

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 693 349**

51 Int. Cl.:

C12P 19/34 (2006.01)

C40B 40/02 (2006.01)

C40B 50/06 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **06.06.2013 PCT/US2013/044537**

87 Fecha y número de publicación internacional: **12.12.2013 WO13184930**

96 Fecha de presentación y número de la solicitud europea: **06.06.2013 E 13800410 (6)**

97 Fecha y número de publicación de la concesión europea: **01.08.2018 EP 2859139**

54 Título: **Aplicación de códigos de barras de ADN de bancos de matrices de cromatinas y mononucleosomas diseñadores para la creación de perfiles de lectores, escritores, borradores y moduladores de cromatina de los mismos**

30 Prioridad:

06.06.2012 US 201261656233 P
10.10.2012 US 201261712148 P

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

11.12.2018

73 Titular/es:

THE TRUSTEES OF PRINCETON UNIVERSITY
(100.0%)
4 New South Building, P.O.Box 36
Princeton, NJ 08544, US

72 Inventor/es:

MUIR, TOM W.;
NGUYEN, UYEN T. T. y
MUELLER, MANUEL M.

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 693 349 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

Aplicación de códigos de barras de ADN de bancos de matrices de cromatinas y mononucleosomas diseñadores para la creación de perfiles de lectores, escritores, borradores y moduladores de cromatina de los mismos

Esta solicitud reivindica el beneficio de la fecha de presentación de la solicitud provisional de EE. UU. 61/656.233, presentada el 6 de junio de 2012, y de la solicitud provisional de EE. UU. 61/712.148, presentada el 10 de octubre de 2012.

LISTADO DE SECUENCIAS

La presente solicitud contiene un Listado de secuencias que se ha enviado en formato ASCII por medio de EFS-Web. Dicha copia ASCII, creada el 5 de junio de 2013, se denomina 32108-348453_SL.txt y tiene 14.653 bytes de tamaño.

INFORMACIÓN SOBRE ANTECEDENTES

En las células eucariotas, el ADN se empaqueta junto con proteínas histonas en un complejo de nucleoproteínas denominado cromatina. Las unidades de repetición mínima de cromatina son los nucleosomas, que permiten el plegamiento de la cromatina en fibras y estructuras de orden superior. La regulación génica en el nivel de cromatina ("epigenética") se logra por naturaleza a través de modificaciones químicas dinámicas ("marcas") tanto de ADN como de histonas, mediadas por enzimas "escritoras de cromatina" y "borradoras de cromatina" especializadas (denominadas conjuntamente "modificadores de cromatina"). Los 'modificadores de histonas' son proteínas que unen ('escritores de histonas') o retiran ('borradores de histonas') una o más marcas a o de proteínas histonas, respectivamente. Los 'modificadores de ADN' son proteínas que unen ('escritores de ADN') o retiran ('borradores de ADN') una o más marcas a o de ADN, respectivamente. Los ejemplos incluyen las histona desacetilasas (HDAC) e histona metiltransferasas (HMT) farmacológicamente relevantes. En combinación, estas modificaciones forman patrones locales (dentro de la fibra de cromatina, dentro de un solo nucleosoma, y/o dentro de una sola histona), que se cree que sirven como plataformas de reclutamiento para factores de proteínas con módulos especializados que reconocen distintas marcas ('lectores de cromatina' o 'interactuantes de cromatina'). Los 'lectores de histona' o 'interactuantes de histona' son proteínas que reconocen, o se unen a, una o más marcas en las proteínas histonas, respectivamente. Los 'lectores de ADN' o 'interactuantes de ADN' son proteínas que reconocen, o se unen a, una o más marcas en el ADN, respectivamente. Las marcas de ADN e histona son importantes en

el desarrollo y la diferenciación celular, y, en consecuencia, las modificaciones anómalas y lecturas combinatorias alteradas están implicadas en enfermedades humanas, en especial cáncer. Como consecuencia, la biología de la cromatina y la epigenética se han vuelto el centro de muchas iniciativas de investigación a nivel universitario y en la industria farmacéutica. Y aún así, existe una discrepancia rápidamente creciente entre la cantidad de información que se genera por los enfoques epigenómicos y proteómicos descendentes y la capacidad para completar sistemáticamente los detalles moleculares de la bioquímica de la cromatina asociada. A pesar de la información genómica e información proteómica en expansión sobre las secuencias de histonas, variaciones, y tipos y abundancia de sus modificaciones naturales, y algunas enzimas responsables de las modificaciones, el conocimiento de los mecanismos epigenéticos altamente complejos permanece incompleto, y existe una falta de herramientas bioquímicas eficaces.

Los patrones de modificación postraduccional anómalos sobre proteínas histonas así como los encontrados en bases de ADN se encuentran a menudo en enfermedades. Existe una necesidad de entender, ensayar y manipular los mecanismos subyacentes como prerrequisito para el diseño racional de fármacos epigenéticos de próxima generación.

