar una nickasa Cas9 inactivando uno o más de los dominios de nucleasa Cas9. En algunas realizaciones, una sustitución de aminoácido en el resto 10 en el dominio RuvC I de Cas9 convierte la nucleasa en una nickasa de ADN. Por ejemplo, el aspartato en el resto de aminoácido 10 se puede sustituir por alanina (Cong *et al.*, Science, 339:819-823).

En algunos ejemplos, la nucleasa dirigida puede ser una nucleasa dirigida catalíticamente inactiva (por ejemplo, nucleasa específica de sitio catalíticamente inactiva). En algunos ejemplos, se puede utilizar una nucleasa dirigida catalíticamente inactiva junto con un dominio efector para modular la unión de un factor de transcripción a una región promotora o potenciadora modificando el grado de metilación de la región promotora o potenciadora. Las mutaciones de aminoácidos que crean una proteína Cas9 catalíticamente inactiva incluyen la mutación en el resto 10 y/o el resto 840. Las mutaciones tanto en el resto 10 como en el resto 840 pueden crear una proteína Cas9 catalíticamente inactiva, a veces denominado en el presente documento dCas9. En algunas realizaciones, dCas9 es un mutante D10A y H840A Cas9 que es catalíticamente inactivo. Como se usa en el presente documento, un "dominio efector" es una molécula (por ejemplo, una proteína) que modula la expresión y/o activación de una secuencia genómica (por ejemplo, un gen). El dominio efector puede tener actividad de metilación (por ejemplo, actividad de metilación del ADN). En algunos casos, el dominio efector se dirige a uno o ambos alelos de un gen. El dominio efector puede introducirse como una secuencia de ácido nucleico y/o como una proteína. En algunos casos, el dominio efector puede ser un dominio efector constitutivo o inducible. En algunos casos, una secuencia de ácido nucleico Cas (por ejemplo, dCas)

o una variante de la misma y una secuencia de ácido nucleico del dominio efector se introducen en la célula como una secuencia quimérica. En algunos casos, el dominio efector se fusiona con una molécula que se asocia con (por ejemplo, se une a) la proteína Cas (por ejemplo, la molécula efectora se fusiona con un anticuerpo o fragmento de unión a antígeno del mismo que se une a la proteína Cas). En ciertas ocasiones, una proteína Cas (por ejemplo, dCas) o una variante de la misma y un dominio efector se fusionan o unen creando una proteína quimérica y se introducen en la célula como la proteína quimérica. En algunos casos, la proteína Cas (por ejemplo, dCas) y el dominio efector se unen como una interacción proteína-proteína. En algunos casos, la proteína Cas (por ejemplo, dCas) y el dominio efector están unidos covalentemente. En algunos casos, el dominio efector se asocia de forma no covalente con la proteína Cas (por ejemplo, dCas). En algunos casos, una secuencia de ácido nucleico Cas (por ejemplo, dCas) y una secuencia de ácido nucleico del dominio efector se introducen como secuencias y/o proteínas separadas. En algunos casos, la proteína Cas (por ejemplo, dCas) y el dominio efector no están fusionados ni anclados.

Una nucleasa o polipéptido específico del sitio (por ejemplo, polipéptido de fusión que comprende una nucleasa específica del sitio y un dominio efector, un polipéptido de fusión que comprende una nucleasa específica del sitio y un dominio efector que tiene actividad de metilación o desmetilación) se puede dirigir a un sitio único en el genoma (por ejemplo, un sitio de unión del factor de transcripción, un sitio de unión YY1) de una célula de mamífero mediante el diseño apropiado de la nucleasa, ARN guía o polipéptido. Un polipéptido, nucleasa y/o ARN guía pueden introducirse en las células introduciendo un ácido nucleico que lo codifica en la célula. Se pueden usar métodos estándar, tales como la transfección de ADN plasmídico, liberación de vectores virales, transfección con ARNm modificado o sintético (por ejemplo, se tapó, ARNm poliadenilado) o microinyección. En algunos ejemplos, el ARNm modificado o sintético comprende una o más modificaciones que estabilizan el ARNm o proporcionan otras mejoras sobre el ARNm natural (por ejemplo, aumento de la captación celular). Se describen ejemplos de ARNm modificado o sintético en Warren *et al.* (Cell Stem Cell 7(5):618-30, 2010, Mandal PK, Rossi DJ. Nat Protoc. 2013 8(3):568-82, la publicación de patente de EE.UU. n.º 20120046346 y/o PCT/US2011/032679 (WO/2011/130624). El ARNm también se analiza en R.E. Rhoads (Ed.), "Synthetic mRNA: Production, Introduction Into Cells, and Physiological Consequences," Series: Methods in Molecular Biology, Vol. 1428. Se encuentran ejemplos adicionales en numerosas solicitudes PCT y de EE. UU. y patentes otorgadas a Moderna Therapeutics, por ejemplo, el documento PCT/US2011/046861; el documento PCT/US2011/054636, el documento PCT/US2011/054617, el documento USSN 14/390,100 (y patentes y solicitudes de patentes adicionales mencionadas en estas). Si se introduce ADN que codifica la nucleasa o el ARN guía, las secuencias de codificación deben estar operativamente vinculadas a los elementos reguladores apropiados para la expresión, tal como un promotor y una señal de terminación. En algunos ejemplos, una secuencia que codifica un ARN guía se une operativamente a un promotor de ARN polimerasa III tal como U6 o promotor de ARNt. En algunos ejemplos, uno o más ARN guía y secuencias codificantes de proteína Cas se transcriben a partir del mismo ácido nucleico (por ejemplo, plásmido). En algunos ejemplos, se transcriben múltiples ARN guía del mismo plásmido o de diferentes plásmidos o se introducen de otro modo en la célula. Los ARN guía múltiples pueden dirigir Cas9 a diferentes secuencias objetivo en el genoma, permitiendo la edición del genoma multiplexado. En algunos ejemplos, una proteína nucleasa (por ejemplo, Cas9) puede comprender o modificarse para comprender una señal de localización nuclear (por ejemplo, SV40 NLS). Se puede introducir una proteína nucleasa en las células, por ejemplo, mediante la transducción de proteínas. Se pueden introducir proteínas de nucleasa, ARN guía, o ambos, mediante microinyección. Se describen métodos de uso de nucleasas específicas del sitio, por ejemplo, para realizar la edición del genoma, en numerosas publicaciones, tales como *Methods in Enzymology*, Doudna JA, Sontheimer EJ. (eds), The use of CRISPR/Cas9, ZFNs, and TALENs in generating site-specific genome alterations. Methods Enzymol. 2014, Vol. 546 (Elsevier); Carroll, D., Genome Editing with Targetable Nucleases, Annu. Rev. Biochem. 2014. 83:409-39 y las referencias en cualquiera de ellos. Véanse también las publicaciones de patente de EE.UU. n.º 20140068797, 20140186919, 20140170753 y/o PCT/US2014/034387 (WO/2014/172470).

En algunos ejemplos, la una o más secuencias guía incluyen secuencias que reconocen el ADN de una manera específica del sitio. Por ejemplo, las secuencias guía pueden incluir secuencias guía de ácido ribonucleico (ARN) utilizadas por un sistema CRISPR o secuencias dentro de un sistema TALEN o de dedos de cinc que reconocen el ADN de una manera específica del sitio. Las secuencias guía comprenden una porción que es complementaria a una porción de cada una o más secuencias genómicas y comprenden un sitio de unión para la nucleasa específica del sitio catalíticamente inactiva. En algunos ejemplos, la secuencia de ARN se denomina ARN guía (ARNg) o ARN guía sencillo (ARNgs).

En algunos casos, una secuencia guía puede ser complementaria a una o más (por ejemplo, todas) de las secuencias genómicas que se están modulando o modificando. En un caso, una secuencia guía es complementaria a una única secuencia genómica diana. En un caso particular en el que se van a modular o modificar dos o más secuencias genómicas diana, se introducen múltiples (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más) secuencias guía en las que cada secuencia guía es complementaria a (específica de) una secuencia genómica diana. En algunos casos, dos o más, tres o más, cuatro o más, cinco o más, o seis o más secuencias guía son complementarias (específicas para) diferentes partes de la misma secuencia diana. En un caso, dos o más secuencias guía se unen a diferentes secuencias de la misma región de ADN. En algunos casos, una sola secuencia guía es complementaria a al menos dos objetivos o más (por ejemplo, todas) de las secuencias genómicas. También será evidente para los expertos en la técnica que la porción de la secuencia guía que es complementaria a una o más de las secuencias genómicas y la porción de la secuencia guía que se une a la nucleasa específica del sitio catalíticamente inactiva puede introducirse como una sola secuencia o como 2 (o más) secuencias separadas en una célula.

Cada secuencia guía puede variar en longitud desde alrededor de 8 pares de bases (pb) hasta aproximadamente 200 pb. En algunos ejemplos, las secuencias de ARN pueden tener de aproximadamente 9 a aproximadamente 190 pb; de aproximadamente 10 a aproximadamente 150 pb; de aproximadamente 15 a aproximadamente 120 pb; de aproximadamente 20 a aproximadamente 100 pb; de aproximadamente 30 a aproximadamente 90 pb; de aproximadamente 40 a aproximadamente 80 pb; de aproximadamente 50 a aproximadamente 70 pb de longitud.

La porción de cada secuencia genómica (por ejemplo, una región promotora o potenciadora, un sitio de unión del factor de transcripción en una región promotora o potenciadora) al que cada secuencia guía es complementaria también puede variar en tamaño. En casos particulares, la porción de cada secuencia genómica a la que la secuencia guía es complementaria puede ser de aproximadamente 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98 o 100 nucleótidos (nucleótidos contiguos) de longitud. En algunos ejemplos, cada secuencia de guía puede ser aproximadamente un 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 100 %, etc. idéntica o similar a la porción de cada secuencia genómica. En algunos ejemplos, cada secuencia guía es completa o parcialmente idéntica o similar a cada secuencia genómica. Por ejemplo, cada secuencia guía puede diferir de la complementariedad perfecta con la porción de la secuencia genómica en aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, etc. nucleótidos. En algunos ejemplos, una o más secuencias guía son perfectamente complementarias (100 %) en al menos aproximadamente 10 a aproximadamente 25 (por ejemplo, aproximadamente 20) nucleótidos de la secuencia genómica.

En algunos ejemplos, una o más secuencias de nucleótidos (por ejemplo, secuencias guía) son complementarias u homólogas a una región de ADN genómico implicada en la unión del factor de transcripción o en la formación del bucle de ADN potenciador-promotor. En algunos ejemplos, una o más secuencias de nucleótidos son complementarias u homólogas a una región potenciadora o promotora de un bucle de ADN potenciador-promotor. En algunos ejemplos, una o más secuencias de nucleótidos son complementarias u homólogas a un sitio de unión del factor de transcripción (por ejemplo, YY1) (por ejemplo, el sitio de unión del factor de transcripción en una región promotora o potenciadora). En algunos ejemplos, una o más secuencias de nucleótidos son complementarias u homólogas a una región de ADN genómico que, según el grado de metilación, modula la unión del factor de transcripción (por ejemplo, la unión de YY1), la formación de bucles de ADN promotores potenciadores y/o la estabilidad del bucle de ADN promotores potenciadores. En algunos ejemplos, la una o más secuencias de nucleótidos son complementarias u homólogas a 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más secuencias de ADN genómico. En algunos ejemplos, una o más secuencias de nucleótidos son complementarias u homólogas a una secuencia de ADN genómico única, y pueden utilizarse para modular la expresión de uno o más genes asociados con un bucle de ADN potenciador-promotor específico.

En algunos ejemplos, una nucleasa específica del sitio (por ejemplo, Talen, proteína en dedo de cinc) se une a una región de ADN genómico involucrada en la unión del factor de transcripción o en la formación de bucles de ADN promotor-potenciador. En algunos ejemplos, una nucleasa específica del sitio (por ejemplo, Talen, proteína en dedo de cinc) se une a una región promotora-potenciadora de un bucle de ADN promotor-potenciador. En algunos ejemplos, una nucleasa específica del sitio (por ejemplo, Talen, proteína en dedo de cinc) se une a un sitio de unión del factor de transcripción (por ejemplo, YY1) (por ejemplo, el sitio de unión del factor de transcripción en una región promotora o potenciadora). En algunos ejemplos, una nucleasa específica del sitio (por ejemplo, Talen, proteína en dedo de cinc) se une a una región del ADN genómico que, según el grado de metilación, modula la unión del factor de transcripción (por ejemplo, la unión de YY1), la formación de bucles de ADN promotores potenciadores y/o la estabilidad del bucle de ADN promotores potenciadores. En algunos ejemplos, una nucleasa específica del sitio (por ejemplo, Talen, proteína en dedo de cinc) se une a 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más secuencias de ADN genómico. En algunos ejemplos, una nucleasa específica del sitio (por ejemplo, Talen, proteína en dedo de cinc) se une a una secuencia de ADN genómico única y se puede utilizar para modular la expresión de uno o más genes asociados con un bucle de ADN potenciador-promotor específico.

En algunos ejemplos, los ácidos nucleicos (por ejemplo, ácidos nucleicos mejorados) (por ejemplo, constructos de ADN, ARN sintéticos, por ejemplo, ARN homólogos o complementarios descritos en el presente documento, ARNm descritos en el presente documento, etc.) se puede introducir en las células de interés a través de transfección, electroporación, agentes catiónicos, polímeros o moléculas de administración basadas en lípidos bien conocidas por los expertos en la materia. Como se usa en el presente documento, un "ácido nucleico potenciado" tiene una propiedad potenciada (por ejemplo, una estabilidad potenciada, captación celular potenciada, unión potenciada, especificidad potenciada) en comparación con un ácido nucleico homólogo natural.

En algunos ejemplos, los métodos de la presente divulgación potencian la liberación de ácido nucleico en una población celular, *in vivo*, *ex-vivo*, o en cultivo. Por ejemplo, un cultivo celular que contiene una pluralidad de células huésped (por ejemplo, células eucariotas tales como levaduras o células de mamíferos) se pone en contacto con una composición que contiene un ácido nucleico potenciado que tiene al menos una modificación de nucleósido y, opcionalmente, una región traducible. En algunos ejemplos, la composición también contiene generalmente un reactivo de transfección u otro compuesto que aumenta la eficacia de la captación mejorada de ácido nucleico en las células huésped. El ácido nucleico potenciado exhibe una retención mejorada en la población celular, en relación con un ácido

nucleico no modificado correspondiente. En algunos ejemplos, la retención del ácido nucleico potenciado es mayor que la retención del ácido nucleico no modificado. En algunas realizaciones, es al menos aproximadamente un 50 %, 75 %, 90 %, 95 %, 100 %, 150 %, 200 %, o más del 200 % mayor que la retención del ácido nucleico no modificado. Tal ventaja de retención puede lograrse mediante una ronda de transfección con el ácido nucleico potenciado, o puede obtenerse después de repetidas rondas de transfección.

Los ARN sintéticos (por ejemplo, ARNm modificado, ácidos nucleicos potenciados) del objeto de la presente descripción pueden combinarse opcionalmente con un gen informador (por ejemplo, corriente arriba o corriente abajo de la región codificante del ARNm) que, por ejemplo, facilita la determinación de la liberación de ARNm modificado a las células o tejidos diana. Los genes indicadores adecuados pueden incluir, por ejemplo, ARNm de proteína fluorescente verde (ARNm GFP), ARNm de luciferasa de Renilla (ARNm de luciferasa), ARNm de luciferasa de luciérnaga, o cualquier combinación de los mismos. Por ejemplo, el ARNm de GFP puede fusionarse con un ARNm que codifica una secuencia de localización nuclear para facilitar la confirmación de la localización del ARNm en las células diana donde tiene lugar el ARN transcrito a partir del al menos un elemento regulador.

En algunos ejemplos, el ARN puede modificarse aún más después de la transcripción, por ejemplo, mediante la adición de un capuchón u otro grupo funcional. En un aspecto, un ARN sintético (ácido nucleico potenciado) comprende una estructura 5' y/o 3'-capuchón. El ARN sintético puede ser monocatenario (por ejemplo, ARNmc) o bicatenario (por ejemplo, ARNbc). La estructura 5' y/o 3'-capuchón puede estar solo en la hebra sentido, la hebra antisentido o en ambas hebras. Por "estructura de capuchón" se entiende modificaciones químicas, que se han incorporado en cualquiera de los extremos del oligonucleótido (véase, por ejemplo, Adamic *et al.*, patente de EE. UU. N.º 5.998.203. Estas modificaciones terminales protegen la molécula de ácido nucleico de la degradación por exonucleasa y pueden ayudar en la liberación y/o localización dentro de una célula. El capuchón puede estar presente en el extremo 5' (capuchón en 5') o en el extremo 3' (capuchón en 3') o puede estar presente en ambos extremos.

Los ejemplos no limitantes del capuchón en 5' incluyen, aunque sin limitación, glicerilo, resto desoxi abásico invertido (resto); nucleótido de 4',5'-metileno; 1-(beta-D-eritrofuranosil)nucleótido, 4'-tio nucleótido; nucleótido carbocíclico; nucleótido de 1,5-anhidrohexitol; L-nucleótidos; alfa-nucleótidos; nucleótido base modificado; enlace fosforoditioato; nucleótido de treo-pentofuranosilo; 3',4'-seco nucleótido acíclico; nucleótido de 3,4-dihidroxibutilo acíclico; nucleótido de 3,5-dihidroxipentilo acíclico, resto nucleotídico invertido 3'-3'; resto abásico invertido 3'-3'; resto nucleotídico invertido 3'-2'; resto abásico invertido 3'-2'; fosfato de 1,4-butanodiol; 3'-fosforamidato; hexilfosfato; fosfato de aminohexilo; 3'-fosfato; 3'-fosforotioato; fosforoditioato; o resto metilfosfonato puente o no puente.

Los ejemplos no limitantes del capuchón en 3' incluyen, aunque sin limitación, glicerilo, resto desoxi abásico invertido (resto), nucleótido de 4',5'-metileno; nucleótido de 1-(beta-D-eritrofuranosil); nucleótido 4'-tio, nucleótido carbocíclico; fosfato de 5'-amino-alquilo; fosfato de 1,3-diamino-2-propilo; fosfato de 3-aminopropilo; fosfato de 6-aminohexilo; fosfato de 1,2-aminododecilo; fosfato de hidroxipropilo; nucleótido de 1,5-anhidrohexitol; L-nucleótido; alfa-nucleótido; nucleótido base modificado; fosforoditioato; nucleótido de treo-pentofuranosilo; 3',4'-seco nucleótido acíclico; nucleótido de 3,4-dihidroxibutilo; nucleótido de 3,5-dihidroxipentilo, resto nucleotídico invertido 5'-5'; resto abásico invertido 5'-5'; 5'-fosforamidato; 5'-fosforotioato; fosfato de 1,4-butanodiol; 5'-amino; 5'-fosforamidato puente y/o no puente, fosforotioato y/o fosforoditioato, restos de metilfosfonato y 5'-mercapto puente o no puente (para obtener más detalles, véase Beaucage e Iyer, 1993, Tetrahedron 49, 1925).

El ARN sintético puede comprender al menos un nucleósido modificado, tal como pseudouridina, m5U, s2U, m6A y m5C, N1-metilguanosina, N1-metiladenosina, N7-metilguanosina, 2'-)-metiluridina y 2'-O-metilcitidina. Las polimerasas que aceptan nucleósidos modificados son conocidas por los expertos en la técnica. Las polimerasas modificadas se pueden utilizar para generar polimerasas sintéticas, ARN modificados. De este modo, por ejemplo, una polimerasa que tolera o acepta un nucleósido modificado particular como sustrato puede usarse para generar un ARN sintético, modificado que incluye ese nucleósido modificado.

En algunos ejemplos, el ARN sintético provoca una respuesta inmunitaria innata reducida (o ausente) *in vivo* o respuesta de interferón reducida *in vivo* por el tejido transfectado o la población celular. ARNm producido en células eucariotas, por ejemplo, células de mamíferos o humanos, está muy modificado, las modificaciones que permiten a la célula detectar ARN no producido por esa célula. La célula responde interrumpiendo la traducción o iniciando de otro modo una respuesta inmunitaria innata o de interferón. De este modo, en la medida en que un ARN agregado exógenamente puede modificarse para imitar las modificaciones que ocurren en los ARN endógenos producidos por una célula diana, el ARN exógeno puede evitar al menos parte de la defensa de la célula diana contra los ácidos nucleicos extraños. De este modo, en algunos ejemplos, los ARN sintéticos incluyen ARN *in vitro* transcritos, incluidas las modificaciones que se encuentran en el ARN eucariota/mamífero/humano *in vivo*. Otras modificaciones que imitan tales modificaciones naturales también pueden ser útiles para producir una molécula de ARN sintético que será tolerada por una célula.

En algunos ejemplos, el ARN sintético tiene una o más modificaciones (por ejemplo, secuencias UTR 5' y/o 3' modificadas, codones optimizados) que pueden mejorar la estabilidad del ARNm y/o la eficiencia de traducción en células de mamíferos (por ejemplo, humanas). Véase la publicación de patente de EE.UU. N.º 20140206753.

Como se usa en el presente documento, los términos "transfectar y "transfección" significan la introducción de un ácido nucleico, por ejemplo, un ARN sintético, por ejemplo, ARNm modificado en una célula o, preferentemente, en una célula diana. El ARN sintético introducido (por ejemplo, ARNm modificado) puede mantenerse de forma estable o transitoria en la célula diana. La expresión "eficiencia de la transfección" se refiere a la cantidad relativa de ARN sintético (por ejemplo, ARNm modificado) absorbido por la célula diana que se somete a transfección. En la práctica, la eficiencia de la transfección puede estimarse mediante la cantidad de un producto de ácido nucleico informador expresado por las células diana después de la transfección. Los ejemplos incluyen composiciones con altas eficacias de transfección y, en particular, aquellas composiciones que minimizan los efectos adversos que están mediados por la transfección de células no diana. En algunos ejemplos, las composiciones de la presente invención que demuestran altas eficacias de transfección mejoran la probabilidad de que las dosis apropiadas del ARN sintético (por ejemplo, ARNm modificado) se administren a la célula diana, al tiempo que minimiza los posibles efectos adversos sistémicos.

En algunos ejemplos, una célula puede modificarse genéticamente (*in vitro* o *en vivo*) (por ejemplo, usando un constructo de ácido nucleico, por ejemplo, un constructo de ADN) para hacer que exprese (i) un agente que modula la multimerización del factor de transcripción (por ejemplo, YY1) o la unión a una región promotora o potenciadora o (ii) un ARNm que codifica dicho agente. Por ejemplo, la presente divulgación contempla generar una célula o línea celular que exprese de forma transitoria o estable un ARN que inhiba la unión a una región promotora o potenciadora o la multimerización del FT o que exprese de forma transitoria y estable un ARNm que codifique un anticuerpo (u otra proteína capaz de unión específica) que inhibe la unión a una región promotora o potenciadora o la multimerización del FT. Las células y constructos genéticamente modificados pueden ser útiles, por ejemplo, en los enfoques de terapia génica. Por ejemplo, en algunos casos, dicho constructo de ácido nucleico se administra a un individuo que lo necesita. En otros ejemplos, células (por ejemplo, autólogas) que se han puesto en contacto *ex-vivo* con tal constructo se puede administrar a un individuo que lo necesite. El constructo puede incluir un promotor ligado operativamente a una secuencia que codifica el agente o ARNm.

El ARN sintético (por ejemplo, ARNm modificado, ácido nucleico potenciado) se puede formular con uno o más reactivos aceptables, que proporcionan un vehículo para administrar dicho ARN sintético a las células diana. Los reactivos apropiados generalmente se seleccionan con respecto a una serie de factores, que incluyen, entre otras cosas, las propiedades biológicas o químicas del ARN sintético, la vía de administración prevista, el entorno biológico previsto al que se expondrá dicho ARN sintético (por ejemplo, ARNm modificado) y las propiedades específicas de las células diana previstas. En algunos ejemplos, los vehículos de transferencia, tales como liposomas, encapsulan el ARN sintético sin comprometer la actividad biológica. En algunas realizaciones, el vehículo de transferencia demuestra una unión preferente y/o sustancial a una célula diana en relación con las células no diana. En un ejemplo preferido, el vehículo de transferencia entrega su contenido a la célula diana de manera que el ARN sintético se entrega al compartimento subcelular apropiado, tal como el citoplasma.

En algunos ejemplos, el vehículo de transferencia en las composiciones de la invención es un vehículo de transferencia liposomal, por ejemplo, una nanopartícula lipídica. En un ejemplo, el vehículo de transferencia se puede seleccionar y/o preparar para optimizar la liberación del ácido nucleico (por ejemplo, ácido nucleico potenciado, ARN sintético (por ejemplo, ARNm modificado)) a una célula diana. Por ejemplo, si la célula diana es un hepatocito, las propiedades del vehículo de transferencia (por ejemplo, el tamaño, la carga y/o el pH) pueden optimizarse para entregar de manera efectiva dicho vehículo de transferencia a la célula diana, reducir la eliminación inmunitaria y/o promover la retención en esa célula diana. Como alternativa, si la célula diana está en el sistema nervioso central (por ejemplo, para el tratamiento de enfermedades neurodegenerativas, el vehículo de transferencia puede apuntar específicamente al tejido cerebral o espinal), la selección y preparación del vehículo de transferencia debe considerar la penetración y la retención dentro de la barrera hematoencefálica y/o el uso de medios alternativos para administrar directamente dicho vehículo de transferencia a dicha célula diana. En un ejemplo, las composiciones de la presente invención se pueden combinar con agentes que facilitan la transferencia de ARN sintético exógeno (por ejemplo, ARNm modificado) (por ejemplo, agentes que interrumpen o mejoran la permeabilidad de la barrera hematoencefálica y, por lo tanto, mejoran la transferencia de ARNm exógeno a las células diana).

La presente divulgación contempla el uso de vehículos de transferencia liposómicos para facilitar la liberación de ácidos nucleicos a las células diana. Los liposomas (por ejemplo, nanopartículas de lípidos liposomales) son generalmente útiles en una variedad de aplicaciones en investigación, la industria y la medicina, particularmente para su uso como vehículos de transferencia de compuestos diagnósticos o terapéuticos *in vivo* (Lasic, Trends Biotechnol., 16: 307-321, 1998; Drummond *et al.*, Pharmacol. Rev., 51: 691-743, 1999) y normalmente se caracterizan como vesículas microscópicas que tienen un espacio acuático interior separado de un medio exterior por una membrana de una o más bicapas. Las membranas bicapa de los liposomas suelen estar formadas por moléculas anfífilas, tales como los lípidos de origen sintético o natural que comprenden dominios hidrofílicos e hidrofóbicos espacialmente separados (Lasic, Trends Biotechnol., 16: 307-321, 1998). Las membranas bicapa de los liposomas también pueden estar formadas por polímeros anfífilos y tensioactivos (por ejemplo, polimerosomas, niosomas, etc.).

En el contexto de la presente divulgación, un vehículo de transferencia liposomal normalmente sirve para transportar el ácido nucleico (por ejemplo, ARNm modificado) a la célula diana. Como se describe en el presente documento, los vehículos de transferencia liposomal se preparan para contener los ácidos nucleicos deseados. El proceso de incorporación de una entidad deseada (por ejemplo, un ácido nucleico) en un liposoma a menudo se denomina "carga"

(Lasic, *et al.*, FEBS Lett., 312: 255-258, 1992). Los ácidos nucleicos incorporados al liposoma pueden estar ubicados total o parcialmente en el espacio interior del liposoma, dentro de la membrana bicapa del liposoma, o asociado con la superficie exterior de la membrana del liposoma. La incorporación de un ácido nucleico en los liposomas también se denomina en el presente documento "encapsulación", en la que el ácido nucleico está completamente contenido dentro del espacio interior del liposoma. El propósito de incorporar un ácido nucleico en un vehículo de transferencia, tal como un liposoma, es a menudo para proteger el ácido nucleico de un entorno que puede contener enzimas o productos químicos que degradan los ácidos nucleicos y/o sistemas o receptores que provocan la excreción rápida de los ácidos nucleicos. En un ejemplo preferido, el vehículo de transferencia seleccionado es capaz de mejorar la estabilidad del ácido nucleico que contiene. El liposoma puede permitir que el ácido nucleico encapsulado (por ejemplo, ARNm modificado) llegue a la célula diana y/o puede permitir preferentemente que el ácido nucleico encapsulado (por ejemplo, ARNm modificado) llegue a la célula diana o, alternativamente, limitar la liberación de dicho ácido nucleico. (por ejemplo, ARNm modificado) a otros sitios o células donde la presencia del ácido nucleico administrado (por ejemplo, ARNm modificado) puede ser inútil o indeseable. Por otra parte, incorporar el ARN sintético (por ejemplo, ARNm modificado) en un vehículo de transferencia, tal como, por ejemplo, un liposoma catiónico, también facilita la liberación de dicho ARN sintético (por ejemplo, ARNm modificado) en una célula diana.

Los vehículos de transferencia liposomal se pueden preparar para encapsular uno o más ARN sintéticos deseados (por ejemplo, ARNm modificado) de modo que las composiciones demuestren una alta eficiencia de transfección y una mayor estabilidad. Si bien los liposomas pueden facilitar la introducción de ácidos nucleicos en las células diana, la adición de policationes (por ejemplo, poli L-lisina y protamina), como copolímero puede facilitar y, en algunos casos, potenciar notablemente la eficiencia de transfección de varios tipos de liposomas catiónicos entre 2 y 28 veces en varias líneas celulares, tanto *in vitro* como *in vivo*. (Véase NJ Caplen, *et al.*, Gene Ther. 1995; 2: 603; S. Li, *et al.*, Gene Ther. 1997; 4, 891.)

En algunos ejemplos, el vehículo de transferencia está formulado como una nanopartícula lipídica. Como se usa en el presente documento, la expresión "nanopartícula lipídica" se refiere a un vehículo de transferencia que comprende uno o más lípidos (por ejemplo, lípidos catiónicos, lípidos no catiónicos y lípidos modificados con PEG). Preferentemente, las nanopartículas de lípidos se formulan para administrar uno o más ARN sintéticos (por ejemplo, ARNm modificado) a una o más células diana.

Los ejemplos de lípidos adecuados incluyen, por ejemplo, los compuestos de fosfatidilo (por ejemplo, fosfatidilglicerol, fosfatidilcolina, fosfatidilserina, fosfatidiletanolamina, esfingolípidos, cerebrósidos y gangliósidos). También se contempla el uso de polímeros como vehículos de transferencia, ya sea solos o en combinación con otros vehículos de transferencia. Los polímeros adecuados pueden incluir, por ejemplo, poliácridatos, polialquilcianoacrilatos, polilactida, copolímeros de polilactida-poliglicolida, policaprolactonas, dextrano, albúmina, gelatina, alginato, colágeno, quitosano, ciclodextrinas, dendrímeros y polietilénimina. En una realización, el vehículo de transferencia se selecciona en función de su capacidad para facilitar la transfección de un ácido nucleico (por ejemplo, ARNm modificado) en una célula diana.

La presente divulgación contempla el uso de nanopartículas lipídicas como vehículos de transferencia que comprenden un lípido catiónico para encapsular y/o mejorar la liberación de ácido nucleico (por ejemplo, ARNm modificado) en la célula diana, por ejemplo, que actuará como un depósito para la producción de un péptido, polipéptido o proteína (por ejemplo, anticuerpo o fragmento de anticuerpo) como se describe en el presente documento. Como se usa en el presente documento, la expresión "lípido catiónico" se refiere a cualquier número de especies de lípido que porta una carga neta positiva a un pH seleccionado, tal como el pH fisiológico. Las nanopartículas de lípidos contempladas se pueden preparar incluyendo mezclas de lípidos de componentes múltiples de proporciones variables empleando uno o más lípidos catiónicos, lípidos no catiónicos y lípidos modificados con PEG. Varios lípidos catiónicos han sido descritos en la literatura, muchos de los cuales están comercializados.

Los lípidos catiónicos adecuados de uso en las composiciones y métodos de el presente documento incluyen los descritos en la publicación de patente internacional WO 2010/053572, por ejemplo, C12-200 descrito en el párrafo [00225] del documento WO 2010/053572. Las composiciones y métodos descritos en el presente documento pueden emplear nanopartículas lipídicas que comprenden un lípido catiónico ionizable descrito en la solicitud de patente provisional de EE. UU. 61/617,468, presentada el 29 de marzo de 2012, tales como, por ejemplo, (15Z, 18Z)-N,N-dimetil-6-(9Z, 12Z)-octadeca-9,12-dien-1-il)tetracosa-15,18-dien-1-amina (HGT5000), (15Z,18Z)-N, N-dimetil-6-((9Z,12Z)-octadeca-9,12-dien-1-il)tetracosa-4,15,18-trien-1-amina (HGT5001) y (15Z,18Z)-N,N-dimetil-6-((9Z,12Z)-octadeca-9,12-dien-1-il)tetracosa-5,15,18-trien-1-amina (HGT5002).

En algunos ejemplos, se utiliza el lípido catiónico cloruro de N-[1-(2,3-dioleiloxy)propil]-N,N,N-trimetilamonio o "DOTMA". (Felgner *et al.* (Proc. Nat'l Acad. Sci. 84, 7413 (1987); la patente de los Estados Unidos n.º 4.897.355). DOTMA se puede formular solo o se puede combinar con el lípido neutro, dioleoilfosfatidil-etanolamina o "DOPE" u otros lípidos catiónicos o no catiónicos en un vehículo de transferencia liposomal o una nanopartícula lipídica, y dichos liposomas se pueden usar para mejorar el suministro de ácidos nucleicos a las células diana. Otros lípidos catiónicos adecuados incluyen, por ejemplo, 5-carboxiespermilglicinadiodecylamida o "DOGS", 2,3-dioleiloxy-N-[2(espermina-carboxamido)etil]-N,N-dimetil-1-propanaminio o "DOSPA" (Behr *et al.* Proc. Nat'l Acad. Sci. 86, 6982 (1989); la patente de los Estados Unidos n.º 5.171.678; la patente de los Estados Unidos n.º 5.334.761), 1,2-Dioleoil-3-Dimetilamonio-

Propano o "DODAP", 1,2-Dioleoil-3-Trimetilamonio-Propano o "DOTAP". Los lípidos catiónicos contemplados también incluyen 1,2-diesteriloxi-N,N-dimetil-3-aminopropano o "DSDMA", 1,2-dioleiloxi-N,N-dimetil-3-aminopropano o "DODMA", 1,2-dilinoleiloxi-N,N-dimetil-3-aminopropano o "DLinDMA", 1,2-dilinoleniloxi-N,N-dimetil-3-aminopropano o "DLenDMA", Cloruro de N-dioleil-N,N-dimetilamonio o "DODAC", bromuro de N,N-diesteril-N,N-dimetilamonio o "DDAB", Bromuro de N-(1,2-dimiristiloxiprop-3-il)-N,N-dimetil-N-hidroxietilamonio o "DMRIE", 3-dimetilamino-2-(colest-5-en-3-beta-oxibutan-4-oxi)-1-(cis,cis-9,12-octadecadienoxi)propano o "CLinDMA", 2-[5'-(colest-5-en-3-beta-oxi)-3'-oxapentoxi]-3-dimetil-1-(cis,cis-9', 1'-2'-octadecadienoxi)propano o "CpLinDMA", N,N-dimetil-3,4-dioleiloxibencilamina o "DMOBA", 1,2-N,N'-dioleilcarbamil-3-dimetilaminopropano o "DOcarbDAP", 2,3-dilinoleoiloxi-N,N-dimetilpropilamina o "DLinDAP", 1,2-N,N'-dilinoleilcarbamil-3-dimetilaminopropano o "DLincarbDAP", 1,2-dilinoleoilcarbamil-3-dimetilaminopropano o "DLinC-DAP", 2,2-dilinoleil-4-dimetilaminometil-[1,3]-dioxolano o "DLin-K-DMA", 2,2-dilinoleil-4-dimetilaminoetil-[1,3]-dioxolano o "DLin-K-XTCC2-DMA", y 2-(2,2-di((9Z,12Z)-octadeca-9,12- dien-1-il)-1,3-dioxolan-4-il)-N,N-dimetiletanamina (DLin-KC2-DMA)) (véase el documento WO 2010/042877; Semple *et al.*, Nature Biotech. 28:172-176 (2010)) o mezclas de los mismos. (Heyes, J., *et al.*, J Controlled Release 107: 276-287 (2005); Morrissey, D V., *et al.*, Nat. Biotechnol. 23(8): 1003-1007 (2005); publicación PCT WO2005/121348A1).

La presente divulgación también contempla el uso de lípidos catiónicos basados en colesterol. Dichos lípidos catiónicos a base de colesterol se pueden utilizar, ya sea solos o en combinación con otros lípidos catiónicos o no catiónicos. Los lípidos catiónicos basados en colesterol adecuados incluyen, por ejemplo, DC-Col (N,N-dimetil-N-etilcarboxamidocolesterol), 1,4-bis(3-N-oleilamino-propil)piperazina (Gao, *et al.* Biochem. Biophys. Res. Comm. 179, 280 (1991); Wolf *et al.* BioTechniques 23, 139 (1997); la patente de EE.UU. No. 5,744,335), o ICE.

El experto en la materia apreciará que varios reactivos están disponibles comercialmente para mejorar la eficacia de la transfección. Los ejemplos adecuados incluyen LIPOFECTIN (DOTMA:DOPE) (Invitrogen, Carlsbad, Calif.), LIPOFECTAMINE (DOSPA:DOPE) (Invitrogen), LIPOFECTAMINE2000. (Invitrogen), FUGENE, TRANSFECTAM (DOGS) y EFFECT-ENE.

También se contemplan lípidos catiónicos tales como los lípidos basados en dialquilamino, basados en imidazol y guanidinio. Por ejemplo, ciertas realizaciones están dirigidas a una composición que comprende uno o más lípidos catiónicos basados en imidazol, por ejemplo, el éster de colesterol de imidazol o lípido "ICE" 3-(1H-imidazol-4-il)propanoato de (3S,10R,13R,17R)-10,13-dimetil-17-((R)-6-metilheptan-2-il)-2,3,4,7,8,9,10,11,12,13,14,15,16,17-tetradecahidro-1H-ciclopenta[a]fenantren-3-ilo. En una realización preferida, un vehículo de transferencia para la liberación de ARN sintético (por ejemplo, ARNm modificado) puede comprender uno o más lípidos catiónicos basados en imidazol, por ejemplo, el éster de colesterol de imidazol o lípido "ICE" 3-(1H-imidazol-4-il)propanoato de (3S,10R,13R,17R)-10,13-dimetil-17-((R)-6-metilheptan-2-il)-2,3,4,7,8,9,10,11,12,13,14,15,16,17-tetradecahidro-1H-ciclopenta[a]fenantren-3-ilo.

Los lípidos catiónicos a base de imidazol también se caracterizan por su toxicidad reducida en relación con otros lípidos catiónicos. Los lípidos catiónicos a base de imidazol (por ejemplo, ICE) se pueden usar como el único lípido catiónico en la nanopartícula lipídica o, alternativamente, se pueden combinar con los lípidos catiónicos tradicionales, lípidos no catiónicos y lípidos modificados con PEG. El lípido catiónico puede comprender una relación molar de aproximadamente el 1 % a aproximadamente el 90 %, de aproximadamente 2 % a aproximadamente 70 %, de aproximadamente 5 % a aproximadamente 50 %, de aproximadamente el 10 % a aproximadamente el 40 % de los lípidos totales presentes en el vehículo de transferencia, o preferentemente de aproximadamente el 20 % a aproximadamente el 70 % de los lípidos totales presentes en el vehículo de transferencia.

En algunos ejemplos, las nanopartículas lipídicas comprenden el lípido catiónico HGT4003 2-((2,3-Bis((9Z,12Z)-octadeca-9,12-dien-1-iloxi)propil)disulfanil)-N,N-dimetiletanamina, como se describe adicionalmente en la publicación de EE.UU. N.º 20140288160.

En otros ejemplos, las composiciones y métodos descritos en el presente documento están dirigidos a nanopartículas de lípidos que comprenden uno o más lípidos escindibles, tales como, por ejemplo, uno o más lípidos catiónicos o compuestos que comprenden un grupo funcional disulfuro escindible (S-S) (por ejemplo, HGT4001, HGT4002, HGT4003, HGT4004 y HGT4005), como se describe adicionalmente en la publicación de EE.UU. N.º 20140288160.

El uso de fosfolípidos modificados con polietilenglicol (PEG) y lípidos derivados como las ceramidas derivadas (PEG-CER), incluyendo N-octanoil-esfingosina-1-[succinil(metoxi polietilenglicol)-2000] (ceramida C8 PEG-2000) también se contempla en la presente invención, ya sea solo o preferentemente en combinación con otros lípidos que comprenden el vehículo de transferencia (por ejemplo, una nanopartícula lipídica). Los lípidos modificados con PEG contemplados incluyen, pero sin limitación, una cadena de polietilenglicol de hasta 5 kDa de longitud unida covalentemente a un lípido con cadena(s) de alquilo de C₆-C₂₀ longitud. La adición de tales componentes puede prevenir la agregación compleja y también puede proporcionar un medio para aumentar la vida útil de la circulación y aumentar la liberación de la composición de lípidos y ácidos nucleicos a la célula diana, (Klibanov *et al.* (1990) FEBS Letters, 268 (1): 235-237), o pueden seleccionarse para cambiar rápidamente fuera de la formulación *in vivo* (véase la patente de EE.UU. N.º 5,885,613). En algunos ejemplos, los lípidos intercambiables comprenden PEG-ceramidas que tienen cadenas de acilo más cortas (por ejemplo, C14 o C18). El fosfolípido modificado con PEG y los lípidos derivados de la presente invención pueden comprender una relación molar de aproximadamente 0 % a aproximadamente 20 %, de

aproximadamente 0,5 % a aproximadamente 20 %, de aproximadamente 1 % a aproximadamente 15 %, de aproximadamente 4 % a aproximadamente 10 % o de aproximadamente 2 % de los lípidos totales presentes en el vehículo de transferencia liposomal.

5 La presente divulgación también contempla el uso de lípidos no catiónicos. Como se usa en el presente documento, la expresión "lípidos no catiónicos" se refiere a cualquier lípido neutro, zwitteriónico o aniónico. Como se usa en el presente documento, la expresión "lípidos aniónicos" se refiere a cualquiera de una serie de especies de lípidos que llevan una carga neta negativa a un pH seleccionado, tal como el pH fisiológico. Los lípidos no catiónicos incluyen, aunque sin limitación, distearoilfosfatidilcolina (DSPC), dioleoilfosfatidilcolina (DOPC), dipalmitoilfosfatidilcolina (DPPC), dioleoilfosfatidilglicerol (DOPG), dipalmitoilfosfatidilglicerol (DPPG), dioleoilfosfatidiletanolamina (DOPE), palmitoiloleoilfosfatidilcolina (POPC), palmitoiloleoilfosfatidiletanolamina (POPE), dioleoilfosfatidiletanolamina 4-(N-maleimidometil)-ciclohexano-1-carboxilato (DOPE-mal), dipalmitoilfosfatidiletanolamina (DPPE), dimiristoilfosfoetanolamina (DMPE), diestearoilfosfatidiletanolamina (DSPE), 16-O-monometil PE, 16-O-dimetil PE, 18-1-trans PE, 1-estearoil-2-oleoilfosfatidiletanolamina (SOPE), colesterol o una mezcla de los mismos. Dichos lípidos no catiónicos pueden usarse solos, pero se utilizan preferentemente en combinación con otros excipientes, por ejemplo, lípidos catiónicos. Cuando se usa en combinación con un lípido catiónico, el lípido no catiónico puede comprender una proporción molar de 5 % a aproximadamente 90 %, o preferentemente de aproximadamente 10 % a aproximadamente 70 % del lípido total presente en el vehículo de transferencia.

20 En algunos ejemplos, el vehículo de transferencia (por ejemplo, una nanopartícula de lípido) se prepara combinando múltiples componentes de lípidos y/o polímeros. Por ejemplo, se puede preparar un vehículo de transferencia utilizando C12-200, DOPE, col, DMG-PEG2K en una relación molar de 40:30:25:5, o DODAP, DOPE, colesterol, DMG-PEG2K en una relación molar de 18:56:20:6 o HGT5000, DOPE, col, DMG-PEG2K en una relación molar de 40:20:35:5 o HGT5001, DOPE, col, DMG-PEG2K en una relación molar de 40:20:35:5. La selección de lípidos catiónicos, lípidos no catiónicos y/o lípidos modificados con PEG que comprenden la nanopartícula lipídica, así como la relación molar relativa de dichos lípidos entre sí, se basa en las características de los lípidos seleccionados, la naturaleza de las células diana previstas, las características del ARN sintético (por ejemplo, ARNm modificado) a administrar. Las consideraciones adicionales incluyen, por ejemplo, la saturación de la cadena alquílica, así como el tamaño, la carga, el pH, el pKa, la fusogenicidad y la toxicidad de los lípidos seleccionados. Por lo tanto, las relaciones molares pueden ajustarse en consecuencia. Por ejemplo, el porcentaje de lípido catiónico en la nanopartícula lipídica puede ser superior al 10 %, superior al 20 %, superior al 30 %, superior al 40 %, superior al 50 %, mayor del 60 % o mayor del 70 %. El porcentaje de lípido no catiónico en la nanopartícula lipídica puede ser superior al 5 %, superior al 10 %, superior al 20 %, mayor del 30 % o mayor del 40 %. El porcentaje de colesterol en la nanopartícula lipídica puede ser superior al 10 %, superior al 20 %, mayor del 30 % o mayor del 40 %. El porcentaje de lípido modificado con PEG en la nanopartícula lipídica puede ser superior al 1 %, superior al 2 %, superior al 5 %, mayor del 10 % o mayor del 20 %.

En determinados ejemplos, las nanopartículas lipídicas de la presente divulgación comprenden al menos uno de los siguientes lípidos catiónicos: C12-200, Dlin-KC2-DMA, DODAP, HGT4003, ICE, HGT5000 o HGT5001. En los ejemplos, el vehículo de transferencia comprende colesterol y/o un lípido modificado con PEG. En algunos ejemplos, los vehículos de transferencia comprenden DMG-PEG2K. En determinados ejemplos, el vehículo de transferencia comprende una de las siguientes formulaciones lipídicas: C12-200, DOPE, col, DMG-PEG2K; DODAP, DOPE, colesterol, DMG-PEG2K; HGT5000, DOPE, col, DMG-PEG2K, HGT5001, DOPE, col, DMG-PEG2K.

Los vehículos de transferencia liposómicos para usar en las composiciones de la descripción se pueden preparar mediante diversas técnicas que se conocen actualmente en la técnica. Las vesículas multilamelares (VML) se pueden preparar con técnicas convencionales, por ejemplo, depositando un lípido seleccionado en la pared interior de un recipiente o recipiente adecuado disolviendo el lípido en un solvente apropiado y luego evaporando el disolvente para dejar una película delgada en el interior del recipiente o mediante secado por aspersión. Luego se puede añadir una fase acuosa al recipiente con un movimiento de vórtice que da como resultado la formación de VML. Las vesículas unilamelares (VUL) pueden formarse luego por homogeneización, sonicación o extrusión de las vesículas multilamelares. Adicionalmente, las vesículas unilamelares pueden formarse mediante técnicas de eliminación de detergentes.

En determinados ejemplos, las composiciones de la presente descripción comprenden un vehículo de transferencia en el que el ARN sintético (por ejemplo, ARNm modificado) está asociado tanto en la superficie del vehículo de transferencia como encapsulado dentro del mismo vehículo de transferencia. Por ejemplo, durante la preparación de las composiciones de la presente invención, los vehículos de transferencia liposómicos catiónicos pueden asociarse con el ARN sintético (por ejemplo, ARNm modificado) a través de interacciones electrostáticas.

60 En determinados ejemplos, las composiciones de la presente divulgación pueden cargarse con radionúclidos de diagnóstico, materiales fluorescentes u otros materiales que son detectables en aplicaciones tanto *in vitro* como *in vivo*. Por ejemplo, materiales de diagnóstico adecuados para su uso en la presente invención pueden incluir rodamina-dioleoilfosfatidiletanolamina (Rh-PE), ARNm de proteína fluorescente verde (ARNm GFP), ARNm de luciferasa de Renilla y ARNm de luciferasa de luciérnaga.

65 La selección del tamaño apropiado de un vehículo de transferencia de liposomas puede tener en cuenta el sitio de la

célula o tejido diana y, hasta cierto punto, la aplicación para la que se fabrica el liposoma. En algunos ejemplos, puede ser deseable limitar la transfección del ARN sintético (por ejemplo, ARNm modificado) a ciertas células o tejidos. Por ejemplo, para dirigirse a los hepatocitos, se puede dimensionar un vehículo de transferencia liposomal de modo que sus dimensiones sean más pequeñas que las fenestraciones de la capa endotelial que recubre los sinusoides hepáticos en el hígado; en consecuencia, el vehículo de transferencia liposomal puede penetrar fácilmente tales fenestraciones endoteliales para alcanzar los hepatocitos diana. Como alternativa, un vehículo de transferencia liposomal puede tener un tamaño tal que las dimensiones del liposoma tengan un diámetro suficiente para limitar o evitar expresamente la distribución en ciertas células o tejidos. Por ejemplo, un vehículo de transferencia liposomal puede tener un tamaño tal que sus dimensiones sean mayores que las fenestraciones de la capa endotelial que recubre los sinusoides hepáticos para limitar así la distribución del vehículo de transferencia liposomal a los hepatocitos. Por lo general, el tamaño del vehículo de transferencia está dentro del rango de aproximadamente 25 a 250 nm, preferentemente menos de aproximadamente 250 nm, 175 nm, 150 nm, 125 nm, 100 nm, 75 nm, 50 nm, 25 nm o 10 nm.