Myrriah S. Chavez *et al.*, Epigenetics & Chromatin, Biomed, Central Ltd., London UK, vol. 5, n.º 1, 27 de abril de 2012, página 5, divulga un estudio sobre la flexibilidad conformacional del extremo C de las variantes de histona H4.

El documento US 5 972 608 A divulga matrices de nucleosomas que se pueden usar en matrices que acoplan la actividad de una enzima remodeladora de cromatina (o complejo enzimático) con la actividad de endonucleasa de restricción para medir cuantitativamente la extensión de la actividad "remodeladora" presentada por una enzima remodeladora o complejo enzimático particular. Las matrices de nucleosomas se pueden usar en procedimientos para cribar los moduladores candidatos, es decir, inhibidores o potenciadores de enzimas remodeladoras de cromatina o complejos enzimáticos.

Dai, J. *et al.*, Cell, vol. 134, n.º 6, 19 de septiembre de 2008, páginas 1066-1078, divulga un banco de mutantes de sustitución y delección de histona H3 y H4 sistemáticos generados en *Saccharomyces cerevisiae*; cada mutante de histona se marcó con códigos de barras moleculares únicos.

Ruthenburg, A. J. *et al.*, Cell, vol. 145, n.º 5, 19 de mayo de 2011, páginas 692-706 ilustra cómo la realización de patrones nucleosómicos por modificación de histonas puede ayudar a entender mejor las asociaciones de cromatina críticas.

5 DESCRIPCIÓN DE LAS FIGURAS

Figura 1: banco de MN codificados por ADN ejemplares. En esta versión particular, se aplica el código de barras a cada miembro del banco de MN en el extremo 5' del ADN nucleosómico y contiene una reacción en cadena de la polimerasa (PCR) directa (FW) y un sitio de cebado de secuenciación para la lectura de secuenciación de ADN posterior. El código de barras de MN codifica las variantes de histona y/o variante de ADN incorporadas en el respectivo MN y/u otras proteínas histonas y distintas de histonas. En general, se pueden incorporar uno o varios códigos de barras en cualquier sitio dentro de, en, o cerca del extremo 5' o 3' del ADN. Blanco: octámero de histona; gris: ADN nucleosómico; protuberancias negras: colas de histona N terminales. Las banderas representan diferentes modificaciones (mostradas aquí: modificaciones en las histonas; las modificaciones se pueden incorporar además, o en su lugar, dentro de la secuencia de ADN). Los diferentes códigos de barras representan diferentes secuencias de ADN que codifican la respectiva variante de MN. H: histona. MN: mononucleosoma. FW: sitio de cebado de secuenciación directa.

Figura 2: banco de CA codificadas por ADN ejemplares que consiste en tri-MN. En esta versión particular, se aplica el código de barras a cada miembro del banco de MN en el extremo 5' del ADN de matriz y contiene una PCR y un sitio de cebado de secuenciación para la lectura de secuenciación de ADN posterior. El código de barras de CA codifica las variantes de histona y/o variantes de ADN, extensiones de ADN, histonas conectoras, y/u otras proteínas distintas de histonas, incorporadas en la respectiva CA. En general, se pueden incorporar uno o varios códigos de barras en cualquier sitio dentro de, en, o cerca del extremo 5' o 3' del ADN. Blanco: octámero de histona; gris: ADN de matriz; protuberancias negras: colas de histona N terminales. Las banderas representan diferentes modificaciones (histona o ADN) como en la figura 1. Los diferentes códigos de barras representan diferentes secuencias de ADN que codifican la respectiva variante de CA. H: histona. MN: mononucleosoma. CA: matriz de cromatina. FW: sitio de cebado de secuenciación directa.

Figura 3: ejemplo para el uso de un banco de MN con código de barras de ADN para la creación de perfiles de interactuantes/modificadores de cromatina. I. En esta versión particular, se aplica el código de barras a cada miembro del banco de MN en el extremo 5' del ADN nucleosómico. El código de barras de MN codifica las variantes de histona y/o variantes de ADN. II. Incubación del/de los interactuante(s) o modificador(es) de cromatina con el banco (recombinante o derivado de extractos celulares nucleares). III. Aislamiento de los sustratos de MN unidos o modificados, tales como por precipitación "pull-down" de anticuerpo. IV. Aislamiento de secuencias de ADN con código de barras usando digestión de proteínas y purificación de ADN. V. Análisis de datos: identificación y cuantificación de ADN después de normalización de entrada. La descodificación del código de barras se puede lograr por hibridación de micromatrices de ADN o bien por secuenciación de ADN después de la adición de sitios de cebado de secuenciación FW y RV, tal como por PCR o fijación de ADN T4 (nota: opcionalmente, el sitio de cebado FW y/o RV también se puede incluir en el ADN nucleosómico antes de la formación de MN). Blanco: octámero de histona; gris: ADN nucleosómico; protuberancias negras: colas de histona N terminales. Las banderas representan diferentes modificaciones mostradas en histonas pero potencialmente también en ADN. Los diferentes códigos de barras representan diferentes secuencias de ADN que codifican la respectiva variante de MN. MN: mononucleosomas. FW: sitio de cebado de secuenciación directa. RV: sitio de cebado de secuenciación inversa.

Figura 4: ejemplos para aplicar códigos de barras en las etapas de manipulación realizadas en un banco de MN. I. En esta versión particular, se aplica el código de barras a cada miembro del banco de MN en el extremo 5' del ADN nucleosómico. El código de barras de MN codifica las variantes de histona y/o variantes de ADN. II. El banco se divide y se somete a diversas manipulaciones bioquímicas o biofísicas, aquí para las manipulaciones ejemplares 1.1 y 1.2. III. Aislamiento de MN manipulados con éxito, tal como por precipitación de anticuerpo. IVa. Preparación de banco de ADN para secuenciación de ADN: Aislamiento de ADN con código de barras usando digestión de proteínas y purificación de ADN, seguido de adición de código de barras de experimento (múltiple) (aquí: MP1.1 y MP1.2) así como sitios de cebado de secuenciación FW y RV por PCR (opcionalmente, el sitio de cebado FW y/o RV también se puede incluir en el ADN nucleosómico antes de la formación de MN). IVb. Adición de códigos de barras de experimento (aquí: MP1 y MP2), por ejemplo al extremo nucleosómico 5' del ADN. Agrupamiento de MN con doble código de barras. División del banco y repetición de las etapas II-IVb como se desee (mostrado aquí: 3 rondas de manipulación con 2 manipulaciones diferentes cada una). Aislamiento de ADN. Adición de sitios de cebado de secuenciación FW y RV (nota: el sitio de cebado FW y/o RV también se puede incluir en el ADN nucleosómico antes de la formación de MN). V. Análisis de datos: Identificación y cuantificación de ADN. Blanco: octámero de histona; gris: ADN nucleosómico; protuberancias negras: colas de histona N terminales. Las banderas representan diferentes modificaciones como en las figuras anteriores. Los

diferentes códigos de barras representan diferentes secuencias de ADN que codifican la respectiva variante de MN. MN: mononucleosomas. MP: código de barras múltiple (experimento). FW: sitio de cebado de secuenciación directa. RV: sitio de cebado de secuenciación inversa.

5 Figura 5: ejemplo para el uso de un banco de MN con código de barras y multiplexado para la creación de
perfiles de interactuantes de cromatina. I. En esta versión particular, se aplica el código de barras a cada
miembro del banco de MN en el extremo 5' del ADN nucleosómico. El código de barras de MN codifica las
variantes de histona y/o variantes de ADN. II. El banco se divide y se incuba con diversos lectores de
10 cromatina, tales como un lector con uno (a) o múltiples (b, c) módulos lectores dentro de una (b) o en
diferentes (c) cadena(s) polipeptídica(s) (recombinante o derivado de extractos celulares nucleares). III.
Aislamiento de MN unidos, tal como por precipitación de anticuerpo. IV. Preparación de banco de ADN para
secuenciación de ADN: IVa. Aislamiento de ADN con código de barras usando digestión de proteínas y
purificación de ADN. IVb. Adición de código de barras de experimento (múltiple) así como sitios de cebado
15 de secuenciación FW y RV por PCR (nota: el sitio de cebado FW y/o RV también se puede incluir en el
ADN nucleosómico antes de la formación de MN). IVc. Agrupamiento de secuencias de ADN con doble
código de barras. V. Secuenciación de ADN para descodificar códigos de barras de MN y múltiples. VI.
Análisis de datos para identificación y cuantificación de ADN. Blanco: octámero de histona; gris: ADN
nucleosómico; protuberancias negras: colas de histona N terminales. Las banderas representan diferentes
20 modificaciones como en las figuras anteriores. Los diferentes códigos de barras representan diferentes
secuencias de ADN que codifican la respectiva variante de MN. MN: mononucleosomas. MP: código de
barras múltiple (experimento). FW: sitio de cebado de secuenciación directa. RV: sitio de cebado de
secuenciación inversa.

25 Figura 6: ejemplo para el uso de un banco de MN con código de barras y multiplexado para la creación de
perfiles de modificadores de cromatina. I. En esta versión particular, se aplica el código de barras a cada
miembro del banco de MN en el extremo 5' del ADN nucleosómico. El código de barras de MN codifica las
variantes de histona y/o variantes de ADN. II. El banco se divide y se incuba con diversos modificadores de
cromatina, por ejemplo que contienen el dominio catalítico (a), la enzima de longitud completa (b), o una
30 enzima incluida en un complejo multisubunitario grande (c) (recombinante o derivado de extractos celulares
nucleares). III. Aislamiento de MN modificados, tal como por precipitación de anticuerpo frente a (a) la
marca unida o (b) la marca retirada (en este caso, la precipitación se realiza para sustraer los no sustratos).
Todas las etapas adicionales son equivalentes al procedimiento descrito en la fig. 5. Blanco: octámero de
histona; gris: ADN nucleosómico; protuberancias negras: colas de histona N terminales. Las banderas
representan diferentes modificaciones mostradas en histonas pero potencialmente también en ADN, como
35 en las figuras anteriores. Los diferentes códigos de barras representan diferentes secuencias de ADN que
codifican la respectiva variante de MN. H: histona. MN: mononucleosomas. MP: código de barras múltiple
(experimento).