Están disponibles una variedad de métodos alternativos conocidos en la técnica para determinar el tamaño de una población de vehículos de transferencia liposómicos. Uno de tales métodos de determinación del tamaño se describe en la patente de EE.UU. N° 4.737.323.

La sonicación de una suspensión de liposomas, ya sea mediante baño o sonicación con sonda, produce una reducción progresiva del tamaño hasta un VUL pequeño de menos de aproximadamente 0,05 micrómetros de diámetro. La homogeneización es otro método que se basa en la energía de corte para fragmentar liposomas grandes en otros más pequeños. En un procedimiento típico de homogeneización, las VML se recirculan a través de un homogeneizador de emulsión estándar hasta que se observan tamaños de liposomas seleccionados, típicamente entre aproximadamente 0,1 y 0,5 micrómetros. El tamaño de las vesículas liposomales puede determinarse mediante dispersión de luz cuasi eléctrica (QELS) como se describe en Bloomfield, Ann. Rev. Biophys. *Bioeng.*, 10:421-450 (1981). El diámetro promedio de los liposomas puede reducirse mediante sonicación de los liposomas formados. Los ciclos de sonicación intermitentes se pueden alternar con la evaluación QELS para guiar la síntesis eficiente de liposomas.

Como se usa en el presente documento, la expresión "célula diana" se refiere a una célula o tejido al que se dirige o dirige una composición de la presente divulgación. Por ejemplo, cuando se desea liberar un ácido nucleico a un hepatocito, el hepatocito representa la célula diana. En algunos ejemplos, las composiciones de la presente descripción transfectan las células diana sobre una base discriminatoria (es decir, no transfectan células no diana). Las composiciones de la presente divulgación también se pueden preparar para dirigirse preferentemente a una variedad de células diana, que incluyen, aunque sin limitación, hepatocitos, células epiteliales, células hematopoyéticas, células epiteliales, células endoteliales, células pulmonares, células óseas, células madre, células mesenquimatosas, células neurales (por ejemplo, meninges, astrocitos, neuronas motoras, células de los ganglios de la raíz dorsal y de las neuronas motoras del asta anterior), células fotorreceptoras (por ejemplo, bastones y conos), células epiteliales pigmentadas de la retina, células secretoras, células cardíacas, adipocitos, células musculares lisas vasculares, cardiomiocitos, células de músculo esquelético, células beta, células hipofisarias, células de revestimiento sinovial, células ováricas, células testiculares, fibroblastos, linfocitos B, linfocitos T, reticulocitos, leucocitos, granulocitos y células tumorales. En algunos ejemplos, las células diana son deficientes o tienen un exceso de una proteína o enzima de interés. En algunos ejemplos, la proteína o enzima de interés está codificada por un gen diana y la composición comprende un agente que modula la expresión del gen diana.

Las composiciones de la presente divulgación pueden prepararse para distribuirse preferentemente a las células diana, como en el corazón, pulmones, riñones, hígado y bazo. En algunos ejemplos, las composiciones de la presente divulgación se distribuyen en las células del hígado para facilitar el suministro y la posterior expresión del ácido nucleico (por ejemplo, ARNm modificado) contenido en las células del hígado (por ejemplo, hepatocitos). Los hepatocitos objetivo pueden funcionar como un "depósito" o "depósito" biológico capaz de producir una proteína o enzima funcional. En un ejemplo, el vehículo de transferencia liposomal puede dirigirse a los hepatocitos y/o distribuirse preferentemente a las células del hígado en el momento de la liberación. Después de la transfección de los hepatocitos diana, el ARN sintético (por ejemplo, ARNm modificado) cargado en el vehículo liposomal se traduce y se produce un producto de proteína funcional. En otros ejemplos, células distintas de los hepatocitos (por ejemplo, pulmón, bazo, corazón, ocular, o células del sistema nervioso central) pueden servir como lugar de depósito para la producción de proteínas.

Los péptidos, polipéptidos o proteínas expresados o traducidos también pueden caracterizarse por la inclusión *in vivo* de modificaciones postraduccionales nativas que a menudo pueden estar ausentes en proteínas o enzimas preparadas de forma recombinante, reduciendo así aún más la inmunogenicidad del péptido, polipéptido o proteína traducidos.

La presente divulgación también contempla el direccionamiento discriminatorio de células y tejidos diana por medios de direccionamiento tanto pasivos como activos. El fenómeno del direccionamiento pasivo explota los patrones de distribución natural de un vehículo de transferencia *in vivo* sin depender del uso de excipientes o medios adicionales para mejorar el reconocimiento del vehículo de transferencia por parte de las células diana. Por ejemplo, los vehículos de transferencia que están sujetos a fagocitosis por las células del sistema reticuloendotelial es probable que se acumulen en el hígado o el bazo y, en consecuencia, pueden proporcionar medios para dirigir pasivamente el

suministro de las composiciones a tales células diana.

La presente divulgación contempla el direccionamiento activo, que implica el uso de excipientes adicionales, denominados en el presente documento "ligandos dirigidos" que pueden unirse (ya sea de forma covalente o no covalente) al vehículo de transferencia para estimular la localización de dicho vehículo de transferencia en ciertas células diana o tejidos diana. Por ejemplo, el direccionamiento puede estar mediado por la inclusión de uno o más ligandos dirigidos endógenos (por ejemplo, apolipoproteína E) en o sobre el vehículo de transferencia para estimular la distribución a las células o tejidos diana. El reconocimiento del ligando dirigido por los tejidos diana facilita activamente la distribución tisular y la captación celular del vehículo de transferencia y/o su contenido en las células y tejidos diana (por ejemplo, la inclusión de un ligando dirigido a la apolipoproteína-E en o sobre el vehículo de transferencia fomenta reconocimiento y unión del vehículo de transferencia a los receptores endógenos de lipoproteínas de baja densidad expresados por los hepatocitos). Como se proporciona en el presente documento, la composición puede comprender un ligando capaz de potenciar la afinidad de la composición por la célula diana. Los ligandos dirigidos pueden unirse a la bicapa externa de la partícula lipídica durante la formulación o después de la formulación. Estos métodos son bien conocidos en la técnica. Adicionalmente, algunas formulaciones de partículas lipídicas pueden emplear polímeros fusogénicos como PEAA, hemaglutinina, otros lipopéptidos (véase la solicitud de patente de EE. UU. N.º 08/835,281 y 60/083,294 y otras características útiles para *in vivo* y/o administración intracelular. En otros algunos ejemplos, las composiciones de la presente divulgación demuestran eficacias de transfección mejoradas y/o demuestran una selectividad mejorada hacia células diana o tejidos de interés. Por lo tanto, se contemplan composiciones que comprenden uno o más ligandos (por ejemplo, péptidos, aptámeros, oligonucleótidos, una vitamina u otras moléculas) que son capaces de potenciar la afinidad de las composiciones y su contenido de ácido nucleico por las células o tejidos diana. Los ligandos adecuados pueden unirse o unirse opcionalmente a la superficie del vehículo de transferencia. En algunas realizaciones, el ligando dirigido puede abarcar la superficie de un vehículo de transferencia o estar encapsulado dentro del vehículo de transferencia. Los ligandos adecuados se seleccionan en función de sus propiedades físicas, químicas o biológicas (por ejemplo, afinidad selectiva y/o reconocimiento de marcadores o características de la superficie celular diana). Los sitios diana específicos de células y su ligando de direccionamiento correspondiente pueden variar ampliamente. Los ligandos dirigidos adecuados se seleccionan de manera que se exploten las características únicas de una célula diana, permitiendo así que la composición discrimine entre células diana y no diana. Por ejemplo, las composiciones de la presente descripción pueden incluir marcadores de superficie (por ejemplo, apolipoproteína-B o apolipoproteína-E) que potencian selectivamente el reconocimiento o la afinidad por los hepatocitos (por ejemplo, mediante el reconocimiento mediado por receptores y la unión a tales marcadores de superficie). Además, cabría esperar que el uso de galactosa (por ejemplo, como un derivado de galactosa tal como N-acetilgalactosamina) como ligando dirigido dirija las composiciones de la presente invención a los hepatocitos parenquimatosos, o alternativamente, el uso de restos de azúcar que contienen manosa como ligando dirigido dirijan las composiciones de la presente invención a las células endoteliales del hígado (por ejemplo, restos de azúcar que contienen manosa que pueden unirse preferentemente al receptor de asialoglicoproteína presente en los hepatocitos). (Véase Hillery A M, *et al.* "Drug Delivery and Targeting: For Pharmacists and Pharmaceutical Scientists" (2002) Taylor y Francis, Inc.) La presentación de dichos ligandos dirigidos que se han conjugado con fracciones presentes en el vehículo de transferencia (por ejemplo, una nanopartícula lipídica), por lo tanto, facilita el reconocimiento y la absorción de las composiciones de la presente divulgación en células y tejidos diana. Los ejemplos de ligandos dirigidos adecuados incluyen uno o más péptidos, proteínas, aptámeros, moléculas pequeñas, vitaminas y oligonucleótidos.

En algunos ejemplos, la unión del factor de transcripción a la región promotora y/o potenciadora del bucle de ADN del promotor potenciador (por ejemplo, un sitio de unión del factor de transcripción en una región promotora o potenciadora) se modula poniendo en contacto la célula con una composición o compuesto que comprende una molécula pequeña, péptido, el polipéptido, ácido nucleico y/u oligonucleótido. En algunos ejemplos, la unión del factor de transcripción se puede reducir en aproximadamente un 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más. En algunas realizaciones, la unión del factor de transcripción se puede aumentar en aproximadamente un 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 110 %, 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, 190 %, 200 %, 210 %, 220 %, 230 %, 240 %, 250 %, 260 %, 270 %, 280 %, 290 %, 300 %, 310 %, 320 %, 330 %, 340 %, 350 %, 360 %, 370 %, 380 %, 390 %, 400 %, 500 %, 600 % o más. En algunos ejemplos, la unión del factor de transcripción se puede aumentar o disminuir aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces o más.

En algunos ejemplos, una composición que modula la unión del factor de transcripción a la región promotora y/o potenciadora (por ejemplo, un sitio de unión del factor de transcripción en una región promotora o potenciadora) del bucle de ADN potenciador-promotor comprende un polipéptido o un ácido nucleico que codifica un polipéptido. En algunos ejemplos, el polipéptido es una variante del factor de transcripción (por ejemplo, variante YY1) con aumento, disminución o ausencia de actividad de unión al promotor y/o región potenciadora (por ejemplo, sitio de unión al factor de transcripción del potenciador o del promotor) del bucle de ADN potenciador-promotor. En algunos ejemplos, el polipéptido (por ejemplo, la variante del factor de transcripción, variante YY1) tiene una mayor afinidad (por ejemplo, afinidad de multimerización) por el factor de transcripción que por el propio factor de transcripción. En algunos ejemplos, el polipéptido (por ejemplo, variante del factor de transcripción, variante YY1) ha disminuido la actividad de unión a la región promotora y/o potenciadora (por ejemplo, sitio de unión del factor de transcripción promotor o

potenciador) del bucle de ADN potenciador-promotor y mayor afinidad (por ejemplo, afinidad de multimerización) por el factor de transcripción.

En algunos ejemplos, el polipéptido es una variante del factor de transcripción (por ejemplo, variante YY1) con aumento, disminución o ausencia de afinidad (por ejemplo, afinidad de multimerización) por un factor de transcripción, y actividad de unión a una región promotora y/o potenciadora (por ejemplo, sitio de unión de la transcripción promotor o potenciador) de un bucle de ADN potenciador-promotor. En algunos ejemplos, el polipéptido tiene disminución o ausencia de afinidad (por ejemplo, afinidad de multimerización) por un factor de transcripción (por ejemplo, YY1) y actividad de unión a una región promotora y/o potenciadora igual o aumentada (por ejemplo, sitio de unión de la transcripción promotor o potenciador) de un bucle de ADN potenciador-promotor. En algunos ejemplos, el polipéptido es una variante del factor de transcripción (por ejemplo, variante YY1) que tiene una afinidad aumentada (por ejemplo, afinidad de multimerización) por el factor de transcripción (por ejemplo, YY1).

En algunos ejemplos, el polipéptido se une a un sitio de unión del factor de transcripción promotor o potenciador y modula la unión del factor de transcripción. En algunos ejemplos, el polipéptido tiene aumento, disminución o la misma afinidad de unión a un sitio de unión del factor de transcripción promotor o potenciador como el factor de transcripción afín.

En algunos ejemplos, una composición que modula la unión del factor de transcripción al promotor y/o región potenciadora del bucle de ADN potenciador-promotor comprende un anticuerpo. En algunos ejemplos, el anticuerpo se une al factor de transcripción (p. ej., YY1) y modula (por ejemplo, disminuye) la unión a un sitio de unión del factor de transcripción (por ejemplo, el sitio de unión de YY1) en la región promotora y/o potenciadora del bucle de ADN potenciador-promotor. En algunos ejemplos, el anticuerpo se une a un sitio de unión del factor de transcripción (por ejemplo, el sitio de unión de YY1) en la región promotora y/o potenciadora del bucle de ADN del promotor potenciador y modula (por ejemplo, disminuye) la unión del factor de transcripción afín (por ejemplo, YY1).

El término "anticuerpo" abarca inmunoglobulinas y derivados de las mismas que contienen un dominio de inmunoglobulina capaz de unirse a un antígeno. Un anticuerpo puede tener su origen en cualquier especie de mamífero o ave, por ejemplo, ser humano, roedores (por ejemplo, ratón, conejo), cabra, pollo, camélido, etc., o se puede generar usando, por ejemplo, presentación en fagos. El anticuerpo puede ser un miembro de cualquier clase de inmunoglobulina, por ejemplo, IgG, IgM, IgA, IgD, IgE o subclases de las mismas tales como IgG1, IgG2, etc. En la presente divulgación, "anticuerpo" se refiere a un fragmento de anticuerpo tal como un Fab', F(ab')₂, scFv (variable de cadena sencilla) u otro fragmento que conserva un sitio de unión a antígeno o un fragmento scFv producido de manera recombinante, incluyendo fragmentos producidos de manera recombinante. Un anticuerpo puede ser monovalente, bivalente o multivalente en varias realizaciones. En algunos ejemplos, un anticuerpo es un anticuerpo de dominio único, por ejemplo, que comprende un dominio variable (V_H) de un anticuerpo de cadena pesada. El anticuerpo puede ser un anticuerpo quimérico o "humanizado", que puede generarse utilizando métodos conocidos en la técnica. Un anticuerpo puede ser policlonal o monoclonal, aunque pueden preferirse los anticuerpos monoclonales. Se conocen en la técnica métodos para producir anticuerpos que se unen específicamente con prácticamente cualquier molécula de interés. En algunos casos, el anticuerpo es un intracuerpo, que puede expresarse intracelularmente. En algunos ejemplos, la composición comprende un anticuerpo monocatenario y un dominio de transducción de proteína (por ejemplo, como un polipéptido de fusión).

En algunos ejemplos, la composición que modula el factor de transcripción (por ejemplo, YY1) al promotor y/o región potenciadora del bucle de ADN potenciador-promotor comprende un ácido nucleico de interferencia. En algunos ejemplos, el ácido nucleico de interferencia se une a una región promotora y/o potenciadora del genoma de la célula e inhibe la unión de un factor de transcripción (por ejemplo, YY1) al promotor o potenciador. En algunos ejemplos, el ácido nucleico de interferencia se une a un sitio de unión del factor de transcripción (por ejemplo, el sitio de unión de YY1) en una región promotora y/o potenciadora e inhibe la unión de un factor de transcripción (por ejemplo, YY1) al sitio de unión del factor de transcripción. En algunos ejemplos, el ácido nucleico de interferencia se une al factor de transcripción (por ejemplo, YY1) e inhibe la unión del factor de transcripción a una región potenciadora o promotora del genoma de una célula. En algunos ejemplos, el ácido nucleico de interferencia se une a un factor de transcripción (por ejemplo, YY1) e inhibe la multimerización (por ejemplo, dimerización) del factor de transcripción.

Un ácido nucleico de interferencia se puede producir en cualquiera de una variedad de formas en diversas realizaciones. Por ejemplo, las hebras de ácido nucleico se pueden sintetizar químicamente (por ejemplo, usando técnicas estándar de síntesis de ácido nucleico) o se pueden producir en células o usando un sistema de transcripción *in vitro*.

En algunos ejemplos, el ácido nucleico de interferencia es una secuencia de ARN natural, una secuencia de ARN modificada (por ejemplo, una secuencia de ARN que comprende una o más bases modificadas), una secuencia de ARN sintética, o una combinación de las mismas. Como se usa en el presente documento, un "ARN modificado" es un ARN que comprende una o más modificaciones del ARN (por ejemplo, ARN que comprende una o más bases y/o modificaciones no estándar y/o no naturales y/o del esqueleto y/o del azúcar). Los métodos para modificar bases de ARN son bien conocidos en la técnica. Los ejemplos de dichas bases modificadas incluyen las contenidas en los nucleósidos 5-metilcitidina (5mC), pseudouridina (ψ), 5-metiluridina, 2'-O-metiluridina, 2-tiouridina, N-6-metiladenosina,

hipoxantina, dihidrouridina (D), inosina (I) y 7-metilguanosa (m7G). Cabe señalar que cualquier número de bases, azúcares o enlaces internucleosídicos en una secuencia de ARN se pueden modificar en diversas realizaciones. Debe entenderse además que pueden usarse combinaciones de diferentes modificaciones.

5 En algunos casos, el ácido nucleico es un morfolino. Los morfolinós son típicamente moléculas sintéticas, de aproximadamente 25 bases de longitud. Los morfolinós tienen bases de ácido nucleico estándar, pero esas bases están unidas a anillos de morfolina en lugar de anillos de desoxirribosa y están unidas a través de grupos fosforodiamidato en lugar de fosfatos.

10 En algunos aspectos de la invención, la formación del bucle de ADN potenciador-promotor se modula modulando la multimerización (por ejemplo, dimerización) de un factor de transcripción YY1 en la célula. Puede usarse cualquier método de modulación de la multimerización (por ejemplo, dimerización) divulgado en el presente documento. En algunos ejemplos, la multimerización (por ejemplo, la dimerización) se puede reducir en aproximadamente un 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más. En algunas realizaciones, la multimerización (por ejemplo, dimerización) se puede aumentar en aproximadamente un 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 110 %, 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, 190 %, 200 %, 210 %, 220 %, 230 %, 240 %, 250 %, 260 %, 270 %, 280 %, 290 %, 300 %, 310 %, 320 %, 330 %, 340 %, 350 %, 360 %, 370 %, 380 %, 390 %, 400 %, 500 %, 600 % o más. En algunos ejemplos, la multimerización (por ejemplo, la dimerización) se puede 20 aumentar o disminuir aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces o más.

En algunas realizaciones, la multimerización (por ejemplo, dimerización) se modula poniendo en contacto la célula con una molécula pequeña, péptido, el polipéptido, ácido nucleico y/u oligonucleótido. En algunos ejemplos, la molécula pequeña, péptido, el polipéptido, el ácido nucleico y/o el oligonucleótido se une al factor de transcripción (por ejemplo, YY1) e inhibe la multimerización (por ejemplo, dimerización). En algunos ejemplos, el factor de transcripción es una 25 proteína dedos de cinc. Según la invención, el factor de transcripción es YY1.

En la presente divulgación, la modulación de la expresión de uno o más genes comprende modular la expresión de un factor de transcripción (p. ej., YY1) que se une a una región potenciadora y promotora del ADN (p. ej., un sitio de 30 unión del factor de transcripción en una región promotora o potenciadora) y forma un bucle de ADN potenciador-promotor. En algunos ejemplos, el método comprende poner en contacto una célula con una molécula pequeña o ácido nucleico que reduce la expresión del factor de transcripción (por ejemplo, YY1). En algunos ejemplos, la expresión del factor de transcripción (por ejemplo, YY1) está disminuida. La expresión del factor de transcripción se puede disminuir en aproximadamente un 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más. En algunas 35 realizaciones, la expresión del factor de transcripción (por ejemplo, YY1) está aumentada. La expresión del factor de transcripción se puede aumentar en aproximadamente un 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 110 %, 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, 190 %, 200 %, 210 %, 220 %, 230 %, 240 %, 250 %, 260 %, 270 %, 280 %, 290 %, 300 %, 310 %, 320 %, 330 %, 340 %, 350 %, 360 %, 370 %, 380 %, 390 %, 400 %, 500 %, 600 % o más. En algunos ejemplos, la expresión del factor de transcripción (por ejemplo, YY1) puede 40 aumentar o disminuir aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces o más.

En algunos casos, el método comprende poner en contacto una célula con un ácido nucleico que reduce la expresión del factor de transcripción (por ejemplo, YY1). El ácido nucleico es un polímero de nucleótidos de ribosa o de 45 nucleótidos de desoxirribosa que tiene más de tres nucleótidos de longitud. El ácido nucleico puede incluir nucleótidos naturales; sintéticos, modificados o pseudonucleótidos, tales como fosforotiolatos; así como nucleótidos que tienen un marcador detectable tal como P³², biotina, colorante fluorescente o digoxigenina. Un ácido nucleico que puede reducir la expresión de un factor de transcripción puede ser completamente complementario al ácido nucleico del factor de transcripción. Como alternativa, puede permitirse cierta variabilidad entre las secuencias.

50 El ácido nucleico de la presente divulgación puede hibridarse con el ácido nucleico del factor de transcripción (por ejemplo, YY1) en condiciones intracelulares o en condiciones de hibridación rigurosas. Los ácidos nucleicos de la invención son suficientemente complementarios a los ácidos nucleicos del factor de transcripción (por ejemplo, YY1) para inhibir la expresión del factor de transcripción en una o ambas condiciones. Las condiciones intracelulares se refieren a condiciones tales como temperatura, pH y concentraciones de sal que normalmente se encuentran dentro 55 de una célula, por ejemplo, una célula de mamífero.

Por lo general, las condiciones rigurosas de hibridación se seleccionan para que sean aproximadamente 5 °C más bajas que el punto de fusión térmica (T_m) para la secuencia específica a una fuerza iónica y un pH definidos. Sin 60 embargo, las condiciones rigurosas abarcan temperaturas en el rango de aproximadamente 1 °C a aproximadamente 20 °C por debajo del punto de fusión térmica de la secuencia seleccionada, dependiendo del grado deseado de rigurosidad como se califica aquí de otro modo. Los ácidos nucleicos que comprenden, por ejemplo, 2, 3, 4 o 5 o más tramos de nucleótidos contiguos que son precisamente complementarios a una secuencia codificante del factor de transcripción, cada uno separado por un tramo de nucleótidos contiguos que no son complementarios a las secuencias 65 de codificación adyacentes, pueden inhibir la función de un ácido nucleico del factor de transcripción. En general, cada tramo de nucleótidos contiguos tiene una longitud de al menos 4, 5, 6, 7 u 8 o más nucleótidos. Las secuencias

intermedias no complementarias pueden tener una longitud de 1, 2, 3 o 4 nucleótidos. Un experto en la técnica puede usar fácilmente el punto de fusión calculado de un ácido nucleico hibridado con un ácido nucleico sentido para estimar el grado de desajuste que será tolerado para inhibir la expresión de un ácido nucleico diana particular. Los ácidos nucleicos de la invención incluyen, por ejemplo, una ribozima o una molécula de ácido nucleico antisentido.

Una molécula de ácido nucleico antisentido puede ser de cadena simple o doble (por ejemplo, un ARN de interferencia pequeño (ARNip)) y puede funcionar de manera dependiente de enzimas o mediante bloqueo estérico. Las moléculas antisentido que funcionan de manera dependiente de enzimas incluyen formas dependientes de la actividad de la RNasa H para degradar el ARNm diana. Estos incluyen ADN monocatenario, moléculas de ARN y fosforotioato, así como el sistema ARNi/ARNip de doble cadena que implica el reconocimiento del ARNm diana a través del apareamiento de cadenas sentido-antisentido seguido de la degradación del ARNm diana por el complejo de silenciamiento inducido por ARN. El antisentido de bloqueo estérico, independiente de la RNasa-H, interfiere con la expresión génica u otros procesos celulares dependientes del ARNm al unirse a un ARNm diana e interferir con otros procesos como la traducción. El antisentido de bloqueo estérico incluye 2'-O alquilo (generalmente en quimeras con antisentido dependiente de RNasa-H), ácido peptidonucleico (PNA, *peptide nucleic acid*), ácido nucleico bloqueado (LNA) y morfolino antisentido.

Los pequeños ARNp de interferencia, por ejemplo, pueden usarse para reducir específicamente el nivel de ARNm que codifica un factor de transcripción (por ejemplo, YY1) y/o reducir la traducción del ARNm que codifica un factor de transcripción (por ejemplo, YY1) de manera que se reduce el nivel de factor de transcripción (por ejemplo, YY1). Los ARNip median el silenciamiento génico postranscripcional de una manera específica de secuencia. Véase, por ejemplo, Carthew *et al.*, "Origins and Mechanisms of miRNAs and siRNAs," Cell, Volume 136, Edición 4, p642-655, 20 de febrero de 2009. Una vez incorporado en un complejo de silenciamiento inducido por ARN, el ARNip media en la escisión del transcrito de ARNm endógeno homólogo al guiar el complejo al transcrito de ARNm homólogo, que luego es escindido por el complejo. El ARNip puede ser homólogo a cualquier región del transcrito de ARNm del factor de transcripción (por ejemplo, YY1). La región de homología puede tener una longitud de 30 nucleótidos o menos, menos de 25 nucleótidos, de aproximadamente 21 a 23 nucleótidos de longitud o menos, por ejemplo, 19 nucleótidos de longitud. El ARNip suele ser de doble cadena y puede tener salientes nucleotídicos en 3'. Los salientes en 3' pueden tener hasta aproximadamente 5 o 6 salientes nucleotídicos en 3', por ejemplo, dos salientes nucleotídicos en 3', tales como, dinucleótidos UU salientes en 3', por ejemplo. En algunas realizaciones, los ARNip pueden no incluir ningún saliente nucleotídico en 3'. Los métodos para diseñar ARNip son conocidos por los expertos en la técnica. Véase, por ejemplo, Elbashir *et al.* Nature 411: 494-498 (2001); Harborth *et al.* Antisense Nucleic Acid Drug Dev. 13: 83-106 (2003). En algunos ejemplos, se selecciona un sitio de destino que comienza con AA, tiene salientes UU en 3' para las hebras de ARNip sentido y antisentido y tiene un contenido de G/C aproximado del 50 %. En algunos ejemplos, se selecciona un sitio objetivo que es único para uno o más ARNm objetivo y no en otros ARNm cuya degradación o inhibición de la traducción no se desea. Los ARNip pueden sintetizarse químicamente, crearse mediante transcripción *in vitro* o expresarse a partir de un vector de expresión de ARNip o un casete de expresión de PCR. Véase, por ejemplo, <http://www.thermofisher.com/us/en/home/life-science/rnai.html>.

Cuando un ARNip se expresa a partir de un vector de expresión o un casete de expresión de PCR, el inserto que codifica el ARNip puede expresarse como un transcrito de ARN que se pliega en una horquilla de ARNip. De este modo, el transcrito de ARN puede incluir una secuencia de ARNip sentido que está unida a su secuencia de ARNip antisentido complementaria inversa mediante una secuencia espaciadora que forma el bucle de la horquilla así como una cadena de U en el extremo 3'. El bucle de la horquilla puede tener cualquier longitud adecuada, por ejemplo, hasta 30 nucleótidos de longitud, por ejemplo, de 3 a 23 nucleótidos de longitud, y puede tener varias secuencias de nucleótidos. También se pueden producir ARNip *in vivo* por escisión de ARN de doble cadena introducido directamente o a través de un transgén o virus. La amplificación por una ARN polimerasa dependiente de ARN puede ocurrir en algunos organismos. El ARNip puede modificarse adicionalmente de acuerdo con cualquier método conocido por los expertos en la materia.

También se puede usar un ácido nucleico inhibidor antisentido para reducir específicamente la expresión del factor de transcripción (por ejemplo, YY1), por ejemplo, inhibiendo la transcripción y/o la traducción. Un ácido nucleico inhibidor antisentido es complementario a un ácido nucleico sentido que codifica un factor de transcripción (por ejemplo, YY1). Por ejemplo, puede ser complementario a la hebra codificante de una molécula de ADNc de doble hebra o complementario a una secuencia de ARNm. Puede ser complementario a una cadena de codificación completa o solo a una parte de la misma. También puede ser complementario a toda o parte de la región no codificante de un ácido nucleico que codifica un factor de transcripción (por ejemplo, YY1). La región no codificante incluye las regiones 5' y 3' que flanquean la región codificante, por ejemplo, las secuencias no traducidas en 5' y 3'. Un ácido nucleico inhibidor antisentido tiene generalmente al menos seis nucleótidos de longitud, pero puede tener una longitud de hasta aproximadamente 8, 12, 15, 20, 25, 30, 35, 40, 45 o 50 nucleótidos. También se pueden usar ácidos nucleicos inhibidores más largos.

Se puede preparar un ácido nucleico inhibidor antisentido utilizando métodos conocidos en la técnica, por ejemplo, por expresión de un vector de expresión que codifica el ácido nucleico inhibidor antisentido o de un casete de expresión. Como alternativa, puede prepararse por síntesis química utilizando nucleótidos naturales, nucleótidos modificados o cualquier combinación de los mismos. En algunos ejemplos, los ácidos nucleicos inhibidores están

hechos de nucleótidos modificados o enlaces no fosfodiéster, por ejemplo, que están diseñados para aumentar la estabilidad biológica del ácido nucleico inhibidor o para aumentar la estabilidad intracelular del dúplex formado entre el ácido nucleico inhibidor antisentido y el ácido nucleico sentido.

5 Los nucleótidos naturales, nucleósidos y las nucleobases incluyen los nucleótidos de ribosa o desoxirribosa adenosina, guanina, citosina, timina y uracilo. Los ejemplos de nucleótidos modificados, nucleósidos y nucleobases incluyen aquellos que comprenden 5-fluorouracilo, 5-bromouracilo, 5-clorouracilo, 5-yodouracilo, hipoxantina, xantina, 4-acetilcitosina, 5-(carboxihidroximetil)uracilo, 5-carboximetilaminometil-2-tiouridina, 5-carboximetilaminometiluracilo, dihidrouracilo, beta-D-galactosilqueosina, inosina, N6-isopenteniladenina, 1-metilguanina, 1-metilinosina, 2,2-dimetilguanina, 2-metiladenina, 2-metilguanina, 3-metilcitosina, 5-metilcitosina, N6-adenina, 7-metilguanina, 5-metilaminometiluracilo, 5-metoxiaminometil-2-tiouracilo, beta-D-manosilqueosina, 5'-metoxicarboximetiluracilo, 5-metoxiuracilo, 2-metil-N6-isopenteniladenina, ácido uracil-5-oxiacético, butoxocina, pseudouracilo, queosina, 2-tiocitosina, 5-metil-2-tiouracilo, 2-tiouracilo, 4-tiouracilo, 5-metiluracilo, éster metílico del ácido uracil-5-oxacético, ácido uracil-5-oxacético, 5-metil-2-tiouracilo, 3-(3-amino-3-N-2-carboxipropil)uracilo, (acp3)w y 2,6-diaminopurina.

15 Por lo tanto, los ácidos nucleicos de la invención pueden incluir nucleótidos modificados, así como nucleótidos naturales tales como combinaciones de nucleótidos de ribosa y desoxirribosa, y un ácido nucleico de la invención puede tener cualquier longitud discutida anteriormente y que sea complementario a las secuencias de ácido nucleico de un factor de transcripción (por ejemplo, YY1).

20 En algunos ejemplos, un ácido nucleico que modula la expresión de un factor de transcripción es un ARN de horquilla pequeña (es decir, ARN de horquilla corta) (ARNhp).

25 El ARNhp es una secuencia de ARN que forma una curva muy cerrada que se puede utilizar para silenciar la expresión génica por medio de la interferencia del ARN. La estructura de horquilla de ARNhp es dividida por la maquinaria celular en un ARNip, que luego se une y escinde el ARNm objetivo. El ARNhp se puede introducir en las células a través de un vector que codifica el ARNhp, donde la región de codificación de ARNhp está unida operativamente a un promotor. El promotor seleccionado permite la expresión del ARNhp. Por ejemplo, el promotor puede ser un promotor U6, que es útil para la expresión continua del ARNhp. El vector puede, por ejemplo, pasar a las células hijas, permitiendo que el silenciamiento del gen sea heredado. Véase, McIntyre G, Fanning G, Design and cloning strategies for constructing shRNA expression vectors, BMC BIOTECHNOL. 6:1 (2006); Paddison *et al.*, Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells, GENES DEV. 16 (8): 948-58 (2002).

35 En algunos ejemplos, un ácido nucleico que modula la expresión de un factor de transcripción (por ejemplo, YY1) es una ribozima. Una ribozima es una molécula de ARN con actividad catalítica y es capaz de escindir un ácido nucleico monocatenario, como un ARNm que tiene una región homóloga. Véase, por ejemplo, Cech, Science 236: 1532-1539 (1987); Cech, Ann. Rev. Biochem. 59:543-568 (1990); Cech, Curr. Opin. Struct. Biol. 2: 605-609 (1992); Couture y Stinchcomb, Trends Genet. 12: 510-515 (1996).

40 Se han desarrollado y descrito en la técnica métodos para diseñar y construir una ribozima que puede escindir una molécula de ARN en trans de una manera muy específica de secuencia. Véase, por ejemplo, Haseloff *et al.*, Nature 334:585-591 (1988). Una ribozima puede dirigirse a un ARN específico modificando por ingeniería genética una región de "hibridación" discreta en la ribozima. La región de hibridación contiene una secuencia complementaria al ARN diana que permite que la ribozima se hibride específicamente con la diana. Véase, por ejemplo, Gerlach *et al.*, documento EP 321201. La secuencia diana puede ser un segmento de aproximadamente 5, 6, 7, 8, 9, 10, 12, 15, 20 o 50 nucleótidos contiguos. Pueden usarse secuencias complementarias más largas para aumentar la afinidad de la secuencia de hibridación por la diana.

50 En algunos ejemplos, la célula de las composiciones y métodos descritos aquí es una célula madre. En algunas realizaciones, la célula es una célula madre embrionaria. El tipo de celda no está limitado y puede ser cualquier celda descrita en el presente documento o conocida en la técnica.

En algunos ejemplos, la modulación de la expresión de uno o más genes por las composiciones y métodos descritos en el presente documento comprende la modulación de la expresión de Oct4, Nanog y/o Sox2. Los genes modulados por los métodos de la invención no están limitados y pueden comprender cualquier gen expresado en un bucle de ADN potenciador-promotor.

60 En algunos ejemplos, los métodos y composiciones descritos en el presente documento pueden usarse para modular la expresión de genes que dependen de la formación de un bucle de ADN potenciador-promotor mediado por un factor de transcripción. En algunos ejemplos, los métodos y composiciones descritos en el presente documento se pueden usar para modular la expresión de genes que dependen de la formación de uno o más bucles de ADN potenciadores-promotores específicos mediados por un factor de transcripción. Por ejemplo, los métodos y composiciones descritos en el presente documento pueden dirigirse específicamente a regiones promotoras y/o potenciadoras exclusivas de uno o más bucles de ADN potenciadores-promotores, modulando así la expresión de genes bajo el control de uno o más bucles de ADN potenciadores-promotores pero no modulando la expresión de genes en otros bucles de ADN potenciadores-promotores dependientes del mismo factor de transcripción.

Tratamiento de enfermedades y afecciones asociadas con la expresión génica aberrante

- En el presente documento se describen métodos para tratar una enfermedad o afección asociada con la expresión génica anómala o la actividad del producto génico anómalo en un sujeto que lo necesite (por ejemplo, un ser humano), que comprende administrar una composición que modula la formación y/o la estabilidad de bucles de ADN potenciadores-promotores, en el que la formación y/o la estabilidad del bucle de ADN del promotor potenciador depende del factor de transcripción (por ejemplo, YY1). Cualquier enfermedad asociada con el aumento o disminución de la expresión o actividad de un gen o producto génico, en la que la expresión del gen está regulada al menos en parte por la formación de un bucle de ADN potenciador-promotor mediado al menos en parte por la multimerización de FT puede tratarse mediante los métodos descritos en el presente documento. En algunos ejemplos, las enfermedades son enfermedades neurodegenerativas, trastornos del neurodesarrollo, enfermedades autoinmunitarias, enfermedades metabólicas, etc. En algunos ejemplos, la enfermedad es cáncer.
- En algunos casos, la formación del bucle de ADN potenciador-promotor se modula modulando la unión de un factor de transcripción (por ejemplo, YY1) a un promotor y/o región potenciadora del bucle de ADN potenciador-promotor (por ejemplo, un sitio de unión del factor de transcripción en un promotor o potenciador región). La unión se puede modular mediante cualquier método o composición descritos en el presente documento. En algunos ejemplos, la unión del factor de transcripción (por ejemplo, YY1) puede reducirse en aproximadamente un 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más. En algunas realizaciones, la unión del factor de transcripción (por ejemplo, YY1) se puede aumentar en aproximadamente un 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 110 %, 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, 190 %, 200 %, 210 %, 220 %, 230 %, 240 %, 250 %, 260 %, 270 %, 280 %, 290 %, 300 %, 310 %, 320 %, 330 %, 340 %, 350 %, 360 %, 370 %, 380 %, 390 %, 400 %, 500 %, 600 % o más. En algunos ejemplos, la unión del factor de transcripción (por ejemplo, YY1) se puede aumentar o disminuir aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces o más.
- En algunos casos, la unión se modula modificando la región promotora y/o potenciadora (por ejemplo, un sitio de unión del factor de transcripción en una región promotora o potenciadora). La modificación de la región promotora o potenciadora puede realizarse mediante cualquier método o composición divulgados en el presente documento.
- En algunos casos, la modificación comprende modificar la metilación de la región promotora y/o potenciadora (por ejemplo, un sitio de unión del factor de transcripción en una región promotora o potenciadora). La modulación de la metilación puede realizarse mediante cualquier método o composición descritos en el presente documento. En algunos ejemplos, la metilación se puede aumentar o disminuir aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces o más. En algunos ejemplos, el grado de metilación se modifica con una nucleasa específica del sitio catalíticamente inactiva y un dominio efector que tiene actividad de metilasa o desmetilasa como se describe en el presente documento o se conoce en la técnica. En algunas realizaciones, el dominio efector tiene actividad de desmetilación del ADN y es TetI, ACID A, MBD4, Apobec1, Apobec2, Apobec3, Tdg, Gadd45a, Gadd45b y/o ROS1. En algunas realizaciones, el dominio efector tiene actividad de metilación del ADN y es Dnmt1, Dnmt3a, Dnmt3b, CpG metiltransferasa M.SssI y/o M.EcoHK31I. En algunos ejemplos, la nucleasa específica del sitio catalíticamente inactiva y un dominio efector se fusionan.
- En algunos casos, la modificación comprende modificar la secuencia de nucleótidos de la región promotora y/o potenciadora (por ejemplo, un sitio de unión del factor de transcripción en una región promotora o potenciadora). El método para modificar la secuencia de nucleótidos de la región promotora y/o potenciadora puede ser cualquier composición o método divulgado en el presente documento.
- En algunos ejemplos, la unión del factor de transcripción (por ejemplo, YY1) a la región promotora y/o potenciadora (por ejemplo, un sitio de unión del factor de transcripción en una región promotora o potenciadora) se modula administrando al sujeto una composición que comprende una molécula pequeña, péptido, el polipéptido, ácido nucleico y/u oligonucleótido. Puede usarse cualquier composición para modular la unión del factor de transcripción descrito en el presente documento. En algunas realizaciones, la composición comprende un ácido nucleico de interferencia.
- En algunos ejemplos, la formación del bucle de ADN potenciador-promotor se modula modulando la multimerización (por ejemplo, dimerización) de un factor de transcripción (por ejemplo, YY1) en el sujeto. La modulación de la multimerización (por ejemplo, dimerización) del factor de transcripción (por ejemplo, YY1) puede realizarse mediante cualquier método o composición descritos en el presente documento. En algunos ejemplos, la multimerización (por ejemplo, dimerización) se modula administrando al sujeto una molécula pequeña, péptido, el polipéptido, ácido nucleico y/u oligonucleótido. En algunas realizaciones, la multimerización (por ejemplo, dimerización) se puede reducir en aproximadamente un 10 %, 15 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o más. En algunos ejemplos, la multimerización (por ejemplo, dimerización) se puede aumentar en aproximadamente un 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 100 %, 110 %, 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, 190 %, 200 %, 210 %, 220 %, 230 %, 240 %, 250 %, 260 %, 270 %, 280 %, 290 %, 300 %, 310 %, 320 %, 330 %, 340 %, 350 %, 360 %, 370 %, 380 %, 390 %, 400 %, 500 %, 600 % o más. En algunas realizaciones, la multimerización (por ejemplo,

dimerización) se puede aumentar o disminuir aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 6 veces o más.

En algunos ejemplos, el factor de transcripción es una proteína de cinc. Según la invención, el factor de transcripción es YY1. En algunas realizaciones, la expresión génica aberrante está disminuida. En algunas realizaciones, la expresión génica aberrante está aumentada.

En algunas realizaciones, la enfermedad o afección asociada con la expresión génica aberrante es cáncer. "Cáncer" se usa generalmente para referirse a una enfermedad caracterizada por uno o más tumores, por ejemplo, uno o más tumores malignos o potencialmente malignos. El término "tumor", como se usa en el presente documento, abarca crecimientos anormales que comprenden células que proliferan de forma aberrante. Como se conoce en la materia, los tumores se caracterizan típicamente por una proliferación celular excesiva que no está debidamente regulada (por ejemplo, que no responde normalmente a las influencias y señales fisiológicas que normalmente limitarían la proliferación) y pueden exhibir una o más de las siguientes propiedades: displasia (por ejemplo, falta de diferenciación celular normal, que da como resultado un aumento en el número o proporción de células inmaduras); anaplasia (por ejemplo, mayor pérdida de diferenciación, más pérdida de organización estructural, pleomorfismo celular, anomalías tales como núcleos grandes, hipercromáticos, alta proporción nuclear a citoplasmática, mitosis atípicas, etc.); invasión de tejidos adyacentes (por ejemplo, ruptura de una membrana basal); y/o metástasis. Los tumores malignos tienen una tendencia al crecimiento sostenido y la capacidad de propagarse, por ejemplo, invadir localmente y/o metastatizar regionalmente y/o a lugares distantes, mientras que los tumores benignos a menudo permanecen localizados en el sitio de origen y suelen ser autolimitados en términos de crecimiento. El término "tumor" incluye tumores sólidos malignos, por ejemplo, carcinomas (cánceres que surgen de las células epiteliales), sarcomas (cánceres que surgen de células de origen mesenquimatoso) y crecimientos malignos en los que puede no haber una masa tumoral sólida detectable (por ejemplo, ciertas neoplasias malignas hematológicas). El cáncer incluye, aunque sin limitación: cáncer de mama; cáncer de las vías biliares; cáncer de vejiga; cáncer de cerebro (por ejemplo, glioblastomas, meduloblastomas); cáncer de cuello de útero; coriocarcinoma; cáncer de colon; cáncer endometrial; cáncer de esófago; cáncer gástrico; neoplasias hematológicas que incluyen leucemia linfocítica aguda y leucemia mielógena aguda; leucemia/linfoma linfoblástico agudo de linfocitos T; tricoleucemia; leucemia linfocítica crónica, leucemia mielógena crónica, mieloma múltiple; leucemia/linfoma de linfocitos T/linfoma en el adulto; neoplasias intraepiteliales, incluyendo la enfermedad de Bowen y la enfermedad de Paget; cáncer de hígado; cáncer de pulmón; linfomas, incluida la enfermedad de Hodgkin y linfomas linfocíticos; neuroblastoma; melanoma, cáncer oral, incluyendo carcinoma de células escamosas; cáncer de ovario, incluido el cáncer de ovario que surge de las células epiteliales, células estromales, células germinales y células mesenquimatosas; neuroblastoma, cáncer de páncreas; cáncer de próstata; cáncer de recto; sarcomas, incluyendo angiosarcoma, tumores estromales gastrointestinales, leiomiomasarcoma, rabdomiosarcoma, liposarcoma, fibrosarcoma y osteosarcoma; cáncer renal que incluye carcinoma de células renales y tumor de Wilms; cáncer de piel (incluyendo cáncer basocelular y cáncer escamocelular; cáncer testicular, incluyendo tumores germinales tales como seminoma, no seminoma (teratomas, coriocarcinomas), tumores estromales y tumores de células germinales; cáncer de tiroides, incluyendo adenocarcinoma de tiroides y el carcinoma medular. Se apreciará que una variedad de diferentes tipos de tumores pueden surgir en ciertos órganos, que pueden diferir con respecto a, por ejemplo, las características clínicas y/o patológicas y/o los marcadores moleculares. Los tumores que surgen en una variedad de órganos diferentes se tratan, por ejemplo, en la serie Clasificación de tumores de la OMS, 4.^a ed. o 3.^a ed. (Pathology and Genetics of Tumours series), por la Agencia Internacional para la Investigación del Cáncer (IARC), WHO Press, Ginebra, Suiza. En algunos ejemplos, el cáncer es cáncer de pulmón, cáncer de mama, cáncer de cuello uterino, cáncer de colon, cáncer gástrico, cáncer de riñón, leucemia, cáncer de hígado, linfoma, (por ejemplo, un linfoma no Hodgkin, por ejemplo, linfoma difuso de células B grandes, linfoma de Burkitt) cáncer de ovario, cáncer de páncreas, cáncer de próstata, cáncer de recto, sarcoma, cáncer de piel, cáncer testicular o cáncer uterino. El tipo de cáncer no está limitado. En algunos ejemplos, el cáncer exhibe una expresión génica aberrante. En algunos ejemplos, el cáncer exhibe actividad de producto génico aberrante. En algunos ejemplos, el cáncer expresa un producto génico a un nivel normal pero alberga una mutación que altera su actividad. En el caso de un oncogén que tiene una actividad anormalmente aumentada, los métodos de la invención se pueden usar para reducir la expresión del oncogén. En el caso de un gen supresor de tumores que tiene una actividad anormalmente reducida (por ejemplo, debido a una mutación), los métodos descritos en el presente documento pueden usarse para aumentar la expresión del gen supresor de tumores.

Un cáncer puede estar asociado con una mayor expresión de un oncogén y/o una menor expresión de un gen de supresión tumoral. En algunos ejemplos, un método de la presente descripción comprende la disminución de la expresión de un oncogén que se sobreexpresa en el cáncer al reducir la formación y/o el mantenimiento de un bucle entre un potenciador y el promotor del oncogén. En algunos ejemplos, un método de la presente descripción comprende aumentar la expresión de un gen supresor de tumores aumentando la formación y/o el mantenimiento de un bucle de ADN entre un potenciador y el promotor del gen supresor de tumores. Los oncogenes y los genes supresores de tumores se conocen en la técnica y se enumeran en bases de datos disponibles públicamente.

En algunos ejemplos, los métodos y composiciones de la presente descripción se usan en combinación con otras composiciones o métodos. En algunos ejemplos, los métodos descritos en el presente documento se usan para tratar enfermedades (por ejemplo, cáncer) en combinación con otros agentes (por ejemplo, agentes anticancerosos) o terapias (por ejemplo, radiación).

En algunos ejemplos, un método de la presente divulgación puede comprender analizar una muestra obtenida de un tumor, identificar uno o más genes que se expresan de forma aberrante o que codifica un producto génico con actividad aberrante en el tumor, y modular la expresión del gen usando una composición o método descrito en el presente documento.

La administración de las composiciones descritas en el presente documento puede ser por cualquier vía (por ejemplo, oral, intravenosa, intraperitoneal, alimentación forzada, tópica, transdérmica, intramuscular, enteral, subcutánea), puede ser sistémica o local, puede incluir cualquier dosis (por ejemplo, desde aproximadamente 0,01 mg/kg hasta aproximadamente 500 mg/kg), puede implicar una sola dosis o múltiples dosis. En algunas realizaciones, la administración se puede realizar mediante administración directa a un tejido u órgano (por ejemplo, piel, corazón, hígado, pulmón, riñón, cerebro, ojo, músculo, hueso, nervio) o tumor. El (los) ácido(s) nucleico(s) o la(s) proteína(s) pueden estar físicamente asociados con, por ejemplo, encapsulados, por ejemplo, en partículas que contienen lípidos, por ejemplo, nanopartículas lipídicas sólidas, liposomas, partículas poliméricas (por ejemplo, partículas de PLGA). En algunas realizaciones, se pueden administrar uno o más ácidos nucleicos usando un vector (por ejemplo, un vector viral como un vector adenoviral, vector lentiviral o vector de virus adenoasociado). En algunos ejemplos, uno o más ácidos nucleicos, proteínas y/o vectores pueden combinarse con un vehículo farmacéuticamente aceptable para producir una composición farmacéutica, que puede administrarse a un sujeto.