40 Figura 7: ejemplo para el uso de un banco de MN con código de barras y multiplexado para la creación de
perfiles de estabilidad de MN. I. En esta versión particular, se aplica el código de barras a cada miembro del
banco de MN en el extremo 5' del ADN nucleosómico. El código de barras de MN codifica las variantes de
histona y/o variantes de ADN. II. El banco se divide y se expone a diversas condiciones experimentales que
desestabilizan los MN, tales como incremento en concentraciones salinas o exposición a extracto celular
45 nuclear que contiene complejos remodeladores de cromatina. III. Aislamiento de ADN nucleosómico con
código de barras liberado, tal como por purificación de gel de agarosa. Todas las etapas adicionales son
equivalentes al procedimiento descrito en la fig. 5. Blanco: octámero de histona; gris: ADN de matriz;
protuberancias negras: colas de histona N terminales. Las banderas representan diferentes modificaciones
como en las figuras anteriores. Los diferentes códigos de barras representan diferentes secuencias de ADN
que codifican la respectiva variante de MN. H: histona. MN: mononucleosomas. MP: código de barras
50 múltiple (experimento). FW: sitio de cebado de secuenciación directa. RV: sitio de cebado de secuenciación
inversa.

55 Figura 8: ejemplo para el uso de un banco de MN con código de barras para la creación de perfiles de
moléculas que modulan la actividad o función de interactuantes y modificadores de cromatina. I. En esta
versión particular, solo se muestra una variante de MN, a la que se le aplica el código de barras en el
extremo 5' del ADN nucleosómico. El código de barras de MN codifica las variantes de histona y/o variantes
de ADN. II. El banco se divide y se expone a diversas moléculas que modulan la función o actividad de los
interactuantes o modificadores de cromatina (recombinante o derivado de extractos celulares nucleares). III.
Aislamiento de resultados positivos, tal como por precipitación frente al lector, la marca unida (escritor), o
60 marca retirada (borrador). IV. Adición de código de barras de inhibidor (por ejemplo por fijación de ADN). V.
Generación del banco de ADN: adición de cebadores FW y RV (por fijación de ADN y/o PCR). Todas las
etapas adicionales son equivalentes al procedimiento descrito en la fig. 5. Blanco: octámero de histona;
gris: ADN nucleosómico; protuberancias negras: colas de histona N terminales. Las banderas representan
diferentes modificaciones como en las figuras anteriores. Los diferentes códigos de barras representan
65 diferentes secuencias de ADN que codifican la respectiva variante de MN. H: histona. MN:

mononucleosomas. IBC: código de barras de inhibidor. FW: sitio de cebado de secuenciación directa. RV: sitio de cebado de secuenciación inversa.

Figura 9: resumen de semisíntesis de histonas y ensamblaje de nucleosomas. (A) Representación esquemática de la estrategia semisintética para la preparación de histonas modificadas, en este caso particular, histonas modificadas de forma N-terminal. (B) Las histonas recombinantes wt o semisintéticas modificadas se combinan en proporciones equimolares y se dializan a partir de desnaturalizantes en alto contenido salino ('dialisis 1'). Sin purificación adicional, se añade una mezcla de BC-601 (0,6 eq con respecto al octámero de histona) y ADN con tampón con MMTV biotinilado (0,4 eq, con respecto al octámero de histona; 'BIO-MMTV') ('dialisis 2', de alto a bajo contenido salino). La purificación de los nucleosomas deseados se logra por precipitación de estreptavidina (SAP) para retirar los complejos ADN-histona formados de forma improductiva. MN: mononucleosoma; NT: extremo N; BIO: biotina; SAP: purificación por afinidad con estreptavidina; PTM: modificación postraduccional. La secuencia de BIO-MMTV en parte inferior de la figura es SEQ ID NO: 13.

Figura 10: generación de secuencias de ADN con doble código de barras multiplexadas para la secuenciación de próxima generación. (A) Generación de moléculas de ADN 601 nucleosómico con código de barras ejemplares para la formación de MN (longitud en este ejemplo: 190 pb) por PCR. (B) La estrategia de fijación de ADN T4 usando salientes no palindrómicos apropiados permite una preparación flexible de la secuencia de ADN 601 nucleosómico con código de barras. Se hibridaron el ADN monocatenario complementario que abarca nt 10-30 del sitio de cebado Ion Torrent directo ('FW-iT₁₀₋₃₀') y los respectivos códigos de barras de MN de 6 pb, se combinaron con una secuencia 601 nucleosómica digerida por Bsal-DrallI, se fosforiló *in situ* usando ADN T4 polinucleótido cinasa (PNK), y se fijó usando ADN ligasa T4. Se unió un saliente AA en el extremo 3' de la hebra inferior para evitar la autofijación en el extremo romo del molde de ADN 'BC-601'. (C) Multiplexación de experimento de ADN. Se muestra un ejemplo para moldes de ADN con código de barras que se originan de 4 variantes de MN diferentes y dos experimentos (EXP1 y EXP2). La multiplexación de PCR para codificar el origen experimental de los MN se logra usando cebadores inversos apropiados (que contiene un sitio de hibridación 601 común, '601-RV', códigos de barras de experimento de multiplexación 'BC-EXP1' o 'BC-EXP2', así como el adaptador inverso IonTorrent® 'RV-iT'). Las moléculas de ADN con doble código de barras se agrupan y se analizan usando el secuenciador de próxima generación IonTorrent®. Para la descodificación y normalización, véase (D). (D) Resumen del análisis de datos después de la secuenciación con IonTorrent®. En este ejemplo, 4 MN diferentes se sometieron a 3 experimentos (EXP1-3), siendo EXP2 la secuenciación de la entrada del banco. En primer lugar, se clasifican las lecturas de secuenciación brutas de acuerdo con sus códigos de barras experimentales, seguido de clasificación de acuerdo con los códigos de barras de MN. Las lecturas clasificadas se normalizan posteriormente frente a la entrada secuenciada para corregir diferencias en las cantidades iniciales de cada MN individual (intermedio). los datos normalizados finales se presentan como % entrada. eq: equivalente; FW: directo; RV: inverso; iT: IonTorrent®; MN: mononucleosoma; BC: código de barras; EXP: experimento; IP: entrada; ss: monocatenario; ds: bicatenario. Las secuencias en la parte inferior de la figura son, leyendo de la parte superior a la inferior, SEQ ID NO: 14, 10, 15, 11 y 16-18, respectivamente.

Figura 11: el banco de nucleosomas codificado permite investigaciones rápidas y detalladas de cientos de interacciones nucleosoma-cromatina reguladoras en poco tiempo. (a) Un banco de nucleosomas, que contiene diferentes patrones de PTM que se codifican en un código de barras de 6 pb ('BC-MN') añadido en el ADN 601 nucleosómico ('BC-601'), se somete a escritor(es), borrador(es), y/o lector(es) de histonas o una combinación de los mismos en forma de un extracto celular nuclear en múltiples experimentos separados (mostrados aquí: EXP1 y EXP2). Usando experimentos de precipitación, el ADN nucleosómico con código de barras de los mejores fijadores se aísla, se codifica además con los códigos de barras de experimento de multiplexación BC-EXP1 y BC-EXP2, se agrupa, se descodifica usando el secuenciador de próxima generación IonTorrent® y se normaliza frente a la entrada. Para detalles sobre diseño y preparación de ADN, véase la fig. 10A-C. (B) El flujo de trabajo simplificado permite el equivalente de miles de experimentos (con un tamaño de banco de 39 sometidos para 120 experimentos de banco) en una semana, comenzando a partir de histonas disponibles.

Figura 12: (A) Representación diagramática de las variantes de MN 1-39. (B) Listado de la composición de histona de las variantes de MN 1-39 y su correspondiente código de barras de ADN nucleosómico. Todas las histonas wt se prepararon de forma recombinante en *E.coli*. Las proteínas histonas semisintéticas se prepararon por NCL (fig. 9a). Las proteínas H4 semisintéticas contenían un grupo acetilo N terminal adicional (no indicado en la tabla). (C) Evaluación de productos de fijación de ADN por geles de poliacrilamida natural con tinción de bromuro de etidio. ub: ubiquitina; ac: acetil; me: metil; wt: natural; MN: mononucleosoma; ID: identificación; H3Kac₅: K9/14/18/23/27ac; H4Kac₅: K5/8/12/16/20ac. Las secuencias en la fig. 12B, leyendo de la parte superior a la inferior, son SEQ ID NO: 19-57, respectivamente.

Figura 13: Análisis de la calidad de miembros del banco de nucleosomas con código de barras individuales por electroforesis en gel de poliacrilamida natural (PAGE) seguido de tinción con bromuro de etidio de ADN.

Todos los nucleosomas formaron una sola banda (pureza > 90 %) y mostraron comportamientos de migración en gel diferenciales, dependiendo de las modificaciones instaladas en las proteínas histonas ensambladas. Se muestra a la derecha un gel de poliacrilamida con tinción de bromuro de etidio natural del banco de MN 39meros agrupados. ub: ubiquitina; ac: acetil; me: metil; wt: natural; MN: mononucleosoma.