En algunos ejemplos, un ácido nucleico, el polipéptido, el anticuerpo o la partícula pueden estar dirigidos a células de un tipo particular, por ejemplo, células cancerosas de un tipo particular o que expresan un marcador de superficie celular particular. Por ejemplo, un ácido nucleico, una proteína o una partícula que comprende un ácido nucleico o vector puede comprender o conjugarse con un resto dirigido que se une a un marcador expresado en la superficie de una célula diana (por ejemplo, se une a un antígeno tumoral o un receptor expresado por la célula diana). Un resto dirigido puede comprender, por ejemplo, un anticuerpo, o parte de unión a antígeno del mismo, una proteína diseñada capaz de unión específica, un aptámero de ácido nucleico, un ligando, etc.

En algunos ejemplos, los ácidos nucleicos que codifican uno o más componentes (por ejemplo, nucleasa específica del sitio, nucleasa específica del sitio catalíticamente inactiva, dominio efector, proteína de fusión de dominio efector-nucleasa específica de sitio catalíticamente inactiva, una o más secuencias guía, uno o más ácidos nucleicos) son liberados por uno o más vectores virales, por ejemplo, un vector retroviral tal como un vector lentiviral o un vector retroviral gamma, o un vector adenoviral o AAV.

Las composiciones divulgadas en el presente documento utilizadas para modular la multimerización de FT y/o la unión de FT a un potenciador o promotor pueden dirigirse a un tipo de célula particular, tejido u órgano de interés. En algunas realizaciones, el agente comprende una fracción dirigida o se administra utilizando un vehículo de administración que comprende un resto dirigido. Un resto dirigido puede, por ejemplo, comprender un anticuerpo o ligando que se une a una proteína (por ejemplo, un receptor) presente en la superficie de una célula diana de interés. En algunos ejemplos, el resto dirigido está presente en la superficie de una célula cancerosa.

Las composiciones y los compuestos descritos en el presente documento pueden administrarse en una composición farmacéutica. Además del agente activo, las composiciones farmacéuticas comprenden normalmente un vehículo farmacéuticamente aceptable. La expresión "vehículo farmacéuticamente aceptable", como se usa en el presente documento, significa uno o más vehículos sólidos o líquidos compatibles, cargas, diluyentes o sustancias encapsulantes que son adecuadas para la administración a un ser humano o un animal no humano. En ejemplos preferidos, un vehículo farmacéuticamente aceptable es un material no tóxico que no interfiere con la eficacia de la actividad biológica de los principios activos. El término "compatible", como se usa en el presente documento, significa que los componentes de las composiciones farmacéuticas pueden combinarse con un agente y entre sí, de tal manera que no haya ninguna interacción que reduzca sustancialmente la eficacia farmacéutica de la composición farmacéutica en situaciones de uso ordinario. Los vehículos farmacéuticamente aceptables deben tener una pureza suficientemente alta y una toxicidad suficientemente baja para hacerlos adecuados para la administración al ser humano o al animal no humano que se esté tratando.

Algunos ejemplos de sustancias que pueden servir como vehículos farmacéuticamente aceptables son agua apirógena; solución salina isotónica; soluciones de tampón fosfato; azúcares, tales como lactosa, glucosa y sacarosa; almidones, tales como almidón de maíz y almidón de patata; celulosa y sus derivados, tales como carboximetilcelulosa sódica, etilcelulosa, acetato de celulosa; tragacanto en polvo; malta; gelatina; talco; ácido esteárico; estearato de magnesio; sulfato de calcio; aceites vegetales, tales como aceite de cacahuete, aceite de semilla de algodón, aceite de sésamo, aceite de oliva, aceite de maíz y aceite de teobroma; polioles, tales como propilenglicol, glicerina, sorbitol, manitol y polietilenglicol; azúcar; ácido alginico; manteca de cacao (base para supositorios); emulsionantes, tales como Tweens; así como otras sustancias compatibles no tóxicas usadas en formulaciones farmacéuticas. También pueden estar presentes agentes humectantes y lubricantes como laurilsulfato de sodio, así como agentes colorantes, agentes saporíferos, excipientes, agentes de formación de comprimidos, estabilizantes, antioxidantes y conservantes. Se apreciará que una composición farmacéutica puede contener múltiples vehículos diferentes farmacéuticamente aceptables.

Un vehículo farmacéuticamente aceptable, empleado junto con los compuestos descritos en el presente documento, se usa a una concentración o cantidad suficiente para proporcionar una relación práctica entre el tamaño y la dosis. Los vehículos farmacéuticamente aceptables, en total, pueden, por ejemplo, comprender desde aproximadamente el 60 % hasta aproximadamente el 99,99999 % en peso de las composiciones farmacéuticas, por ejemplo, de aproximadamente el 80 % a aproximadamente el 99,99 %, por ejemplo, de aproximadamente el 90 % a aproximadamente el 99,95 %, de aproximadamente el 95 % a aproximadamente el 99,9 % o de aproximadamente el 98 % a aproximadamente el 99 %.

En la técnica se conocen bien vehículos farmacéuticamente aceptables, adecuados para la preparación de formas farmacéuticas unitarias para administración oral y aplicación tópica. Su selección dependerá de cuestiones secundarias tales como el sabor, el coste y/o la estabilidad de almacenamiento, que no son críticas para los fines de invención en cuestión y un experto en la materia puede elaborarlas sin dificultad.

Las composiciones farmacéuticamente aceptables pueden incluir diluyentes, cargas, sales, tampones, estabilizantes, solubilizantes y otros materiales que son bien conocidos en la técnica. La elección del vehículo farmacéuticamente aceptable que se utilizará junto con los compuestos de la presente invención, se determina básicamente por la forma en la que se va a administrar el compuesto. Los vehículos farmacéuticamente aceptables de ejemplo para péptidos en particular se describen en la Patente de EE. UU. No. 5,211,657. Dichas preparaciones pueden contener habitualmente sal, agentes tamponantes, conservantes, vehículos compatibles y, opcionalmente, otros agentes terapéuticos. Cuando se usan en medicina, las sales deben ser farmacéuticamente aceptables, pero en ciertas realizaciones pueden usarse convenientemente sales no farmacéuticamente aceptables para preparar sales de las mismas farmacéuticamente aceptables. Dichas sales farmacológica y farmacéuticamente aceptables incluyen, aunque sin limitación, las preparadas a partir de los siguientes ácidos: clorhídrico, bromhídrico, sulfúrico, nítrico, fosfórico, maleico, acético, salicílico, cítrico, fórmico, malónico, succínico y similares. Además, se pueden preparar sales farmacéuticamente aceptables como sales de metales alcalinos o alcalinotérreos, tales como sales de sodio, potasio o calcio. También se entenderá que se puede proporcionar un compuesto como un profármaco farmacéuticamente aceptable o se puede usar un metabolito activo. Por otro lado, se apreciará que los agentes pueden modificarse, por ejemplo, con restos dirigidos, restos que aumenten su absorción, la semivida biológica (por ejemplo, pegilación), etc.

Los agentes pueden administrarse en soluciones farmacéuticamente aceptables, que habitualmente pueden contener concentraciones farmacéuticamente aceptables de sal, agentes tamponantes, conservantes, vehículos compatibles, adyuvantes y, opcionalmente, otros ingredientes terapéuticos.

Los agentes pueden formularse en preparaciones en estado sólido, semisólido, formas líquidas o gaseosas, tales como comprimidos, cápsulas, polvos, gránulos, pomadas, soluciones, depósitos, inhalantes e inyecciones, y las formas habituales de administración oral, administración parenteral o quirúrgica. La invención también incluye composiciones farmacéuticas que están formuladas para administración local, tal como por implantes.

Las composiciones adecuadas para la administración oral pueden presentarse como unidades individuales, tales como cápsulas, comprimidos, pastillas para chupar, conteniendo cada una de ellas una cantidad predeterminada del agente activo. Otras composiciones incluyen suspensiones en líquidos acuosos o líquidos no acuosos, tales como un jarabe, elixir o una emulsión.

En algunos ejemplos, los agentes pueden administrarse directamente a un tejido, por ejemplo, un tejido en el que se encuentran las células cancerosas o uno en el que es probable que surja un cáncer. La administración directa al tejido puede lograrse mediante inyección directa. Los agentes pueden administrarse una vez o, alternativamente, pueden administrarse en una pluralidad de administraciones. Si se administran varias veces, los péptidos pueden administrarse a través de diferentes rutas. Por ejemplo, las primeras (o las primeras) administraciones pueden realizarse directamente en el tejido afectado, mientras que las administraciones posteriores pueden ser sistémicas.

Para la administración oral, las composiciones se pueden formular fácilmente combinando el o los agentes activos con vehículos farmacéuticamente aceptables bien conocidos en la técnica. Tales vehículos posibilitan que los agentes se formulen como comprimidos, píldoras, grageas, cápsulas, líquidos, geles, jarabes, pastas, suspensiones y similares, para ingestión oral por un sujeto a tratar. Pueden obtenerse preparaciones farmacéuticas para su uso oral como excipiente sólido, moliendo opcionalmente una mezcla resultante y procesando la mezcla de gránulos, después de añadir adyuvantes adecuados, si se desea, para obtener comprimidos o núcleos de grageas. Son excipientes adecuados, en particular, cargas, tales como azúcares, incluyendo lactosa, sacarosa, manitol o sorbitol; preparaciones de celulosa tales como, por ejemplo, almidón de maíz, almidón de trigo, almidón de arroz, almidón de patata, gelatina, goma de tragacanto, metilcelulosa, hidroxipropilmetilcelulosa, carboximetilcelulosa sódica y/o polivinilpirrolidona (PVP). Si se desea, pueden añadirse agentes disgregantes, tales como polivinilpirrolidona reticulada, agar o ácido alginico o una sal del mismo, tal como alginato de sodio. Opcionalmente, las formulaciones orales también se pueden formular en solución salina o tampones para neutralizar las condiciones ácidas internas o se pueden administrar sin ningún vehículo.

Los núcleos de grageas se proporcionan con recubrimientos adecuados. Para este fin, se pueden usar soluciones de azúcar concentradas, que pueden contener opcionalmente goma arábiga, talco, polivinilpirrolidona, gel de carbopol,

polietilenglicol y/o dióxido de titanio, soluciones de laca y disolventes orgánicos adecuados o mezclas de disolventes. Pueden añadirse colorantes o pigmentos a los comprimidos o recubrimientos de grageas para la identificación o para caracterizar distintas combinaciones de dosis de compuestos activos.

5 Las preparaciones farmacéuticas que se pueden usar por vía oral incluyen cápsulas de ajuste a presión hechas de gelatina, así como cápsulas blandas, selladas hechas de gelatina y un plastificante, tal como glicerol o sorbitol. Las cápsulas de ajuste por presión pueden contener los principios activos mezclados con una carga, tal como lactosa, aglutinantes, tales como almidones, y/o lubricantes, tales como talco o estearato de magnesio y, opcionalmente, estabilizantes. En las cápsulas blandas, los compuestos activos pueden estar disueltos o suspendidos en líquidos
10 adecuados, tales como aceites grasos, parafina líquida o polietilenglicoles líquidos. Adicionalmente, pueden añadirse estabilizantes. También se pueden usar microesferas formuladas para administración oral. Dichas microesferas han sido bien definidas en la técnica. Todas las formulaciones para administración oral deben estar en dosis adecuadas para dicha administración. Para la administración bucal, las composiciones pueden adoptar la forma de comprimidos o grageas formulados de forma convencional.

15 Los compuestos, cuando es deseable suministrarlos de manera sistémica, pueden formularse para administración parenteral mediante inyección, por ejemplo, mediante inyección intravenosa rápida o infusión continua. Las formulaciones para inyección pueden presentarse en una forma farmacéutica unitaria, por ejemplo, en ampollas o recipientes multidosis, con un conservante añadido. Las composiciones pueden adoptar formas tales como
20 suspensiones, soluciones o emulsiones en vehículos oleosos o acuosos y pueden contener agentes de formulación, tales como agentes de suspensión, estabilizantes y/o dispersantes.

25 Las preparaciones para administración parenteral incluyen soluciones acuosas o no acuosas estériles, suspensiones y emulsiones. Los ejemplos de disolventes no acuosos son propilenglicol, polietilenglicol, aceites vegetales tales como aceite de oliva y ésteres orgánicos inyectables tales como oleato de etilo. Los vehículos acuosos incluyen agua, soluciones alcohólicas/acuosas, emulsiones o suspensiones, que incluyen solución salina y medio tamponado. Los vehículos parenterales incluyen solución de cloruro sódico, dextrosa de Ringer, dextrosa y cloruro sódico, lactato de Ringer o aceites fijos. Los vehículos intravenosos incluyen reponedores de líquidos y nutrientes, reponedores de electrolitos (tales como los basados en dextrosa de Ringer) y similares. También pueden estar presentes conservantes
30 y otros aditivos tales como, por ejemplo, antimicrobianos, antioxidantes, agentes quelantes y gases inertes y similares. Las dosis más bajas resultarán de otras formas de administración, tal como administración intravenosa. En caso de que la respuesta en un sujeto sea insuficiente a las dosis iniciales aplicadas, las dosis más altas (o efectivamente dosis más altas por un diferente, vía de administración más localizada) pueden emplearse en la medida en que lo permita la tolerancia del paciente. Se contemplan dosis múltiples por día en algunas realizaciones para lograr niveles sistémicos apropiados de compuestos.

35 En algunos ejemplos, un método comprende la disminución de la expresión de un oncogén que se sobreexpresa o tiene una actividad anormalmente aumentada en el cáncer mediante la reducción de la formación y/o la estabilidad de un bucle entre un potenciador y el promotor del oncogén. En algunos ejemplos, un método comprende aumentar la
40 expresión de un gen supresor de tumores aumentando la formación y/o el mantenimiento de un bucle de ADN entre un potenciador y el promotor del gen supresor de tumores. En algunos ejemplos, un método puede comprender analizar una muestra obtenida de un tumor, identificar uno o más genes que se expresan de manera aberrante o codifican un producto génico con actividad aberrante en el tumor, y modular la expresión del gen utilizando una composición del método descrito en el presente documento (por ejemplo, reducir la expresión de un oncogén que se sobreexpresa de manera aberrante o tiene actividad relativa a la expresión o actividad del homólogo normal en células normales).

45 En algunos ejemplos, las composiciones y métodos de la presente invención que comprenden la reducción del bucle de ADN potenciador-promotor pueden usarse para tratar cualquier enfermedad asociada con una mayor expresión de un gen o una mayor actividad de un producto génico. En algunos ejemplos, las composiciones y métodos de la presente invención que comprenden el aumento del bucle de ADN potenciador-promotor se pueden usar para tratar cualquier enfermedad asociada con la disminución de la expresión de un gen o la disminución de la actividad de un producto génico en relación con los niveles normales. Dichas enfermedades incluyen, por ejemplo, enfermedades neurodegenerativas, trastornos del neurodesarrollo, enfermedades autoinmunitarias, enfermedades metabólicas, etc.
50 Cualquier enfermedad asociada con el aumento o disminución de la expresión o actividad de un gen o producto génico, en la que la expresión del gen está regulada al menos en parte por la formación de un bucle de ADN promotor potenciador mediado al menos en parte por la multimerización de FT podría tratarse.

Método de detección

60 Algunos aspectos de la invención están dirigidos a un *in vitro* método de detección de un compuesto que modula la expresión de uno o más genes en una célula, que comprende poner en contacto la célula con un agente de prueba (por ejemplo, una molécula pequeña, ácido nucleico, anticuerpo o polipéptido), y medir los bucles de ADN del promotor potenciador dependiente de YY1 en la célula, en el que el agente de prueba se identifica como un modulador de la
65 expresión génica si el nivel de bucle de ADN potenciador-promotor en la célula en contacto con el agente de prueba es diferente del nivel de formación de bucle de ADN potenciador-promotor en una célula de control que no se puso en

contacto con el agente de prueba. En algunos ejemplos, el factor de transcripción es capaz de homomultimerización (por ejemplo, homodimerización). En algunos ejemplos, el método comprende medir el bucle de ADN entre un potenciador y un promotor de un gen particular de interés en una célula, en el que el agente de prueba se identifica como un modulador de la expresión del gen particular de interés si el nivel de bucle de ADN entre un potenciador y el promotor del gen en la célula en contacto con el agente de prueba es diferente al nivel de bucle de ADN entre un potenciador y el promotor del gen en una célula de control no en contacto con el agente de prueba. En algunos ejemplos, cualquiera de los métodos descritos en el presente documento comprende además la medición de la expresión del gen o genes en células en contacto con el agente de prueba. En algunos ejemplos, el método comprende además comparar la expresión de los genes por células que se ponen en contacto con el agente de prueba con la expresión del gen por células que no se ponen en contacto con el agente de prueba.

Los métodos para medir la formación de bucles de ADN potenciadores-promotores en una célula son conocidos en la técnica. Véase Hepelev, *et al.*, (2012) "Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization," *Cell Res.* 22, 490-503.

En el presente documento se describen métodos de detección de moduladores de la formación y/o la estabilidad del bucle de ADN (por ejemplo ej., formación y/o estabilidad del bucle de ADN potenciador-promotor) que comprenden poner en contacto un ADN lineal que comprende 2 o más sitios de unión de FT con un factor de transcripción afín capaz de multimerización y un agente de prueba y medir el grado de circularización del ADN. En algunos ejemplos, el agente de prueba se pone en contacto con el ADN lineal al mismo tiempo que se pone en contacto el FTi. En algunos ejemplos, el agente de prueba se pone en contacto con el ADN lineal después de que el FTi se ponga en contacto. En algunas realizaciones, el agente de prueba se pone en contacto con el ADN lineal antes de que se ponga en contacto con el FT. La actividad del agente de prueba para modular la formación de bucles de ADN y/o la estabilidad puede evaluarse por comparación con un control que comprenda el ADN y el factor de transcripción pero no el agente de prueba.

Algunos aspectos de la invención están dirigidos a métodos para identificar uno o más genes con expresión dependiente de un potenciador en una célula, que comprende identificar uno o más bucles de ADN promotor-potenciador dependientes de YY1 que comprenden el potenciador en la célula, e identificar uno o más genes expresados en el bucle de ADN promotor-potenciador, en el que uno o más genes expresados en el bucle de ADN potenciador-promotor se identifican como genes cuya expresión depende del potenciador. En algunos ejemplos, el factor de transcripción es capaz de homomultimerización (por ejemplo, homodimerización). En algunas realizaciones, la etapa de identificar uno o más bucles de ADN promotor-potenciador que comprenden el potenciador comprende realizar un ensayo ChIP-MS.

Se ha encontrado que los factores de transcripción pueden unirse a diferentes potenciadores y múltiples factores de transcripción pueden unirse al mismo potenciador. Mediante el método anterior, pueden identificarse genes con expresión controlada por un potenciador particular. En el caso de potenciadores asociados a una enfermedad o afección, el método anterior se puede usar para identificar genes expresados que pueden ser objetivos para estudios adicionales o para tratamientos.

Un experto en la materia aprecia fácilmente que la presente invención está bien adaptada para conseguir los objetivos y obtener los fines y ventajas mencionados, así como los inherentes a la misma.

Los artículos "un" y "uno/a", como se usan en el presente documento en la memoria descriptiva y en las reivindicaciones, a menos que se indique claramente lo contrario, deben entenderse que incluyen los referentes en plural. Las reivindicaciones o descripciones que incluyen "o" entre uno o más miembros de un grupo se consideran satisfechas si uno, más de uno, o todos los miembros del grupo están presentes en, se emplean en, o son de otro modo relevantes para un producto o un proceso dado a menos que se indique lo contrario o sea evidente de otro modo a partir del contexto. La invención incluye realizaciones en las que exactamente un miembro del grupo está presente en, se emplean en, o son de otro modo relevantes, para un producto o proceso dado. La invención incluye también realizaciones en las que más de uno, o todos los miembros del grupo están presentes en, se emplean en, o son de otro modo relevantes, para un producto o proceso dado. Por otra parte, debe entenderse que la invención proporciona todas las variaciones, combinaciones y permutaciones en las que una o más limitaciones, elementos, cláusulas, términos descriptivos, etc., de una o más de las reivindicaciones enumeradas se introduce en otra reivindicación dependiente de la misma reivindicación base (o, o según corresponda, cualquier otra reivindicación) a menos que se indique lo contrario o a menos que sea evidente para un experto en la materia que surja una contradicción o incoherencia. Se contempla que todas las realizaciones descritas en el presente documento son aplicables a todos los diferentes aspectos de la invención cuando sea apropiado. También se contempla que cualquiera de las realizaciones o aspectos puede combinarse libremente con una u otras más de tales realizaciones o aspectos cuando sea apropiado. Cuando los elementos se presentan como listados, por ejemplo, en formato de grupo Markush o similar, ha de entenderse que también se desvela cada subgrupo de los elementos y cualquier elemento o elementos pueden retirarse del grupo. Debe entenderse que, en general, cuando la invención o aspectos de la invención, se refiere, o refieren, a comprender elementos particulares, características, etc., determinadas realizaciones de la invención o aspectos de la invención consisten, o consisten esencialmente en, dichos elementos, características, etc. Por motivos de simplicidad, estas realizaciones no se han expuesto específicamente en todos los casos en el presente documento

con tantas palabras. También debe entenderse que cualquier realización o aspecto de la invención puede excluirse explícitamente de las reivindicaciones, independientemente de si la exclusión específica se menciona en la memoria descriptiva. Por ejemplo, uno cualquiera o más ácidos nucleicos, polipéptidos, células, especies o tipos de organismos, trastornos, sujetos o combinaciones de los mismos, puede ser excluido.

5 Cuando las reivindicaciones o la descripción se refieran a una composición de materia, por ejemplo, un ácido nucleico, el polipéptido, célula o animal transgénico no humano, se debe entender que los métodos para hacer o usar la composición de materia de acuerdo con cualquiera de los métodos descritos en el presente documento, y los métodos para usar la composición de materia para cualquiera de los fines descritos en el presente documento son aspectos de la invención, a menos que se indique lo contrario o a menos que sea evidente para un experto habitual en la materia que surgiría una contradicción o incoherencia. Cuando las reivindicaciones o la descripción se refieran a un método, por ejemplo, debe entenderse que los métodos para hacer composiciones útiles para realizar el método, y los productos producidos de acuerdo con el método, son aspectos de la invención, a menos que se indique lo contrario o a menos que sea evidente para un experto habitual en la materia que surgiría una contradicción o incoherencia.

15 Cuando se proporcionan intervalos en el presente documento, la invención incluye realizaciones en las que se incluyen valores extremos, realizaciones en las que se excluyen los dos valores extremos y realizaciones en las que se incluye un valor extremo y se excluye el otro. Ha de suponerse que, a menos que se indique de otro modo, se incluyen los dos valores extremos. Por otra parte, ha de entenderse que, a menos que se indique lo contrario o que sea evidente de otro modo por el contexto y la comprensión de un experto en la materia, los valores que se expresan como intervalos pueden asumir cualquier valor específico o subintervalo dentro de los intervalos indicados en realizaciones diferentes de la invención, hasta la décima parte de la unidad del límite inferior del intervalo, a menos que el contexto dictamine claramente otra cosa. También se entiende que cuando, en el presente documento, se indica una serie de valores numéricos, la invención incluye casos que se refieren análogamente a cualquier valor o intervalo intermedio definido por dos valores cualesquiera de la serie, y que el valor más bajo puede tomarse como mínimo y el valor más alto como máximo. Valores numéricos, como se usa en el presente documento, incluye los valores expresados como porcentajes. Para cualquier realización de la invención en la que un valor numérico esté precedido por "alrededor de" o "aproximadamente", la invención incluye una realización en la que se indica el valor exacto. Para cualquier realización de la invención en la que un valor numérico no esté precedido por "alrededor de" o "aproximadamente", la invención incluye una realización en la que el valor está precedido por "alrededor de" o "aproximadamente". "Aproximadamente" o "alrededor de" generalmente incluye números que se encuentran en un intervalo de un 1 % o, en algunos casos dentro de un intervalo de un 5 % de un número, o en algunos casos dentro de un intervalo de un 10 % de un número en cualquier dirección (mayor o menor que el número) a menos que se indique lo contrario o sea evidente de otra manera a partir del contexto (excepto cuando dicho número supere inadmisiblemente el 100 % de un valor posible). Debe entenderse que, a menos que se indique claramente lo contrario, en cualquiera método reivindicado en el presente documento, que incluya más de una actuación, el orden de las actuaciones del método no se limita necesariamente al orden en el que se enumeran las actuaciones del método, aunque la invención incluya realizaciones en las que el orden está limitado. También debe entenderse que, a menos que se indique lo contrario o sea evidente por el contexto, cualquier producto o composición que se describa en el presente documento puede considerarse "aislado(a)".

Los ejemplos específicos de estos métodos se exponen a continuación en los Ejemplos.

Ejemplos

45 Se cree que la estructura de los cromosomas desempeña un papel importante en el control de los genes, pero se dispone de una comprensión limitada de las proteínas que contribuyen a estructurar las interacciones potenciador-promotor. En el presente documento se informa de que Yin Yang 1 (YY1) contribuye a las interacciones potenciador-promotor de una manera análoga al bucle de ADN mediado por CTCF. YY1 y CTCF comparten muchas características: 50 ambos son esenciales, se expresan de forma ubicua, son proteínas coordinadoras de cinc que se unen a secuencias de ADN hipometiladas, forman homodímeros y así facilitan la formación de bucles. Las dos proteínas difieren en que YY1 ocupa preferentemente potenciadores y promotores que interactúan, mientras que CTCF ocupa preferentemente sitios distales a estos elementos reguladores que tienden a formar bucles más grandes y participar en el aislamiento. La eliminación de los sitios de unión o el agotamiento de YY1 pueden alterar los contactos potenciador-promotor y la expresión génica normal. De este modo, la estructuración mediada por YY1 de bucles potenciadores-promotores es análoga a la estructuración mediada por CTCF de TAD, dominios de contacto de CTCF y vecindades aisladas. Este modelo de estructuración mediada por YY1 de bucles potenciadores-promotores da cuenta de diversas funciones informadas anteriormente para YY1, incluidas las contribuciones tanto a la activación como a la represión de genes y a la regulación alterada de los genes en el cáncer. Por lo tanto, se propone que YY1 es un regulador estructural de los bucles potenciadores-promotores y que los bucles potenciadores-promotores estructurados YY1 pueden ser un mecanismo general de control de genes de mamíferos.

Análisis de interacciones de bucles de ADN promotor potenciador

65 Se buscó identificar un factor de proteína que pudiera contribuir a las interacciones potenciador-promotor de una manera análoga a la de CTCF en los aisladores. Se esperaba que tal proteína se una a potenciadores y promotores

activos, sea esencial para la viabilidad celular, muestre expresión ubicua y sea capaz de dimerización. Para identificar proteínas que se unen a potenciadores y promotores activos, se buscaron candidatos de inmunoprecipitación de cromatina con espectrometría de masas (ChIP-MS), usando anticuerpos dirigidos contra histonas con modificaciones características de la cromatina potenciadora y promotora (H3K27ac y H3K4me3, respectivamente) (Creyghton *et al.*, 2010), realizaron previamente en células madre embrionarias murinas (células MEr) (Ji *et al.*, 2015). De 26 factores de transcripción que ocupan tanto potenciadores como promotores (FIG. 1A), cuatro (CTCF, YY1, NRF1 y ZBTB11) son esenciales según una pantalla de esencialidad celular CRISPR (FIG. 1B) (Wang *et al.*, 2015) y dos (CTCF, YY1) se expresan en > 90 % de los tejidos examinados (FIG. 1C). YY1 y CTCF comparten características adicionales: como el CTCF, YY1 es un factor de transcripción de dedo de cinc (Klenova *et al.*, 1993; Shi *et al.*, 1991), esencial para la viabilidad de las células embrionarias y adultas (Donohoe *et al.*, 1999; Brezo *et al.*, 2008) y capaz de formar homodímeros (López-Perrote *et al.*, 2014; Saldaña-Meyer *et al.*, 2014)(Table S1). YY1, sin embargo, tiende a ocupar potenciadores y promotores activos, así como algunos aisladores, mientras que CTCF ocupa preferentemente elementos aisladores (FIG. 1D, FIG. 18A-C).

Tabla S1. Comparación de YY1 y CTCF

Características compartidas	Cita de CTCF	Cita de YV1
Dominio de FT de unión a ADN coordinador de cinc FT	(Klenova <i>et al.</i> , 1993)	(Shi <i>et al.</i> , 1991)
Expresado de forma ubicua en células de mamíferos	(Mele <i>et al.</i> , 2015)	(Mele <i>et al.</i> , 2015)
Funciones esenciales en el desarrollo normal	(Heath <i>et al.</i> , 2008; Moore <i>et al.</i> , 2012; Splinter <i>et al.</i> , 2006)	(Donohoe <i>et al.</i> , 1999)
Notificado como activador	(Klenova <i>et al.</i> , 1993; Lobanenko <i>et al.</i> , 1990)	(Seto <i>et al.</i> , 1991)
Notificado como represor	(Baniahmad <i>et al.</i> , 1990; Köhne <i>et al.</i> , 1993)	(Shi <i>et al.</i> , 1991)
Se une al ARN	(Saldaña-Meyer <i>et al.</i> , 2014)	(Jeon y Lee, 2011; Sigova <i>et al.</i> , 2015)
Puede formar dímeros	(Saldaña-Meyer <i>et al.</i> , 2014; Yusufzai <i>et al.</i> , 2004)	(Lopez-Perrote <i>et al.</i> , 2014; Wu <i>et al.</i> , 2007)
Implicado en la recombinación de V(D)J	(Liu <i>et al.</i> , 2007)	(Guo <i>et al.</i> , 2011)
Notificado que pliega el ADN	(Arnold <i>et al.</i> , 1996)	(Natesan y Gilman, 1993)
Enriquecido en anclajes de bucle	(Heidari <i>et al.</i> , 2014; Rao <i>et al.</i> , 2014)	(Heidari <i>et al.</i> , 2014; Rao <i>et al.</i> , 2014)
Mal expresado en el cáncer	(Filippova <i>et al.</i> , 1998, 2002)	(Castellano <i>et al.</i> , 2009)
Implicado en XCI	(Xu <i>et al.</i> , 2007)	(Jeon y Lee, 2011)
Características distintivas	Cita de CTCF	Cita de YV1
Se une predominantemente a los aisladores	Figura 1C, este estudio	
Se une predominantemente a potenciadores y promotores activos		

Si YY1 contribuye a las interacciones potenciador-promotor, luego el análisis de interacción de la cromatina por secuenciación de etiquetas de extremos emparejados (ChIA-PET) (Fullwood *et al.*, 2009) para YY1 debería mostrar que YY1 se asocia preferentemente con estas interacciones. CTCF CHIA-PET, por el contrario, debería mostrar que CTCF se asocia preferentemente con interacciones de ADN aislador. Se generaron datos de ChIA-PET para YY1 y CTCF en células MEr y se compararon estos dos conjuntos de datos. Los resultados mostraron que la mayoría de las interacciones asociadas con YY1 conectan elementos reguladores activos (potenciador-potenciador, potenciador-promotor y promotor-promotor, que en adelante se denominarán interacciones potenciador-promotor), mientras que la mayoría de las interacciones asociadas a CTCF conectan elementos aisladores (FIG. 1E, FIG. 18D). Algunas interacciones YY1-YY1 involucraron contactos simples potenciador-promotor, como se ve en el locus Raf1 (FIG. 1F) y otras implicaron contactos más complejos entre los constituyentes del superpotenciador y sus promotores objetivo, como se ve en el locus geométrico Klf9 (FIG. 11E). Los superpotenciadores generalmente estaban ocupados por YY1 a densidades relativamente altas y exhibieron frecuencias de interacción YY1-YY1 relativamente altas (FIG. 18E-H). Tanto para YY1 como para CTCF, también hubo evidencia de interacciones potenciador-aislador y promotor-aislador, pero estas fueron más pronunciadas para CTCF (FIG. 18D).

Estudios previos han notificado que YY1 puede formar dímeros (López-Perrote *et al.*, 2014). Para confirmar que se produce la dimerización de YY1, las versiones marcadas con FLAG y marcadas con HA de la proteína YY1 se expresaron en células, se aislaron los núcleos y las proteínas YY1 marcadas en extractos nucleares se

inmunoprecipitaron con anticuerpos anti-FLAG o anti-HA. Los resultados muestran que las proteínas YY1 marcadas con FLAG y marcadas con HA interactúan (FIG. 1G, H, FIG. 18I, J), consistente con informes previos de que las proteínas YY1 se oligomerizan (López-Perrote *et al.*, 2014). Otras proteínas nucleares altamente expresadas, como OCT4, no coprecipitaron, indicativo de que el ensayo era específico (FIG. 18J). Anteriormente se informó de que YY1 puede unirse tanto al ADN como al ARN de forma independiente, y que la unión de YY1 a elementos de ADN reguladores activos se ve reforzada por la unión de especies de ARN que se transcriben en estos loci (Sigova *et al.*, 2015). Por lo tanto, es posible que las interacciones YY1-YY1 se vean potenciadas por la capacidad de cada una de las proteínas YY1 para unirse a especies de ARN. De hecho, cuando se repitió el experimento descrito anteriormente con extractos nucleares que contenían las proteínas YY1 etiquetadas, y una parte de la muestra se trató con RNasa A antes de la inmunoprecipitación con anticuerpos anti-marcador, se produjo una reducción de ~60 % en la cantidad de proteína asociada YY1 coimmunoprecipitada (FIG. 1G, H). Estos resultados sugieren que las interacciones estables YY1-YY1 pueden ser facilitadas por el ARN.

YY1 está asociado con elementos reguladores activos

Si YY1 es una proteína estructurante potenciadora-promotora, esperaríamos que YY1 ocupara y conectara elementos reguladores activos. Se examinó la unión global de YY1 en células madre embrionarias murinas (CMEr) y descubrimos que YY1 se localiza predominantemente en promotores y potenciadores activos (Fig. 1c, 1e). A continuación, se realizó un análisis de interacción de cromatina mediante secuenciación de etiquetas de extremos emparejados (ChIA-PET) para YY1 y CTCF. Se descubrió que la mayoría de las interacciones asociadas con YY1 conectan elementos reguladores activos, lo cual contrasta marcadamente con las interacciones asociadas a CTCF donde la mayoría conecta elementos aisladores (Fig. 1d). La Figura 1e muestra un ejemplo de un gen con bucles potenciadores-promotores asociados con la unión a YY1. Estos resultados muestran que YY1 está asociado con elementos reguladores activos y sugiere que puede estar involucrado en la estructuración de bucles potenciadores-promotores.

YY1 es fundamental para la expresión génica adecuada

Se cree que la expresión génica está controlada por bucles entre potenciadores y promotores, por lo que si YY1 está implicado en la estructuración de estos bucles, la perturbación de YY1 debería perturbar la expresión génica. Para probar esto, se usó un ARN de horquilla pequeña inducible (ARNhp) para eliminar YY1 y analizamos la expresión génica con hibridación *in situ* fluorescente de ARN de una sola molécula (FISH de ARNms). Se observó que había una disminución en las transcripciones de los reguladores clave de células MA Oct4 y Sox2 (Fig. 2). Los promotores y potenciadores de estos genes están todos densamente ocupados por YY1. Todos estos genes también tienen bucles potenciadores-promotores asociados con YY1 (Fig. 2). Esto muestra que YY1 es fundamental para la expresión génica adecuada y respalda la idea de que YY1 regula la expresión génica conectando potenciadores y promotores.

*YY1 puede potenciar las interacciones del ADN *in vitro**

Las proteínas CTCF pueden formar homodímeros y oligómeros más grandes y, por lo tanto, cuando se unen a dos sitios de ADN diferentes, pueden formar un bucle con el ADN intermedio (Saldaña-Meyer *et al.*, 2014). La observación de que YY1 está ligado a potenciadores y promotores que interactúan, junto con la evidencia de que pueden ocurrir interacciones YY1-YY1 *in vitro* y en extractos celulares, es consistente con la idea de que las interacciones YY1-YY1 pueden contribuir a la formación de bucles entre potenciadores y promotores. Para obtener evidencia de que YY1 puede tener un efecto directo en las interacciones del ADN, se usó un ensayo *in vitro* de circularización de ADN para determinar si YY1 purificado puede mejorar la tasa de interacción del ADN *in vitro*. La tasa de circularización del ADN catalizada por la ADN ligasa de T4 se ha utilizado previamente para medir la longitud de persistencia y otras propiedades físicas del ADN (Shore *et al.*, 1981). Se razonó que si YY1 unido al ADN es capaz de dimerizarse y, por lo tanto, formar bucles de ADN, a continuación, la incubación de una plantilla de ADN lineal que contiene sitios de unión a YY1 con proteína YY1 purificada debería acercar los extremos y aumentar la tasa de circularización (FIG. 3A, D). La proteína YY1 recombinante se purificó y se demostró que tenía actividad de unión al ADN usando un ensayo de cambio de movilidad (FIG. 17A, B). Posteriormente, este YY1 recombinante se analizó en el ensayo de circularización de ADN; los resultados mostraron que YY1 aumentaba la tasa de circularización y que esto dependía de la presencia de motivos YY1 en el ADN (FIG. 3B, 3C). La adición de un exceso de un fragmento de ADN competitivo de 200 pares de bases que contenía la secuencia de unión consenso YY1 anuló la circularización de la molécula de ADN más grande (FIG. 3D-F). La adición de seroalbúmina bovina (BSA) no incrementó la tasa de unión del ADN (FIG. 3C, F). Estos resultados respaldan la idea de que YY1 puede facilitar directamente las interacciones del ADN. Estos resultados sugieren que YY1 puede multimerizarse y que esta multimerización es capaz de formar bucles en el ADN.

La interrupción del bucle YY1 perturba la expresión génica

Habiendo demostrado que el agotamiento global de YY1 perturba la expresión génica y que YY1 es capaz de oligomerización, a continuación, se deseaba probar directamente el papel de la unión de YY1 en los potenciadores y promotores. Se usó el sistema CRISPR/Cas9 para generar una pequeña delección en el sitio de unión de YY1 en el potenciador Zfp518a y luego caracterizamos el efecto sobre la expresión génica, la unión de YY1 y la formación de bucles (Fig. 4). Se descubrió que la mutación resultó en una disminución en la expresión de Zfp518a (Fig. 4c), pérdida de unión de YY1 (Fig. 4d) y una disminución en el bucle entre el potenciador y el promotor (Fig. 4b, e). Estos resultados

indican que la unión de YY1 a los potenciadores es necesaria para el bucle normal a los promotores, y que la interrupción de este bucle perturba la expresión génica, lo que sugiere que YY1 es fundamental para el control génico adecuado.

5 *Las interacciones potenciador-promotor dependen de YY1 en las células vivas*

Para probar más a fondo si las interacciones potenciador-promotor en las células vivas dependen de los sitios de unión de YY1 en estos elementos, se usó un sistema CRISPR/Cas9 para generar una pequeña delección de un motivo de unión a YY1 en las regiones reguladoras de dos genes (FIG. 6A). La eliminación del motivo de unión al ADN óptimo para YY1 en el promotor del gen *Raf1* resultó en una disminución de la unión de YY1 en el promotor, una reducción de la frecuencia de contacto entre el potenciador y el promotor y una disminución en los niveles de ARNm de *Raf1* (FIG. 6B, FIG. 13A). La delección del motivo óptimo de unión al ADN para YY1 en el promotor del gen *Etv4* también resultó en una disminución de la unión de YY1 y una disminución de la frecuencia de contacto entre el potenciador y el promotor, aunque no afectó significativamente a los niveles de ARNm de *Etv4* (FIG. 6C, FIG. 13B). Estos resultados sugieren que los sitios de unión de YY1 contribuyen a la unión de YY1 y a las frecuencias de contacto del potenciador-promotor tanto en *Raf1* como en *Etv4*, aunque la reducción en las frecuencias de formación de bucles en *Etv4* no fue suficiente para tener un impacto significativo en los niveles de ARNm de *Etv4*. La falta de un efecto sobre los niveles de ARNm de *Etv4* puede ser una consecuencia del YY1 residual que está unido a la región promotora de *Etv4*, donde se observan motivos CCAT adicionales (FIG. 6C). De hecho, cuando se agota la proteína YY1 (véase más adelante; FIG. 14E), los niveles de ARNm de *Raf1* y *Etv4* disminuyen.

En estudios previos se ha indicado que YY1 es un activador de algunos genes y un represor de otros, pero no se ha descrito un análisis global de las dependencias de YY1 con un agotamiento completo de YY1 en las células MEr (Gordon *et al.*, 2006; Shi *et al.*, 1997; Thomas y Seto, 1999). Se usó un sistema de degradación inducible (Erb *et al.*, 2017; Huang *et al.*, 2017; Winter *et al.*, 2015) para agotar completamente los niveles de proteína YY1 y se midió el impacto en la expresión génica en células MEr en todo el genoma a través del análisis RNA-seq (FIG. 7A, B). El agotamiento de YY1 condujo a cambios significativos (valor de *p* ajustado <0,05) en la expresión de 8234 genes, dividida casi por igual entre genes con expresión aumentada y genes con expresión disminuida (FIG. 7C, Tabla S3). Los genes que experimentaron los mayores cambios en la expresión con el agotamiento de YY1 fueron generalmente ocupados por YY1 (FIG. 7D).

Estudios previos han demostrado que se requiere YY1 para el desarrollo embrionario normal (Donohoe *et al.*, 1999). Se investigó si la pérdida de YY1 conducía a defectos en la diferenciación de células madre embrionarias (ES) en las tres capas germinales (Figura 7E). Las células ME murinas y las células isogénicas que se sometieron a degradación inducible de YY1, fueron estimuladas para formar cuerpos embrioides (FIG. 7F) y las células de estos cuerpos se sometieron a tinción inmunohistoquímica y RNA-seq de una sola célula para controlar la expresión de factores específicos de diferenciación. Los resultados mostraron que las células que carecían de YY1 mostraban defectos pronunciados en la expresión de los factores de transcripción maestros que impulsan la diferenciación normal (FIG. 7G, H; FIG. 15).

A continuación, se investigó si se producían cambios en la formación de bucles de ADN tras el agotamiento global de YY1 en células MEr. HiChIP para H3K27ac, una modificación de histona presente tanto en los potenciadores como en los promotores, se realizó antes y después del agotamiento de YY1 para detectar diferencias en las frecuencias de interacción potenciador-promotor. Antes del agotamiento de YY1, los resultados del experimento HiChIP mostraron interacciones entre los diversos elementos que eran similares a los resultados anteriores de YY1 ChIA-PET (FIG. 14A, B). Después del agotamiento de YY1, las interacciones entre los potenciadores y promotores ocupados por YY1 disminuyeron significativamente (FIG. 8A, B). La mayoría (60 %) de los genes conectados por bucles potenciadores-promotores YY1 mostraron cambios significativos en la expresión génica (FIG. 8C; FIG. 14D). El examen de los perfiles de interacción del ADN de HiChIP en genes específicos confirmó estos efectos. Por ejemplo, con el agotamiento de YY1, el promotor *Slc7a5* y su potenciador mostraron una reducción del -50 % en la frecuencia de interacción, y los niveles de expresión de *Slc7a5* se redujeron en un -27 % (FIG. 8D). De forma similar, después del agotamiento de YY1, el promotor *Klf9* y su superpotenciador mostraron una reducción del -40 % en la frecuencia de interacción y los niveles de expresión de *Klf9* se redujeron en un -50 % (FIG. 8E).

55 *Rescate de interacciones potenciador-promotor en células*

La capacidad de una proteína YY1 unida artificialmente para rescatar defectos asociados con una mutación del sitio de unión de YY1 sería una prueba sólida del modelo de que YY1 media en las interacciones potenciador-promotor (FIG. 9A). Dicha prueba se realizó con una proteína de fusión dCas9-YY1 dirigida a un sitio adyacente a una mutación del sitio de unión a YY1 en la región proximal del promotor de *Etv4* (FIG. 7B, C). Se descubrió que unir artificialmente la proteína YY1 al promotor conducía a una mayor frecuencia de contacto entre el promotor *Etv4* y su potenciador y provocaba un aumento de la transcripción del gen (FIG. 9D). Estos resultados respaldan el modelo de que YY1 está directamente involucrado en la estructuración de bucles potenciadores-promotores.

Para probar de manera más global si YY1 puede rescatar la pérdida de interacciones potenciador-promotor después de la degradación de YY1, las células MEr se sometieron a la degradación de YY1 con el método dTAG y luego lavaron

el compuesto dTAG y permitieron que YY1 se restaurara a niveles normales (FIG. 9E; FIG. 16A, B). Las frecuencias del potenciador-promotor se monitorizaron con H3K27ac HiChIP. De acuerdo con un experimento anterior (FIG. 8), la pérdida de YY1 provocó una pérdida en las interacciones potenciador-promotor, pero la recuperación de los niveles de YY1 se acompañó de un aumento sustancial en las interacciones potenciador-promotor (FIG. 9F). Estos resultados fueron comparables a los efectos observados con el rescate de interacciones CTCF-CTCF en un experimento similar descrito recientemente (FIG. 9F; FIG. 16C) (Nora *et al.*, 2017) y respaldan el modelo de que YY1 contribuye a la estructuración de una gran fracción de bucles potenciadores-promotores en todo el genoma.

Discusión

En el presente documento se describen evidencias de que el factor de transcripción YY1 contribuye a las interacciones estructurales potenciador-promotor. Para un amplio espectro de genes, YY1 se une a potenciadores y promotores activos y es necesario para niveles normales de interacción potenciador-promotor y transcripción génica. YY1 se expresa de forma ubicua, ocupa potenciadores y promotores en todos los tipos de células examinados, está asociado con sitios de bucles de ADN en células donde se han realizado tales estudios, y es esencial para la viabilidad de células adultas y embrionarias, por lo que es probable que las interacciones potenciador-promotor mediadas por YY1 sean una característica general del control de genes de mamíferos.

Las evidencias de que las interacciones CTCF-CTCF desempeñan un papel importante en las estructuras de bucle cromosómico, pero solo ocasionalmente participan en interacciones potenciador-promotor, llevó a considerar la posibilidad de que una proteína puente análoga a CTCF podría participar generalmente en interacciones potenciador-promotor. CTCF y YY1 comparten muchas características: son factores de dedo de cinc que se unen al ADN (Klenova *et al.*, 1993; Shi *et al.*, 1991) que se unen selectivamente a secuencias de ADN hipometiladas (Bell y Felsenfeld, 2000; Yin *et al.*, 2017), se expresan de forma ubicua (FIG. 11) (Mele *et al.*, 2015), son esenciales para la viabilidad embrionaria (Donohoe *et al.*, 1999; Brezo *et al.*, 2008) y tienen capacidad de dimerización (FIG. 12) (López-Perrote *et al.*, 2014; Saldaña-Meyer *et al.*, 2014). Las dos proteínas difieren en varios aspectos importantes. Las interacciones CTCF-CTCF ocurren predominantemente entre sitios que pueden actuar como aisladores y en menor grado entre potenciadores y promotores (FIG. 1D). Las interacciones YY1-YY1 ocurren predominantemente entre potenciadores y promotores y, en menor medida, entre aisladores (FIG. 1D). En aisladores, CTCF se une a un motivo de secuencia relativamente grande y conservado (en comparación con los unidos por otros FT); estos mismos sitios tienden a unirse en muchos tipos de células diferentes, lo que puede contribuir a la observación de que los límites de TAD tienden a conservarse entre los tipos de células. En los potenciadores y promotores, YY1 se une a un motivo de secuencia relativamente pequeño y mal conservado dentro de estas regiones, donde se producen especies de ARN que pueden facilitar la unión estable al ADN de YY1 (Sigova *et al.*, 2015). La actividad específica del tipo de célula de los potenciadores y promotores contribuye así a la observación de que las interacciones YY1-YY1 tienden a ser específicas del tipo de célula.

El modelo en el que YY1 contribuye a la estructuración de los bucles potenciador-promotor puede dar cuenta de las muchas funciones diversas previamente informadas para YY1, incluyendo activación y represión, diferenciación y proliferación celular. Por ejemplo, tras su descubrimiento a principios de la década de 1990 (Hariharan *et al.*, 1991; Park y Atchison, 1991; Shi *et al.*, 1991), YY1 se estudió intensamente y se informó que actúa como represor de algunos genes y activador de otros; estos efectos específicos del contexto se han atribuido a muchos mecanismos diferentes (revisados en (Gordon *et al.*, 2006; Shi *et al.*, 1997; Thomas y Seto, 1999)). Hay muchos informes similares de activación y represión específicas del contexto por parte de CTCF (revisado en (Olsson *et al.*, 2001; Phillips y Corces, 2009)). Aunque es razonable suponer que YY1 y CTCF pueden actuar directamente como activadores o represores en algunos genes, la evidencia de que estas proteínas contribuyen a la estructuración de los bucles de ADN hace probable que las diversas funciones activas y represivas que se les han atribuido sean a menudo una consecuencia de sus funciones en la estructuración del ADN. En este modelo, la pérdida de CTCF o YY1 podría tener efectos positivos o negativos debido a que otros reguladores ya no estaban adecuadamente posicionados para realizar sus actividades reguladoras.

Estudios previos han insinuado un papel para YY1 en las interacciones de ADN a larga distancia. CTCF, YY1 y la cohesina se han implicado en la formación de bucles de ADN necesarios para el reordenamiento de V(D)J en el locus de inmunoglobulina durante el desarrollo de los linfocitos B (Degner *et al.*, 2011; Guo *et al.*, 2011; Liu *et al.*, 2007). La delección específica de linfocitos B de YY1 provoca una disminución en la contracción del locus IgH, se cree que está mediado por bucles de ADN y un bloqueo en el desarrollo de las células B (Liu *et al.*, 2007). También se ha demostrado que la eliminación de YY1 reduce las interacciones intracromosómicas entre Th2 LCR y el promotor de IL4 (Hwang *et al.*, 2013). A medida que se completaba este manuscrito, apareció un artículo que informaba que YY1 está presente en la base de las interacciones entre los potenciadores y los genes específicos de las células precursoras neuronales y que la eliminación de YY1 provoca una pérdida de estas interacciones (Beagan *et al.*, 2017). Los resultados descritos aquí argumentan que YY1 es más un regulador estructural general de las interacciones potenciador-promotor para una gran población de genes, tanto específicas del tipo de célula como de otro tipo, en todas las celdas. De este modo, la tendencia de YY1 a participar en bucles específicos del tipo de célula es un reflejo de la especificidad del tipo de célula de los potenciadores y, en consecuencia, sus interacciones con genes que pueden expresarse en una forma específica de célula o más general.

YY1 desempeña un papel importante en la enfermedad humana; La haploinsuficiencia de YY1 se ha implicado en un síndrome de discapacidad intelectual y la sobreexpresión de YY1 ocurre en muchos cánceres. Se ha descrito que una cohorte de pacientes con varias mutaciones en un alelo y que presentan discapacidad intelectual tienen un "Síndrome YY1", y las líneas celulares linfoblastoides de estos pacientes muestran una ocupación reducida de las regiones reguladoras y pequeños cambios en la expresión génica en un subconjunto de genes asociados con enlace YY1 (Gabriele *et al.*, 2017). Estos resultados son consistentes con el modelo que describimos para YY1 en la estructuración del promotor-potenciador global, y con la idea de que las funciones neurológicas superiores son especialmente sensibles a tal regulación alterada de genes. YY1 se sobreexpresa en un amplio espectro de células tumorales, y se ha propuesto que esta sobreexpresión provoca una proliferación celular descontrolada, tumorigénesis, potencial metastásico, resistencia a los estímulos apoptóticos inmunomediados y resistencia a los quimioterapéuticos (Gordon *et al.*, 2006; Zhang *et al.*, 2011). Los mecanismos que se ha informado que median estos efectos incluyen la regulación a la baja mediada por YY1 de la actividad de p53, interferencia con la poli-ADP-ribosa polimerasa, alteración en la expresión de c-Myc y NF- κ B, regulación de genes de muerte y productos génicos, unión diferencial de YY1 en presencia de mediadores inflamatorios y unión de YY1 al factor de transcripción oncogénico c-Myc (Gordon *et al.*, 2006; Zhang *et al.*, 2011). Aunque es posible que YY1 realice todas estas funciones, su papel como factor de estructuración potenciador-promotor general es una explicación más parsimoniosa de estos fenotipos pleotrópicos.

Muchos factores de transcripción coordinadores con cinc son capaces de homodimerizar y heterodimerizar (Amoutzias *et al.*, 2008; Lamb y McKnight, 1991) y debido a que estos comprenden la clase más grande de factores de transcripción en mamíferos (Weirauch y Hughes, 2011), sugerimos que una combinación de factores de transcripción específicos del tipo de célula y ubicuos de la célula hace una contribución sustancial y subestimada a estructuras de bucle potenciador-promotor. Hay estudios convincentes de factores de transcripción bacterianos y bacteriófagos que contribuyen a la formación de bucles de elementos reguladores del ADN a través de la oligomerización (Adhya, 1989; Schleif, 1992), e informes de varios factores eucariotas con capacidades similares (Matthews, 1992). No obstante, el estudio más reciente de las interacciones potenciador-promotor eucariótico se ha centrado en los cofactores que carecen de capacidades de unión al ADN y puentean los factores de transcripción unidos al potenciador y el aparato de transcripción unido al promotor (Allen y Taatjes, 2015; Deng *et al.*, 2012; Jeronimo *et al.*, 2016; Kagey *et al.*, 2010; Malik y Roeder, 2010; Petrenko *et al.*, 2016), con la notable excepción de las propuestas de que algunas interacciones potenciador-promotor están determinadas por la naturaleza de los factores de transcripción unidos en los dos sitios (Muerdter y Stark, 2016). Se prevé que estudios futuros revelarán factores de transcripción adicionales que pertenecen a la clase de proteínas de unión al ADN cuyo papel predominante es contribuir a la estructura cromosómica.

En conclusión, se ha demostrado que YY1 es responsable de estructurar bucles potenciadores-promotores en células de mamíferos. YY1 ocupa y conecta potenciadores y promotores activos. YY1 se dimeriza y esta dimerización es capaz de enlazar dos fragmentos de ADN. La pérdida de YY1 provoca la pérdida del bucle potenciador-promotor. Se propone que YY1 es una proteína estructurante potenciadora-promotora global. La regulación génica depende de bucles potenciadores-promotores, y la regulación génica es fundamental para un desarrollo adecuado; por lo tanto, comprender la base mecánica de los bucles potenciador-promotor es fundamental para comprender el desarrollo. Por otra parte, la enfermedad a menudo es causada por la mala regulación de la expresión génica, por lo que los hallazgos aquí ayudarán a comprender la patogénia.

Métodos STAR:

Condiciones de los cultivos celulares:

Células madre embrionarias

Se cultivaron células madre embrionarias murinas (MEr) V6.5 en fibroblastos embrionarios murinos (FEM) irradiados. Las células se cultivaron en condiciones estándar de células MEr como se ha descrito anteriormente (1). Las células se cultivaron en placas de cultivo de tejidos gelatinizados al 0,2 % (Sigma, G1890) en medio ESC; DMEM-KO (Invitrogen, 10829-018) suplementado con suero fetal bovino al 15 % (Hyclone, caracterizado SH3007103), LIF 1000 U/ml (ESGRO, ESG1106), aminoácidos no esenciales 100 mM (Invitrogen, 11140-050), L-glutamina 2 mM (Invitrogen, 25030-081), 100 U/ml de penicilina, estreptomycin 100 mg/ml (Invitrogen, 15140-122) y 8 ul/ml de 2-mercaptoetanol (Sigma, M7522).

Producción y purificación de proteínas:

La proteína YY1 se purificó usando métodos establecidos por Lee Lab (Jeon y Lee, 2011) y descritos previamente en (Sigova *et al.*, 2015). Un plásmido que contenía la secuencia de codificación YY1 humana marcada con His6 N-terminal (un regalo del Dr. Yang Shi) se transformó en células BL21-CodonPlus (DE3)-RIL (Stratagene, 230245). Se inoculó una colonia bacteriana fresca en medio LB que contenía ampicilina y cloranfenicol y se cultivó durante la noche a 37 °C. Estas bacterias se diluyeron 1:10 en 500 ml de LB precalentado con ampicilina y cloranfenicol y se cultivaron durante 1,5 horas a 37 °C. Después de la inducción de la expresión de YY1 con IPTG 1 mM, las células se cultivaron durante otras 5 horas, se recolectaron y almacenaron congeladas a -80 °C hasta que estuvieran listos para usar.

Los sedimentos de 500 ml de células se resuspendieron en 15 ml de tampón A (GuHCl 6 M, Tris 25 mM, NaCl 100 mM,

pH 8,0) que contiene imidazol 10 mM, 2-mercaptoetanol 5 mM, inhibidores de proteasa completos (Roche, 11873580001) y sonicados (diez ciclos de 15 segundos encendido, 60 segundos apagado). El lisado se aclaró mediante centrifugación a 12.000 g durante 30 minutos a 4 °C y se añadió a 1 ml de agarosa Ni-NTA (Invitrogen, R901-15) preequilibrada con volúmenes 10X de tampón A. Los tubos que contenían esta suspensión de lisado de agarosa se rotaron a temperatura ambiente durante 1 hora. La suspensión se vertió en una columna y la agarosa empaquetada se lavó con 15 volúmenes de tampón A que contenía imidazol 10 mM. La proteína se eluyó con 4 x 2 ml de tampón A que contenía imidazol 500 mM.

Las fracciones se extrajeron mediante electroforesis en gel SDS-PAGE y se tiñeron con Coomassie Brilliant Blue (datos no mostrados). Las fracciones que contenían proteína del tamaño correcto y alta pureza se combinaron y diluyeron 1:1 con tampón de elución. Se añadió DTT a una concentración final de 100 mM y se incubó a 60 °C durante 30 minutos. La proteína se replegó por diálisis frente a 2 cambios de 1 litro de Tris-HCl 25 mM pH 8,5, NaCl 100 mM, ZnCl₂ 0,1 mM y DTT 10 mM a 4 °C, seguido de 1 cambio del mismo tampón de diálisis con glicerol al 10 %. La proteína se almacenó en alícuotas a -80 °C.

Caracterización de YY1

La pureza del YY1 recombinante se evaluó mediante electroforesis en gel SDS-PAGE seguida de tinción con azul brillante de Coomassie y transferencia Western (FIG. 17A). La actividad de la proteína recombinante se evaluó mediante EMSA (FIG. 17B).

EMSA se realizó utilizando el kit EMSA quimioluminiscente LightShift (Thermo Scientific n.º 20148) siguiendo las recomendaciones del fabricante. Resumiendo, la proteína recombinante se incubó con una sonda biotinilada en presencia o ausencia de un competidor frío. Las reacciones se separaron usando un gel nativo y se transfirieron a una membrana. El ADN marcado se detectó usando quimioluminiscencia.

Para generar la sonda marcada con biotina, se hibridaron oligonucleótidos monocatenarios (IDT) biotinilados de 30 nucleótidos de largo en 5' en Tris 10 mM pH 7,5, NaCl 50 mM y EDTA 1 mM a una concentración de 50 uM. Se utilizó el mismo protocolo para generar el competidor frío. La sonda se diluyó en serie a una concentración de 10 fmol/μl y el competidor frío a una concentración de 2 pmol/μl. Se usaron 2 μl de sonda diluida y competidor frío para cada reacción de unión para una cantidad final de 20 fmol de sonda marcada y 4 pmol de competidor frío (exceso de 200 veces) en cada reacción.

Las reacciones de unión se prepararon en un volumen de 20 μl que contenía tampón de unión 1x (Tris 10 mM, KCl 50 mM, DTT 1 mM; pH 7,5), glicerol al 2,5 %, 5 mM MgCl₂, 50 ng/μl de Poli dI dC, Np-40 al 0,05 %, ZnCl₂ 0,1 mM, Hepes 10 mM y 2 ug de proteína YY1 recombinante. Las reacciones de unión se preincubaron durante 20 minutos a temperatura ambiente con o sin el competidor frío. A continuación, se añadió la sonda marcada a las reacciones de unión y se incubó durante 80 minutos a temperatura ambiente. Después de la incubación de 80 minutos, se añadió tampón de carga 5x (Thermo Scientific #20148) a la reacción y se procesó en un gel de TBE al 4-12 % usando TBE 0,5x a 40 mA durante 2,5 horas a 4 °C. El gel TBE se corrió previamente durante 1 hora a 4 °C. A continuación, el ADN se transfirió electroforéticamente a una membrana de nailon Biotodyne B (previamente empapada en TBE 0,5x frío durante 10 minutos) a 380 mA durante 30 minutos a 4 °C. A continuación, el ADN se entrecruzó con la membrana colocando la membrana en un transiluminador Dark Reader durante 15 minutos. La membrana se dejó secar al aire a temperatura ambiente durante la noche y se detectó la quimioluminiscencia al día siguiente.

La detección de ADN marcado con biotina se realizó como sigue. La membrana se bloqueó durante 20 minutos utilizando Blocking Buffer (Thermo Scientific #20148). A continuación, la membrana se incubó en tampón de bloqueo/conjugado (Thermo Scientific #20148) durante 15 minutos. Luego, la membrana se lavó cuatro veces con tampón de lavado 1x (Thermo Scientific n.º 20148) durante 5 minutos. A continuación, la membrana se incubó en tampón de equilibrio de sustrato (Thermo Scientific n.º 20148) durante 5 minutos y, a continuación, se incubó en solución de trabajo de sustrato (Thermo Scientific n.º 20148) durante 5 minutos. Luego se tomaron imágenes de la membrana usando una cámara CCD usando una exposición de 120 segundos. Todas estas etapas se realizaron a temperatura ambiente.

Edición genómica:

El sistema CRISPR/Cas9 se utilizó para diseñar genéticamente líneas CME. Los oligonucleótidos específicos de la diana se clonaron en un plásmido que llevaba una versión de Cas9 con codones optimizados con GFP (obsequio de R. Jaenisch). Los oligos utilizados para la clonación se incluyen en la Tabla S5.

Tabla S5. Oligos utilizados en el estudio, relacionados con los métodos STAR				
Nombre	Secuencia (5'-3')	SEQ ID NO:	Uso	
RAF1_prom_F	caccGACTCCCGCATCCAAGATGG	SEQ ID NO: a	Motivo objetivo YY1 en el promotor Raf1	
RAF1_prom_R	aaacCCATCTTGGATGGCGGGAGTC	SEQ ID NO: 9	Motivo objetivo YY1 en el promotor Raf1	
ETV4_prom_F	caccGAGCTACTTGAAAACAAAATGG	SEQ ID NO: 10	Motivo objetivo YY1 en el promotor ETV4	
ETV4_prom_R	aaacCCATTGTTTTCAAAGTAGCTC	SEQ ID NO: 11	Motivo objetivo YY1 en el promotor ETV4	
yy1_sg1_F	CACCGctctctctctctctttcac	SEQ ID NO: 12	Objetivo YY1 para activación	
yy1_sg1_R	AAACgtgaaaagaagagagaagac	SEQ ID NO: 13	Objetivo YY1 para activación	
YY1_gPCR_3F	ctgtgcagtgattgggtcct	SEQ ID NO: 14	Genotipado de activación	
YY1_gPCR_3R	TGCGCGCTCTGCACCTTAAGT	SEQ ID NO: 15	Genotipado de activación	
Raf1_negative_F	GCTTCCTCACATTGAAACAGAA	SEQ ID NO: 16	ChIP-qPCR	
Raf1_negative_R	GGGAAGCTCTGAGAGTCCTTAT	SEQ ID NO: 17	ChIP-qPCR	
Raf1_ROI_F	CGCCACCAGGATGACAG	SEQ ID NO: 18	ChIP-qPCR	
Raf1_ROI_R	GAATGTGACCGCAACCAAC	SEQ ID NO: 19	ChIP-qPCR	
Etv4_negative_F	CATTTTACCTGCCCCAGTA	SEQ ID NO: 20	ChIP-qPCR	
Etv4_negative_R	CAGCCTTAAACAGCCTGGAA	SEQ ID NO: 21	ChIP-qPCR	
Etv4_ROI_F	TTTCAAAGCCACCAAGTCT	SEQ ID NO: 22	ChIP-qPCR	
Etv4_ROI_R	CAAGTAGCTCGGGGTCTCAG	SEQ ID NO: 23	ChIP-qPCR	
Puente_enlazador_F	/5Phos/CGCGATATC//BIodT/TATCTGACT	SEQ ID NO: 24	ChIA-PET	
Puente_enlazador_R	/5Phos/GTCAGATAAGATATCGCGT	SEQ ID NO: 25	ChIA-PET	
ligadura_R	GTCTGGATCCTCGTCTTTGAGCC	SEQ ID NO: 26	Amplificar el molde para la ciclación de ADN mediada por ligadura	
ligadura_F	CCAAAGGATCCGTAAGCTAGGCT	SEQ ID NO: 27	Amplificar el molde para la ciclación de ADN mediada por ligadura	

Tabla S5. Oligos utilizados en el estudio, relacionados con los métodos STAR

Nombre	Secuencia (5'-3')	Uso
competidor_ADN	GAGCGACACACACAAACGAGCGGTTCGACCTCCCGCGGCATCTTTTCGACC TCCCGGGGCATCTTTTCGACCTCCCGGGGCATCTTTTCGACCTCCCGGGGCAT CTTTTCGACCTCCCGGGGCATCTTTTCGACCTCCCGGGGCATCTTTTCGACCTC GCGGCATCTTTTCGACCTCCCGGGGCATCTTTTCGACCTCCCGGGGCAT SEQ ID NO: 28	ADN competidor en ciclación de ADN mediada por ligadura SEQ ID NO: 28
Raf1_4C_forward	CAAGGGCAAGTAACCCGATC	Cebador sin cola utilizado para amplificar bibliotecas de Raf1 4C
Raf1_4C_reverse	AATAGATACATCCCGACCT	Cebador sin cola utilizado para amplificar bibliotecas de Raf1 4C
Etv4_4C_forward	CAAGGGCAAGTAACCCGATC	Cebador sin cola utilizado para amplificar bibliotecas de Etv4 4C
Etv4_4C_reverse	AATAGATACATCCCGACCT	Cebador sin cola utilizado para amplificar bibliotecas de Etv4 4C
EMSA_directo	TCGCTCCCGGGCCATCTTGGCGGCTGGTGT	Sonda utilizada en EMSA
EMSA_inverso	ACACGAGCCGCCAAGATGGCCGGGAGCGCA	Sonda utilizada en EMSA
etv4_p_sgT1_F	caccgAAGTAGCTCGGGGTCTCAGA	Diana dCas9/dCas9-YY1 al promotor Etv4
etv4_p_sgT1_R	aaacTCTGAGACCCCCGAGCTACTTc	Diana dCas9/dCas9-YY1 al promotor Etv4
etv4_p_sgT2_F	caccGGTGTCTCAGTAAATGTAAAC	Diana dCas9/dCas9-YY1 al promotor Etv4
etv4_p_sgT2_R	aaacGTTTACATTTTACTGAGCAACC	Diana dCas9/dCas9-YY1 al promotor Etv4

Las secuencias del ADN objetivo (el motivo adyacente al protoespaciador está subrayado) se enumeran a continuación:

Locus	ADN diana
Raf1_promoter	5'-ACTCCCGCCATCCAAGATGG <u>C</u> GG-3' -SEQ ID NO: 39
Etv4_promoter	5'-GAGCTACTTGAAAACAAATGG <u>A</u> GG-3'-SEQ ID NO: 40
YY1_stop_codon	5'-GTCTTCTCTCTTCTTTTCACT <u>I</u> GG-3'-SEQ ID NO: 41

- 5 Para las supresiones de motivo, se transfectaron quinientas mil células MEr con 2,5 µg de plásmido y se clasificaron 48 horas más tarde en cuanto a la presencia de GFP. Treinta mil células clasificadas positivas para GFP se sembraron en una placa de seis pocillos en una dilución en serie 1:2 (primer pocillo 15.000 células, segundo pocillo 7.500 células, etc.). Las células se cultivaron durante aproximadamente una semana en 2i LIF. Se recogieron colonias individuales utilizando un estereoscopio en una placa de 96 pocillos. Las células se expandieron y genotiparon mediante PCR y
- 10 secuenciación de Sanger. Los clones con deleciones que abarcaban el motivo se ampliaron aún más y se usaron para experimentos.

- 15 Para la generación de líneas marcadas endógenamente, quinientas mil células MEr se transfectaron con 2,5 ug de plásmido Cas9 y 1,25 ug de plásmido de reparación no linealizado 1 (pAW62.YY1.FKBP.knock-in.mCherry) y 1,25 ug de plásmido de reparación no linealizado 2 (pAW63.YY1.FKBP. knock-in.BFP). Las células se clasificaron después de 48 horas por la presencia de GFP. Las células se expandieron durante cinco días y luego se clasificaron de nuevo para células mCherry y BFP doblemente positivas. Treinta mil células clasificadas con mCherry/BFP se sembraron en una placa de seis pocillos en una dilución en serie 1:2 (primer pocillo 15.000 células, segundo pocillo 7.500 células, etc.). Las células se cultivaron durante aproximadamente una semana en medio 2i y luego se recogieron colonias
- 20 individuales utilizando un estereoscopio en una placa de 96 pocillos. Las células fueron expandidas y genotipadas mediante PCR (YY1_gPCR_3F/3R, Tabla S3). Los clones con un marcador de activación homocigótica se expandieron aún más y se usaron para experimentos.

Tabla S3. Análisis GO, relacionado con la figura 7

- 25 Tipo de análisis: Prueba de sobrerepresentación PANTHER (versión 20170413)
- Versión de la anotación y fecha de publicación: PANTHER versión 11.1 Lanzamiento 24/10/2016
- Lista analizada: Entrada de cuadro de texto del cliente (Mus musculus)
- Lista de referencias: Mus musculus (todos los genes en la base de datos)
- Corrección de Bonferroni: VERDADERO
- Recuento de Bonferroni: 241

Proceso biológico PANTHER GO	N.º de genes de ratón en la categoría	N.º de genes dif. expresados	N.º de genes esperados	valor de p
proceso metabólico (GO:0008152)	6955	3136	2436,52	1.74E-60
proceso metabólico de compuestos nitrogenados (GO:0006807)	2001	1034	701	3.96E-33
proceso biosintético (GO:0009058)	1520	792	532,5	1.35E-25
organización de componentes celulares o biogénesis (GO:0071840)	1767	824	619,03	3.14E-14
procesamiento de ARNm (GO:0006397)	247	171	86,53	1.16E-13

transcripción del promotor de ARN polimerasa II (GO:0006366)	1212	585	424,6	4.25E-12
corte y empalme de ARNm, mediante espliceosoma (GO:0000398)	185	127	64,81	1.07E-09
organización de componentes celulares (GO:0016043)	1627	729	569,98	3.84E-09

(continuación)

proceso metabólico del ADN (GO:0006259)	344	197	120,51	1,75E-08
proceso metabólico de compuestos que contienen fosfato (GC:0006796)	1251	563	438,26	5.19E-07
proceso metabólico del ARNt (GO:0006399)	108	79	37,84	7.58E-07
Corte y empalme de ARN, mediante reacciones de transesterificación (GO:0000375)	144	97	50,45	8.20E-07
mitosis (GO:0007067)	355	194	124,37	8.23E-07
reparación de ADN (GC:0006281)	162	104	56,75	2,55E-06
direccionamiento a proteínas (GO:0006605)	151	95	52,9	2,59E-05
localización de proteínas (GO:0008104)	236	134	82,68	2,73E-05
segregación cromosómica (GO:0007059)	111	74	38,89	7.87E-05
regulación del ciclo celular (GO:0051726)	107	72	37,48	8.07E-05
proceso catabólico del ARN (GO:0006401)	58	46	20,32	1.60E-04
proceso metabólico del ARNr (GO:0016072)	118	76	41,34	1.95E-04
regulación de la transcripción del promotor de la ARN polimerasa II (GC:0006357)	952	420	333,51	4.16E-04
transporte nuclear (GO:0051169)	99	65	34,68	6.33E-04
iniciación de la transcripción del promotor de la ARN polimerasa II (GC:0006367)	50	37	17,52	7.84E-03
localización (GO:0051179)	2177	868	762,66	1.05E-02
poliadenilación de ARNm (GO:0006378)	25	23	8,76	1.07E-02
procesamiento del extremo 3' del ARNm (GO:0031124)	29	25	10,16	1.43E-02
plegamiento de proteínas (GO:0006457)	119	69	41,69	1.53E-02
organización del citoesqueleto (GO:0007010)	154	83	53,95	3.31E-02
proceso metabólico de aminoácidos celulares (GO:0006520)	253	124	88,63	4.96E-02

ChIP:

- 5 ChIP se realizó como se describe en (Lee *et al.*, 2006) con algunas adaptaciones. Las células MEF se agotaron de MEF dividiéndolas dos veces en placas recién gelatinizadas sin MEF. Aproximadamente 50 millones de células MEF se entrecruzaron durante 15 minutos a temperatura ambiente mediante la adición de una décima parte del volumen de solución fresca de formaldehído al 11 % (formaldehído al 11 %, HEPES 50 mM pH 7,3, NaCl 100 mM, EDTA 1 mM
- 10 pH 8,0, EGTA 0,5 mM pH 8,0) al medio de cultivo seguido de 5 minutos de inactivación con glicina 125 mM. Las células se enjuagaron dos veces con PBS 1X y se recolectaron usando un raspador de silicona y se congelaron rápidamente en nitrógeno líquido. Se entrecruzaron células Jurkat durante 10 minutos en medio a una concentración de 1 millón de células/ml. Las células reticuladas congeladas se almacenaron a -80 °C.
- 15 Se lavaron 100 µl de Protein G Dynabeads (Life Technologies #10009D) 3 veces durante 5 minutos con BSA al 0,5 % (p/v) en PBS. Las perlas magnéticas se unieron con 10 µg de anticuerpo anti-YY1 (Santa Cruz, sc-281X) durante la noche a 4 °C y luego se lavaron 3 veces con BSA al 0,5 % (p/v) en PBS.
- 20 Las células se prepararon para ChIP de la siguiente manera. Todos los tampones contenían inhibidores de proteasa completos 1 x recién preparados (Roche, 11873580001). Las células reticuladas congeladas se descongelaron en hielo y luego se resuspendieron en tampón de lisis I (HEPES-KOH 50 mM, pH 7,5, NaCl 140 mM, EDTA 1 mM, glicerol al 10 %, NP-40 al 0,5 %, Triton X-100 al 0,25 %, inhibidores de la proteasa 1x) y se rota durante 10 minutos a 4 °C, luego se centrifugó a 1350 rcf durante 5 minutos a 4 °C. El sedimento se resuspendió en tampón de lisis II (Tris-HCl 10 mM, pH 8,0, NaCl 200 mM, EDTA 1 mM, EGTA 0,5 mM, 1 x inhibidores de la proteasa) y rotar durante 10 minutos
- 25 a 4 °C y centrifugar a 1350 rcf durante 5 minutos a 4 °C. El sedimento se resuspendió en tampón de sonicación (Tris-HCl 20 mM, pH 8,0, NaCl 150 mM, EDTA 2 mM, pH 8,0, SDS al 0,1 % y Triton X-100 al 1 %, 1 x inhibidores de la proteasa) y luego sonificado en un sonicador Misonix 3000 durante 10 ciclos de 30 segundos cada uno en hielo (18-21 W) con 60 segundos encendidos. hielo entre ciclos. Los lisados sonicados se aclararon una vez mediante centrifugación a 16.000 rcf durante 10 minutos a 4 °C. Se reservaron 50 µl para la entrada y luego el resto se incubó

durante la noche a 4 °C con perlas magnéticas unidas con anticuerpo para enriquecer los fragmentos de ADN unidos por el factor indicado.

Las perlas se lavaron dos veces con cada uno de los siguientes tampones: tampón de lavado A (HEPES-KOH 50 mM pH 7,5, NaCl 140 mM, EDTA 1 mM pH 8,0, desoxicolato sódico al 0,1 %, 1 % Tritón X-100, 0,1 % SDS), tampón de lavado B (HEPES-KOH 50 mM pH 7,9, NaCl 500 mM, EDTA 1 mM pH 8,0, desoxicolato sódico al 0,1 %, 1 % Tritón X-100, 0,1 % SDS), tampón de lavado C (Tris-HCl 20 mM pH 8,0, LiCl 250 mM, EDTA 1 mM pH 8,0, desoxicolato sódico al 0,5 %, 0,5 % IGEPAL C-630 0,1 % SDS), tampón de lavado D (TE con 0,2 % Tritón X-100) y tampón TE. El ADN se eluyó de las perlas mediante incubación a 65 °C durante 1 hora con agitación vorticial intermitente en 200 µl de tampón de elución (Tris-HCl 50 mM, pH 8,0, EDTA 10 mM, SDS al 1 %). Los enlaces cruzados se revirtieron durante la noche a 65 °C. Para purificar el ADN eluido, se añadieron 200 µl de TE y luego se degradó el ARN mediante la adición de 2,5 µl de 33 mg/ml de ARNasa A (Sigma, R4642) y se incubó a 37 °C durante 2 horas. La proteína se degradó mediante la adición de 10 µl de 20 mg/ml de proteinasa K (Invitrogen, 25530049) y la incubación a 55 °C durante 2 horas. Se realizó una extracción con fenol:cloroformo:alcohol isoamílico seguida de una precipitación con etanol. Luego, el ADN se resuspendió en 50 µl de TE y se usó para qPCR o secuenciación. Para experimentos ChIP-qPCR, la qPCR se realizó con Power SYBR Green mix (Life Technologies n.º 4367659) en un sistema QuantStudio 5 o QuantStudio 6 (Life Technologies). Los valores mostrados en las figuras se normalizaron a la entrada, una región de control negativo y valores de tipo salvaje de acuerdo con las siguientes fórmulas:

$$\text{Norma de entrada} = 2^{(Ct_{\text{entrada}} - Ct_{\text{chip}})}$$

$$\text{norma neg} = \frac{\text{Factor de cambio}_{\text{RDI}}}{\text{Factor de cambio}_{\text{neg}}}$$

$$\text{norma SV} = \frac{\text{Neg norma}_{\text{mut}}}{\text{Neg norma}_{\text{TS}}}$$

Las qPCR se realizaron por triplicado técnico y las ChIP por triplicado biológico. Los valores fueron comparables entre los duplicados. Se muestran los valores normales de SV promedio y la desviación estándar (FIG. 6A, 6B). Los cebadores utilizados se enumeran en la Tabla S5.

Para experimentos ChIP-seq, el ADN de ChIP purificado se usó para preparar bibliotecas de secuenciación multiplexadas de Illumina. Las bibliotecas para la secuenciación de Illumina se prepararon siguiendo el kit de preparación de muestras de ADN Illumina TruSeq v2. Las bibliotecas amplificadas se seleccionaron por tamaño utilizando un casete de gel al 2 % en el sistema Pippin Prep de Sage Science para capturar fragmentos entre 200 y 400 pb. Las bibliotecas se cuantificaron mediante qPCR utilizando el kit Illumina Library Quantification de KAPA Biosystems de acuerdo con los protocolos del kit. Las bibliotecas se secuenciaron en Illumina HiSeq 2500 para 40 bases en modo de lectura única.

ChIA-PET

ChIA-PET se realizó utilizando una versión modificada (Tang *et al.*, 2015) de un protocolo previamente descrito (Fullwood *et al.*, 2009). células MEr (~ 500 millones de células, cultivadas hasta -80 % de confluencia) se entrecruzaron con formaldehído al 1 % a temperatura ambiente durante 15 minutos y luego se neutralizaron con glicina 125 mM. Las células entrecruzadas se lavaron tres veces con PBS enfriado con hielo, se ultracongelaron en nitrógeno líquido y se almacenaron a -80 °C antes de su posterior procesamiento. Los núcleos se aislaron como se describió anteriormente y la cromatina se fragmentó utilizando un sonicador Misonix 3000. Se usaron anticuerpos CTCF o YY1 para enriquecer los fragmentos de cromatina unidos a proteínas exactamente como se describe en la sección ChIP-seq. Se eluyó una porción de ADN de ChIP de perlas recubiertas de anticuerpos para la cuantificación de la concentración y para el análisis de enriquecimiento mediante qPCR. Para la construcción de la biblioteca ChIA-PET, los fragmentos de ADN de ChIP se repararon en los extremos utilizando la ADN polimerasa T4 (NEB n.º M0203) seguido de la cola A con Klenow (NEB M0212). Los oligos del enlazador puente (Tabla S5) se recogieron para generar un enlazador puente de doble hebra con salientes en T. Se añadieron 800 ng de conector puente y la ligadura de proximidad se realizó durante la noche a 16 °C en un volumen de 1,5 ml. Luego, el ADN no ligado se digirió con exonucleasa y nucleasa lambda (NEB M0262S, M0293 S). El ADN se eluyó de las perlas mediante incubación a 65 °C durante 1 hora con agitación vorticial intermitente en 200 µl de tampón de elución (Tris-HCl 50 mM, pH 8,0, EDTA 10 mM, SDS al 1 %). Los enlaces cruzados se revirtieron durante la noche a 65 °C. Para purificar el ADN eluido, se añadieron 200 µl de TE y luego se degradó el ARN mediante la adición de 2,5 µl de 33 mg/ml de ARNasa A (Sigma, R4642) y se incubó a 37 °C durante 2 horas. La proteína se degradó mediante la adición de 10 µl de 20 mg/ml de proteinasa K (Invitrogen, 25530049) y la incubación a 55 °C durante 2 horas.

Se realizó una extracción con fenol:cloroformo:alcohol isoamílico seguida de una precipitación con etanol. El ADN precipitado se resuspendió en tampón de resuspensión de ADN Nextera (Illumina FC-121-1030). A continuación, se marcó el ADN con el kit Nextera Tagmentation (Illumina FC-121-1030). Se usaron 5 µl de transposón por 50 ng de ADN. La biblioteca etiquetada se purificó con un Zymo DNA Clean & Concentrator (Zymo D4003) y se unió a perlas de estreptavidina (Life Technologies n.º 11205D) para enriquecer las uniones de ligadura (que contienen el enlazador

de puente biotinilado). Se realizaron 12 ciclos de la reacción en cadena de la polimerasa para amplificar la biblioteca utilizando cebadores Nextera estándar (Illumina FC-121-1030). La biblioteca amplificada se seleccionó por tamaño (350-500 pb) y se secuenció mediante secuenciación de extremos emparejados en una plataforma Illumina Hi-Seq 2500.

5

HiChIP

HiChIP se realizó como se describe en (Mumbach *et al.*, 2016) con algunas modificaciones. Diez millones de células se entrecruzaron durante 10 minutos a temperatura ambiente con formaldehído al 1 % en medios de cultivo y se extinguieron en glicina 0,125 M. Después de lavar dos veces con PBS helado, el sobrenadante se aspiró y el sedimento celular se congeló rápidamente en nitrógeno líquido y se almacenó a -80 °C.

10

15

20

Los sedimentos de células reticuladas se descongelaron en hielo, se resuspendió en 800 µl de tampón de lisis Hi-C enfriado con hielo (Tris-HCl 10 mM, pH 8,0, NaCl 10 mM e IGEPAL CA-630 al 0,2 % con 1x inhibidor de proteasa completo (Roche, 11697498001)), y se incubó a 4 °C durante 30 minutos con rotación. Los núcleos se sedimentaron mediante centrifugación a 2500 rcf durante 5 min a 4 °C y se lavaron una vez con 500 µl de tampón de lisis Hi-C enfriado con hielo. Después de eliminar el sobrenadante, los núcleos se resuspendieron en 100 µl de SDS al 0,5 % y se incubaron a 62 °C durante 10 minutos. El SDS se inactivó añadiendo 335 µl de Triton X-100 al 1,5 % e incubando durante 15 minutos a 37 °C. Después de añadir 50 µl de 10X NEB Tampón 2 (NEB, B7002) y 375 U de enzima de restricción MboI (NEB, R0147), la cromatina se digirió a 37 °C durante 2 horas con rotación. Después de la digestión, la enzima MboI se inactivó con calor incubando los núcleos a 62 °C durante 20 minutos.

25

30

Para rellenar los fragmentos de restricción que sobresalen y marcar los extremos del ADN con biotina, se añadieron 52 µl de mezcla maestra de relleno, que contiene 37,5 µl de biotina-dATP 0,4 mM (Invitrogen, 19524016), 1,5 µl de dCTP 10 mM (Invitrogen, 18253013), 1,5 µl de dGTP 10 mM (Invitrogen, 18254011), 1,5 µl de dTTP 10 mM (Invitrogen, 18255018) y 10 µl de 5 U/µl de ADN polimerasa I. Se añadió fragmento grande (Klenow) (NEB, M0210) y los tubos se incubaron a 37 °C durante 1 hora con rotación. La ligadura de proximidad se realizó mediante la adición de 947 µl de mezcla maestra de ligadura, que contiene 150 µl de tampón de ADN ligasa de T4 NEB 10X (NEB, B0202), 125 µl de Triton X-100 al 10 %, 7,5 µl de 20 mg/ml de BSA (NEB, B9000), 10 µl de 400 U/µl de ADN ligasa DE T4 (NEB, M0202), y 655,5 µl de agua, e incubación a temperatura ambiente durante 4 horas con rotación.

35

40

45

50

55

Después de la ligadura de proximidad, los núcleos se sedimentaron por centrifugación a 2500 rcf durante 5 minutos y se resuspendieron en 1 ml de tampón de sonicación ChIP (HEPES-KOH 50 mM pH 7,5, NaCl 140 mM, EDTA 1 mM pH 8,0, EGTA 1 mM pH 8,0, Triton X-100 al 1 %, desoxicolato de sodio al 0,1 % y SDS al 0,1 % con inhibidor de la proteasa). Los núcleos se sonicaron usando un Covaris S220 durante 6 minutos con los siguientes ajustes: llenar el nivel 8, ciclo de trabajo 5, pico de potencia de incidencia 140, ciclos por ráfaga 200. La cromatina sonicada se aclaró mediante centrifugación a 16.100 rcf durante 15 minutos a 4 °C y el sobrenadante se transfirió a un tubo. Se lavaron 60 µl de perlas magnéticas de proteína G tres veces con tampón de sonicación, se resuspendieron en 50 µl de tampón de sonicación. Luego se añadieron perlas lavadas a la cromatina sonicada y se incubaron durante 1 hora a 4 °C con rotación. A continuación, las perlas se separaron en un soporte magnético y el sobrenadante se transfirió a un tubo nuevo. Se añadieron al tubo 7,5 µg de anticuerpo H3K27ac (Abcam, ab4729) o 7,5 µg de anticuerpo YY1 (Abcam, ab109237) y el tubo se incubó durante la noche a 4 °C con rotación. Para YY1 se llevaron a cabo seis reacciones y se agruparon antes del marcaje. Al día siguiente, se lavaron 60 µl de perlas magnéticas de proteína G tres veces en BSA al 0,5 % en PBS y se lavaron una vez con tampón de sonicación antes de resuspenderse en 100 µl de tampón de sonicación y agregarse a cada tubo de muestra. Las muestras se incubaron durante 2 horas a 4 °C con rotación. Luego, las perlas se separaron en un soporte magnético y se lavaron tres veces con 1 ml de tampón de sonicación con alto contenido de sal (HEPES-KOH 50 mM, pH 7,5, NaCl 500 mM, EDTA 1 mM pH 8,0, EGTA 1 mM pH 8,0, Triton X-100 al 1 %, desoxicolato de sodio al 0,1 %, SDS al 0,1 %) seguido de tres veces de 1 ml de tampón de lavado LiCl (Tris-HCl 20 mM pH 8,0, 1 mM EDTA pH 8,0, LiCl 250 mM, 0,5 % de IGEPAL CA-630, 0,5 % de desoxicolato de sodio, SDS al 0,1 %) y una vez con 1 ml de TE con sal (Tris-HCl 10 mM pH 8,0, EDTA 1 mM pH 8,0, NaCl 50 mM). A continuación, las perlas se resuspendieron en 200 µl de tampón de elución (Tris-HCl 50 mM pH 8,0, EDTA 10 mM pH 8,0, SDS al 1 %) y se incubaron a 65 °C durante 15 minutos para eluir. Para purificar el ADN eluido, el ARN se degradó mediante la adición de 2,5 µl de 33 mg/ml de ARNasa A (Sigma, R4642) e incubación a 37 °C durante 2 horas. La proteína se degradó mediante la adición de 10 µl de 20 mg/ml de proteinasa K (Invitrogen, 25530049) e incubación a 55 °C durante 45 minutos. Luego, las muestras se incubaron a 65 °C durante 5 horas para revertir los enlaces cruzados. A continuación, el ADN se purificó utilizando columnas Zymo DNA Clean and Concentrate 5 (Zymo, D4013) según el protocolo del fabricante y se eluyó en 14 µl de agua. La cantidad de ADN eluido se cuantificó mediante el kit Qubit dsDNA HS (Invitrogen, Q32854).

60

65

El marcaje del ADN de ChIP se realizó con el kit de preparación de bibliotecas de ADN Illumina Nextera (Illumina, FC-121-1030). En primer lugar, se lavaron 5 µl de perlas magnéticas de estreptavidina C1 (Invitrogen, 65001) con 1 ml de tampón de lavado tween (Tris-HCl 5 mM pH 7,5, EDTA 0,5 mM pH 8,0, NaCl 1 M, Tween-20 al 0,05 %) y se resuspendieron en 10 µl de tampón de unión a biotina 2X (Tris-HCl 10 mM, pH 7,5, EDTA 1 mM, pH 8,0, NaCl 2 M). Se añadieron 54,19 ng de ADN purificado en un volumen total de 10 µl de agua a las perlas y se incubaron a temperatura ambiente durante 15 minutos con agitación cada 5 minutos. Después de la captura, las perlas se separaron con un imán y se descartó el sobrenadante. Luego, las perlas se lavaron dos veces con 500 µl de tampón

de lavado Tween, incubando a 55 °C durante 2 minutos con agitación para cada lavado. Las perlas se resuspendieron en 25 µl de tampón de ADN Nextera Tagment. Para marcar el ADN capturado, se añadieron 3,5 µl de Nextera Tagment DNA Enzyme 1 con 21,5 µl de Nextera Resuspension Buffer y las muestras se incubaron a 55 °C durante 10 minutos con agitación. Las perlas se separaron en un imán y se descartó el sobrenadante. Las perlas se lavaron con 500 µl de EDTA 50 mM a 50 °C durante 30 minutos, luego se lavaron tres veces con 500 µl de tampón de lavado tween a 55 °C durante 2 minutos cada vez, y finalmente se lavaron una vez con 500 µl de Tris-HCl 10 mM pH 7,5 durante 1 minuto a temperatura ambiente. Las perlas se separaron en un imán y se descartó el sobrenadante.

Para generar la biblioteca de secuenciación, se realizó amplificación por PCR del ADN marcado mientras el ADN aún estaba unido a las perlas. Las perlas se resuspendieron en 15 µl de Nextera PCR Master Mix, 5 µl de cóctel de cebadores de PCR Nextera, 5 µl de Nextera Index Primer 1, 5 µl de Nextera Index Primer 2 y 20 µl de agua. El ADN se amplificó con 8 ciclos de PCR. Después de la PCR, las perlas se separaron en un imán y el sobrenadante que contenía la biblioteca amplificada por PCR se transfirió a un tubo nuevo, se purificó con el kit Zymo DNA Clean and Concentrate-5 (Zymo D4003T) según el protocolo del fabricante y eluido en 14 µl de agua. Las bibliotecas HiChIP purificadas se seleccionaron en tamaño de 300 a 700 pb con un instrumento Sage Science Pippin Prep de acuerdo con el protocolo del fabricante y se sometieron a secuenciación de extremos emparejados en un Illumina HiSeq 2500. Las bibliotecas se secuenciaron inicialmente con secuenciación de extremos emparejados de 100 x 100 pb. Se realizó una segunda ronda de secuenciación en las mismas bibliotecas con secuenciación de extremos emparejados de 50x50 pb.

4C-seq:

Se desarrolló una versión modificada de 4C-seq (van de Werken *et al.*, 2012; Van De Werken *et al.*, 2012). El principal cambio fue que la ligadura de proximidad se realiza en núcleos intactos (*in situ*). Este cambio se incorporó porque trabajos anteriores han observado que la ligadura *in situ* reduce drásticamente la tasa de ligaduras quiméricas y las interacciones de fondo (Nagano *et al.*, 2015; Rao *et al.*, 2014).

Aproximadamente 5 millones de células MEr se digirieron con tripsina y luego se resuspendieron en 5 ml de FBS/PBS al 10 %. Se añadieron 5 ml de formaldehído al 4 % en FBS/PBS al 10 % y las células se entrecruzaron durante 10 minutos. Se añadió glicina hasta una concentración final de 0,125 M y las células se centrifugaron a 300 rcf durante 5 minutos. Las células se lavaron dos veces con PBS, se transfirieron a un tubo Eppendorf de 1,5 ml, se ultracongelaron y se almacenaron a -80 °C.

Los sedimentos se resuspendieron suavemente en tampón de lisis Hi-C (Tris-HCl 10 mM, pH 8, NaCl 10 mM, 0,2 % Igepal) con 1x inhibidores completos de la proteasa (Roche 11697498001). Las células se incubaron en hielo durante 30 minutos y luego se lavaron una vez con 500 µl de tampón de lisis Hi-C enfriado con hielo sin inhibidores de proteasa. Los sedimentos se resuspendieron en 50 µl de SDS al 0,5 % y se incubaron a 62 °C durante 7 minutos. Se añadieron 145 µl de H₂O y 25 µl de Triton X-100 al 10 % y los tubos se incubaron a 37 °C durante 15 minutos. Se añadieron 25 µl del tampón de enzimas de restricción 10X New England Biolabs apropiado y 200 unidades de enzima y se incubó la cromatina a 37 °C grados en termomezcladora a 500 RPM durante cuatro horas. Se añadieron 200 unidades más de enzima y la reacción se incubó toda la noche a 37 °C grados en termomezcladora a 500 RPM, luego se añadieron 200 unidades más y la reacción se incubó otras cuatro horas a 37 °C grados en termomezcladora a 500 RPM. DpnII (NEB) se utilizó como cortador principal tanto para Raf1 como para ETV4. La enzima de restricción se inactivó calentando a 62 °C durante 20 minutos mientras se agitaba a 500 rpm. La ligadura de proximidad se realizó en un total de 1200 µl con 2000 unidades de ADN ligasa T4 (NEB M020) durante seis horas a temperatura ambiente. Después de la ligadura, las muestras se centrifugaron durante 5 minutos a 2500 rcf y se resuspendieron en 300 µl de Tris-HCl 10 mM, SDS al 1 % y NaCl 0,5 mM con 1000 unidades de proteinasa K. Los entrecruzamientos se revirtieron mediante incubación durante la noche a 65 °C.

A continuación, las muestras se extrajeron con fenol-cloroformo y se precipitaron con etanol y la segunda digestión se realizó durante la noche en 450 µl con 50 unidades de enzima de restricción. Se usó BfaI (NEB R0568S) para ETV4 y CviQI (NEB R0639S) para Raf1. Las muestras se extrajeron con fenol-cloroformo y se precipitaron con etanol y la segunda ligadura se realizó en 14 ml en total con 6700 unidades de ADN ligasa T4 (NEB M020) a 16 °C durante la noche. Las muestras se precipitaron con etanol, se resuspendieron en 500 µl de tampón Qiagen EB y se purificaron con un kit de purificación Qiagen PCR.

La amplificación por PCR se realizó con 16 reacciones de PCR de 50 µl utilizando la polimerasa Roche Expand Long Template (Roche 11759060001). Las condiciones de reacción son las siguientes: 11,2 µl de Roche Expand Long Template Polymerase, 80 µl de 10 x Roche Tampón 1, 16 µl de dNTP 10 mM (Promega PAU1515), 112 µl de cebador directo 10 uM, 112 µl de cebador inverso 10 uM (Tabla S5), 200 ng de molde y agua milli-q hasta 800 µl en total. Las reacciones se mezclaron y luego se distribuyeron en 16 reacciones de 50 µl para la amplificación. Las condiciones de ciclo fueron una "PCR de contacto" basada en informes de que esto disminuye la amplificación no específica de bibliotecas 4C (Ghavi-Helm *et al.*, 2014). Las condiciones son: 2' a 94 °C, 10" a 94 °C, 1' a 63 °C, 3' a 68 °C, repetir las etapas 2-4 pero disminuir la temperatura de hibridación un grado, hasta que se alcanzan los 53 °C, momento en el cual la reacción se cicla 15 veces más a 53 °C, después de realizar un total de 25 ciclos, la reacción se mantiene durante 5' a 68 °C y luego a 4 °C. Las bibliotecas se limpiaron utilizando un kit de purificación de PCR de Roche (Roche

11732676001) utilizando 4 columnas por biblioteca. Luego, las reacciones se purificaron adicionalmente con perlas Ampure XP (Agencourt A63882) con una proporción de 1:1 de solución de perlas a biblioteca siguiendo las instrucciones del fabricante. A continuación, las muestras se cuantificaron con Qubit y el kit Illumina Library Quantification de KAPA Biosystems de acuerdo con los protocolos del kit. Las bibliotecas se secuenciaron en Illumina HiSeq 2500 para 40 bases en modo de lectura única.

Aislamiento de ARN, qRT-PCR y secuenciación

El ARN se aisló utilizando el RNeasy Mini Kit (Qiagen, 74136) de acuerdo con las instrucciones del fabricante.

Para ensayos de RT-qPCR, se realizó transcripción inversa utilizando SuperScript III Reverse Transcriptase (Invitrogen, 18080093) con cebadores oligo-dT (Promega, C1101) de acuerdo con las instrucciones del fabricante. La PCR cuantitativa en tiempo real se realizó en Applied Biosystems 7000, Instrumentos QuantStudio 5 y QuantStudio 6 que utilizan sondas TaqMan para *Raf1* (Applied Biosystems, Mm00466513_m1) y *Etv4* (Applied Biosystems, Mm00476696_m1) junto con TaqMan Universal PCR Master Mix (Applied Biosystems, 4304437) según las instrucciones del fabricante.

Para experimentos de RNA-seq, las bibliotecas seleccionadas de poliA de cadena se prepararon con el kit de preparación de bibliotecas de ARNm de cadena TruSeq (Illumina, RS-122-2101) de acuerdo con el protocolo estándar del fabricante. Las bibliotecas se sometieron a secuenciación de un solo extremo de 40 pb en un instrumento Illumina HiSeq 2500.

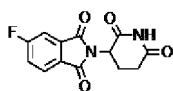
Degradación de YY1

Para los experimentos de degradación se usó una línea homocigota clonal que expresaba YY1 marcado con FKBP. Las células se cultivaron en dos pases de MEF y luego se trataron con dTAG-47 a una concentración de 500 nM durante 24 horas.

Experimentos de lavado dTAG-47

La línea de activación homocigótica que expresaba YY1 marcado con FKBP se cultivó en medio 2i LIF. Las células se trataron con dTAG-47 a una concentración de 500 nM durante 24 horas. Después de 24 horas de tratamiento farmacológico, las células se lavaron tres veces con PBS y se pasaron a una placa nueva. Luego, las células se alimentaron diariamente y se pasaron a una nueva placa cada 48 horas hasta que se restauraron los niveles de proteína YY1 (5 días después de la retirada del fármaco). Luego, las células se recolectaron para la extracción de proteínas o ARN o se entrecruzaron para ChIP o HiChIP.