Figura 14. Creación de perfiles del banco de MN usando anticuerpos específicos para modificación y el lector de histona 'factor de transcripción de los dedos de homeodominio vegetal (PHD) del bromodominio BPTF'. (A) Se sometió a prueba la estabilidad del banco de MN con código de barras por inmunoprecipitación del banco usando un anticuerpo α -H3K4me3 (ab8580). No se observó ningún reordenamiento de ADN (entramado) incluso después de almacenamiento prolongado a 4 °C (blanco). (B) Construcciones BPTF con marca GST N terminal recombinantes empleadas en este estudio. Se expresaron de forma recombinante construcciones con marca GST N terminal en *E.coli* y se purificaron por afinidad con glutatión, intercambio iónico y cromatografía de exclusión por tamaño. Se evaluó la pureza de las proteínas por SDS-PAGE seguido de tinción CBB. (C, D) Precipitación de GST del banco de nucleosomas con código de barras usando construcciones BPTF. Se sondaron BPTF-PHD unido a resina (C), BPTF-PHD-BD (D, izquierda) y BPTF-PHD (D, derecha) frente a las 39 variantes de nucleosoma. Se realizó el procesamiento de los datos como se muestra en la fig. 10C, D, y se normalizaron además las lecturas de secuenciación normalizadas con la entrada frente a la variante H3K4me3 (fijada a 1 e indicada como *; (media \pm D.E.) % de valor de entrada se muestra en el recuadro). IP: entrada; GST: glutatión-S-transferasa; PHD: homeodominio vegetal; BD: bromodominio; CBB: azul brillante de Coomassie.

Figura 15. Creación de perfiles del banco de MN usando el lector y escritor de histona p300. (A) Construcciones p300 con marca GST N terminal recombinantes empleadas en el experimento de unión al banco. Se expresaron de forma recombinante construcciones con marca GST N terminal en *E.coli* y se purificaron por afinidad con glutatión, intercambio iónico y cromatografía de exclusión por tamaño. Se evaluó la pureza de las proteínas por SDS-PAGE seguido de tinción CBB. (B, C, D) Precipitación de GST del banco de nucleosomas con código de barras usando construcciones p300 con marca GST. Se sondaron GST-p300-BD-PHD unido a resina (B, izquierda), GST-p300-BD (B, derecha) y GST-p300-PHD (C) frente a las 39 variantes de nucleosoma. Esto se siguió de aislamiento, purificación, multiplexación de ADN usando los códigos de barras múltiples experimentales inversos, secuenciación usando la tecnología IonTorrent®, descodificación y normalización de entrada como se describe en la fig. 10C, D. Se normalizaron además las lecturas de secuenciación normalizadas con la entrada frente a la variante H4Kac₅ (fijada a 1 e indicada como *; (media \pm D.E.) % de valor de entrada se muestra en el recuadro). IP: entrada; GST: glutatión-S-transferasa; PHD: homeodominio vegetal; BD: bromodominio; CBB: azul brillante de Coomassie.

Figura 16. Creación de perfiles de actividad de escritura de histona de p300 y de un extracto celular nuclear usando el banco de MN con código de barras. (A, B) El banco 39-mero con código de barras se sometió a acetilación de p300 en ausencia (entramado) o presencia (blanco) de acetil-coenzima A (A). Se evaluó la extensión de la acetilación específica de sitio por inmunoprecipitación de los productos de reacción usando anticuerpos dirigidos hacia H3K18ac (A) o H4K5ac (B), seguido de aislamiento, purificación, multiplexación de ADN usando los códigos de barras múltiples experimentales inversos, secuenciación usando la tecnología IonTorrent®, descodificación, y normalización de entrada como se describe en la fig. 10C, D. Las eficacias de precipitación, que representan la extensión de la acetilación en ese sitio particular, se expresan en % de entrada. (C) Detección de la actividad de escritura de histona de un lisado celular nuclear derivado de células 293T humanas. Se incubó el banco con código de barras con un extracto celular nuclear en presencia de AcCoA 20 μ M, S-adenosina-metionina (SAM) 10 μ M, y trifosfato de adenosina (ATP) 10 μ M, y se aisló el producto de reacción usando un anticuerpo específico para la marca H3K14ac. Esto se siguió de aislamiento, purificación, multiplexación de ADN usando los códigos de barras múltiples experimentales inversos, secuenciación usando la tecnología IonTorrent®, descodificación, y normalización de entrada como se describe en la fig. 10C, D.

La figura 17 representa una arquitectura ejemplar para implementar un dispositivo informático de acuerdo con uno o más modos de realización.

DESCRIPCIÓN

Esta divulgación se refiere, por ejemplo, a la aplicación de códigos de barras de ADN de bancos de matrices de cromatinas y mononucleosomas diseñadores para la creación de perfiles de lectores, escritores, borradores y moduladores de cromatina de los mismos. Proporciona componentes y procedimientos para la bioquímica de cromatina cuantitativa masivamente paralelizada, incluyendo un banco con código de barras de nucleosomas químicamente definidos, y un banco de polinucleosomas químicamente definidos con código de barras (a veces denominado en el presente documento banco de matrices de cromatina diseñadoras, o "CA").