Síntesis de dTAG-47

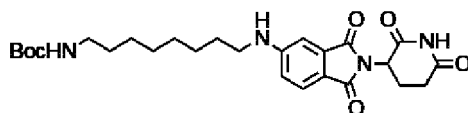


2-(2,6-dioxopiperidin-3-yl)-5-fluoroisoindolin-1,3-dione

Se disolvieron anhídrido 4-fluoroftálico (3,32 g, 20 mmol, 1 eq) y sal clorhidrato de 3-aminopiperidin-2,6-diona (3,620 g, 22 mmol, 1,1 eq) en AcOH (50 ml) seguido de acetato de potasio (6,08 g, 62 mmol, 3,1 eq.). La mezcla se equipó con un condensador de aire y se calentó a 90 °C. Después de 16 horas, la mezcla se diluyó con 200 ml de agua y se enfrió sobre hielo. A continuación, la suspensión se centrifugó (4000 rpm, 20 minutos, 4 °C) y se decantó. El sólido restante se resuspendió en agua, se centrifugó y se decantó nuevamente. Luego, el sólido se disolvió en MeOH y se filtró a través de un lecho de sílice (que se había humedecido previamente con MeOH), se lavó con MeOH al 50 %/DCM y se concentró a presión reducida para producir el producto deseado como un sólido gris (2,1883 g, 7,92 mmol, 40 %).

RMN ¹H (500 MHz, DMSO-*d*₆) δ 11,13 (s, 1H), 8,01 (dd, *J* = 8,3, 4,5 Hz, 1H), 7,85 (dd, *J* = 7,4, 2,2 Hz, 1H), 7,72 (ddd, *J* = 9,4, 8,4, 2,3 Hz, 1H), 5,16 (dd, *J* = 12,9, 5,4 Hz, 1H), 2,89 (ddd, *J* = 17,2, 13,9, 5,5 Hz, 1H), 2,65 - 2,51 (m, 2H), 2,07 (dtd, *J* = 12,9, 5,3, 2,2 Hz, 1H).

LCMS 277,22 (M+H).

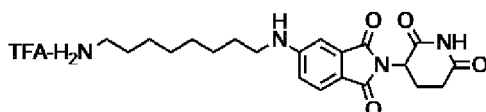


(8-((2-(2,6-dioxopiperidin-3-il)-1,3-dioxoisindolin-5-il)amino)octil)carbamato de *terc*-butilo

2-(2,6-dioxopiperidin-3-il)-5-fluoroisindolin-1,3-diona (294 mg, 1,06 mmol, 1 eq) y (8-aminoctil)carbamato de *terc*-butilo (286 mg, 1,17 mmol, 1,1 eq) se disolvieron en NMP (5,3 ml, 0,2 M). Se añadió DIPEA (369 µl, 2,12 mmol, 2 eq) y la mezcla se calentó a 90 °C. Después de 19 horas, la mezcla se diluyó con acetato de etilo y se lavó con agua y tres veces con salmuera. La capa orgánica se secó sobre sulfato de sodio, se filtró y se concentró a presión reducida. La purificación por cromatografía en columna (ISCO, columna de 12 g, MeOH al 0-10 %/DCM, gradiente de 30 minutos) dio el producto deseado como un sólido marrón (0,28 g, 0,668 mmol, 63 %).

RMN ¹H (500 MHz, Cloroformo-*d*) δ 8,12 (s, 1H), 7,62 (d, *J* = 8,3 Hz, 1H), 7,02 (s, 1H), 6,81 (d, *J* = 7,2 Hz, 1H), 4,93 (dd, *J* = 12,3, 5,3 Hz, 1H), 4,51 (s, 1H), 3,21 (t, *J* = 7,2 Hz, 2H), 3,09 (d, *J* = 6,4 Hz, 2H), 2,90 (dd, *J* = 18,3, 15,3 Hz, 1H), 2,82 - 2,68 (m, 2H), 2,16 - 2,08 (m, 1H), 1,66 (p, *J* = 7,2 Hz, 2H), 1,37 (d, *J* = 62,3 Hz, 20H).

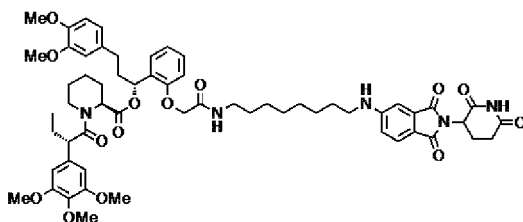
LCMS 501,41 (M+H).

**Trifluoroacetato de 5-((8-aminoctil)amino)-2-(2,6-dioxopiperidin-3-il)isindolin-1,3-diona**

Se disolvió (8-((2-(2,6-dioxopiperidin-3-il)-1,3-dioxoisindolin-5-il)amino)octil)carbamato de *terc*-butilo (334,5 g, 0,668 mmol, 1 eq) en TFA (6,7 ml) y se calentó a 50 °C. Después de 1 hora, la mezcla se enfrió a temperatura ambiente, se diluyó con DCM y se concentró a presión reducida. El material bruto se trituró con éter dietílico y se secó al vacío para dar una espuma de color amarillo oscuro (253,1 mg, 0,492 mmol, 74 %).

RMN ¹H (500 MHz, metanol-*d*₄) δ 7,56 (d, *J* = 8,4 Hz, 1H), 6,97 (d, *J* = 2,1 Hz, 1H), 6,83 (dd, *J* = 8,4, 2,2 Hz, 1H), 5,04 (dd, *J* = 12,6, 5,5 Hz, 1H), 3,22 (t, *J* = 7,1 Hz, 2H), 2,94 - 2,88 (m, 2H), 2,85 - 2,68 (m, 3H), 2,09 (ddd, *J* = 10,4, 5,4, 3,0 Hz, 1H), 1,70 - 1,61 (m, 4H), 1,43 (d, *J* = 19,0 Hz, 8H).

LCMS 401,36 (M+H).

**1-((S)-2-(3,4,5-trimetoxifenil)butanoil)piperidin-2-carboxilato de (2S)-(1R)-3-(3,4-dimetoxifenil)-1-(2-((8-((2-(2,6-dioxopiperidin-3-il)-1,3-dioxoisindolin-5-il) amino)octil)amino)-2-oxoetoxi)fenil)propilo (dTAG47)**

Se añadió sal trifluoroacetato de 5-((8-aminoctil)amino)-2-(2,6-dioxopiperidin-3-il)isindolin-1,3-diona (10,3 mg, 0,020 mmol, 1 eq.) a ácido 2-(2-((R)-3-(3,4-dimetoxifenil)-1-(((S)-1-((S)-2-(3,4,5-trimetoxifenil)butanoil)piperidin-2-carbonil)oxi)propil)fenoxi)acético (13,9 mg, 0,020 mmol, 1 eq.) como solución 0,1 M en DMF (200 microlitros) a temperatura ambiente. A continuación se añadieron DIPEA (10,5 microlitros, 0,060 mmol, 3 eq.) y HATU (7,6 mg, 0,020 mmol, 1 eq.). Después de 29,5 horas, la mezcla se diluyó con EtOAc y se lavó con ácido cítrico al 10 % (ac.), salmuera, bicarbonato de sodio saturado, agua y salmuera. La capa orgánica se secó sobre sulfato de sodio, se filtró y se condensó. La purificación por cromatografía en columna (ISCO, columna de 4 g de sílice, MeOH al 0-10 %/DCM, gradiente de 25 minutos) dio el producto deseado como un sólido amarillo (14,1 mg, 0,0131 mmol, 65 %).

RMN ¹H (500 MHz, metanol-*d*₄) δ 7,55 (d, *J* = 8,4 Hz, 1H), 7,26 - 7,20 (m, 1H), 6,99 - 6,93 (m, 1H), 6,89 (t, *J* = 7,7 Hz, 2H), 6,82 (dd, *J* = 8,4, 2,3 Hz, 2H), 6,77 (d, *J* = 7,5 Hz, 1H), 6,74 (d, *J* = 1,9 Hz, 1H), 6,63 (d, *J* = 9,6 Hz, 2H), 6,12 (dd, *J* = 8,1, 6,0 Hz, 1H), 5,40 (d, *J* = 4,3 Hz, 1H), 5,03 (dd, *J* = 13,1, 5,5 Hz, 1H), 4,57 (d, *J* = 14,9 Hz, 1H), 4,46 - 4,39 (m, 1H), 4,11 (d, *J* = 13,6 Hz, 1H), 3,86 (t, *J* = 7,3 Hz, 1H), 3,80 - 3,76 (m, 7H), 3,71 - 3,65 (m, 8H), 3,14 (ddt, *J* = 17,2, 13,3, 7,1 Hz, 4H), 2,90 - 2,80 (m, 1H), 2,77 - 2,40 (m, 6H), 2,24 (d, *J* = 13,8 Hz, 1H), 2,12 - 1,97 (m, 3H), 1,92 (dc, *J* = 14,0, 7,8 Hz, 1H), 1,67 (ddt, *J* = 54,1, 14,7, 7,1 Hz, 5H), 1,50 (dd, *J* = 46,1, 14,1 Hz, 3H), 1,38 (dt, *J* = 14,5, 7,1 Hz, 4H), 1,28 - 1,17 (m, 6H), 0,87 (t, *J* = 7,3 Hz, 3H).

RMN ¹³C (126 MHz, MeOD) δ 174,78, 174,69, 172,53, 171,71, 170,50, 169,66, 169,31, 156,22, 155,41, 154,62, 150,36, 148,83, 138,05, 136,90, 136,00, 134,93, 130,54, 128,40, 126,21, 123,14, 121,82, 117,94, 116,62, 113,58, 113,05,

112,73, 106,59, 70,69, 68,05, 61,06, 56,59, 56,51, 56,45, 53,42, 50,99, 50,31, 45,01, 44,09, 40,07, 37,44, 32,22, 32,17, 30,38, 30,32, 30,18, 29,84, 29,32, 28,05, 27,80, 27,58, 26,38, 23,87, 21,95, 12,57.

LCMS: 1077,35 (M+H)

5

Ensayo *in vitro* de circularización de ADN

En primer lugar, se generaron dos plásmidos (pAW49, pAW79). pAW49 contiene sitios de unión a YY1 separados por ~3,5 kb de ADN intermedio. pAW79 es idéntico excepto que contiene ADN de relleno en lugar de los motivos YY1. El ADN intermedio se eligió en función de observar YY1 ChIP-seq y la distribución de motivos en células MEr para identificar regiones que carecían de ocupación YY1 y motivos de unión YY1. Los motivos de unión de YY1 se eligieron en base a EMSA exitosos (Sigova *et al.*, 2015). Se añadieron aproximadamente 200 pb de secuencia entre los motivos de unión y los extremos para proporcionar flexibilidad para que los extremos se ligaran. El plásmido se construyó utilizando el ensamblaje de Gibson.

15

Seguidamente, se ejecutó una PCR utilizando un plásmido como plantilla para generar una pieza lineal de ADN (Tabla S5). Este producto de PCR se purificó por PCR (Qiagen 28104) y luego se digirió con BamHI (NEB R3136) y se purificó por PCR. El molde digerido con BamHI se usó en el ensayo de ligadura. El ensayo de unión se llevó a cabo del modo siguiente. Las reacciones se prepararon en hielo en 66 µl con los siguientes componentes:

20

BSA de control: ADN 0,25 nM, 1 tampón de ADN ligasa de T4 (NEB B0202S), H₂O, 0,12 µg/µl de BSA

YY1: ADN 0,25 nM, 1 tampón de ADN ligasa de T4 (NEB B0202S), H₂O, 0,12 µg/µl de YY1

25

YY1 + competidor: ADN 0,25 nM, 1 tampón de ADN ligasa de T4 (NEB B0202S), H₂O, 0,12 µg/µl de YY1, ADN competidor 100 nM (Tabla S5)

Suponiendo un coeficiente de extinción para YY1 de 19940 M⁻¹ cm⁻¹ y 75 % de pureza, eso da una concentración molar YY1 aproximada de ~ 3 µM.

30

Las reacciones se incubaron a 20 °C durante 20 minutos para permitir la unión de YY1 al ADN. Para cada punto de tiempo, se retiraron 6 µl de la reacción y se extinguieron en un volumen total de 9 µl con una concentración final de EDTA 30 mM, 1 colorante de carga NEB (NEB, B7024S), 1 µg/µl de proteinasa K, y se calentó a 65 °C durante 5 minutos. Se tomó el punto de tiempo 0 y luego se añadieron 600 unidades de T4 ADN ligasa (NEB M0202) y la reacción se llevó a cabo a 20 °C. Se tomaron los puntos de tiempo indicados y luego las muestras se procesaron en un gel con gradiente de TBE al 4-20 % durante tres horas a 120 V. El gel se tiñó con SYBR Gold (Life Technologies S11494) y se tomaron imágenes con una cámara CCD.

35

La cuantificación se realizó utilizando Image Lab versión 5.2.1 (Bio-Rad Laboratories). En primer lugar, se midió la densidad de banda del producto de partida y el producto de ligadura. Luego se calculó el porcentaje circularizado: (producto de ligadura)/(producto de ligadura + banda inicial)* 100. En la Figura 3 para facilitar la visualización se muestran geles sobreexpuestos. Para la cuantificación se utilizaron exposiciones que no presentaban píxeles sobreexpuestos.

40

45 Coinmunoprecipitación

Las CMEr V6.5 se transfectaron con pcDNA3_FLAG_YY1 y pcDNA3_FLAG_HA usando Lipofectamine 3000 (Life Technologies #L3000001) de acuerdo con las instrucciones del fabricante. Resumiendo, las células se dividieron y 8 millones de células se sembraron en una placa gelatinizada de 15 cm. Se mezclaron 7,5 µg de cada plásmido con 30 µl de reactivo P3000 y 75 µl de reactivo Lipofectamine 3000 (Life Technologies #L3000001) en 1250 µl de DMEM (Life technologies #11995-073). Después de -12-16 horas se cambió el medio.

50

Las células se recogieron 48 horas después de la transfección lavando dos veces con PBS helado y se recogieron raspando en PBS helado. Las células recolectadas se centrifugaron a 1000 rcf durante 3 minutos para sedimentar las células. Se descartó el sobrenadante y los sedimentos celulares se congelaron rápidamente y se almacenaron a -80 °C hasta que estuvieran listos para preparar el extracto nuclear. Por cada placa de 15 cm de células, los sedimentos de células congeladas se resuspendieron en 5 ml de tampón de lisis hipotónico enfriado con hielo (HEPES-KOH 20 mM, pH 7,5, 20 % de glicerol, NaCl 10 mM, Tritón X-100 al 0,1 %, MgCl₂ 1,5 mM, DTT 0,5 mM e inhibidor de proteasa (Roche, 11697498001)) y se incubaron en hielo durante 10 minutos para extraer los núcleos. Los núcleos se sedimentaron por centrifugación a 14.000 rcf durante 10 minutos a 4 °C. Se descartó el sobrenadante y los núcleos se resuspendieron en 0,5 ml de tampón de extracción nuclear enfriado con hielo (HEPES-KOH 20 mM pH 7,5, 20 % de glicerol, NaCl 250 mM, Tritón X-100 al 0,1 %, MgCl₂ 1,5 mM e inhibidor de proteasa) y se incubó durante 1 hora a 4 °C con rotación. Los lisados se clarificaron mediante centrifugación a 14.000 rcf durante 10 minutos a 4 °C. El extracto nuclear, el sobrenadante, se transfirió a un tubo nuevo y se diluyó con 1 ml de tampón de dilución enfriado con hielo (20 mM HEPES-KOH pH 7,5, 10 % de glicerol, NaCl 100 mM, Tritón X-100 al 0,1 %, MgCl₂ 1,5 mM, EDTA 0,2 mM, DTT 0,5 mM e inhibidor de proteasa). La concentración de proteína de los extractos se cuantificó mediante

55

60

65

ensayo BCA (Thermo Scientific, 23225) y la concentración de proteína se ajustó a 400 µg/ml mediante la adición del volumen apropiado de tampón de extracción nuclear:tampón de dilución 1:2. Para experimentos de extractos nucleares tratados con ARNasa A, Se trataron 250 µl de extracto nuclear (100 µg) mediante la adición de 7,5 µl de 33 mg/ml de ARNasa A (Sigma, R4642) o 18,75 µl de 20 U/µl de SUPERasa en inhibidor de ARNasa (Invitrogen, AM2696) seguido de incubación a 37 °C durante 10 minutos. Para todos los experimentos, se guardó una alícuota del extracto y se almacenó a -80 °C para su uso como muestra de entrada después de la inmunoprecipitación.

Para preparar perlas para la inmunoprecipitación de YY1 marcado con FLAG y marcado con HA del extracto nuclear, Se lavaron 50 µl de perlas magnéticas de proteína G por inmunoprecipitación tres veces con 1 ml de tampón de bloqueo (BSA al 0,5 % en PBS), rotando durante 5 minutos a 4 °C para cada lavado. Después de la separación en un imán, las perlas se resuspendieron en 250 µl de tampón de bloqueo. Después de la adición de 5 µg de anti-FLAG (Sigma, F7425), anticuerpo anti-HA (Abcam, ab9110) o IgG normal (Millipore, 12-370), se permitió que las perlas se incubaran durante al menos 1 hora a 4 °C con rotación para unir el anticuerpo. Después de la incubación, las perlas se lavaron tres veces con 1 ml de tampón de bloqueo, rotando durante 5 minutos a 4 °C para cada lavado.

Las perlas lavadas se separaron en un imán y el sobrenadante se descartó antes de resuspender en 250 µl de extracto nuclear (100 µg). Se permitió que las perlas se incubaran con el extracto durante la noche a 4 °C con rotación. A la mañana siguiente, las perlas se lavaron cinco veces con 1 ml de tampón de lavado enfriado con hielo, rotando durante 5 minutos a 4 °C para cada lavado. Las perlas lavadas se resuspendieron en 100 µl de tampón de muestra 1X XT (Biorad, 1610791) con DTT 100 mM y se incubaron a 95 °C durante 10 minutos. Las perlas se separaron en un imán y el sobrenadante que contenía material inmunoprecipitado se transfirió a un tubo nuevo.

Para ensayar los resultados de inmunoprecipitación por transferencia Western, se pasaron 10 µl de cada muestra en un gel Bis-Tris al 4-20 % (Bio-rad, 3450124) usando el tampón de carrera XT MOPS (Bio-rad, 1610788) a 80 V durante 20 minutos, seguido de 150 V hasta que el frente del tinte llegó al final del gel. Luego, la proteína se transfirió en húmedo a una membrana de PVDF de 0,45 µm (Millipore, IPVH00010) en tampón de transferencia enfriado con hielo (25 mM Tris, glicina 192 mM, metanol al 20 %) a 250 mA durante 2 horas a 4 °C. Después de la transferencia, la membrana se bloqueó con leche descremada al 5 % en TBS durante 1 hora a temperatura ambiente, con agitación. A continuación, la membrana se incubó con anticuerpo anti-FLAG-HRP 1:50.000 (Sigma, A8592), 1:25:000 anti-HA-HRP (Cell Signaling, 2999), o anti-OCT3/4 (C-10, Santa Cruz sc-5279) diluidos a 1:2000 en leche desgrasada al 5 % en TBST y se incubó durante la noche a 4 °C, con agitación. Por la mañana, la membrana se lavó tres veces con TBST durante 5 min a temperatura ambiente agitando para cada lavado. Las membranas se desarrollaron con sustrato ECL (Thermo Scientific, 34080) y se tomaron imágenes con una cámara CCD o se expusieron con película.

Formación del cuerpo embrioide

Antes de la diferenciación, las CMEr de activación marcadas con YY1-FKBP se cultivaron en suero LIF en MEF irradiados. Comenzando 48 horas antes de la diferenciación y continuando durante todo el experimento, el YY1 se expusieron a dTAG-47 500 nM. 4000 células (ya sea YY1⁺ o YY1) luego se sembraron en cada pocillo de una placa de 96 pocillos (Nunclon Sphera, ThermoFisher) en medio de formación de cuerpos embrioides (suero - LIF). Se generaron tres placas para cada condición. Las EB se cultivaron en placas de 96 pocillos durante 4 días y luego se agruparon y cultivaron en placas de cultivo de unión ultrabaja (Costar, Corning). Después de tres días, las células se recolectaron para RNA-seq de una sola célula (día 7 de diferenciación). Las células se recolectaron para RNA-seq de una sola célula por disociación con Accutase durante 30 minutos a 37 °C. Luego, las células se resuspendieron en PBS con BSA al 0,04 % y luego se prepararon para la secuenciación (consulte la sección sobre RNA-seq de una sola célula). La inmunohistoquímica se realizó después de cuatro días (día 8 de diferenciación).

Inmunohistoquímica

Las células se fijaron en paraformaldehído al 4 % en PBS y se incluyeron en parafina. Las células se seccionaron y tiñeron de acuerdo con los protocolos estándar usando TUJ1 (Biolegend 801201, 1: 1000), GFAP (Dako Z0344, 1:200) y Gata-4 (Abcam ab84593 1:100). anticuerpos primarios y anticuerpos secundarios apropiados conjugados con colorante Alexa Fluor (1:1000, ThermoFisher) y DAPI. Los portaobjetos se montaron con Fluoro-mount G (Electron Microscopy Science) y se tomaron imágenes usando un microscopio confocal de escaneo láser Zeiss LSM 710. En todas las imágenes, las barras de escala son de 50 µm.

Preparación de la biblioteca de RNA-seq de una sola célula

Se prepararon bibliotecas de RNA-seq de una sola célula utilizando Chromium Controller (10X Genomics). Resumiendo, las células individuales en BSA al 0,04 % en PBS se separaron en gotitas y luego se realizó la transcripción inversa y la construcción de bibliotecas de acuerdo con la Guía del usuario del kit de reactivos 3' de células individuales de cromo 10X y se secuenciaron en un Illumina Hi-seq 2500.

Anclaje de dCas9-YY1

Las dos primeras construcciones lentivirales se generaron modificando lenti dCAS-VP64_Blast (lenti dCAS-

VP64_Blast fue un regalo de Feng Zhang (plásmido Addgene n.º 61425), (Konermann *et al.*, 2014)). Se eliminó VP64 para generar dCas9 solo (pAW91) o se insertó el ADNc de YY1 humano en el extremo C para generar dCas9-YY1 (pAW90).

- 5 Para la producción de virus, las células HEK293T se cultivaron hasta un 50-75 % de confluencia en una placa de 15 cm y luego se transfectaron con 15 ug de pAW90 o pAW91, 11,25 ug de psPAX (Addgene 12260) y 3,75 ug de pMD2.G (Addgene 12259). psPAX y pMD2.G fueron amables obsequios de Didier Trono. Después de 12 horas, se reemplazaron los medios. El sobrenadante viral se recogió 24 horas después de la sustitución del medio (36 horas después de la transfección) y se añadió medio fresco. El sobrenadante viral se recogió de nuevo 48 horas después de la sustitución del medio (60 horas después de la transfección). El sobrenadante viral se eliminó de células mediante centrifugación a 500 x g durante 10 minutos. El virus se concentró con el concentrador Lenti-X (Clontech 631231) según las instrucciones del fabricante. El virus concentrado se resuspendió en medio MEr (suero LIF) y se añadió a 5 millones de células en presencia de polibreno (Millipore TR-1003) a 8 ug/ml. Después de 24 horas, se eliminó el medio vírico y se utilizó medio fresco que contenía blastidina (Invitrogen ant-bl-1) a 10 µg/ml. Las células se seleccionaron hasta que todas las células de las placas no transducidas murieron.

- Se generaron dos construcciones lentivirales adicionales (pAW12.lentiguide-GFP, pAW13.lentiguide-mCherry) modificando lentiGuide-puro (lentiGuide-Puro fue un regalo de Feng Zhang (plásmido Addgene n.º 52963) (Sanjana *et al.*, 2014)) para eliminar la puromicina y reemplazarla con GFP o mCherry. Los ARN guía de anclaje (Tabla S5, etv4_p_sgT1_F&R, etv4_p_sgT2_F&R) luego se clonaron en pAW12 y pAW13. Se generó el virus como se ha descrito anteriormente y se transdujeron las células MEr. Las células positivas dobles se identificaron y recolectaron mediante citometría de flujo y se expandieron. Estas líneas celulares expandidas fueron analizadas por 4C-seq, ChIP-qPCR (anti-Cas9, CST 14697) y RT-qPCR exactamente como se describe en otra parte de los métodos.

25 **Análisis bioinformático**

Análisis de datos ChIP-MS

- Se descargaron datos de ChIP-ms publicados anteriormente (Ji *et al.*, 2015). Para cada marca se calculó la relación log2 de la inmunoprecipitación sobre la entrada y sobre IgG. Luego, se identificó un conjunto de proteínas de alta confianza filtrando todas las proteínas que tenían un cambio de log2 veces menor o igual a uno en la entrada o en el control de IgG. Luego, se filtró por factores de transcripción usando la anotación provista en la tabla original para terminar con los 26 candidatos.

35 **Análisis de expresión específica de tejido**

- Con el fin de identificar factores de estructuración candidatos que se expresen ampliamente en muchos tejidos, los datos de expresión específica de tejido de RNA-seq se descargaron del proyecto Genotype-Tissue Expression (GTEx) (versión V6p). Se consideró que los genes se expresaban en tejidos particulares si la mediana de lecturas por millón por kilobase para ese tejido era superior a 5 (RPKM > 5). Los genes ampliamente expresados se identificaron como genes que se expresaron en más del 90 % de los 53 tejidos encuestados por GTEx.

Definición de regiones reguladoras

- 45 A lo largo de la divulgación, múltiples análisis se basan en superposiciones con diferentes regiones reguladoras, a saber, potenciadores, promotores y aisladores. A continuación se explica cómo se definieron estas regiones reguladoras.

Promotores

- 50 Los promotores se definieron como +/- 2 kilobases desde el sitio de inicio de la transcripción.

Promotores activos

- 55 Los promotores activos se definieron como +/- 2 kilobases desde el sitio de inicio de la transcripción que se superponía con un pico H3K27ac.

Potenciadores

- 60 Los potenciadores se definieron como picos de H3K27ac que no se superponían con un promotor.

Aisladores

- 65 Los aisladores se definieron descargando las llamadas vecindades aisladas de (Hnisz *et al.*, 2016a) (disponible en: <http://younglab.wi.mit.edu/insulatedneighborhoods.htm>). Cada fila representa un vecindario aislado (definido como una interacción SMC1 cohesina ChIA-PET con ambos anclajes superpuestos a un pico CTCF). El archivo contiene seis

columnas, las columnas 1-3 contienen las coordenadas de los anclajes de interacción izquierdos de las vecindades aisladas, y las columnas 4-6 contienen las coordenadas de los anclajes de interacción derechos de las vecindades aisladas. Las columnas 1-3 y 4-6 se concatenaron y luego se filtraron para identificar los anclajes únicos. Las regiones de anclajes de bucle únicas corresponden a los picos SMC1 ChIA-PET. Los elementos aisladores se identificaron como el subconjunto de picos CTCF ChIP-seq que se superponían a los anclajes únicos.

Superpotenciadores

Los superpotenciadores y constituyentes Oct4/Sox2/Nanog/Med1 se descargaron de (Whyte *et al.*, 2013)

Componentes potenciadores típicos

Los componentes potenciadores típicos de Oct4/Sox2/Nanog/Med1 se descargaron de (Whyte *et al.*, 2013)

15 Análisis de datos ChIP-seq

Alineación

Las lecturas de los experimentos ChIP-seq se alinearon con la revisión mm9 del genoma de referencia del ratón utilizando solo los cromosomas anotados 1-19, crX, crY y crM o a la revisión hg19 del genoma humano utilizando solo los cromosomas anotados 1-22, crX, crY y crM. La alineación se realizó con bowtie (Langmead *et al.*, 2009) con los parámetros -best -k 1 -m 1 -sam y -1 configurados para leer la longitud.

Leer la pila para visualizar

Los archivos Wiggle que representan recuentos de lecturas de ChIP-Seq en el genoma de referencia se crearon utilizando MACS (Zhang *et al.*, 2008) con parámetros -w -S -space=50 -nomodel -shiftsize=200. Los archivos de movimiento resultantes se normalizaron para secuenciar la profundidad dividiendo los recuentos de lectura en cada contenedor por los millones de lecturas asignadas en cada muestra y se visualizaron en el navegador de genoma UCSC (Kent *et al.*, 2002).

Lista de genes y lista de promotores

Para el análisis de datos de ratones, el 1 de febrero de 2017 se descargaron 36.796 transcripciones de RefSeq en formato GTF desde el navegador de genomas de UCSC. Para el análisis de datos humanos, se descargaron 39.967 transcritos RefSeq el 7 de diciembre de 2016 en formato GTF desde el navegador del genoma de UCSC el 1 de febrero de 2017. Para cada transcrito, se creó un promotor que es una ventana de 4000 pb centrada en el sitio de inicio de la transcripción.

40 Llamada de pico

Las regiones con una cobertura excepcionalmente alta de lecturas de ChIP-Seq (es decir, picos) se identificaron mediante MACS con parámetros -keep-dup=auto -p1e-9 y con el control de entrada correspondiente.

45 Mapas de calor y Metagenes

Se crearon perfiles de señal de ChIP-seq y GRO-seq en regiones individuales de interés mediante la cuantificación de la señal en lecturas por millón por par de bases (rpm/pb) en contenedores que dividen equitativamente cada región de interés utilizando bamToGFF (<https://github.com/BradnerLab/pipeline>) con los parámetros -m 200 -r -d. Las lecturas utilizadas para la cuantificación se eliminaron de las supuestas lecturas duplicadas de PCR mediante samtools v0.1.19-44428cd rmdup (Li *et al.*, 2009). Promotores con el mismo id de gen, cromosoma, coordenadas de inicio y fin se unieron en una sola instancia.

Se usaron mapas de calor de los perfiles de ChIP-seq para mostrar la señal de ChIP-seq en los promotores potenciadores y activos. Cada fila de un mapa de calor representa una región individual de interés con el perfil de señal de ChIP-seq en esa región que se muestra en rpm/pb en una región de ± 2 kb centrada en la región de interés. Para cada mapa de calor, el número de regiones de interés se muestra entre paréntesis en el panel de figuras. Véase la figura. 10A-B. Para mapas de calor de células ES murinas, La señal de ChIP-seq se cuantificó en 200 contenedores por región de interés. Para tejidos humanos y tejidos murinos de células no ME, los mapas de calor se generaron cuantificando la señal de ChIP-seq en 50 contenedores por región de interés.

Se usaron gráficos de Metagene para mostrar la señal promedio de ChIP-seq en las regiones de interés relacionadas. Se generaron gráficos de Metagene para los elementos potenciadores, promotores y aisladores, por separado. El perfil promedio (metagene) se calculó calculando los perfiles de señal medios de ChIP-seq o GRO-seq en las regiones de interés relacionadas. Para cada gráfico metagénico, el perfil promedio se muestra en rpm/pb en una región de ± 2 kb centrada en las regiones de interés. El número de potenciadores, promotores y aisladores examinados se indican

entre paréntesis. Para facilitar las comparaciones de la señal ChIP-seq de un solo factor entre diferentes conjuntos de regiones, la señal total de ChIP-seq para cada análisis de metagene se cuantificó y se muestra en la esquina superior derecha de cada gráfico de metagene. Se observó que diferentes anticuerpos tienen diferentes eficiencias de inmunoprecipitación que dan como resultado diferentes intensidades de señal. Por tanto, se cree que las comparaciones cuantitativas deben hacerse entre diferentes sitios en el mismo ChIP en lugar de entre diferentes ChIP en el mismo sitio.

Análisis de datos RNA-seq

Análisis RNA-seq

Los datos de RNA-seq se alinearon y cuantificaron usando kallisto (versión 0.43.0) (Bray *et al.*, 2016) con los siguientes parámetros: -b 100 --sencillo -1 180 -s 20 usando el transcriptoma mm9 RefSeq (descargado el 1 de febrero de 2017). Los archivos de salida representan los recuentos de transcripciones estimados.

El análisis de expresión génica diferencial se realizó utilizando deseq2 (versión 1.14.1) (Love *et al.*, 2014). El análisis se realizó a nivel de genes. Para calcular los recuentos de lectura a nivel de genes, los recuentos de transcritos estimados se sumaron en todas las isoformas del gen. Esto luego se ingresó en deseq2 y los valores de p ajustados se calcularon usando la configuración predeterminada. Los Log2 de los factores de cambio y los valores de p ajustados se incluyen en la Tabla S2. Se utilizó un valor de FDR de 0,05 como punto de corte para la expresión diferencial significativa. Para la FIG. 5C, los valores en el eje y son los valores del log2 del factor de cambio calculados por deseq2. Los valores en el eje x son los valores baseMean calculados deseq2.

Para la FIG. 7D, el valor absoluto del factor de cambio log2 calculado de deseq2 se representa en el lado izquierdo. En el lado derecho se representa gráficamente la densidad YY1 en el promotor. Debido a que el análisis se realiza a nivel genético, se promedió la señal del promotor YY1 para genes con múltiples isoformas.

Para el análisis GO, la lista de genes expresados diferencialmente (Tabla S3) se ingresó en la herramienta web de análisis PANTHER GO (<http://pantherdb.org/>, Versión 11.1) (Mi *et al.*, 2013, 2017) y se realizó una prueba de sobrerrepresentación estadística utilizando la configuración predeterminada.

Visualización de RNA-seq

Para mostrar pistas de RNA-seq, los datos de RNA-seq se asignaron con Tophat al transcriptoma mm9 RefSeq (descargado el 1 de febrero de 2017) utilizando los siguientes parámetros: -n 10 tophat -p 10 --no-novel-juncs -o. Los archivos Wiggle que representan recuentos de lecturas de RNA-Seq en el genoma de referencia se crearon utilizando MACS (Zhang *et al.*, 2008) con parámetros -w -S -space=50 -nomodel -shiftsize=200. Los archivos de movimiento resultantes se normalizaron para secuenciar la profundidad dividiendo los recuentos de lectura en cada contenedor por los millones de lecturas asignadas en cada muestra y se visualizaron en el navegador de genoma UCSC (Kent *et al.*, 2002).

Análisis de RNA-seq de una sola célula

Los datos de secuenciación se desmultiplexaron utilizando el software 10X Genomics Cell Ranger (versión 2.0.0) y se alinearon con el transcriptoma mm10. Los identificadores moleculares únicos se colapsaron en una matriz de códigos de barras de genes que representan los recuentos de moléculas por célula determinados y filtrados por Cell Ranger utilizando parámetros predeterminados. Los valores de expresión normalizados se generaron utilizando Cell Ranger utilizando los parámetros predeterminados. Para la FIG. 7H se contó el número de células con un valor de expresión normalizado >1 para el transcrito especificado. Para la FIG. 15C, las células se organizaron mediante análisis de componentes principales usando los parámetros predeterminados de Cell Ranger. En la FIG. 15D, las células se dividieron en los dos paneles en función de la condición de la que procedían. La disposición es la misma que en la fig. 15C. A continuación, las células individuales se colorean según el nivel de expresión normalizado.

Análisis de datos 4C-seq

Análisis 4C-seq

Las muestras de 4C-seq se procesaron primero eliminando sus secuencias de cebadores de lectura asociadas (Tabla S5) del extremo 5' de cada lectura FASTQ. Para mejorar la eficiencia del mapeo de las lecturas recortadas haciendo que la lectura sea más larga, el sitio de digestión con enzimas de restricción se mantuvo en la lectura recortada. Después de recortar las lecturas, las lecturas se mapearon usando corbatín con opciones -k 1 -m 1 contra el ensamblaje del genoma mm9. Todas las lecturas no mapeadas o mapeadas repetitivamente se descartaron del análisis posterior. Luego, el genoma mm9 se "digirió" in silico de acuerdo con el par de enzimas de restricción utilizado para esa muestra para identificar todos los fragmentos que podrían generarse mediante un experimento 4C dado un par de enzimas de restricción. Todas las lecturas mapeadas se asignaron a su fragmento correspondiente en función de dónde se mapearon en el genoma. La digestión de una muestra en un experimento 4C crea una serie de fragmentos

"ciegos" y "no ciegos" como se describe en los laboratorios de Tanay y De Laat (van de Werken *et al.*, 2012). En resumen, los fragmentos "ciegos" carecen de un sitio de enzima de restricción secundario, mientras que los fragmentos "no ciegos" contienen un sitio de enzima de restricción secundario. Debido a esto, se espera observar solo lecturas derivadas de fragmentos no ciegos.

Los experimentos se realizaron por triplicado biológico y las muestras mutantes y WT se normalizaron por cuantiles entre sí.

Si no se detectaron lecturas en un fragmento no ciego para una muestra determinada cuando se detectaron lecturas en al menos otra muestra, se asignó un "0" a ese fragmento no ciego para las lecturas faltantes de la muestra.

Visualización 4C-seq

Para mostrar pistas de cobertura genómica 4C-seq, primero se suavizó la señal 4C-seq normalizada utilizando una media móvil de 5 kb en etapas de 50 pb a lo largo del genoma para cada muestra. Los duplicados individuales se muestran en la Fig. 13. Seguidamente, se combinaron duplicados biológicos de la misma condición y se calculó la media y el intervalo de confianza del 95 % de la señal 4C-seq para cada contenedor en el genoma. En la figura 6 y la figura 9, las pistas de la señal de 4C-seq muestran la señal media de 4C-seq a lo largo del genoma como una línea y el intervalo de confianza del 95 % como el área sombreada alrededor de la línea. Para cada pista de señal 4C-seq, el punto de vista utilizado en el experimento 4C-seq se indica como una flecha marcada como VP.

Para cuantificar el cambio en la señal 4C-seq en una región específica de interés, se contó la señal 4C-seq normalizada (no suavizada) para cada muestra y se calcularon la media y la desviación estándar de la señal cuantificada para duplicados biológicos de la misma condición. La desviación media y estándar de la señal cuantificada se normalizó a la condición de control adecuada (ya sea WT o dCas9) antes de trazar. Debajo de cada pista de señal 4C-seq, la región cuantificada se indica como una barra roja etiquetada como "región cuantificada". Las coordenadas de la región cuantificada para Raf1 son cr6:115598005-115604631, y para Etv4 son cr11:101644625-101648624.

Análisis de datos de ChIA-PET

Procesamiento de lectura ChIA-PET

Para cada conjunto de datos de ChIA-PET, las lecturas sin procesar se procesaron para identificar un conjunto de supuestas interacciones que conectan los anclajes de interacción para un posterior modelado y análisis estadístico. En primer lugar, marcadores de extremo emparejado (PET), cada uno contiene dos lecturas emparejadas, se analizaron para determinar la presencia de la secuencia puente-enlazador y se recortaron para facilitar el mapeo de lectura. Los PET que contenían al menos una instancia de la secuencia del enlazador puente en cualquiera de las dos lecturas se mantuvieron para su posterior procesamiento y las lecturas que contenían la secuencia del enlazador puente se recortaron inmediatamente antes de la secuencia del enlazador usando cutadapt con las opciones "-n 3 -O 3 -m 15 -a directo=ACGCGATATCTTATCTGACT (SEQ ID NO: 42) -a inverso=AGTCAGATAAGATATCGCGT" (SEQ ID NO: 43) (<http://cutadapt.readthedocs.io/en/stable/>). Las PET que no contenían una instancia de la secuencia puente-enlazador no se procesaron más. Las lecturas recortadas se asignaron individualmente al genoma de referencia del ratón mm9 utilizando Bowtie con las opciones "-n 1 -m 1 -p 6" (Langmead *et al.*, Genome Biology, 2009). Después de la alineación, las lecturas emparejadas se volvieron a vincular con un script interno utilizando identificadores de lectura. Para evitar posibles artefactos derivados del sesgo de la PCR, las PET redundantes con coordenadas de mapeo genómico idénticas e información de cadena se colapsaron en una sola PET. Los anclajes de interacción potencial se determinaron mediante la identificación de regiones de enriquecimiento local en las lecturas mapeadas individualmente usando MACS con opciones "-g mm -p 1e-9 -nolambda --nomodel --shiftsize=100" (Zhang *et al.*, Genome Biology, 2008). Se usaron PET con dos lecturas mapeadas en las que cada una se superponía a un anclaje de interacción potencial diferente en al menos 1 pb para identificar interacciones putativas entre los anclajes de interacción superpuestos. Cada supuesta interacción representa una conexión entre dos anclajes de interacción y está respaldada por el número de PET (recuento de PET) que conectan los dos anclajes de interacción.

Descripción general del análisis estadístico de ChIA-PET

Al procesar nuestros datos de interacción de la cromatina, se buscaba identificar las supuestas interacciones que representan contactos estructurados de cromatina, definidos como contactos de cromatina que están estructurados por fuerzas distintas a la dinámica de la fibra que resulta de la distancia genómica lineal entre las dos regiones de contacto. Por el contrario, se buscaba filtrar las supuestas interacciones que probablemente resultan de PET que surgen de contactos de cromatina no estructurados, definidos como contactos resultantes de la estrecha proximidad genómica lineal de las dos regiones de contacto, o de artefactos técnicos del protocolo ChIA-PET. Se espera que las supuestas interacciones que representan contactos de cromatina estructurada se detecten con mayor frecuencia, o recuento de PET, de lo esperado dada la distancia genómica lineal entre las dos regiones en contacto, lo que permite distinguir entre estas dos clases de interacciones.

Con este fin, se desarrolló Origami, un método estadístico para identificar interacciones de alta confianza que

probablemente representen contactos de cromatina estructurados. Conceptualmente, Origami utiliza un modelo semi-bayesiano de mezcla de dos componentes para estimar la probabilidad de que una supuesta interacción corresponda a uno de dos grupos: contactos de cromatina estructurados, o contactos de cromatina no estructurados y artefactos técnicos. Origami estima esto como una puntuación de probabilidad para cada supuesta interacción al modelar la

5 relación entre el conteo de PET, distancia genómica lineal entre anclajes de interacción y profundidad de lectura en los anclajes de interacción. Luego, las interacciones de alta confianza se identifican como el subconjunto de interacciones putativas que probablemente representen contactos de cromatina estructurada, al requerir que las interacciones de alta confianza tengan una puntuación de probabilidad $> 0,9$.

10 Todos los métodos a continuación se desarrollaron dentro del software de origami que está disponible en <https://github.com/younglab/origami>. La versión utilizada fue la versión 1.1 (etiquetada en el repositorio de GitHub como v1.1). El siguiente software se ejecutó con los siguientes parámetros: iteraciones=10000 --burn-in=100 --prune=0 --min-dist=4000 --peak-count-filter=5.

15 Modelo estadístico de origami

Se desarrolló Origami, un método para analizar los datos de ChIA-PET, con el fin de identificar interacciones putativas que probablemente representen contactos de cromatina estructurada, y para filtrar interacciones putativas que probablemente representen contactos de cromatina no estructurados que ocurren como resultado de la estrecha

20 proximidad genómica lineal de regiones de contacto e interacciones que representan artefactos técnicos de la ChIA - Protocolo PET. Esto incluye el modelado de la relación entre el número de PET observados para respaldar cada interacción (l_i), distancia genómica lineal entre anclajes de interacción (d_i), y la profundidad de secuenciación en los anclajes de interacción, para estimar la probabilidad de que cada supuesta interacción (i) represente un contacto de cromatina estructurada dado el recuento de PET observado (l_i).

25 Inicialmente se asumió que las supuestas interacciones se clasificaban en uno de dos grupos, $j \in \{0,1\}$, tal que cada supuesta interacción, $i \in \{1 \dots N\}$, tenía una identidad de grupo latente Z_i que corresponde a un valor de j . El Grupo 1 se designa como el conjunto de supuestas interacciones resultantes de contactos de cromatina estructurada que esperamos detectar con frecuencias mayores de lo esperado dada la distancia genómica lineal entre las regiones de

30 contacto. El grupo 0 se designa como el conjunto de supuestas interacciones resultantes de contactos de cromatina no estructurados debido a la estrecha proximidad genómica lineal de las regiones de contacto, o de artefactos técnicos del protocolo ChIA-PET.

Se desarrolló un modelo de mezcla de dos componentes semi-bayesiano para estimar la probabilidad de que cada

35 supuesta interacción represente un contacto de cromatina estructurada. Para cada grupo, se modeló la probabilidad de observar el recuento de PET (l_i) bajo ese grupo como un proceso de Poisson con dos factores subyacentes. Estos factores son el número de PET observadas como consecuencia de formar parte del grupo (G_{ij}), y el número de PET observados como resultado de la distancia genómica lineal entre los anclajes dado el grupo (D_{ij}). Se modeló el número de PET observados como resultado de formar parte del grupo (G_{ij}) como un proceso de Poisson con media, λ_j . Se

40 modeló el número de PET observados como resultado de la distancia genómica lineal entre los anclajes dado el grupo (D_{ij}) como un proceso de Poisson con media, μ_{ij} . Dado que se piensa que estos dos factores son independientes (Phanstiel *et al.*, Bioinformatics, 2015), el proceso total de Poisson es la suma de estos dos factores subyacentes.

Se modelaron las variables de datos bajo las siguientes distribuciones:

$$45 \quad l_i \sim \sum_{j \in \{0,1\}} w_{ij} * (G_{ij} + D_{ij})$$

$$G_{ij} \sim \text{Poisson}(\lambda_j)$$

$$50 \quad D_{ij} \sim \text{Poisson}[\mu_{ij}(d_i)]$$

$$l_i \sim \text{Poisson}[\lambda_j + \mu_{ij}(d_i)]$$

Se modelaron los parámetros con las siguientes distribuciones previas:

$$55 \quad \lambda_j \sim \text{Gamma}(1,1) \\ w_{11} \sim \text{Beta}(1 + \alpha_1, 1 + \beta_1)$$

Dado que w_{11} es una probabilidad binomial, $w_{10} = 1 - w_{11}$.

60 A partir de estas distribuciones previas y de probabilidad, las distribuciones posteriores de estos parámetros son las siguientes:

$$65 \quad \lambda_j \sim \text{Gamma}[1 + \sum_{i=j} G_{ij}, 1 + \#(Z_i = j)]$$

$$w_{11} \sim \text{Beta}[1 + \alpha_1 + Z_i, 1 + \beta_1 + (1 - Z_i)]$$

Aparte de D_{ij} y μ_{ij} , se estimaron los parámetros usando el proceso iterativo Markov Chain Monte Carlo (MCMC) con Gibbs Sampling con el posterior apropiado para muestrear (Gelman *et al.*, 2004).

- 5 Para estimar μ_{ij} , se modeló la función entre D_{ij} y la distancia genómica lineal (d_i) en la escala log10 usando una spline cúbica suavizada (a través de spline suave en R), tomando μ_{ij} como el número esperado de PET que se observarán debido a la distancia (D_{ij}) dada la distancia genómica lineal (d_i), para cada supuesta interacción (i).

- 10 Las constantes α_i y β_i se establecieron para ser lo menos informativo posible. La constante α_i se fijó igual al número de interacciones putativas que comparten un ancla con i que tienen PET cuenta menos que y_{oi} . El constante β_i se fijó igual al número de interacciones putativas que comparten un ancla con i que tienen recuentos de PET mayores que y_{oi} más la relación de la puntuación de profundidad (s_i) a la puntuación de profundidad mediana con todos los valores <1 reducidos a 0. La puntuación de profundidad (s_i) para cada interacción putativa se define como el producto del número de lecturas que se asignan a sus anclas de interacción.

15 Implementación de Origami

- Se implementó el modelo descrito anteriormente mediante la simulación Markov Chain Monte Carlo. Estimando iterativamente la identidad del grupo (Z_i) de cada supuesta interacción, se buscó explorar el espacio de probabilidad para Z_i y determinó una puntuación de probabilidad (p_i) para cada interacción putativa que refleja la probabilidad de que la interacción resulte de un contacto de cromatina estructurada (pertenecer al grupo 1). Las etapas de la implementación de los autores son las siguientes.

- 25 Para cada supuesta interacción, se registró el número de PET observados que apoyan la interacción (l_i), la distancia genómica lineal de la interacción entre los pares de bases más externos de los dos anclajes de la supuesta interacción (d_i), y una puntuación de profundidad (s_i), que se define como el producto de

- 30 Para sembrar los parámetros del modelo para la primera iteración, se realizó lo siguiente. Los pesos de mezcla (w_{ij}) se fijaron en 0,5 para cada interacción. El proceso de grupo significa (λ_i) se les asignó valores de 5 y 1 para el grupo 1 y 0, respectivamente. La media del proceso de distancia (μ_{ij}) se estableció inicialmente en 0 para todas las interacciones.

- 35 Adicionalmente se calcularon los valores de α_i y β_i para cada interacción, pero no se usó en la primera iteración. En todas las iteraciones posteriores, α_i y β_i se utilizan para actualizar los valores de los pesos de mezcla (w_{ij}). El parámetro α_i se fijó igual al número de interacciones putativas que comparten un ancla con i que tienen PET cuenta menos que y_{oi} . El parámetro β_i se fijó igual al número de interacciones putativas que comparten un ancla con i que tienen recuentos de PET mayores que y_{oi} más la relación de la puntuación de profundidad (s_i) sobre la puntuación mediana de profundidad para todas las supuestas interacciones, donde cuando esta relación es menor que 1 se reduce a 0.