Esta divulgación cumple la necesidad de una plataforma sólida para la bioquímica y biofísica de cromatina de alto rendimiento. Específicamente, se ensamblaron histonas recombinantes y sintéticas (que portan modificaciones

postraduccionales específicas; PTM) con secuencias de ADN con código de barras (que portan modificaciones epigenéticas específicas tales como metilación e hidroximetilación y/u otras modificaciones no naturales) y/o proteínas distintas de histonas y/o histonas fijadoras adicionales en bancos de matrices de cromatinas (CA) y mononucleosomas (MN) diseñadores. Las modificaciones de histonas y/o ADN se pueden denominar en general modificaciones nucleosómicas o modificaciones de nucleosomas. A veces en el presente documento, se usa el término modificaciones de cromatina. Este es un modelo *in vitro* que es representativo de un estado de cromatina existente en la naturaleza (por ejemplo, en una célula de interés). Usando técnicas de aislamiento apropiadas, tales como experimentos de precipitación, estos bancos se pueden usar para crear perfiles de, entre otros, (a) lectores de cromatina mono o multivalentes para investigar su patrón de reconocimiento; (b) escritores o borradores de cromatina para investigar interferencias de modificaciones de ADN y PTM de histonas potenciales; (c) modificaciones de ADN e histonas que modulan la actividad de los factores de proteínas o enzimas que interactúan con cromatina; y (d) moléculas que modulan la actividad de factores de proteínas o enzimas que interactúan con y/o modifican la cromatina. Los procedimientos y composiciones y dispositivos divulgados son adaptables en un alto grado de paralelización. Los códigos de barras adicionales, que codifican manipulaciones bioquímicas específicas de los bancos de cromatinas diseñadoras, se pueden unir a las moléculas de ADN. Estas secuencias de ADN multiplexadas (con código tanto para (a) modificaciones de nucleosomas específicas, la naturaleza y tipo de ADN, y la composición de histona de los miembros del banco así como para (b) el experimento específico) se procesan simultáneamente por tecnologías de secuenciación de próxima generación y otras tecnologías de descodificación de ADN. El análisis de los datos de secuenciación puede revelar especificidades del sustrato e interferencias potenciales (escritores y borradores) y afinidades de unión relativas (lectores). Adicionalmente, estos experimentos permiten estudios mecánicos y pueden servir como herramienta de diagnóstico para la actividad de complejos asociados a cromatina grandes encontrados *in vivo*, que a menudo combinan lectores, escritores y borradores de cromatina, tales como los derivados de extractos celulares nucleares de pacientes sanos y con cáncer. Estos procedimientos y composiciones proporcionan el diseño racional y la creación de perfiles de fármacos epigenéticos de próxima generación.

La materia objeto de la invención se define por las reivindicaciones adjuntas.

Un aspecto de la invención es un banco de mononucleosomas sintéticos (aislados, producidos sintéticamente, sin componentes que se encuentren asociados de forma natural con mononucleosomas en una célula, purificados antes de ponerse en el banco), en el que el banco comprende dos o más (por ejemplo, al menos 10, 50, 100, 200, 500, 1000, 2000, 3000, 4000, 5000 hasta aproximadamente 10.000 o más) tipos de mononucleosomas (miembros de un conjunto de mononucleosomas). El menor número de miembros de banco es de 2. El límite superior del tamaño de banco se define por la combinatoria de variantes de histonas (de cientos a miles), variantes de ADN (de cientos a miles) y proteínas asociadas a cromatina distintas de histonas (de cientos a miles). Un ejemplo es un banco que contiene una de cada una de las siguientes modificaciones: modificaciones postraduccionales de histona (aproximadamente 100) (ENCODE Project Consortium *et al.*, 2012), isoformas de histona (aproximadamente cientos), modificaciones de ADN (aproximadamente cientos), y proteínas asociadas a cromatina (aproximadamente cientos), lo que da como resultado un banco de cientos a miles de nucleosomas. En otro ejemplo, un banco contiene todos los estados de cromatina biológicamente relevantes (un estado de cromatina es una molécula de cromatina con una combinación natural y definida de modificaciones postraduccionales de histona, isoformas de histona, modificaciones de ADN y proteínas asociadas a cromatina), lo que da como resultado un banco de cientos, potencialmente miles de nucleosomas. En otro ejemplo, se puede dirigir un banco a un experimento específico. Por ejemplo, el papel de las trimetilisinas en la unión/reconocimiento de nucleosomas se puede estudiar usando solo una fracción de variantes de nucleosomas naturales, a saber, el subconjunto que contiene todas las variantes de histonas que contienen trimetilisinas conocidas, lo que da como resultado un banco de decenas a cientos de nucleosomas. Otro ejemplo comprende miembros del banco de los tres tipos de banco descritos anteriormente.