- 40 Para cada supuesta interacción, se estimó la probabilidad (l_{ij}) que la supuesta interacción se observa con el recuento de PET (l_i), dado que la supuesta interacción pertenece al grupo 1 y al grupo 0, como se indica a continuación.

$$l_{ij} = \text{dPoisson}(l_i; \lambda_i + \mu_{ij})$$

- 45 Cuando dPoisson es la función de densidad de la distribución de Poisson para la media $\lambda_i + \mu_{ij}$ y evaluado en l_i .

- Se calculó la probabilidad ponderada relativa (r_i) de cada supuesta interacción perteneciente al grupo 1. Para hacer esto, se multiplicó cada una de las dos probabilidades calculadas para cada supuesta interacción por sus respectivos pesos de mezcla (w_{ij}) y evaluado de la siguiente manera.

$$r_i = \frac{w_{i1} * L_{i1}}{(w_{i1} * L_{i1}) + (w_{i0} * L_{i0})}$$

- 50 Se actualiza la identidad del grupo (Z_i) de cada interacción extrayendo de la distribución binomial con una probabilidad de r_i como sigue.

$$Z_i = \text{rBinomial}(1, r_i)$$

Donde rBinomial significa que se extrae aleatoriamente 1 o 0 con la probabilidad de r_i para la extracción 1.

- 60 Se actualizan los pesos de la mezcla (w_{ij}) utilizando nuestras identidades de grupo recientemente actualizadas (Z_i), extrayendo de la distribución Beta de la siguiente manera.

$$W_{i1} = \text{rBeta}[1 + \alpha_i + Z_i, 1 + \beta_i + (1 - Z_i)]$$

- 65 Donde rBeta significa que extraemos aleatoriamente de la distribución beta con los parámetros anteriores. Dado que

w_{i1} es una probabilidad binomial, $w_{i0} = 1 - w_{i1}$.

- 5 Para estimar los recuentos de PET para G_{ij} y D_{ij} muestreamos aleatoriamente el número de PET para G_{ij} y D_{ij} aprovechando el hecho de que cuando se sabe que dos variables de Poisson suman un conteo dado, entonces la distribución de cualquiera de las variables sigue una distribución binomial con probabilidad $\lambda_j/(\lambda_j + \mu_{ij})$. En consecuencia, se estimó los recuentos de PET para G_{ij} y D_{ij} de la siguiente manera:

$$G_{ij} = \text{rBinomial}\left(I_i, \frac{\lambda_j}{\lambda_j + \mu_{ij}}\right)$$

$$10 \quad D_{ij} = I_i - G_{ij}$$

Donde rBinomial significa que dibujamos aleatoriamente hasta I_i PET con la probabilidad $\lambda_j/(\lambda_j + \mu_{ij})$ de extraer cada PET.

- 15 Se actualiza la media del proceso de grupo (λ_j) usando la siguiente identidad, requiriendo que $\lambda_1 > \lambda_0$ para mantener la identificabilidad de los dos grupos (aunque durante nuestras ejecuciones esta restricción no fue necesaria).

$$\lambda_j = \text{rGamma}(1 + \sum_{z_i=j} G_{ij}, 1 + \#(Z_i = j))$$

- 20 Donde rGamma significa que se extrae aleatoriamente de la distribución Gamma con los parámetros anteriores.

Para actualizar el proceso de distancia significa (μ_{ij}), se calcula la función entre D_{ij} y el $\log_{10}(d_i + 1)$ utilizando una spline cúbica suavizada (a través de smooth.spline en R). Para simplificar la estimación de μ_{ij} , se elige tomar la estimación de máxima verosimilitud de este proceso.

- 25 Se repitieron las etapas 4-10 de la siguiente manera. Se realizaron 1.000 iteraciones iniciales como *burn-in*, que se desecharon. Luego se realizaron 10.000 iteraciones.

- 30 Se estimó la probabilidad de que cada supuesta interacción pertenezca al grupo 1 calculando una puntuación de probabilidad (p_i) para cada interacción putativa que es igual al valor medio de Z_i en las 10.000 iteraciones. Las interacciones de alta confianza se identificaron como supuestas interacciones con $p_i > 0,9$;

$$= \frac{1}{n \cdot \#(\text{iteraciones})} \sum Z_i \approx P(Z_i = 1)$$

35 **Análisis de datos HiChIP**

Procesamiento HiChIP

- 40 Las muestras de HiChIP se procesaron identificando primero las lecturas con una unión de fragmentos de restricción (es decir, un sitio donde se produjo la ligadura). Las lecturas que contenían la unión del fragmento de restricción se recortaron de manera que se mantuviera la información 5' con respecto a la unión. Las lecturas sin uniones de fragmentos de restricción se dejaron sin recortar. Luego se mapearon las lecturas usando bowtie con opciones -k 1 -m 1 contra el ensamblaje del genoma mm9. Todas las lecturas no mapeadas o mapeadas repetitivamente se descartaron del análisis posterior. Las lecturas se volvieron a unir en pares por su identificador de lectura. El genoma se reunió y para cada par de contenedores se calculó el número de PET que los unían. Luego, estos datos se usaron como entrada en la canalización de Origami descrita anteriormente para identificar pares de interacción significativos de contenedor a contenedor.

Análisis HiChIP

- 50 El análisis cuantitativo de los datos de HiChIP y Hi-C (Figuras 8, 9) se realizó como sigue. Mediante Origami se identificaron interacciones de alta confianza. Luego se creó una unión de interacciones de alta confianza para cada experimento.

Experimento	Figura	Condición	Duplicado
Degron	8, 14	SinFármaco	1
Degron	8, 14	SinFármaco	2
Degron	8, 14	SinFármaco	3
Degron	8, 14	síFármaco	1
Degron	8, 14	síFármaco	2

(continuación)

Experimento	Figura	Condición	Duplicado
Degron	8, 14	síFármaco	3
Lavado	9	Sin tratar (ST)	1
Lavado	9	Sin tratar (ST)	2
Lavado	9	Sin tratar (ST)	3
Lavado	9	Tratado (TR)	1
Lavado	9	Tratado (TR)	2
Lavado	9	Lavado (LV)	1
Lavado	9	Lavado (LV)	2
Lavado	9	Lavado (LV)	3
Lavado CTCF	9	Sin tratar (ST)	1
Lavado CTCF	9	Sin tratar (ST)	2
Lavado CTCF	9	Tratado (TR)	1
Lavado CTCF	9	Tratado (TR)	2
Lavado CTCF	9	Lavado (LV)	1
Lavado CTCF	9	Lavado (LV)	2

Por ejemplo, el conjunto de grado de alta confianza consistiría en la unión de las muestras de 6 grados enumeradas anteriormente. Luego, los recuentos de PET se normalizaron entre sí utilizando deseq2 (Love *et al.*, 2014). Luego se calculó la media de cada grupo y luego se calculó el factor de cambio tomando la relación de la condición perturbada a la condición no perturbada (es decir, síFármaco a noFármaco o TR/ST;LV/ST) con un pseudorecuento de 0,5 añadido a ambos. Este conjunto completo de interacciones significativas es lo que se muestra en la FIG. 8B como "Todas las interacciones".

Para el análisis de subconjuntos, el ancla de cada interacción se clasificó mediante la superposición con características genómicas conocidas, como se definió anteriormente. Esto dio como resultado una puntuación binaria para si un ancla se superponía con un potenciador, un promotor, aislador, YY1 o CTCF. Luego, las interacciones se dividieron en subconjuntos para identificar los siguientes grupos:

YY1 no presente (Fig. 8): no hay YY1 en ninguno de los extremos de la interacción.

Interacciones potenciador-promotor YY1 (Fig. 8, Fig. 9): YY1 en ambos extremos Y un potenciador o promotor en ambos extremos.

Interacción CTCF-CTCF: CTCF en ambos extremos.

El log2 del factor de cambio para estos grupos se representa en la FIG. 8B, 9F.

El análisis de la fig. 6C se realizó mediante la identificación del gen al final de los bucles potenciador-promotor YY1. Esto se hizo cruzando promotores (como se define anteriormente) con los anclajes de bucle significativos. Los genes con múltiples promotores se colapsaron después de la intersección para generar una lista de genes al final de los bucles potenciador-promotor YY1. A continuación, se representa gráficamente el log2 del factor de cambio calculado de deseq2 de estos genes en la FIG. 8C. Los genes están coloreados en función del valor de p ajustado calculado de deseq2 (como en la FIG. 7).

Visualización HiChIP

Las matrices de interacción HiChIP mostradas en la FIG. 8D y 8E. Para estas matrices de interacción, se muestran todas las supuestas interacciones y la intensidad de cada píxel representa la media de la frecuencia de interacción normalizada deseq2 de todas las duplicados biológicos de esa condición. En la Fig. 8D y 8E el píxel delineado, que

refleja la frecuencia de interacción entre los sitios en la base de las diagonales, se utilizó para cuantificar el cambio en la frecuencia de interacción normalizada tras la degradación de YY1.

En la figura 10, las interacciones HiChIP de alta confianza se muestran como arcos. Para mostrar, las interacciones mostradas se filtraron para eliminar los contactos a contenedor a contenedor adyacente y las interacciones no potenciadoras-promotoras. Los arcos se centraron en la característica genómica relevante dentro del contenedor (por ejemplo, una cima de pico ChIP-seq o un sitio de inicio de transcripción).

Clasificación de interacción

Las interacciones ChIA-PET y HiChIP de alta confianza se clasificaron en función de la presencia de elementos potenciadores, promotores y aisladores en los anclajes de cada interacción como se ha definido anteriormente. En el caso de que un ancla de interacción se superpusiera tanto a un potenciador como a un aislador o a un promotor y un aislador, una jerarquía en la que las anclas se consideraban primero como promotores, luego potenciadores, luego aisladores. Por ejemplo, si hay una interacción en la que el ancla izquierda es aislador/promotor y el ancla derecha es potenciador/aislador, se contaría como una interacción potenciador-promotor y no como una interacción aislador-aislador.

Para mostrar resúmenes de las clases de interacciones de alta confianza, cada clase de interacciones se muestra como un arco entre los elementos potenciadores, promotores y aisladores relevantes. El grosor de los arcos refleja aproximadamente el porcentaje de interacciones de esa clase en relación con el número total de interacciones que se clasificaron. En las cifras principales, se muestran las clases de interacción potenciador-potenciador, potenciador-promotor, promotor-promotor y aislador-aislador. Los resúmenes extendidos que además incluyen interacciones potenciador-aislador y promotor-aislador se muestran en las figuras complementarias.

Visualización de figuras

En ciertos paneles de figuras que muestran pistas del genoma, los elementos potenciadores se indican como recuadros rojos con la etiqueta "Potenciador". Estas regiones representan nuestra interpretación de los datos de ChIP-seq y son distintas de los potenciadores definidos algorítmicamente utilizados en el análisis cuantitativo de todo el genoma.

Análisis estadístico

Para usar la prueba t no pareada, se realizaron dos suposiciones.

1) Las poblaciones se distribuyen según una distribución gaussiana. Para la mayoría de los experimentos, se usaron tres réplicas, por lo que los tamaños de muestra eran demasiado pequeños para calcular de manera confiable la desviación de la normalidad (es decir, con una prueba de D'Agostino)

2) Las dos poblaciones tienen la misma varianza. No se realizó una prueba de varianza.

Los valores p completos se enumeran aquí*:

Figura	Subpanel	Prueba	Duplicados biológicos	valor de P
6B	4C-seq	Test de la t de Student	3	0,011
6B	ChIP-qPCR	Test de la t de Student	3	0,0066
6B	RT-qPCR	Test de la t de Student	6	< 0,0001
6C	4C-seq	Test de la t de Student	3	0,0013
6C	ChIP-qPCR	Test de la t de Student	3	0,0048
6C	RT-qPCR	Test de la t de Student	6	0,0394
8B		Prueba T de dos muestras de Welch	3	< 2,2e-16
8D	HiChIP	Test de la t de Student	3	0,0162
8D	RNA-seq	Wald	2	7.22E-13
8E	HiChIP	Test de la t de Student	3	0,0446
8E	RNA-seq	Wald	2	1.25E-58

(continuación)

Figura	Subpanel	Prueba	Duplicados biológicos	valor de P
9D	4c-seq	Test de la t de Student	3	0,004717003
9D	RT-qPCR	Test de la t de Student	6	< 0,0001
14D	Raf1	Wald	2	1.63E-53
14D	Etv4	Wald	2	2.88E-34
*Tenga en cuenta que la prueba T de Student se realizó con GraphPad Prism, que establece un límite inferior en 0,0001, la prueba T de dos muestras de Welch se realizó usando R, que establece un límite inferior en 2,2e-16, La prueba de Wald se realizó utilizando deseq2 en R, que no tiene un límite inferior en el valor de p.				

DISPONIBILIDAD DE DATOS Y SOFTWARE

5 Todos los conjuntos de datos utilizados se resumen en la Tabla S4 a continuación.

Especie	Célula	Tipo	Nombre	GEO	Este estudio o publicación anterior
Ser humano	KBM7	Selección CRISPR	Wang et al 2015	N/A	Publicación anterior
Ser humano	54 tipos	RNA-seq	GTEX	N/A	Publicación anterior
Ratón	V6.5	ChIP-seq	CTCF_merged	GSM747534, GSM747535, GSM747536	Publicación anterior
Ratón	V6.5	ChIP-seq	CTCF_rep1	GSM747534	Publicación anterior
Ratón	V6.5	ChIP-seq	CTCF_rep2	GSM747535	Publicación anterior
Ratón	V6.5	ChIP-seq	CTCF_rep3	GSM747536	Publicación anterior
Ratón	V6.5	ChIP-seq	CTCF_input_merged	GSM747545, GSM747546	Publicación anterior
Ratón	V6.5	ChIP-seq	CTCF_input_rep1	GSM747545	Publicación anterior
Ratón	V6.5	ChIP-seq	CTCF_input_rep2	GSM747546	Publicación anterior
Ratón	V6.5	ChIP-seq	YY1	GSM2645432	Este artículo
Ratón	V6.5	ChIP-seq	YY1_input	GSM2645433	Este artículo
Ratón	V6.5	ChIP-seq	H3K27Ac	GSM1526287	Este artículo
Ratón	V6.5	ChIP-seq	H3K27AC_input	GSM1526285	Este artículo
Ratón	V6.5	ChIA-PET	YY1	GSM2645440	Este artículo
Ratón	V6.5	ChIA-PET	CTCF	GSM2645441	Este artículo

(continuación)

Especie	Célula	Tipo	Nombre	GEO	Este estudio o publicación anterior
Ratón	V6.5	RNA-sec	Ohr_rep1	GSM2645362	Este artículo
Ratón	V6.5	RNA-sec	Ohr_rep2	GSM2645363	Este artículo
Ratón	V6.5	Hi-ChIP	H3K27Ac_0hr_rep1	GSM2645434	Este artículo
Ratón	V6.6	Hi-ChIP	H3K27Ac_0hr_rep2	GSM2645435	Este artículo
Ratón	V6.5	Hi-ChIP	H3K27Ac_0hr_rep3	GSM2645436	Este artículo
Ratón	V6.5	RNA-sec	mES_24hr_rep1	GSM2645364	Este artículo
Ratón	V6.5	RNA-sec	mES_24hr_rep2	GSM2645365	Este artículo
Ratón	V6.5	Hi-ChIP	H3K27Ac_24hr_rep1	GSM2645435	Este artículo
Ratón	V6.5	Hi-ChIP	H3K27Ac_24hr_rep2	GSM2645436	Este artículo
Ratón	V6.5	Hi-ChIP	H3K27Ac_24hr_rep3	GSM2645437	Este artículo
Ser humano	GM12878	ChIP-seq	H3K27AC_input	GSM733742	Publicación anterior
Ser humano	GM12878	ChIP-seq	H3K27Ac	GSM733771	Publicación anterior
Ser humano	GM12878	ChIP-seq	CTCF_input	GSM749669	Publicación anterior
Ser humano	GM12878	ChIP-seq	CTCF	GSM749704	Publicación anterior
Ser humano	GM12878	ChIP-seq	YY1	GSM803406	Publicación anterior
Ser humano	K562	ChIP-seq	CTCF_input	GSM749719	Publicación anterior
Ser humano	K562	ChIP-Seq	CTCF_1	GSM749690	Publicación anterior
Ser humano	K562	ChIP-Seq	H3k27Ac_input	GSM733780	Publicación anterior
Ser humano	K562	ChIP-Seq	H3K27Ac	GSM733656	Publicación anterior
Ser humano	K562	ChIP-Seq	YY1	GSM803470	Publicación anterior
Ser humano	K562	Hi-ChIP	YY1	GSM2114002	Este artículo
Ser humano	CME	ChIP-Seq	CTCF	GSM1705263	Publicación anterior
Ser humano	CME	ChIP-Seq	HSK27Ac	GSM1105260	Publicación anterior
Ser humano	CME	ChIP-Seq	CTCF_H3K27Ac_input	GSM1705264	Publicación anterior
Ser humano	CME	ChIP-Seq	YY1	GSM803513	Publicación anterior
Ser humano	HEPG2	ChIP-Seq	CTCF	GSM803486	Publicación anterior

(continuación)

Especie	Célula	Tipo	Nombre	GEO	Este estudio o publicación anterior
Ser humano	HEPG2	ChIP-Seq	YY1	GSM803381	Publicación anterior
Ser humano	HEPG2	ChIP-Seq	CTCF_YY1_input	GSM803463	Publicación anterior
Ser humano	HEPG2	ChIP-Seq	H3K27Ac	GSM733743	Publicación anterior
Ser humano	HEPG2	ChIP-Seq	H3K27Ac_input	GSM733732	Publicación anterior
Ser humano	HCT-116	ChIP-Seq	YY1	GSM803354	Publicación anterior
Ser humano	HCT-116	ChIP-Seq	CTCF	GSM1022652	Publicación anterior
Ser humano	HCT-116	ChIP-Seq	CTCF_input	GSM749774	Publicación anterior
Ser humano	HCT-116	ChIP-Seq	H3K27Ac	GSM945853	Publicación anterior
Ser humano	HCT-116	ChIP-Seq	H3K27Ac_input	GSM817344	Publicación anterior
Ser humano	HCT-116	Hi-ChIP	YY1	GSM2774000	Este artículo
Ser humano	Jurkat	ChIP-seq	YY1	GSM2773998	Este artículo
Ser humano	Jurkat	ChIP-seq	YY1_input	GSM2773999	Este artículo
Ser humano	Jurkat	ChIP-seq	H3K27ac	GSM1697882	Publicación anterior
Ser humano	Jurkat	ChIP-seq	H3K27ac_input	GSM1697880	Publicación anterior
Ser humano	Jurkat	ChIP-seq	CTCF	GSM1689152	Publicación anterior
Ser humano	Jurkat	ChIP-seq	CTCF_input	GSM1689151	Publicación anterior
Ser humano	Jurkat	Hi-ChIP	YY1	GSM2774001	Este artículo
Ratón	NPC	ChIP-Seq	YY1	GSM628032	Publicación anterior
Ratón	NPC	ChIP-Seq	CTCF	GSM2259909	Publicación anterior
Ratón	NPC	ChIP-Seq	CTCF_input	GSM2259910	Publicación anterior
Ratón	NPC	ChIP-Seq	H3K27Ac	GSM594585	Publicación anterior
Ratón	NPC	ChIP-Seq	H3K27AC_entrada*	GSM2259910	Publicación anterior

(continuación)

Especie	Célula	Tipo	Nombre	GEO	Este estudio o publicación anterior
Ratón	Linfocito B	ChIP-Seq	YY1	GSM1897387	Publicación anterior
Ratón	Linfocito B	ChIP-Seq	CTCF	GSM546526	Publicación anterior
Ratón	Linfocito B	ChIP-Seq	CTCF_input	GSM546540	Publicación anterior
Ratón	Linfocito B	ChIP-Seq	H3K27Ac	GSM594592	Publicación anterior
Ratón	Linfocito B	ChIP-Seq	H3K27Ac_entrada*	GSM546540	Publicación anterior
Ratón	V6.5	GRO-sec	MEr_GRO-sec	GSM1665566	Publicación anterior
Ratón	V6.5	RNA-seq de una sola célula	SinFármaco	GSM2774584	Este artículo
Ratón	V6.5	RNA-seq de una sola célula	síFármaco	GSM2774585	Este artículo
Ratón	V6.5	Hi-ChIP	H3K27Ac_UT_rep1	GSM2774003	Este artículo
Ratón	V6.6	Hi-ChIP	H3K27Ac_UT_rep2	GSM2774004	Este artículo
Ratón	V6.5	Hi-ChIP	H3K27Ac_UT_rep3	GSM2774005	Este artículo
Ratón	V6.5	Hi-ChIP	H3K27Ac_TR_rep1	GSM2774006	Este artículo
Ratón	V6.6	Hi-ChIP	H3K27Ac_TR_rep2	GSM2774007	Este artículo
Ratón	V6.5	Hi-ChIP	H3K27Ac_WO_rep1	GSM2774008	Este artículo
Ratón	V6.6	Hi-ChIP	H3K27Ac_WO_rep2	GSM2774009	Este artículo
Ratón	V6.5	Hi-ChIP	H3K27Ac_WO_rep3	GSM2774010	Este artículo
Ratón	129/Ola, XY	Hi-C	CTCF_UT_rep1	GSM2644945	Publicación anterior
Ratón	129/Ola, XY	Hi-C	CTCF_UT_rep2	GSM2644946	Publicación anterior
Ratón	129/Ola, XY	Hi-C	CTCF_TR_rep1	GSM2644947	Publicación anterior
Ratón	129/Ola, XY	Hi-C	CTCF.TR.rep2	GSM2644948	Publicación anterior
Ratón	129/Ola, XY	Hi-C	CTCF_WO_rep1	GSM2644949	Publicación anterior
Ratón	129/Ola, Xi	Hi-C	CTCF_WO_rep2	GSM2644950	Publicación anterior
Ratón	V6.5	4C	7938357_Etv4_Prom_B3_mES_situ_re	GSM2645350	Este artículo
Ratón	V6.6	4C	7938357_Etv4_Prom_B3_mES_situ_re	GSM2645351	Este artículo

(continuación)

Especie	Célula	Tipo	Nombre	GEO	Este estudio o publicación anterior
Ratón	V6.5	4C	7938357_Erv4_Prom_B3_mES_situ_re	GSM2645352	Este artículo
Ratón	V6.5	4C	DB7938357_WT_mES_situ_rep1	GSM2645353	Este artículo
Ratón	V6.6	4C	DB7938357_WT_mES_situ_rep2	GSM2645354	Este artículo
Ratón	V6.5	4C	DB7938357_WT_mES_situ_rep3	GSM2645355	Este artículo
Ratón	V6.6	4C	4688024_Raf1_Prom_G2_mES_situ.re	GSM2645356	Este artículo
Ratón	V6.5	4C	4688024_Raf1_Prom_G2_mES_situ.re	GSM2645357	Este artículo
Ratón	V6.5	4C	4688024_Raf1_Prom_G2_mES_situ.re	GSM2645358	Este artículo
Ratón	V6.6	4C	DC4688024_WT_mES_situ_rep1	GSM2645359	Este artículo
Ratón	V6.5	4C	DC4688024_WT_mES_situ_rep2	GSM2645360	Este artículo
Ratón	V6.5	4C	DC4688024_WT_mES_situ_rep3	GSM2645361	Este artículo
Ratón	V6.6	4C	8357_Etv4_Prom_B3_dCas9_mES_sit	GSM2773992	Este artículo
Ratón	V6.5	4C	8357_Etv4_Prom_B3_dCas9_mES_sit	GSM2773993	Este artículo
Ratón	V6.6	4C	8357_Etv4_Prom_B3_dCas9_mES_sit	GSM2773994	Este artículo
Ratón	V6.5	4C	57_Etv4_Prom_B3_dCas9_YY1_mES_	GSM2773995	Este artículo
Ratón	V6.5	4C	57_Etv4_Prom_B3_dCas9-YY1_mES_	GSM2773996	Este artículo
Ratón	V6.6	4C	57_Etv4_Prom_B3_dCas9-YY1_mES_	GSM2773997	Este artículo

Origami: <https://github.com/younglab/origami> usando la versión v1.1-alpha-2.

Los datos asociados con este estudio se han depositado en Gene Expression Omnibus (GEO) con el código de identificación GSE99521.

REFERENCIAS

- Adhya, S. (1989). Multipartite genetic control elements: communication by DNA loop. *Annu. Rev. Genet.* 23, 227-250.
- Allen, B.L., y Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. *Nat. Rev. Mol. Cell Biol.* 16, 155-166.
- Amoutzias, G.D., Robertson, D.L., Van de Peer, Y. y Oliver, S.G. (2008). Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.* 33, 220-229.
- Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C. V, Gillis, D.R. y Phillips-cremins, J.E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.*

- Bell, A.C. y Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405, 2-5.
- Bonev, B. y Cavalli, G. (2016). Organization and function of the 3D genome. *Nat. Rev. Genet.* 17, 772-772. Buecker, C. y Wysocka, J. (2012). Enhancers as information integration hubs in development: Lessons from genomics. *Trends Genet.* 28, 276-284.
- Bulger, M. y Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell.* Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S. A.* 107, 21931-21936.
- Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K. y Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* 19, 24-32.
- Degner, S.C., Verma-Gaur, J., Wong, T.P., Bossen, C., Iverson, G.M., Torkamani, A., Vettermann, C., Lin, Y.C., Ju, Z., Schulz, D., *et al.* (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the *Igh* locus and antisense transcription in pro-B cells. *Proc. Natl. Acad. Sci.* 108, 9566-9571.
- Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A. y Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149, 1233-1244.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. y Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376-380.
- Donohoe, M.E., Zhang, X., McGinnis, L., Biggers, J., Li, E. y Shi, Y. (1999). Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. *Mol. Cell. Biol.* 19, 7237-7244.
- Downen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., *et al.* (2014). Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. *Cell* 159, 374-387.
- Erb, M.A., Scott, T.G., Li, B.E., Xie, H., Paulk, J., Seo, H. S., Souza, A., Roberts, J.M., Dastjerdi, S., Buckley, D.L., *et al.* (2017). Transcription control by the ENL YEATS domain in acute leukaemia. *Nature* 543, 270-274.
- Fraser, J., Williamson, I., Bickmore, W. a y Dostie, J. (2015). An Overview of Genome Organization and How We Got There: from FISH to Hi-C. *Microbiol. Mol. Biol. Rev.* 79, 347-372.
- Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y. Bin, Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., *et al.* (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462, 58-64.
- Gabriele, M., Vulto-van Silfhout, A. T., Germain, P.-L., Vitriolo, A., Kumar, R., Douglas, E., Haan, E., Kosaki, K., Takenouchi, T., Rauch, A., *et al.* (2017). YY1 Haploinsufficiency Causes an Intellectual Disability Syndrome Featuring Transcriptional and Chromatin Dysfunction. *Am. J. Hum. Genet.* 100, 907-925.
- Gibcus, J.H. y Dekker, J. (2013). The hierarchy of the 3D genome. *Mol. Cell* 49, 773-782. Gordon, S., Akopyan, G., Garban, H. y Bonavida, B. (2006). Factor de transcripción YY1: estructura, función, and therapeutic implications in cancer biology. *Oncogene* 25, 1125-1142. Gorkin, D.U., Leung, D. y Ren, B. (2014). The 3D Genome in Transcriptional Regulation and Pluripotency. *Cell Stem Cell* 14, 762-775.
- Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.-L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., *et al.* (2011). CTCF-binding elements mediate control of V(D)J recombination. *Nature* 477, 424-430.
- Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., *et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162, 900-910.
- Hariharan, N., Kelley, D.E. y Perry, R.P. (1991). Delta, a transcription factor that binds to downstream elements in several polymerase II promoters, is a functionally versatile zinc finger protein. *Proc. Natl. Acad. Sci. U.S. A.* 88, 9799-9803.
- Heard, E. y Bickmore, W. (2007). The ins and outs of gene regulation and chromosome territory organisation. *Curr. Opin. Cell Biol.* 19, 311-316.
- Heath, H., Ribeiro de Almeida, C., Sleutels, F., Dingjan, G., van de Nobelen, S., Jonkers, I., Ling, K.-W., Gribnau, J., Renkawitz, R., Grosveld, F., *et al.* (2008). CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *EMBO J.* 27, 2839-2850.
- Hnisz, D., Day, D.S. y Young, R.A. (2016a). Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* 167, 1188-1200.
- Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.-L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., *et al.* (2016b). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* (80-.). 351, 1454-1458.
- Huang, H., Seo, H., Zhang, T., Wang, Y., Jiang, B., Li, Q., Buckley, D.L., Nabet, B., Roberts, J.M., Paulk, J., *et al.* (2017). MELK is not necessary for the proliferation of basal-like breast cancer cells. 1-29.
- Hwang, S.S., Kim, Y.U., Lee, S., Jang, S.W., Kim, M.K., Koh, B.H., Lee, W., Kim, J., Souabni, A., Busslinger, M., *et al.* (2013). Transcription factor YY1 is essential for regulation of the Th2 cytokine locus and for Th2 cell differentiation. *Proc. Natl. Acad. Sci.* 110, 276-281.
- Jeronimo, C., Langelier, M.F., Bataille, A.R., Pascal, J.M., Pugh, B.F. y Robert, F. (2016). Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function *In Vivo*. *Mol. Cell* 64, 455-466.
- Ji, X., Dadon, D.B., Abraham, B.J., Ihn, T., Jaenisch, R., Bradner, J.E. y Young, R.A. (2015). Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc. Natl. Acad. Sci.* 112, 3841-3846.
- Ji, X., Dadon, D., Powell, B., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., Hnisz, D., Pegoraro, G.,

- Lee, T.I., *et al.* (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* 18, 1-14.
- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D. a, van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., *et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430-435.
- Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D.D., Zhang, M.Q., Lobanenko, V. V. y Ren, B. (2007). Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. *Cell* 128, 1231-1245.
- Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., Goodwin, G.H., Neiman, P.E. y Lobanenko, V. V (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol. Cell. Biol.* 13, 7612-7624.
- de Laat, W. y Duboulé, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502, 499-506.
- Lamb, P. y McKnight, S.L. (1991). Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. *Trends Biochem. Sci.* 16, 417-422.
- Levine, M., Cattoglio, C. y Tjian, R. (2014). Looping back to leap forward: Transcription enters a new era. *Cell* 157, 13-25.
- Liu, H., Schmidt-suppran, M., Shi, Y., Hobeika, E., Barteneva, N., Jumaa, H., Pelanda, R., Reth, M., Skok, J., Rajewsky, K., *et al.* (2007). Yin Yang 1 is a critical regulator of B-cell development. 1179-1189.
- Lopez-Perrote, A., Alatwi, H.E., Torreira, E., Ismail, A., Ayora, S., Downs, J.A. y Llorca, O. (2014). Structure of Yin Yang 1 oligomers that cooperate with RuvBL1-RuvBL2 ATPases. *J. Biol. Chem.* 289, 22614-22629.
- Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., *et al.* (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* 1-14.
- Malik, S. y Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* 11, 761-772.
- Matthews, K.S. (1992). DNA looping. *Microbiol. Rev.* 56, 123-136.
- Mele, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., *et al.* (2015). The human transcriptome across tissues and individuals. *Science* (80-.). 348, 660-665.
- Merkenschlager, M. y Nora, E.P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu. Rev. Genomics Hum. Genet.* 17, 1-27.
- Muerdter, F. y Stark, A. (2016). Gene Regulation: Activation through Space. *Curr. Biol.* 26, R895-R898.
- Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J. y Chang, H. Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* 13, 919-922.
- Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O. y Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. 347, 1017-1022.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., *et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381-385.
- Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A. y Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930-944.e22.
- Ohlsson, R., Renkawitz, R. y Lobanenko, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* 17, 520-527.
- Ong, C. y Corces, V. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* 12, 283-293.
- Park, K. y Atchison, M.L. (1991). Isolation of a candidate repressor/activator, NF-E1 (YY1, d), that binds to the immunoglobulin k 3' enhancer and the immunoglobulin heavy-chain mE1 site. *Proc. Natl. Acad. Sci.* 88, 9804-9808.
- Petrenko, N., Jin, Y., Wong, K.H. y Struhl, K. (2016). Mediator Undergoes a Compositional Change during Transcriptional Activation. *Mol. Cell* 64, 443-454.
- Phillips, J.E. y Corces, V.G. (2009). CTCF: master weaver of the genome. *Cell* 137, 1194-1211.
- Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T. a, Guo, C., Sun, Y., *et al.* (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281-1295.
- Pombo, A. y Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.* 12.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., *et al.* (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 1-16.
- Ren, B. y Yue, F. (2016). Transcriptional enhancers: Bridging the genome and phenome. *En Cold Spring Harbor Symposia on Quantitative Biology*, pp. 17-26.
- Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F. y Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* 28, 723-734.
- Schleif, R. (1992). DNA looping. *Annu. Rev. Biochem.* 61, 199-223.
- Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P. y Odom, D.T.

- (2010). A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* 20, 578-588.
- Shi, Y., Seto, E., Chang, L.S. y Shenk, T. (1991). Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* 67, 377-388.
- Shi, Y., Lee, J.S. y Galvin, K.M. (1997). Everything you have ever wanted to know about Yin Yang 1... *Biochim. Biophys. Acta - Rev. Cancer* 1332.
- Shore, D., Langowski, J. y Baldwin, R.L. (1981). DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. U.S. A.* 78, 4833-4837.
- Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N. M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P. a. y Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. *Science* (80-.). 350, 978-981.
- Spitz, F. (2016). Gene regulation at a distance: From remote enhancers to 3D regulatory ensembles. *Semin. Cell Dev. Biol.* 57, 57-67.
- Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N. y de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 2349-2354.
- Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczyski, B., *et al.* (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* 163, 1611-1627.
- Thomas, M.J. y Seto, E. (1999). Unlocking the mechanisms of transcription factor YY1: Are chromatin modifying enzymes the key? *Gene* 236, 197-208.
- Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S. y Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. *Science* (80-.). 350, 1096-1101.
- Weirauch, M.T. y Hughes, T.R. (2011). A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. En *A Handbook of Transcription Factors, Subcellular Biochemistry*, pp. 25-73.
- Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., *et al.* (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 796-801.
- Winter, G.E., Buckley, D.L., Paulk, J., Roberts, J.M., Souza, A., Dhe-Paganon, S. y Bradner, J.E. (2015). Phthalimide conjugation as a strategy for *in vivo* target protein degradation. *Science* 348, 1376-1381.
- Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., *et al.* (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* (80-.). 356, eaaj2239.
- Zhang, Q., Stovall, D.B., Inoue, K. y Sui, G. (2011). The Oncogenic Role of Yin Yang 1.

REIVINDICACIONES

1. Un método *in vitro* para modular la expresión de uno o más genes en una célula, que comprende modular la multimerización de YY1 y, por lo tanto, modular la expresión del uno o más genes.
2. Un método *in vitro* para modular la expresión de uno o más genes en una célula, que comprende modular la formación de un bucle de ADN potenciador-promotor en el genoma de la célula, en donde la formación depende de YY1 y en donde la formación es modulada por (a) la modulación de la unión de YY1 a la región promotora y/o potenciadora del bucle de ADN promotor-potenciador modificando la región promotora y/o potenciadora o (b) la modulación de la multimerización de YY1 en la célula.
3. Una composición que modula la formación de bucles de ADN potenciador-promotor, en donde la formación del bucle de ADN promotor potenciador depende de YY1, para su uso en el tratamiento de una enfermedad o afección asociada con la expresión génica aberrante en un sujeto, en donde la formación es modulada por (a) la modulación de la unión de YY1 a la región promotora y/o potenciadora del bucle de ADN promotor-potenciador modificando la región promotora y/o potenciadora, o (b) modulando la multimerización de YY1 en la célula.
4. El método o la composición para su uso de una cualquiera de las reivindicaciones 1-3, en donde la expresión de uno o más genes disminuye o aumenta.
5. El método de una cualquiera de las reivindicaciones 1, 2 o 4, en el que la célula es una célula madre; en donde, opcionalmente, la célula es una célula madre embrionaria.
6. El método de una cualquiera de las reivindicaciones 1, 2, 4 o 5, en donde el uno o más genes comprenden Oct4 y/o Sox2.
7. El método de la reivindicación 1, en donde YY1 se une a un potenciador ya una región promotora del genoma de la célula; opcionalmente, en donde las regiones potenciadora y promotora están ubicadas ambas en la misma vecindad aislada del genoma de la célula.
8. El método de una cualquiera de las reivindicaciones 1, 4, 5, 6 o 7, en donde la multimerización de YY1 está (a) modulada, modulando así la formación de bucles de ADN potenciador-promotor en el genoma de la célula; y/o (b) modulada con una composición que comprende un ácido nucleico y/o una molécula pequeña; y/o (c) disminuida o aumentada.
9. El método de la reivindicación 2, en donde la formación se modula modulando la unión de YY1 a la región promotora y/o potenciadora del bucle de ADN promotor-potenciador modificando la región promotora y/o potenciadora; opcionalmente en donde la modificación comprende (a) modificar el grado de metilación de la región promotora y/o potenciadora o (b) modificar la secuencia de nucleótidos de la región promotora y/o potenciadora; opcionalmente en donde la región promotora y/o potenciadora comprende un sitio de unión a YY1.
10. El método de una cualquiera de las reivindicaciones 2, 4, 5, 6 o 9, en donde la unión se modula poniendo en contacto la célula con una composición que comprende una molécula pequeña y/o un ácido nucleico.
11. El método de la reivindicación 2, 4, 5 o 6, en donde la formación del bucle de ADN potenciador-promotor se modula modulando la multimerización de YY1 en la célula; opcionalmente, en donde la multimerización se modula poniendo en contacto la célula con un ácido nucleico y/o una molécula pequeña y/o en donde la multimerización aumenta o disminuye.
12. La composición para su uso de la reivindicación 3, en donde la formación del bucle de ADN potenciador-promotor se modula modulando la unión de YY1 a una región promotora y/o potenciadora del bucle de ADN potenciador-promotor modificando la región promotora y/o potenciadora; opcionalmente en donde la modificación comprende (a) modificar la metilación de la región promotora y/o potenciadora; o (b) modificar la secuencia de nucleótidos de la región promotora y/o potenciadora; opcionalmente, en donde la unión se modula mediante la administración al sujeto de una composición que comprende una molécula pequeña y/o un ácido nucleico.
13. La composición para el uso de la reivindicación 3 o 4, en donde la formación del bucle de ADN potenciador-promotor se modula modulando la multimerización de YY1 en el sujeto; opcionalmente, en donde la multimerización se modula administrando al sujeto un ácido nucleico y/o una molécula pequeña y/o en donde la multimerización aumenta o disminuye.
14. La composición para su uso de una cualquiera de las reivindicaciones 3, 4, 12 o 13, en donde la enfermedad o afección asociada con la expresión génica aberrante es cáncer.
15. Un método *in vitro* de detección de un compuesto que modula la expresión de uno o más genes en una célula, que comprende

poner en contacto la célula con un agente de prueba y
 medir la formación de bucles de ADN potenciador-promotor dependientes de YY1 en la célula,
 en donde el agente de prueba se identifica como un modulador de la expresión génica si el nivel de bucle de ADN
 5 potenciador-promotor en la célula en contacto con el agente de prueba es diferente del nivel de formación de bucle
 de ADN potenciador-promotor en una célula de control que no se puso en contacto con el agente de prueba.

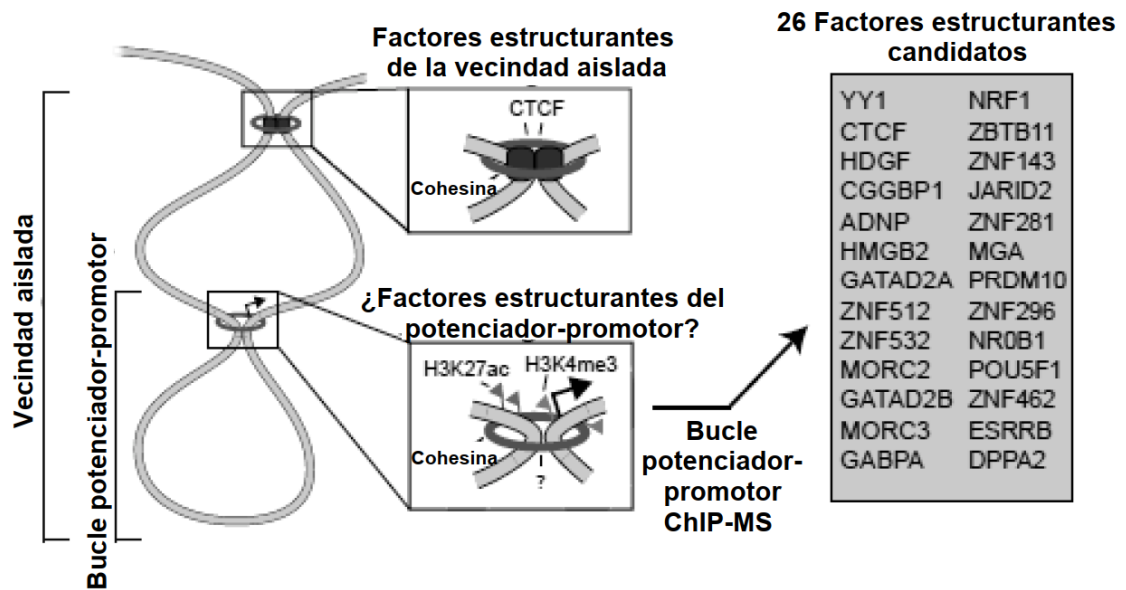
16. Un método *in vitro* para identificar uno o más genes con expresión dependiente de un potenciador en una célula,
 que comprende

10 identificar uno o más bucles de ADN promotor-potenciador dependientes de YY1 que comprenden el potenciador
 en la célula e
 identificar el uno o más genes expresados en el bucle de ADN potenciador-promotor,

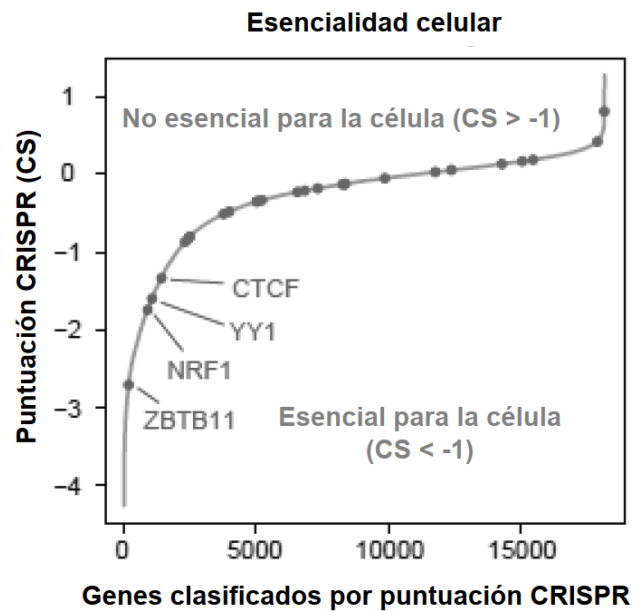
15 en donde el uno o más genes expresados en el bucle de ADN potenciador-promotor se identifican como genes cuya
 expresión depende del potenciador; opcionalmente, en donde la etapa de identificar uno o más bucles de ADN
 promotor-potenciador que comprenden el potenciador comprende realizar un ensayo ChIP-MS; y/o en donde el
 potenciador es un potenciador asociado a una enfermedad.

FIGS. 1A-1B

1A



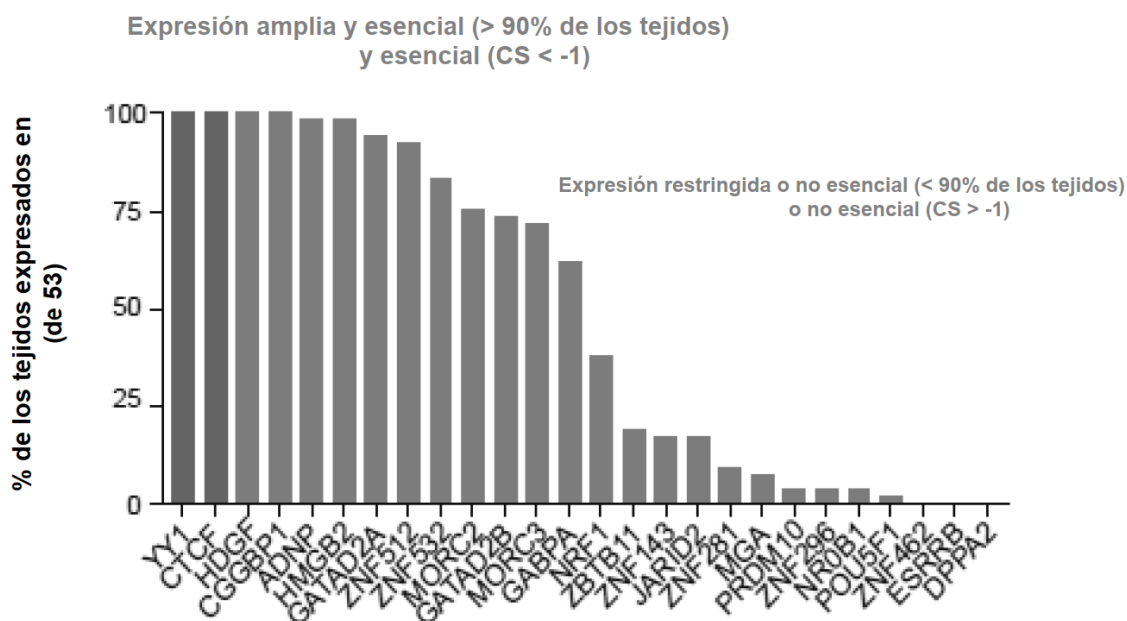
1B



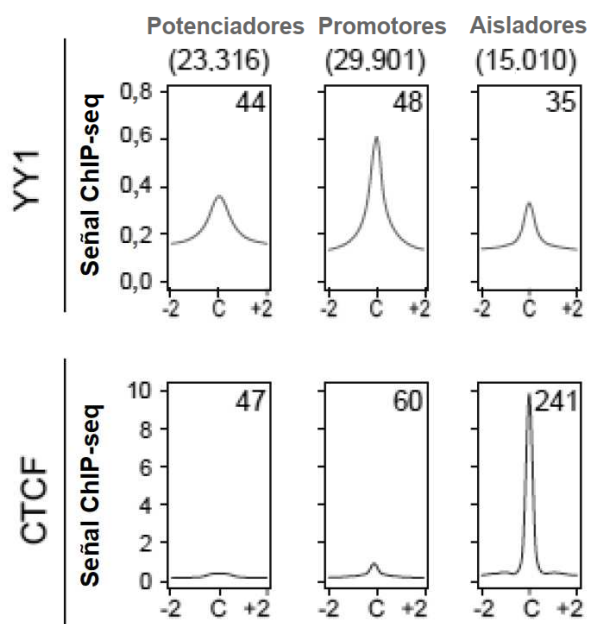
FIGS. 1C-1D

1C

Expresión específica de tejido

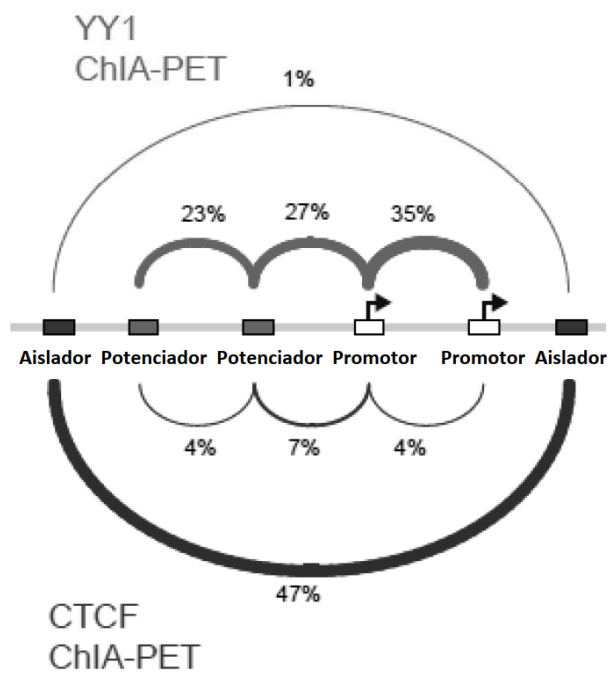


1D

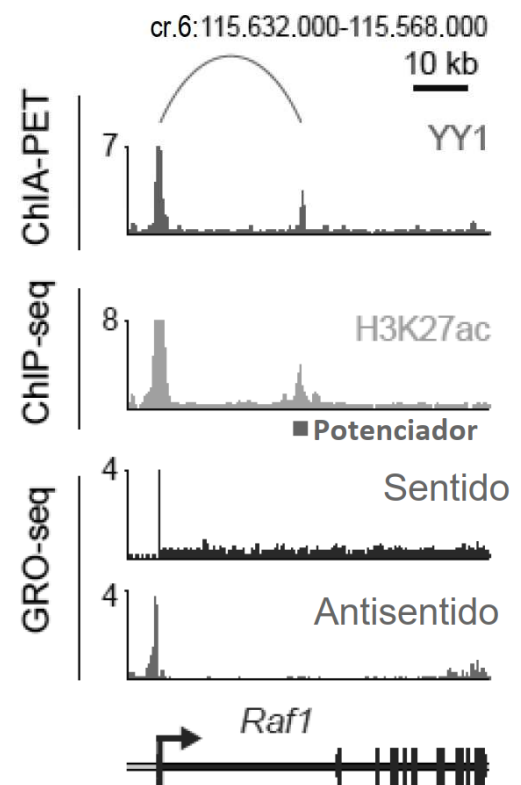


FIGS. 1E-1F

1E

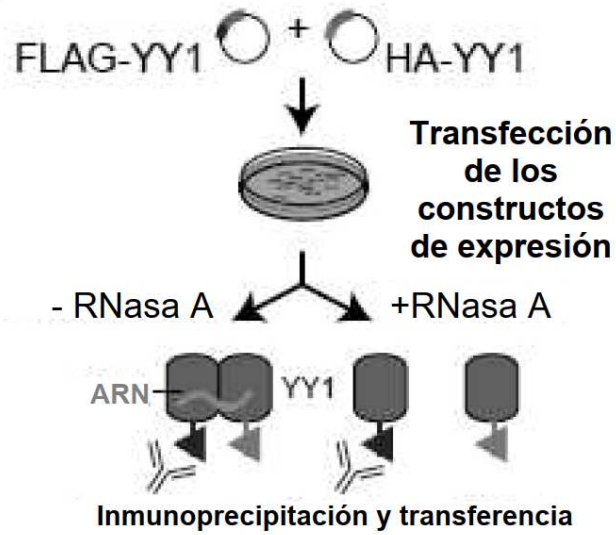


1F

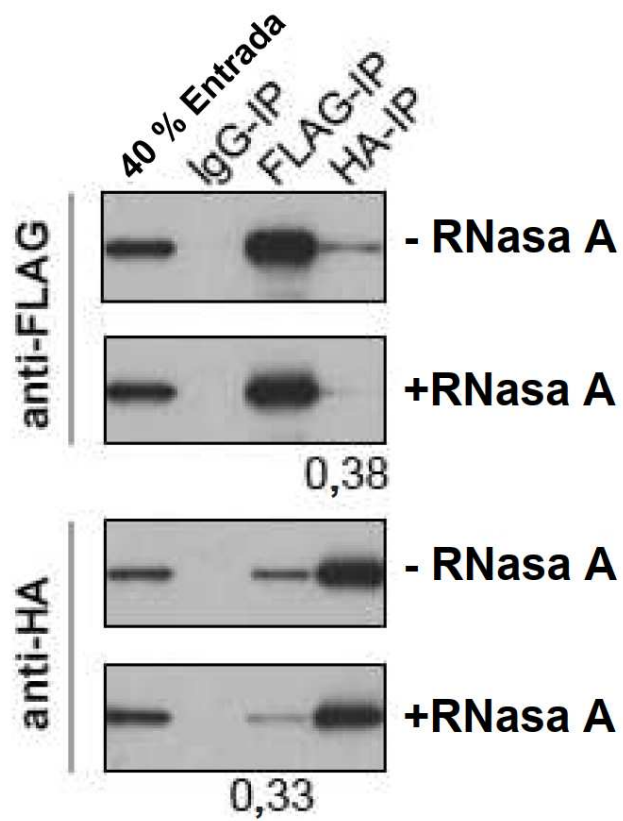


FIGS. 1G-1H

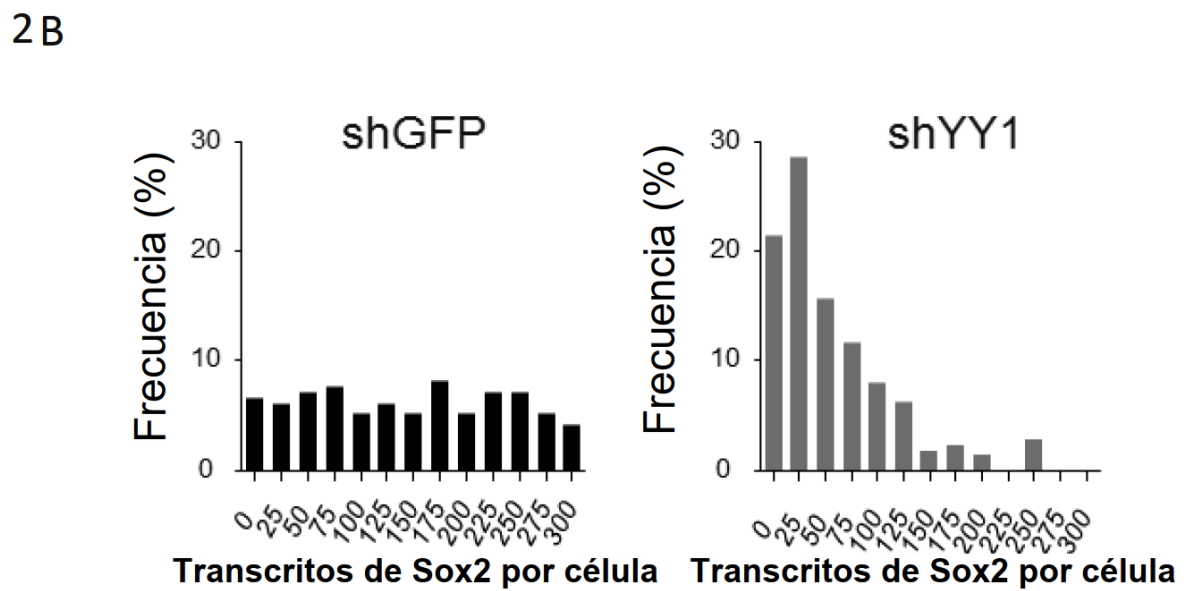
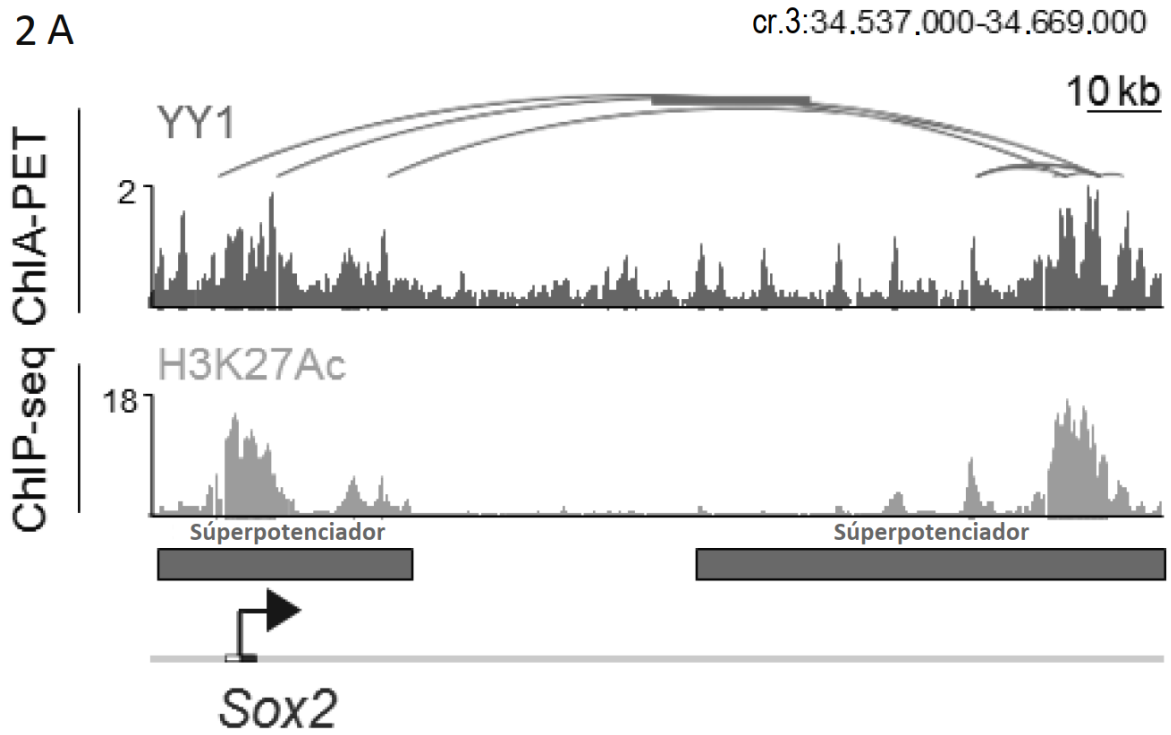
1G



1H

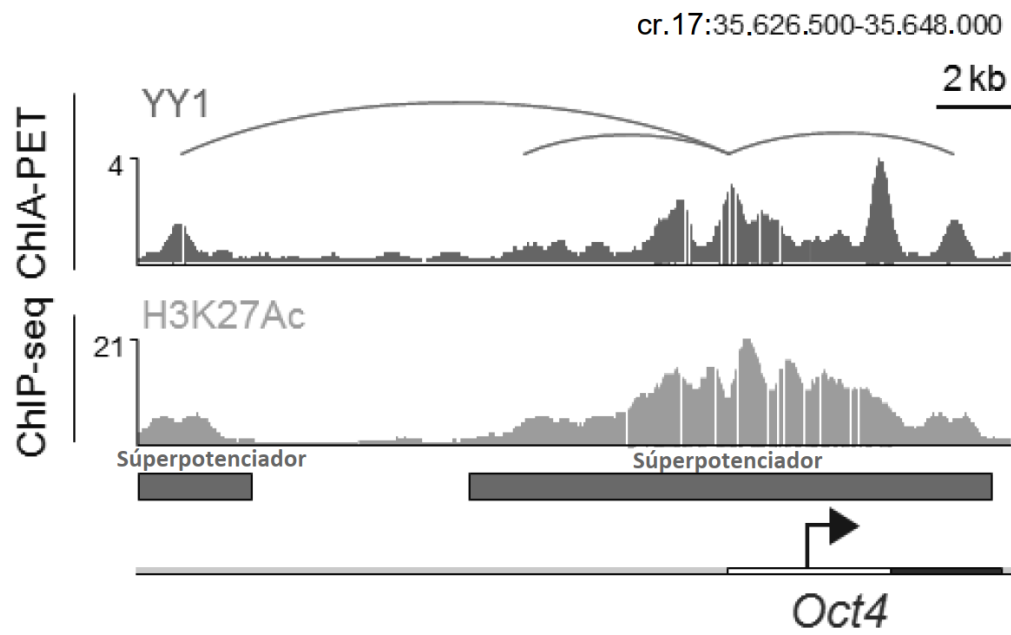


FIGS. 2A-2B

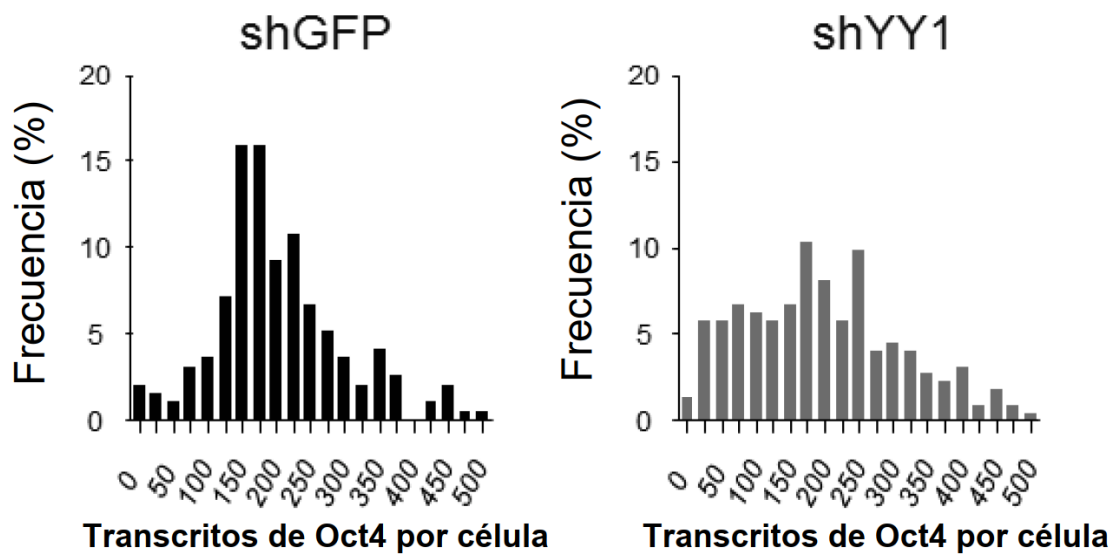


FIGS. 2C-2D

2C

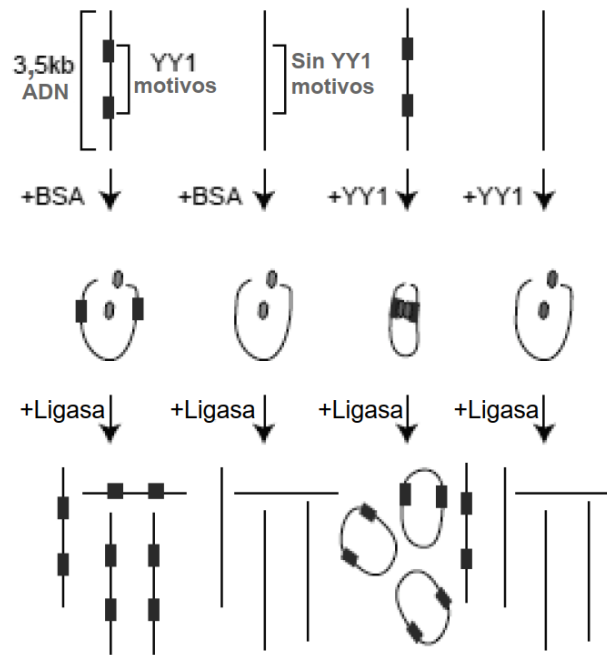


2D



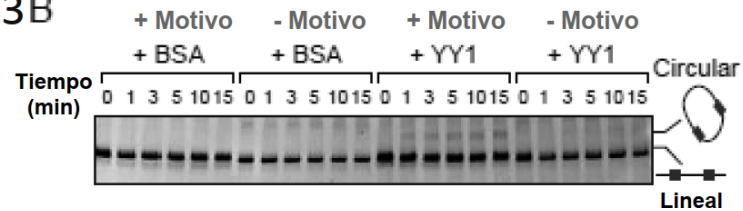
FIGS. 3A-3C

3A

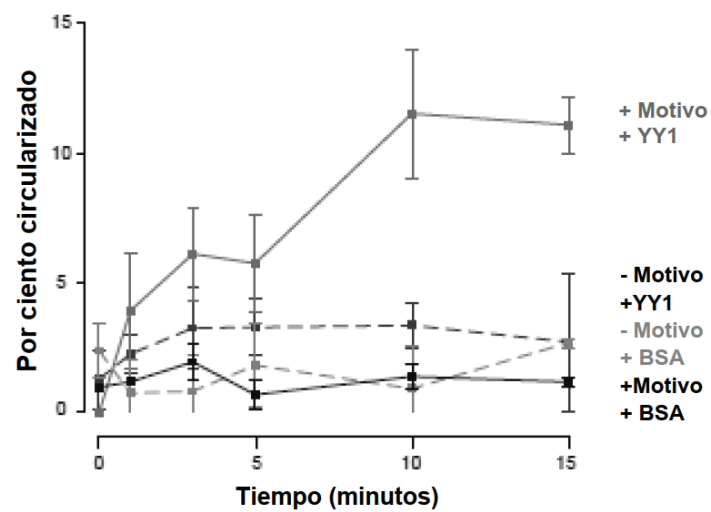


Separar ADN circular del lineal con electroforesis en gel

3B

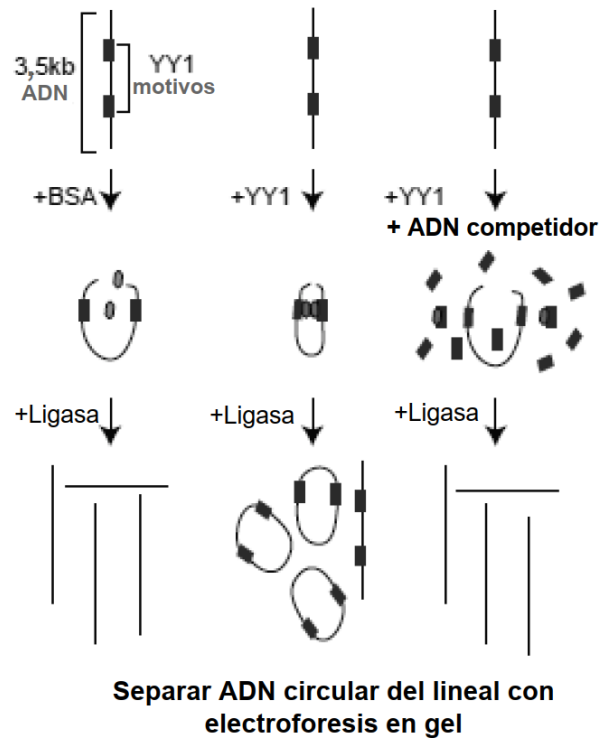


3C

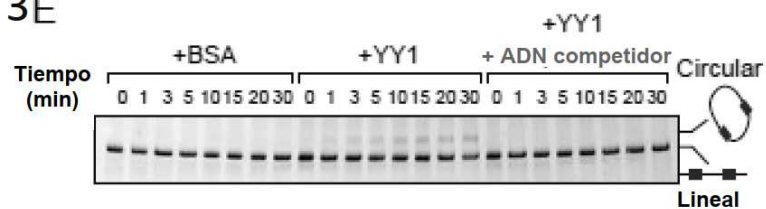


FIGS. 3D-3F

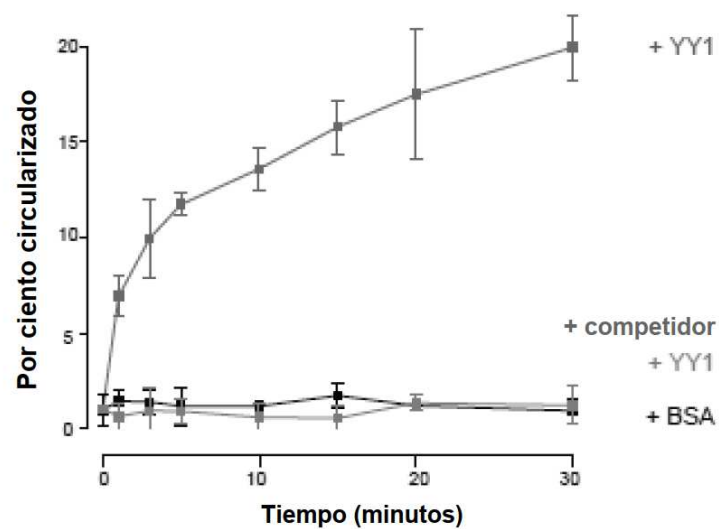
3D



3E



3F



FIGS. 4A-4E

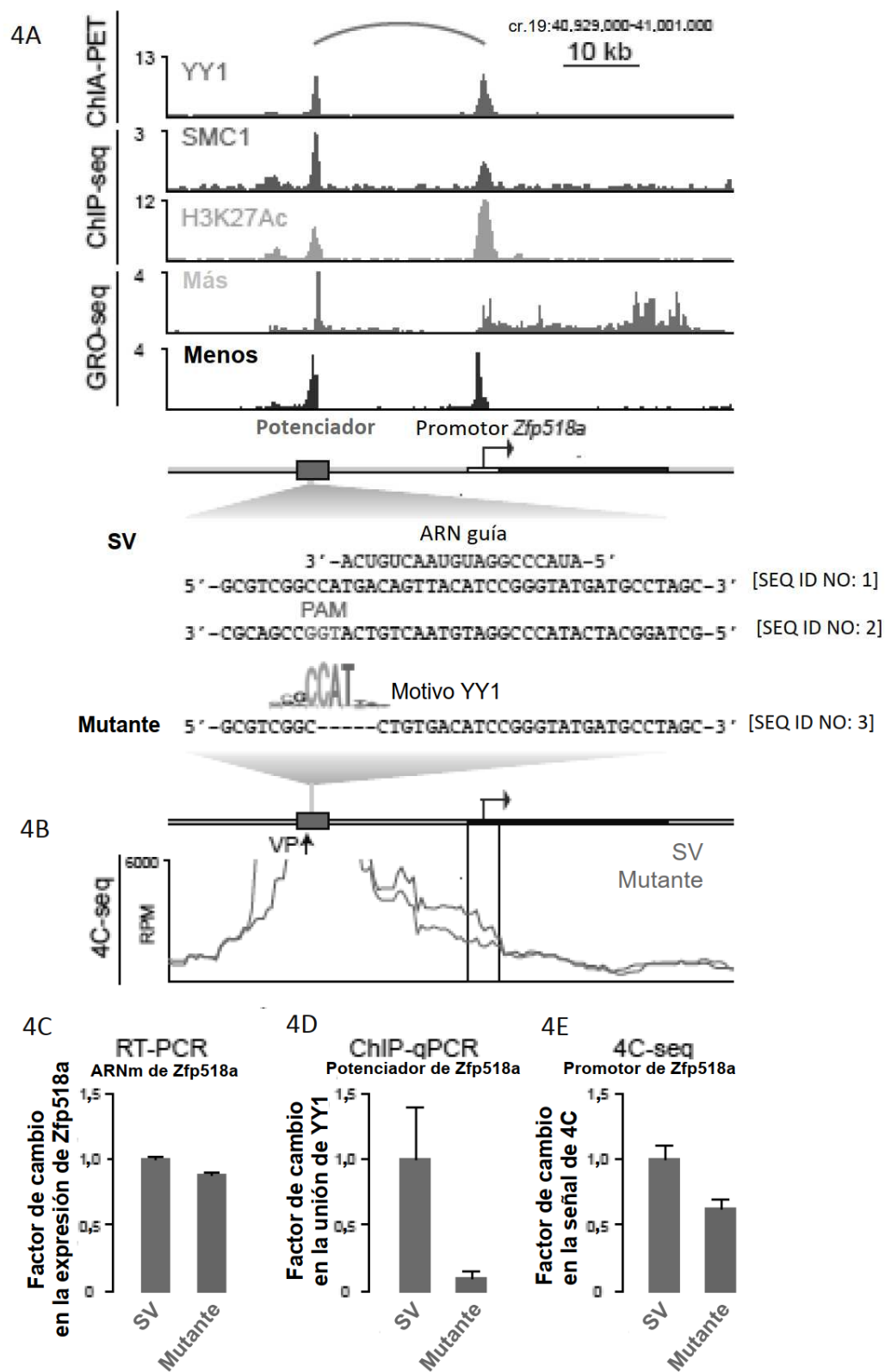
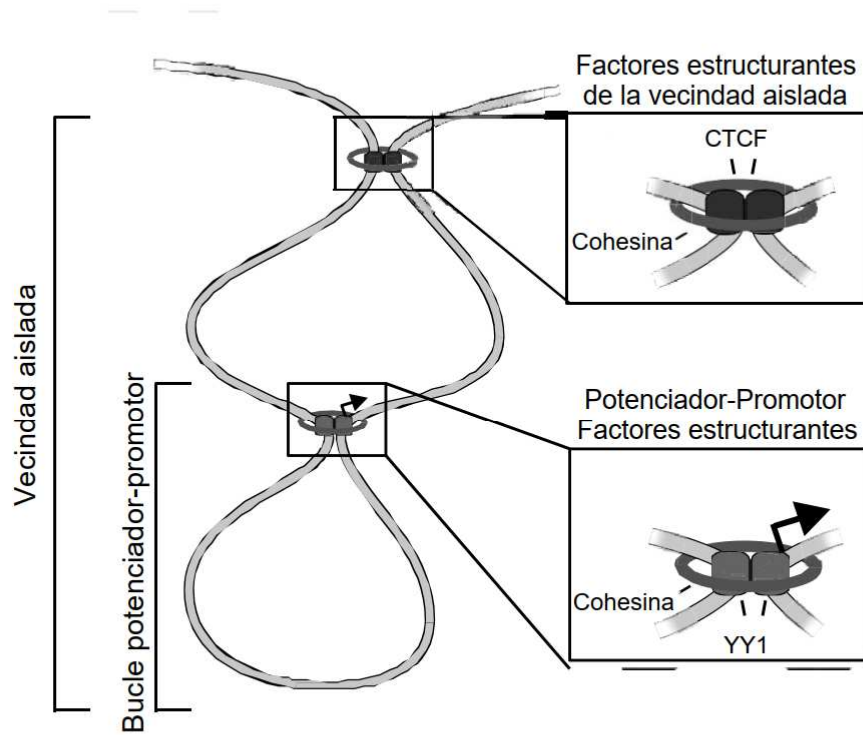
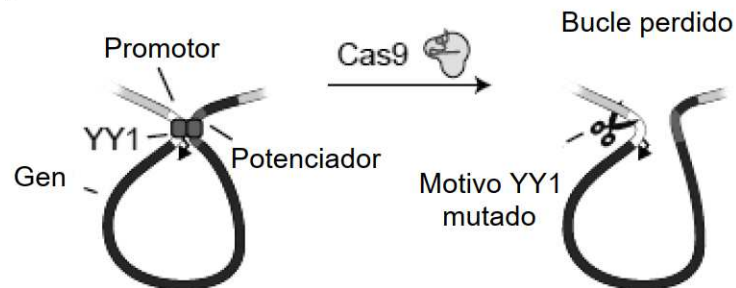


FIG. 5



FIGS. 6A-6B

6A



6B

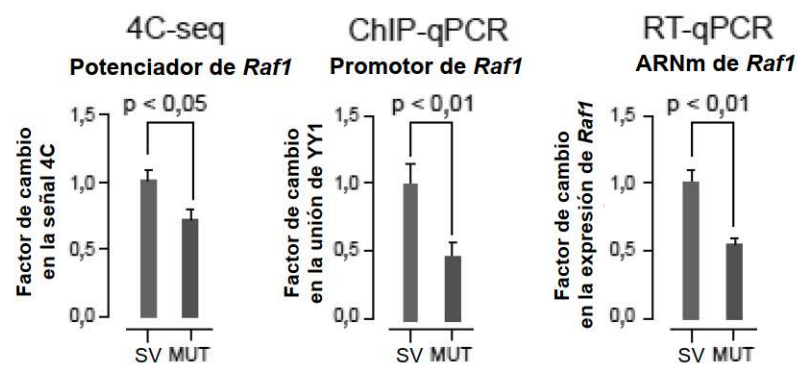
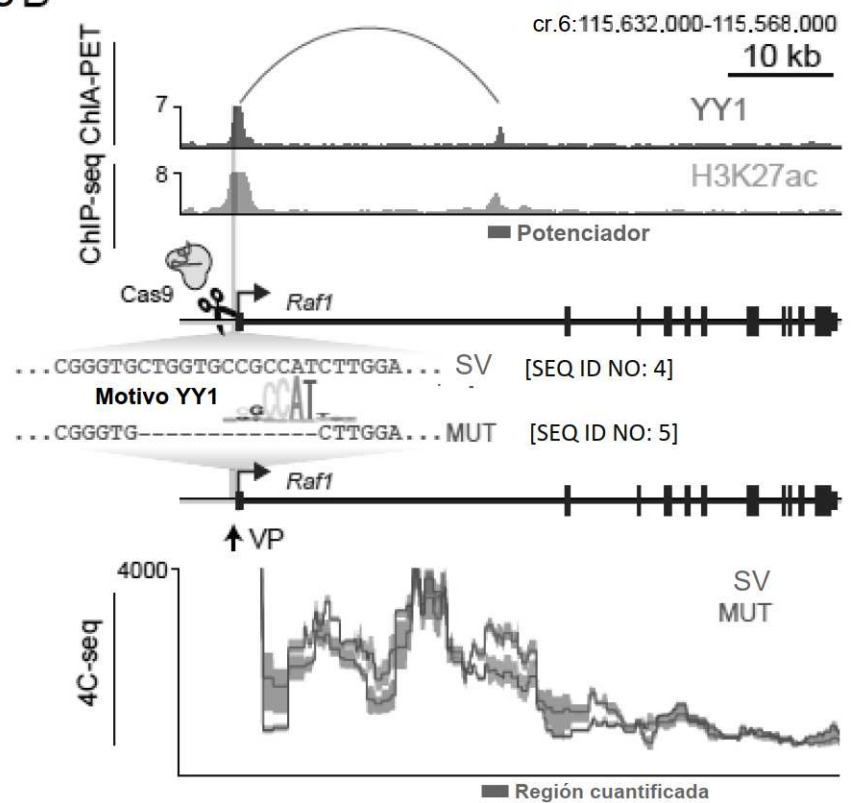
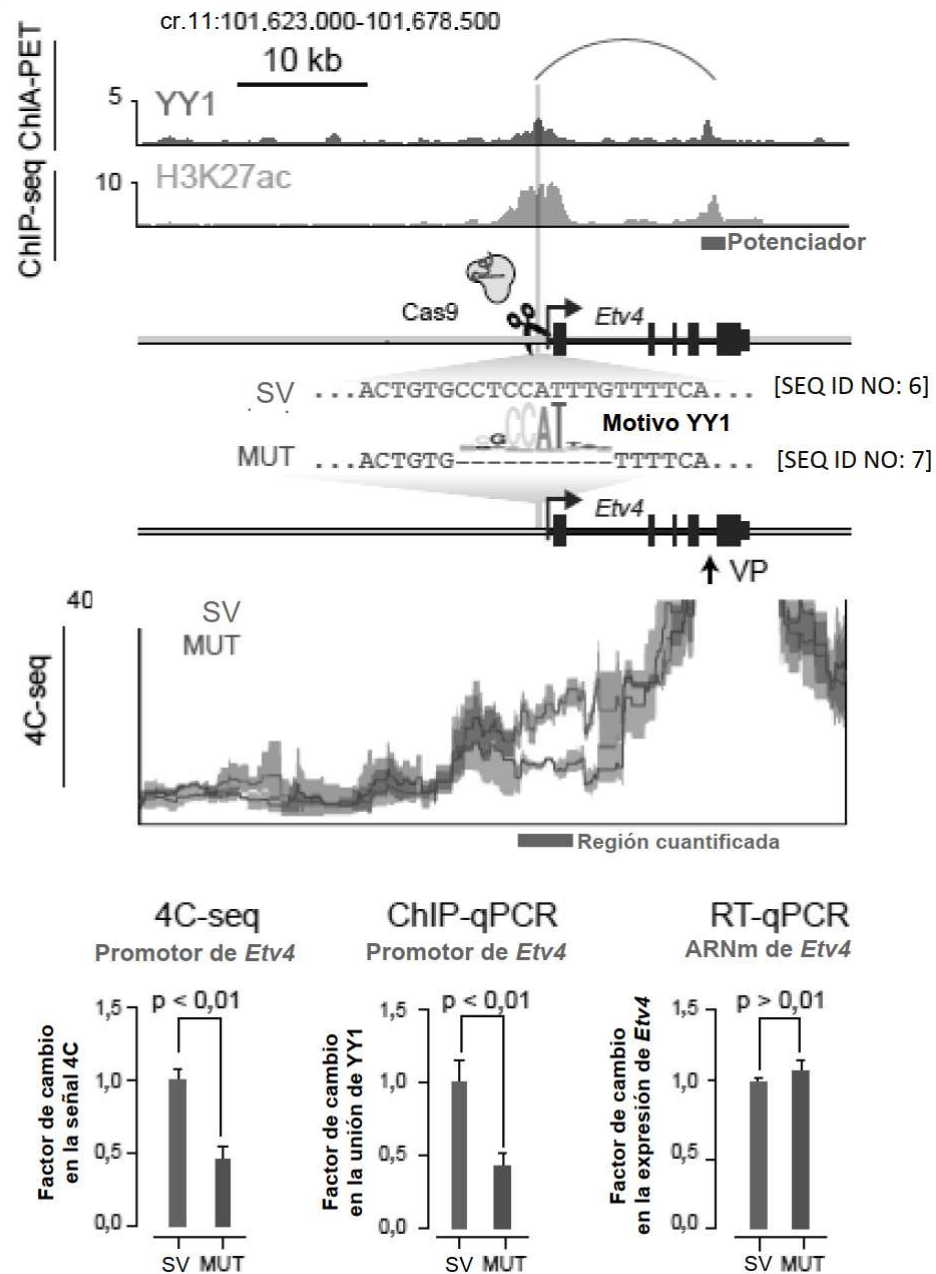
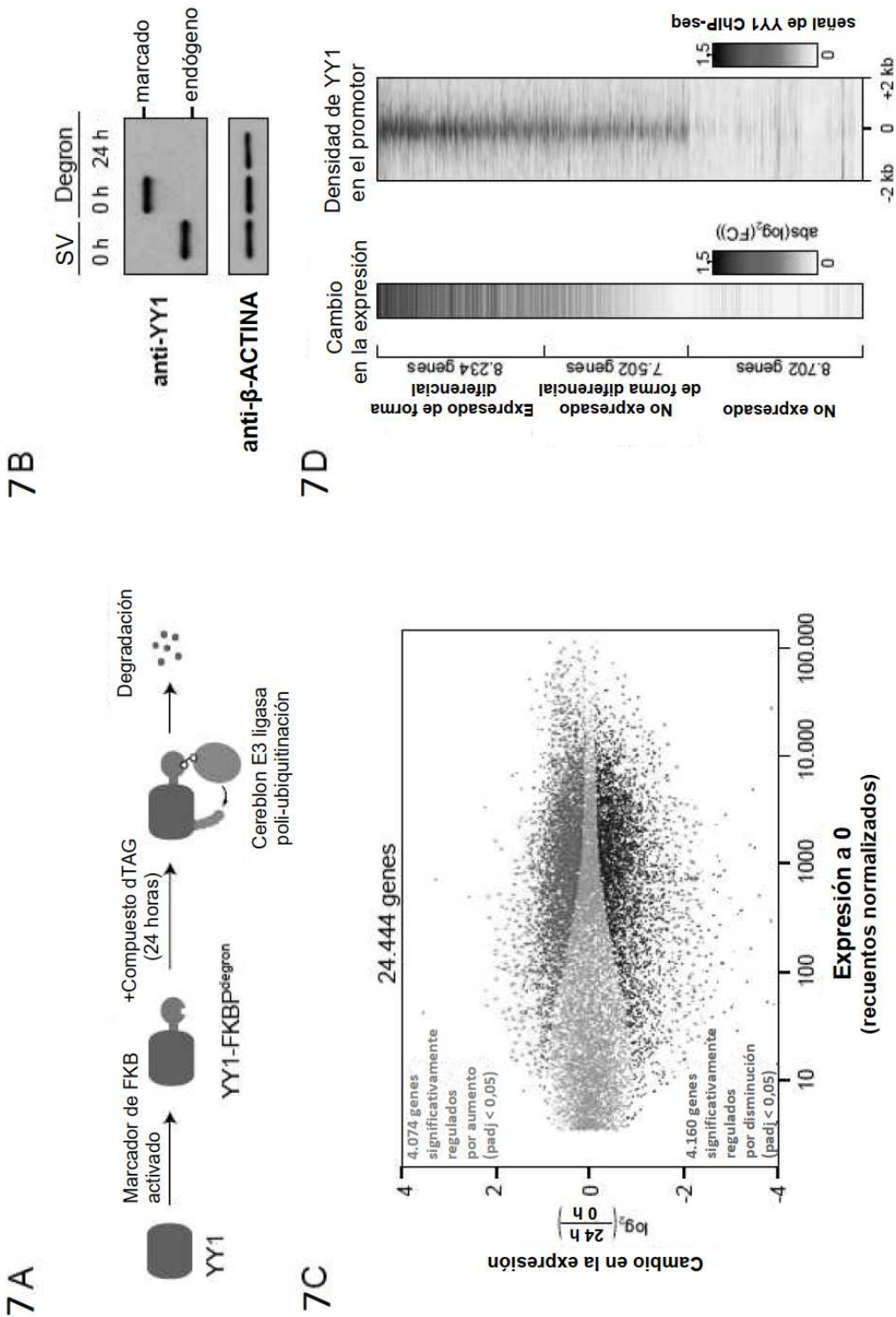


FIG. 6C

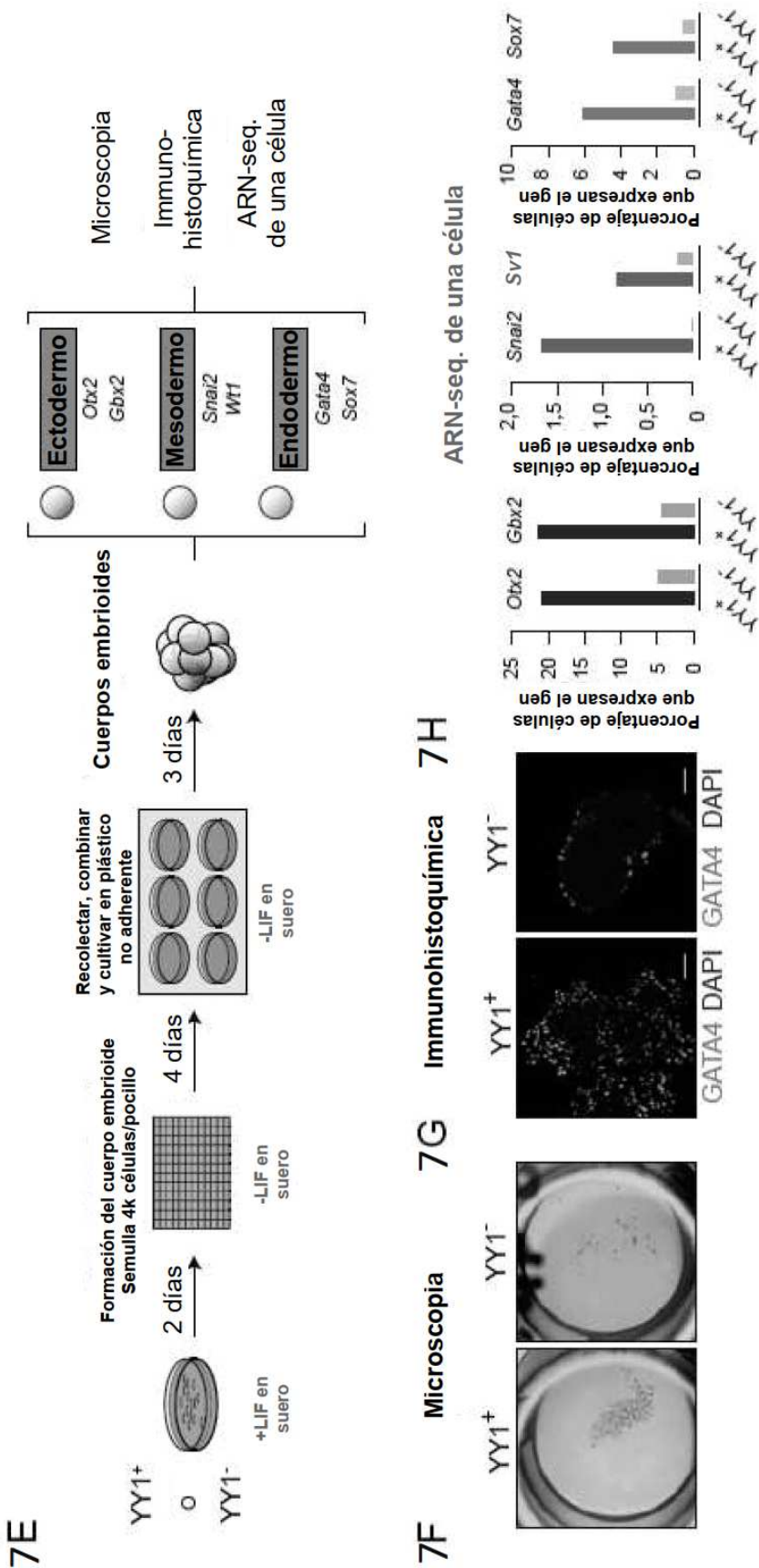
6C



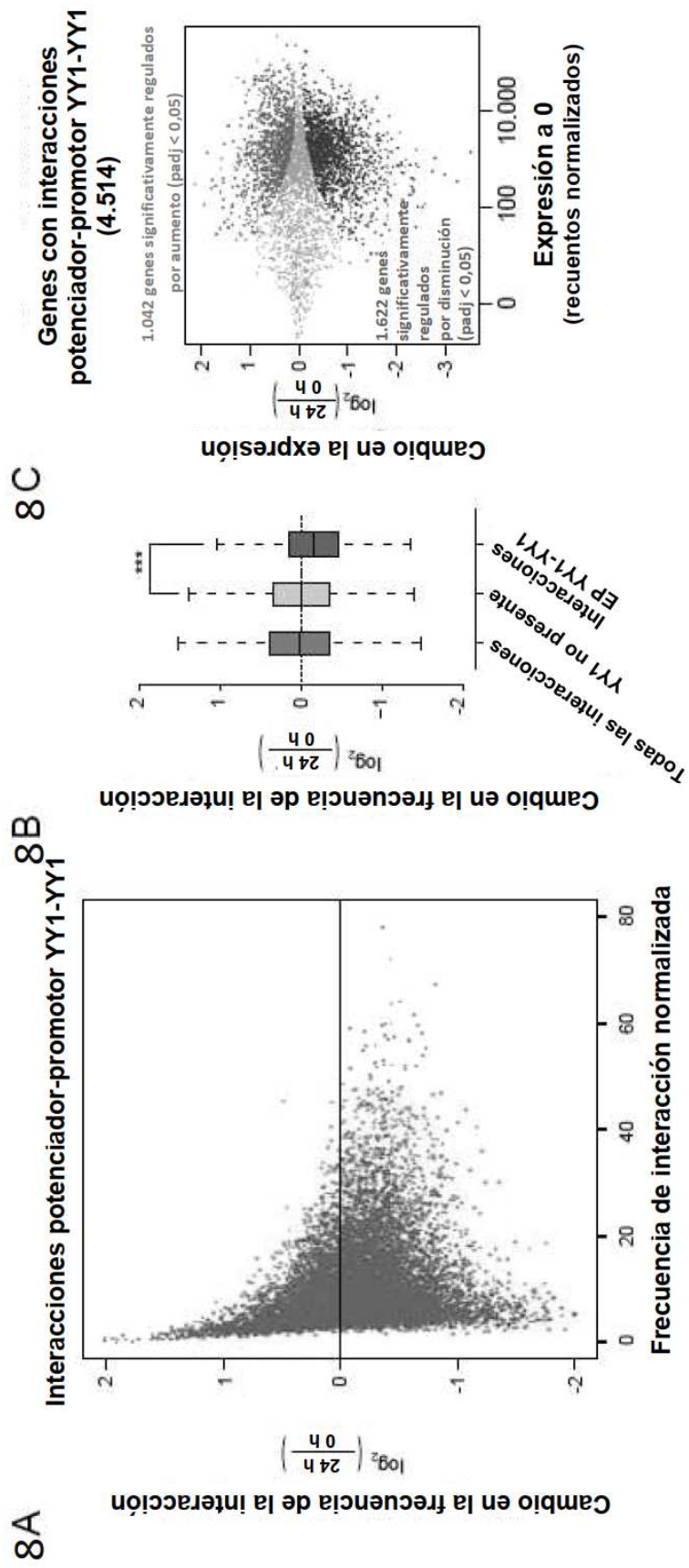
FIGS. 7A-7D



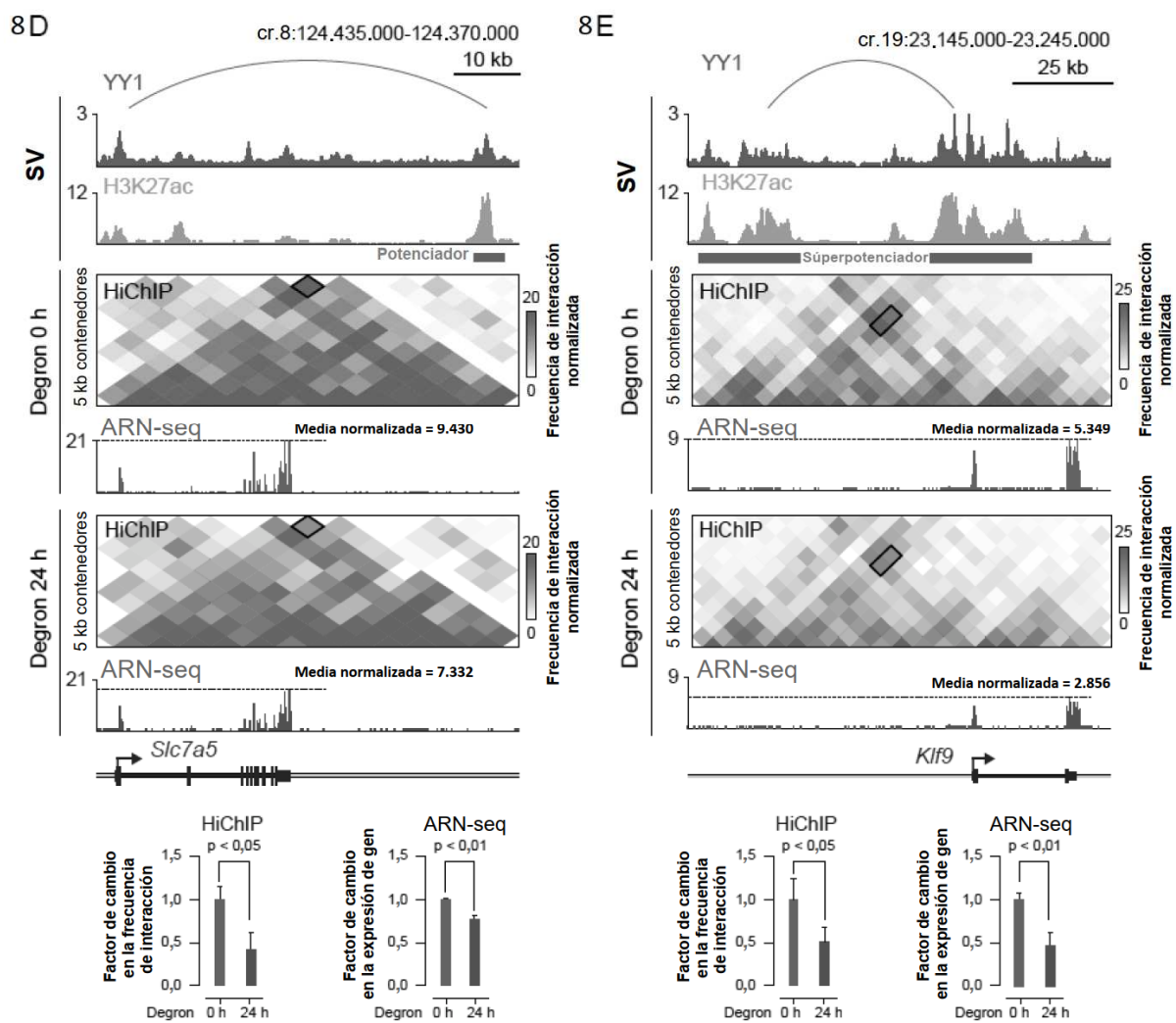
FIGS. 7E-7H



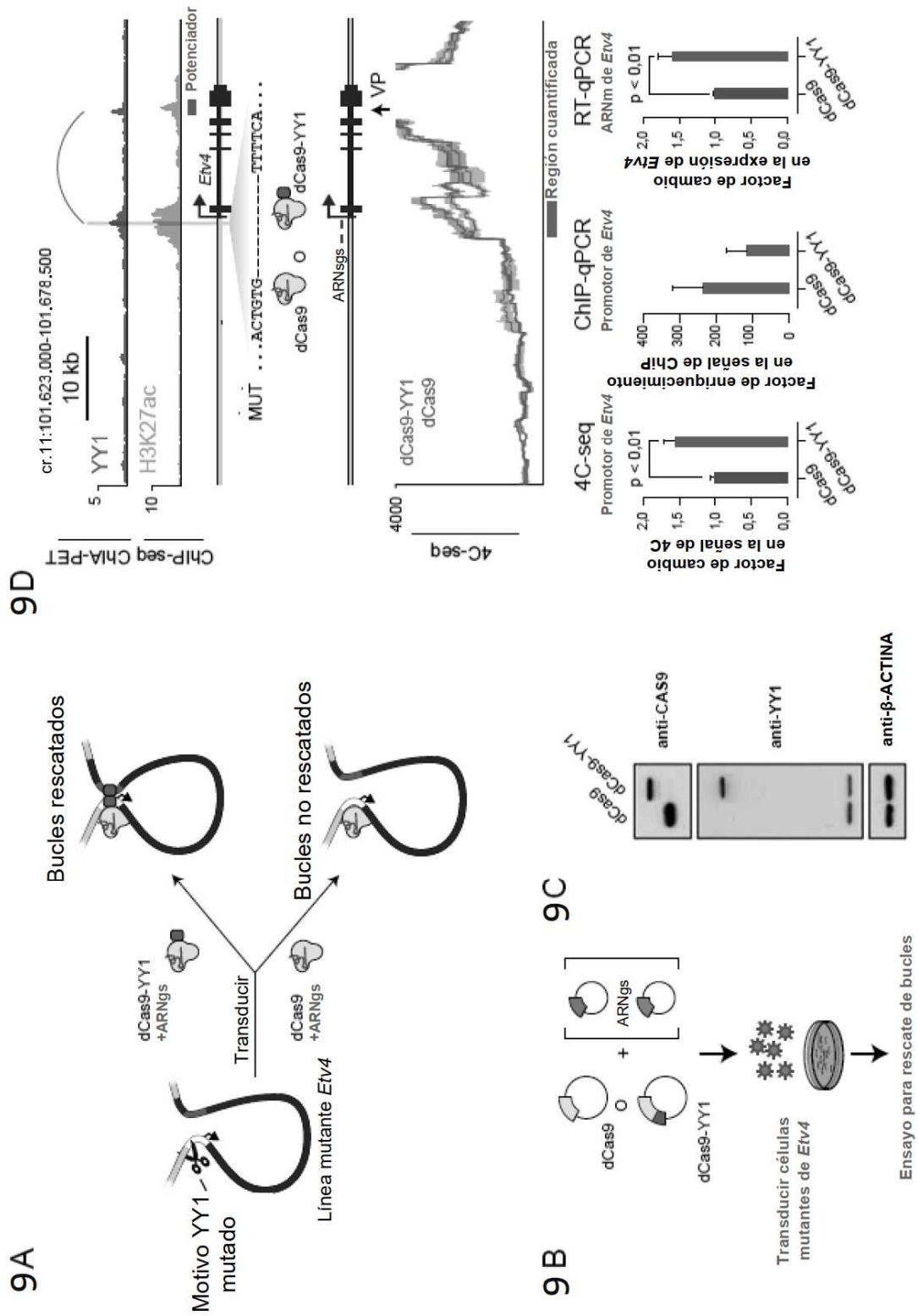
FIGS. 8A-8C



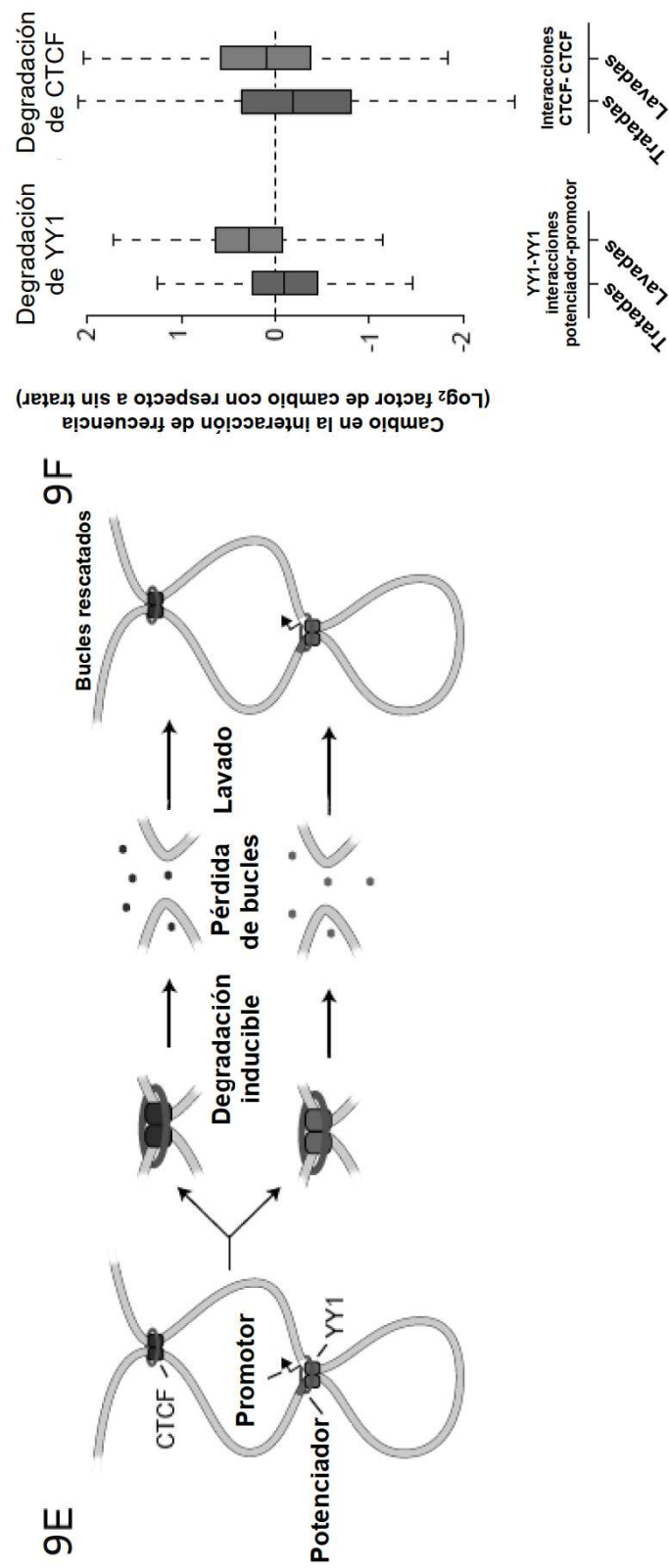
FIGS. 8D-8E



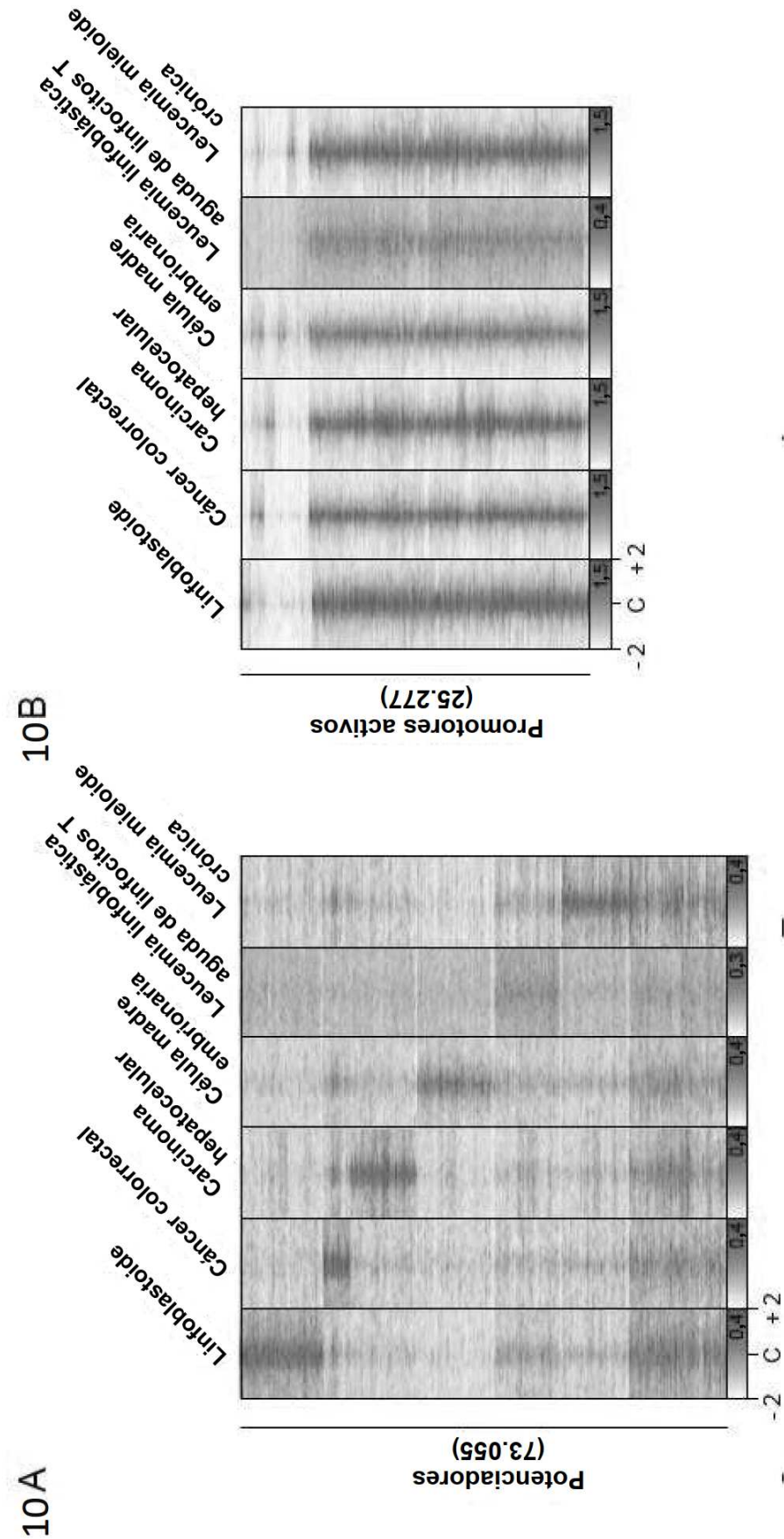
FIGS. 9A-9D



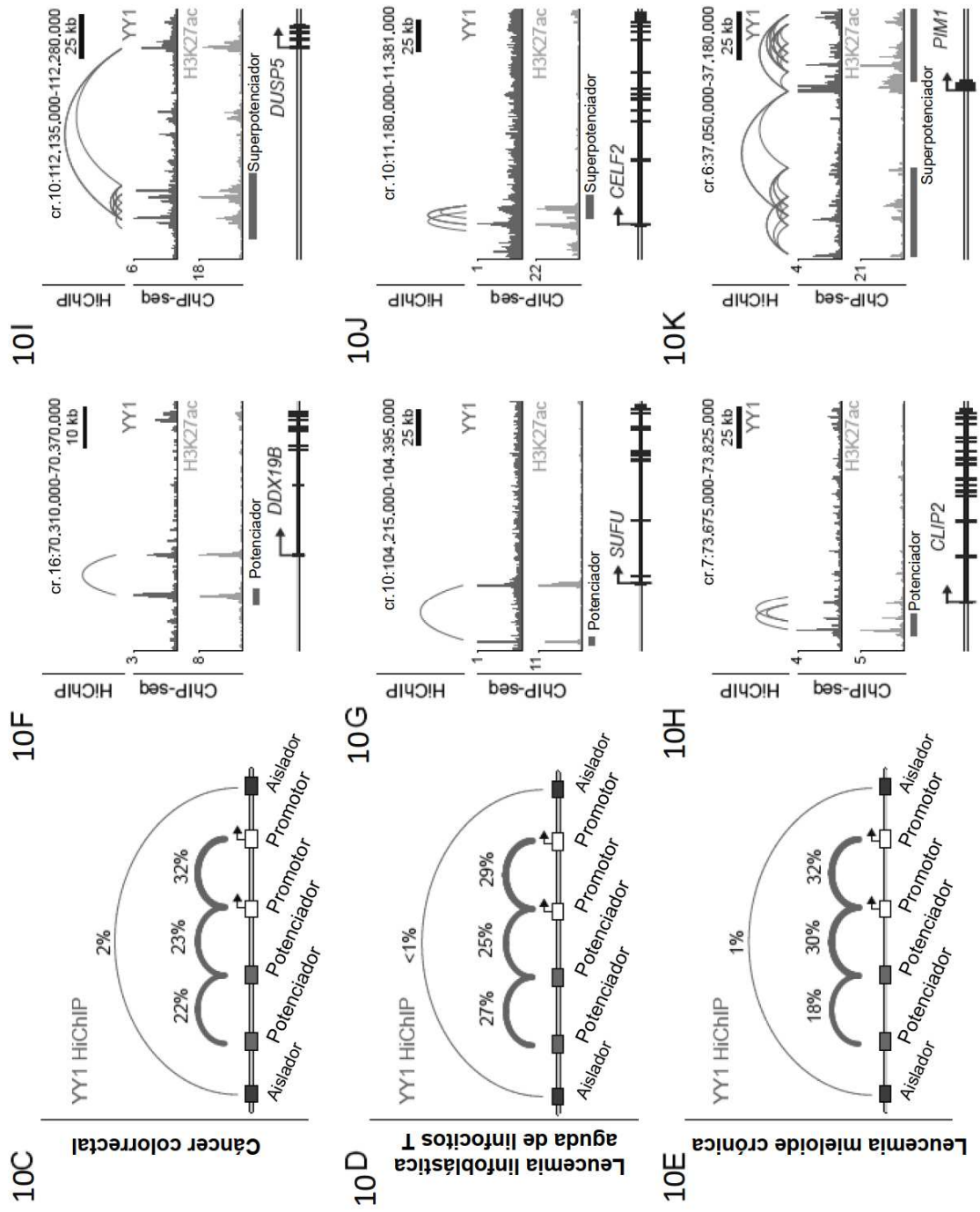
FIGS. 9E-9F



FIGS. 10A-10B



FIGS. 10C-10K



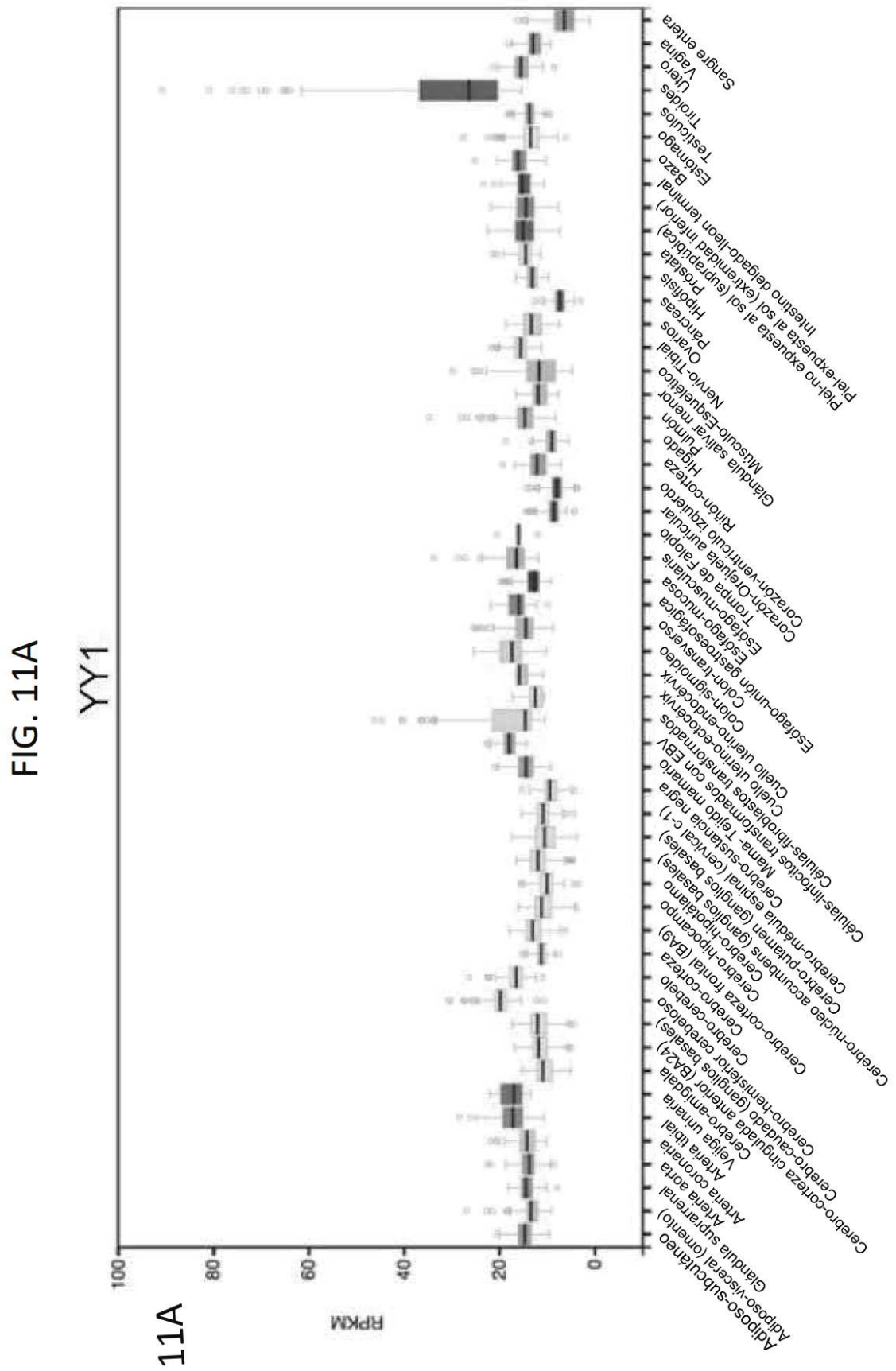


FIG. 11B

CTCF

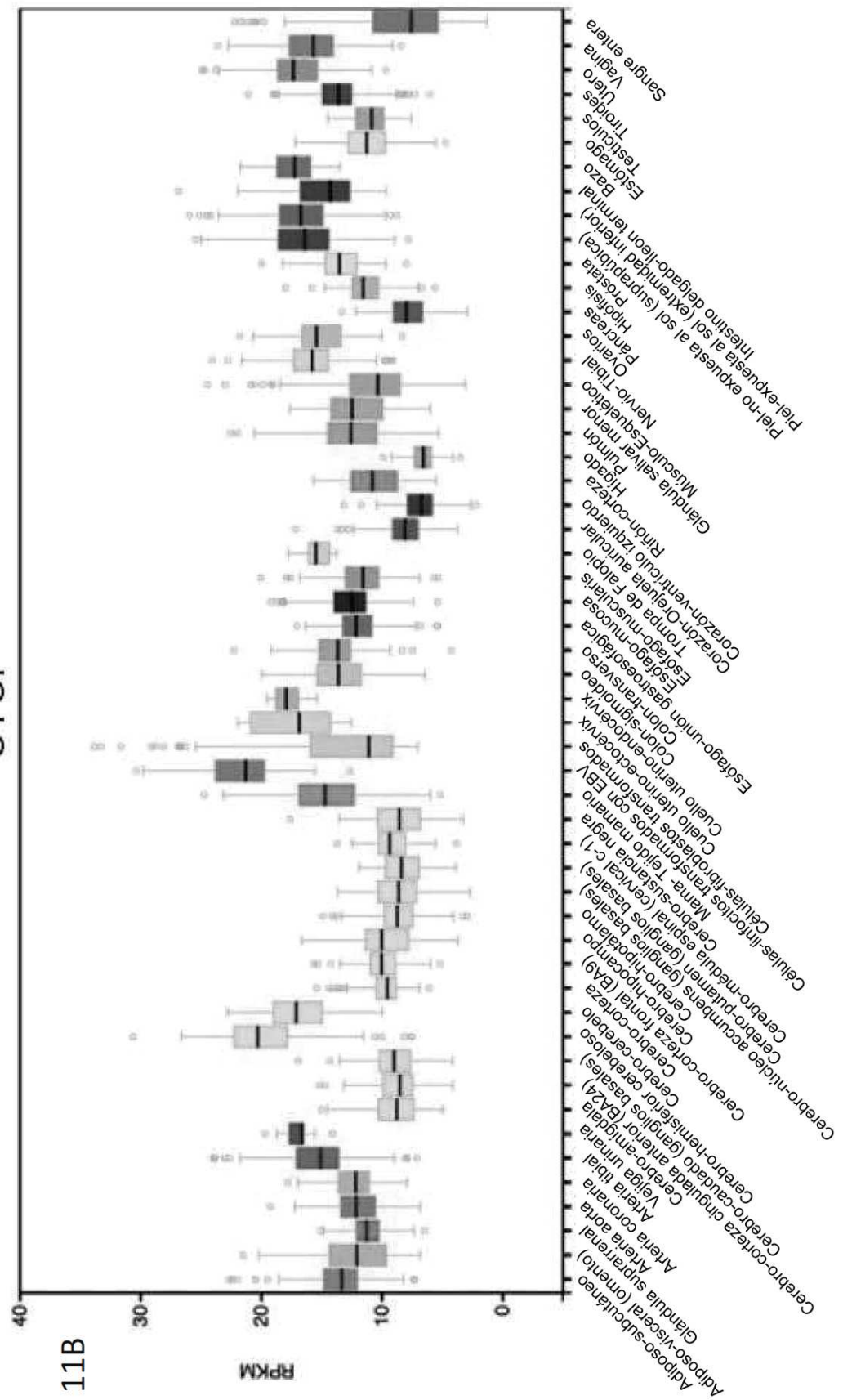


FIG. 11C

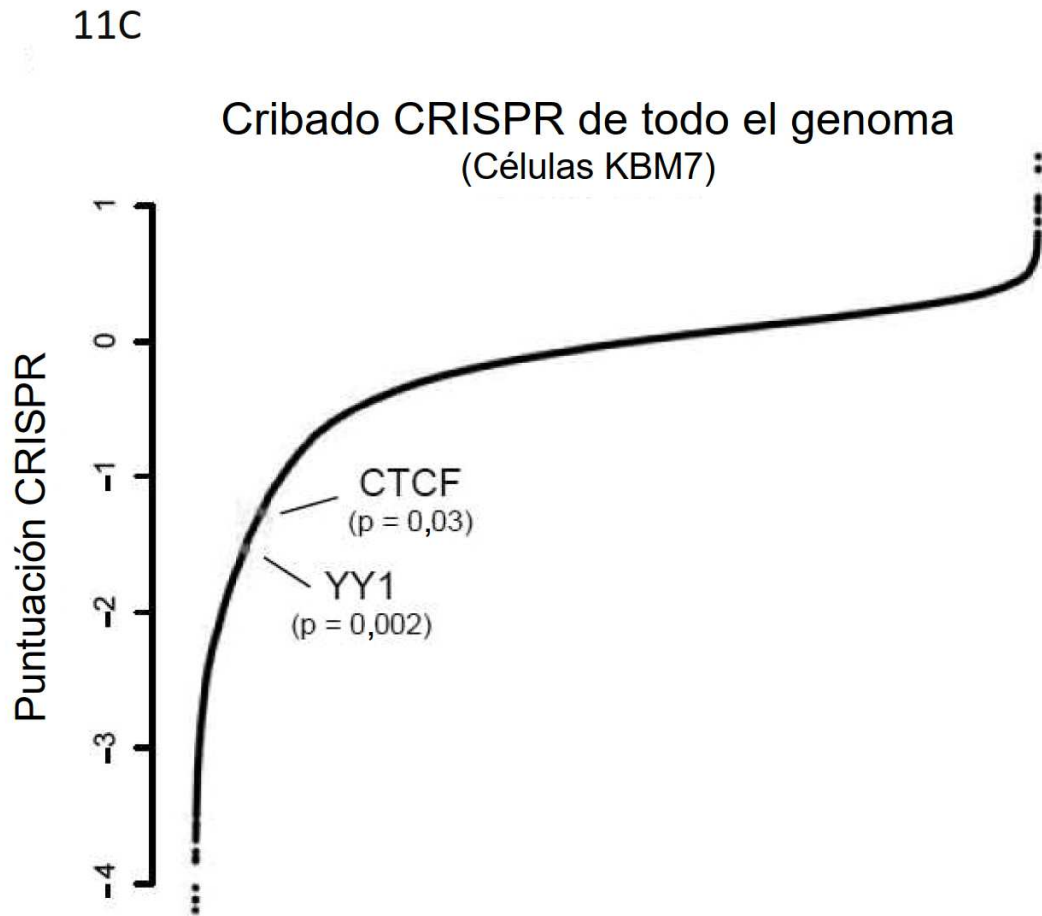
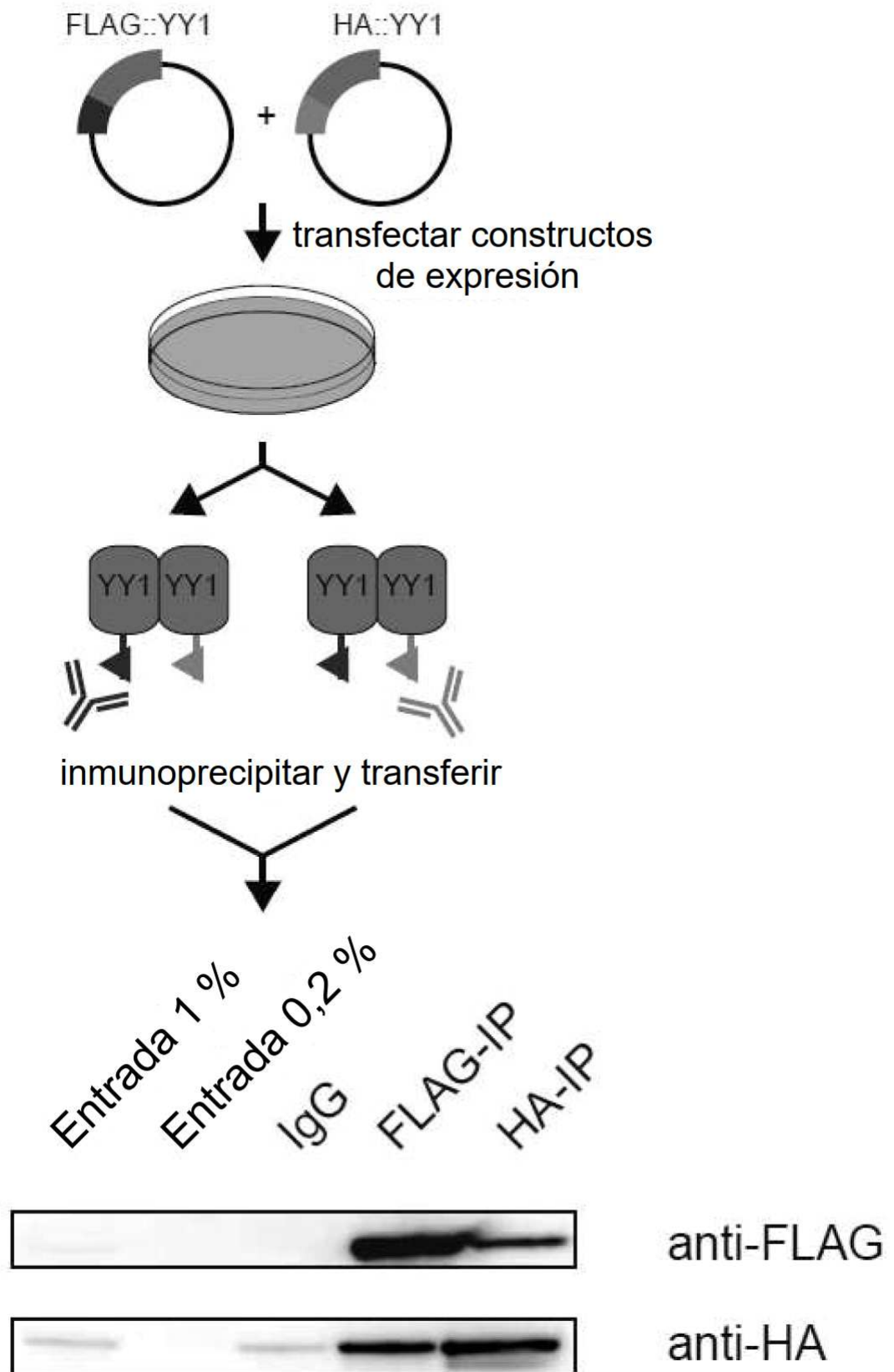
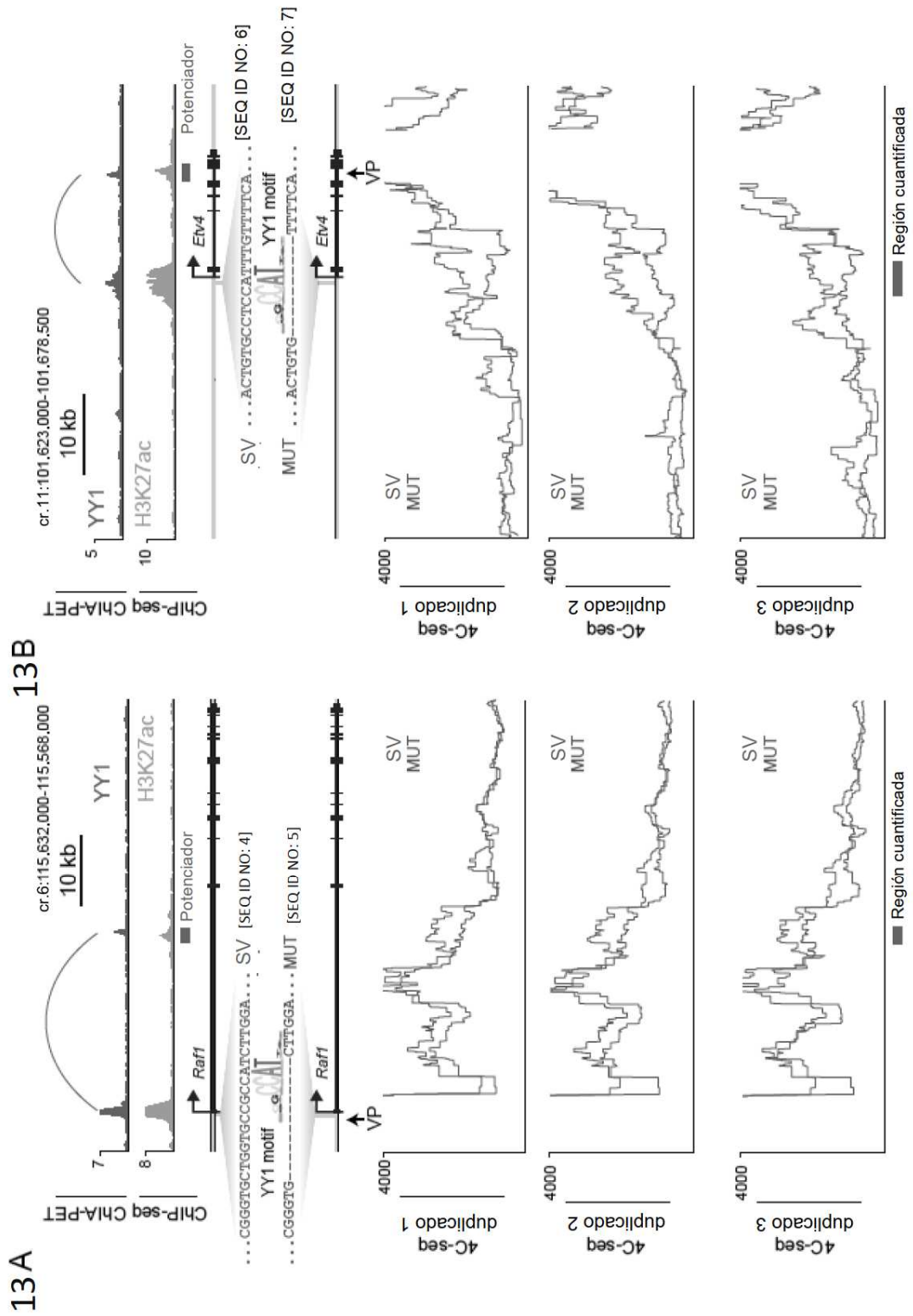


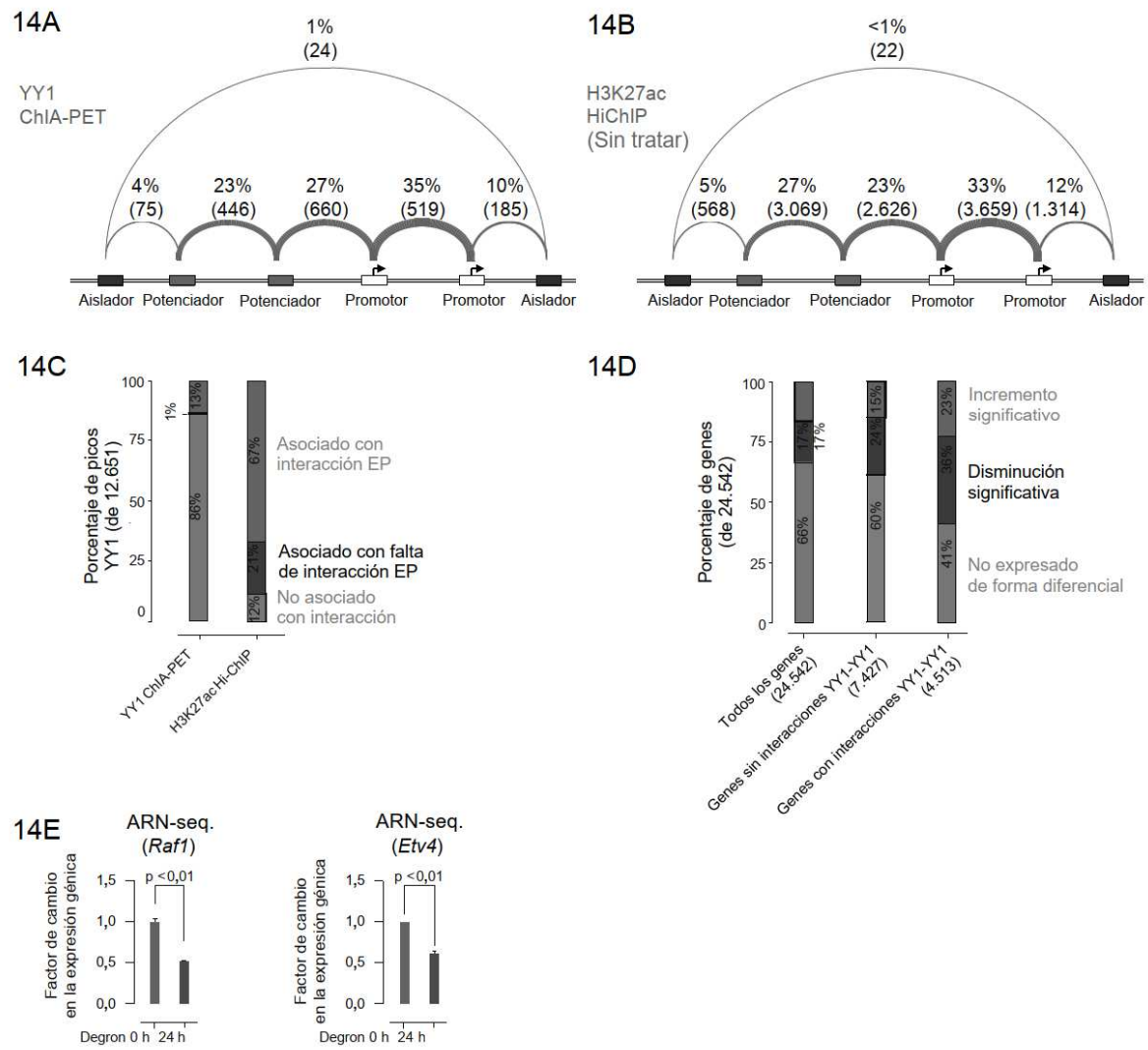
FIG. 12



FIGS. 13A-13B



FIGS. 14A-14E



FIGS. 15A-15C

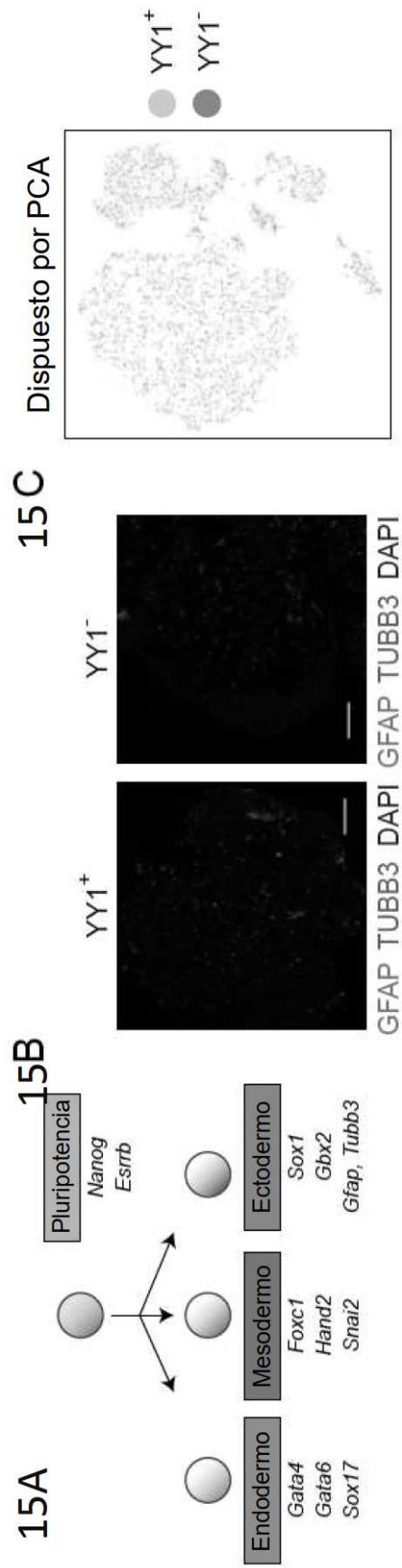
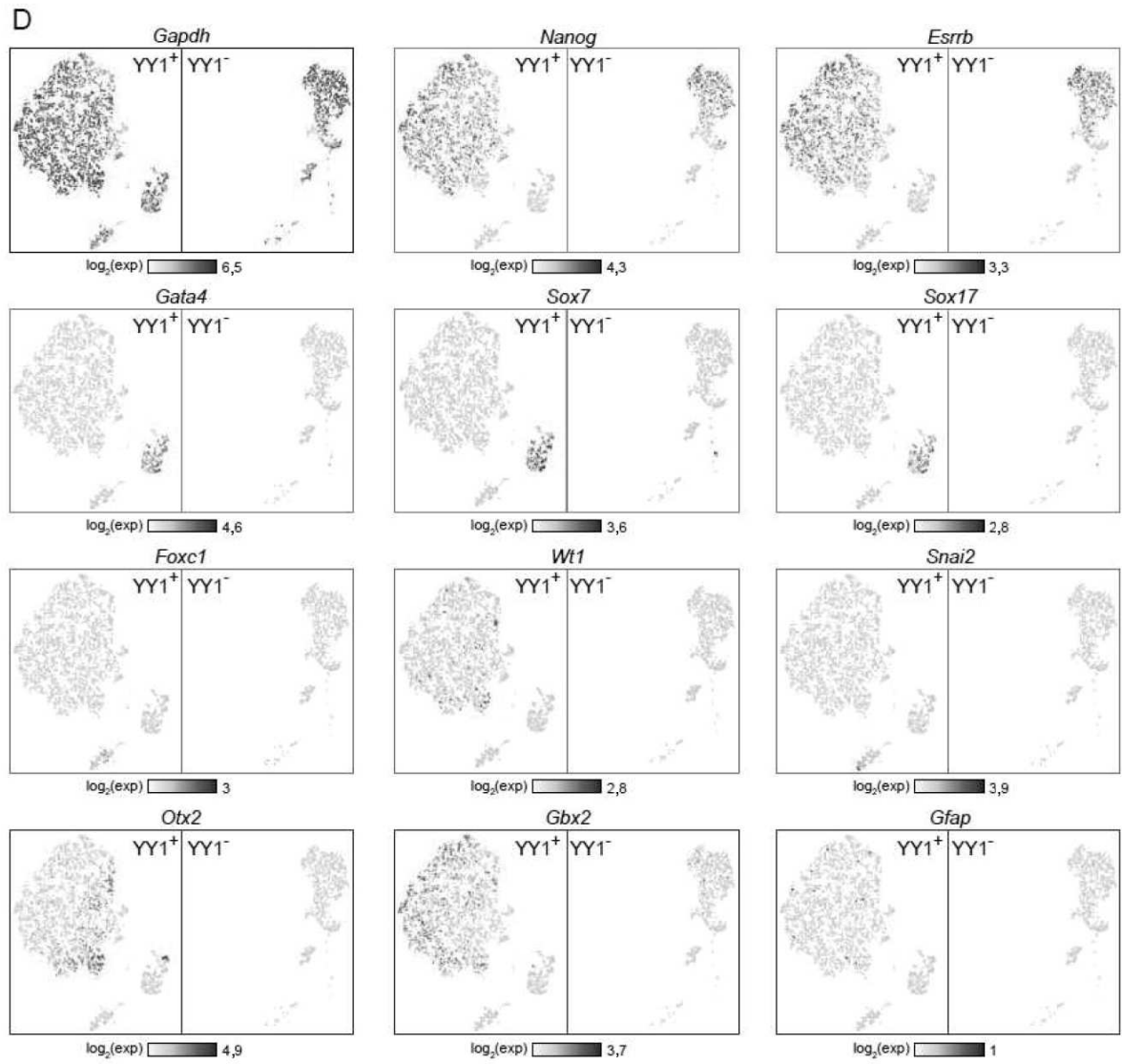
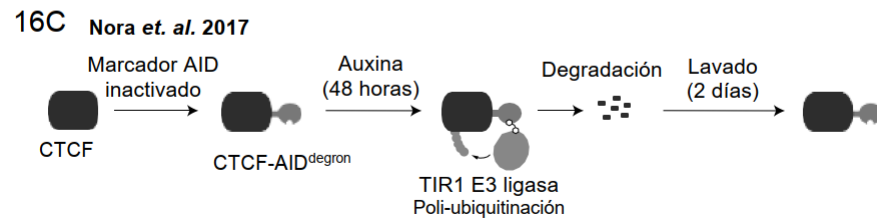
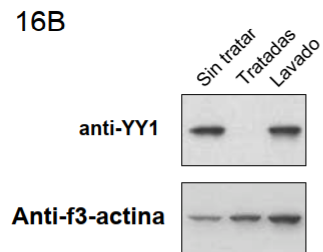
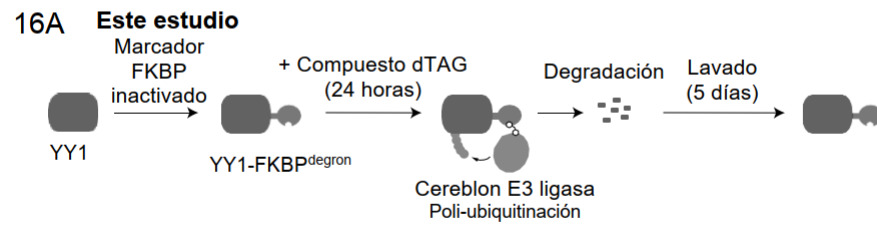


FIG. 15D

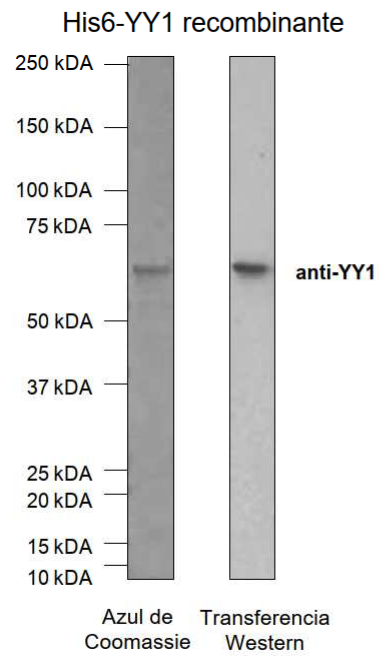


FIGS. 16A-16C

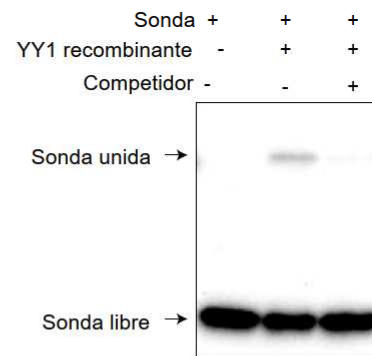


FIGS. 17A-17B

17A



17B



FIGS. 18A-18C

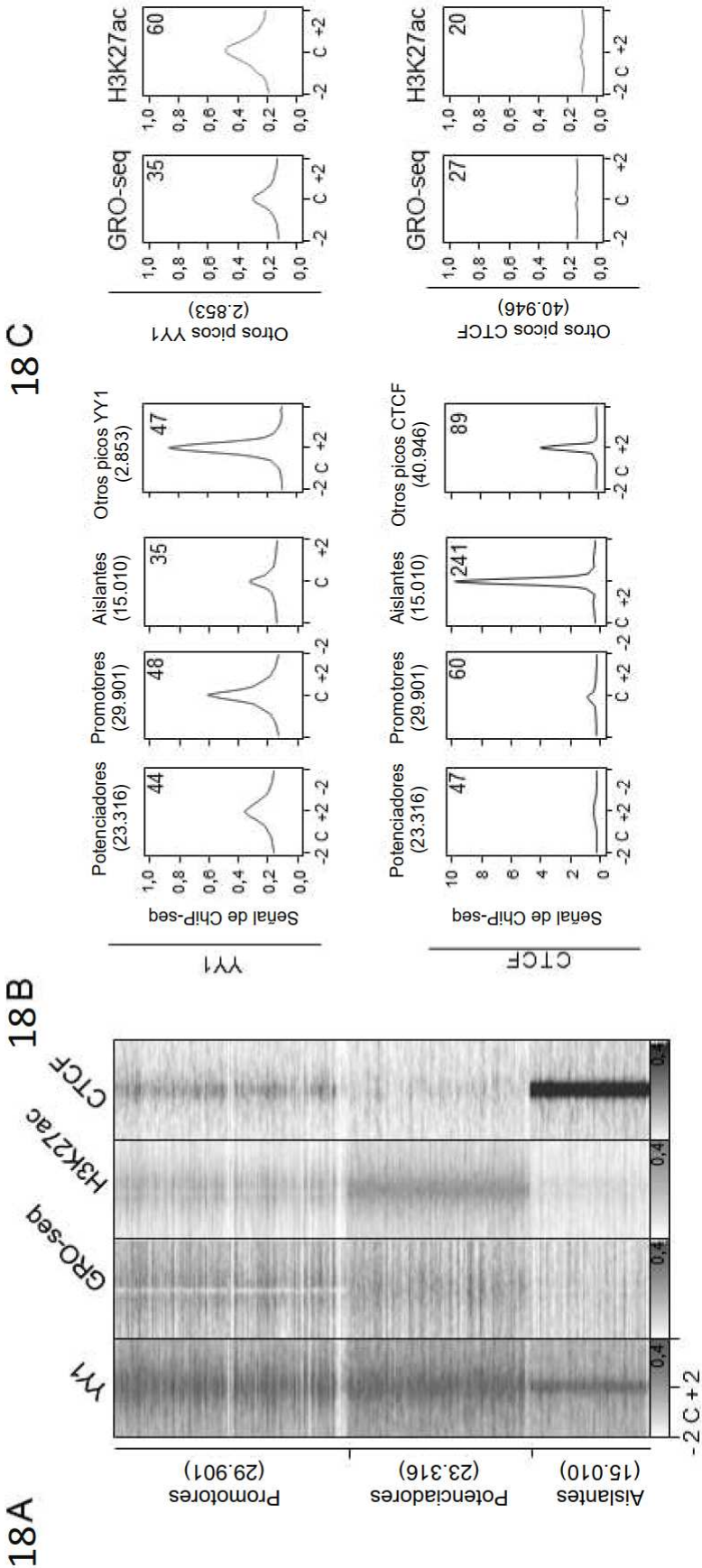
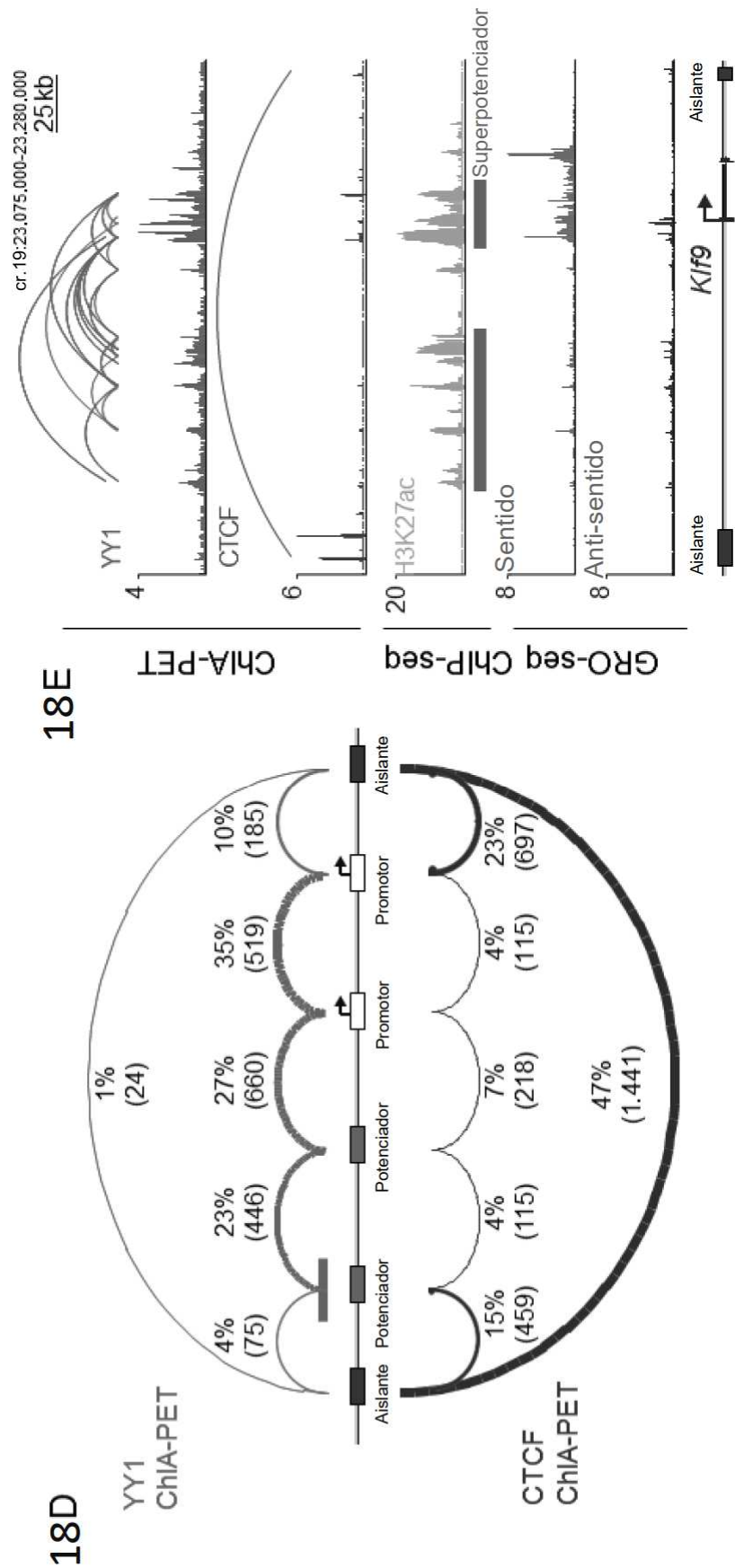


FIG. 18D-18E



FIGS. 18F-18J