Cada mononucleosoma comprende un complejo de:

- (a) un octámero de proteína, que contiene 2 copias cada una de las histonas H2A, H2B, H3 y H4, y opcionalmente, la histona fijadora H1, en el que al menos una de las histonas no está modificada y/o en el que al menos una de las histonas se modifica para formar un patrón de modificaciones de histona, (las modificaciones de histonas pueden ser isoformas de histona, PTM y/o aminoácidos no naturales) y
- (b) un molécula de ADN nucleosómico (una molécula de ADN que es parte de un mononucleosoma, por ejemplo, se enrolla en el octámero de las proteínas histonas. Un ADN nucleosómico presente en un mononucleosoma a veces se denomina en el presente documento "ADN nucleosómico o de mononucleosoma". Un ADN nucleosómico se asocia con un nucleosoma). El ADN nucleosómico comprende (como se usa en el presente documento, los términos "un" y "el" se refieren a uno o más de uno, a menos que el contexto lo indique claramente de otro modo):
 - (i) una secuencia de posicionamiento de nucleosoma (NPS) fuerte (por ejemplo, que se puede unir con suficiente fuerza para evitar el reordenamiento de los mononucleosomas sintéticos, por ejemplo que se unen a octámeros de histonas aproximadamente 50 veces, 70 veces, 80 veces, 90 veces, 100 veces, 125 veces, 150 veces, 200 veces, 250 veces, o más, más fuertemente que el ADN bruto),

(ii) uno o más códigos de barras de ADN situados en las posiciones definidas en el ADN nucleosómico (por ejemplo, situados dentro de, o en o cerca de un extremo del ADN nucleosómico, tal como a una distancia específica de la NPS u otro punto fijo en el ADN), y, opcionalmente,

5 (iii) extensiones de ADN, incluyendo fijadores de ADN, en el extremo 5' y/o 3' de la NPS y/o dentro de la NPS. Estos incluyen moléculas distintas de ADN artificiales y secuencias de ADN unidas covalentemente.

10 La molécula de ADN nucleosómico puede no estar modificada y/o al menos uno de los nucleótidos en el ADN se puede modificar para formar un patrón único de modificaciones de ADN.

15 Opcionalmente, el mononucleosoma puede comprender

(c) una o más proteínas asociadas a cromatina distintas de histona.

20 En un banco de mononucleosomas sintéticos de la invención, cada mononucleosoma del banco tiene un patrón único de modificaciones de histonas y/o un patrón único de modificaciones de ADN, formando de este modo un patrón único de modificaciones de nucleosomas. La molécula de ADN comprende uno o varios códigos de barras únicos con una secuencia y posición en el ADN nucleosómico que son indicativas de (se correlacionan con, se asocian con, tienen una relación predeterminada con) el patrón único de modificaciones de nucleosomas.

25 Otro aspecto de la invención es un polinucleosoma sintético (a veces denominado en el presente documento cromatina sintética, o una matriz de cromatina (CA) sintética), que comprende dos o más mononucleosomas sintéticos (por ejemplo, al menos 3, 5, 7, 9, 12, 15, o 20) unidos entre sí (enlazados entre sí) por una molécula de ADN definida (por ejemplo, cada una de esas moléculas de ADN definidas puede tener la misma secuencia o diferente), teniendo los mononucleosomas una conectividad definida (orientación espacial uno con respecto al otro).

30 Cada uno de estos mononucleosomas comprende un complejo de

(a) un octámero de proteína, que contiene 2 copias cada una de las histonas H2A, H2B, H3 y H4, y opcionalmente, la histona fijadora H1, en el que al menos una de las histonas no está modificada, y/o al menos una de las histonas se modifica para formar un patrón de modificaciones de histonas mononucleosómicas; y

35 (b) una molécula de ADN nucleosómico como se define en las reivindicaciones adjuntas; y opcionalmente

(c) una o más proteínas asociadas a cromatina distintas de histona.

40 En un polinucleosoma sintético de la invención, el patrón de modificaciones de nucleosomas mononucleosómicos de los mononucleosomas en el polinucleosoma puede ser uniforme o puede ser diferente (único), lo que da como resultado un patrón único de modificaciones de nucleosomas polinucleosómicos. El polinucleosoma comprende un (uno o más) código(s) de barras situado(s) en una posición definida del ADN polinucleosómico (por ejemplo, situados internamente dentro del ADN polinucleosómico, o en o cerca del extremo 5' o 3' del ADN polinucleosómico).

45 Un ADN nucleosómico presente en un polinucleosoma a vece se denomina en el presente documento "ADN polinucleosómico". La posición definida del código de barras puede estar, por ejemplo, a una distancia específica de una secuencia de posicionamiento de nucleosoma (NPS) u otro punto fijo en el ADN polinucleosómico. La combinación de la secuencia del código de barras y la posición en el ADN polinucleosómico es indicativa del patrón único de modificaciones de nucleosomas polinucleosómicos.

50 Otro aspecto de la invención es un banco de polinucleosomas sintéticos (a veces denominados en el presente documento cromatina sintética, o una matriz de cromatina (CA) sintética), que comprende dos o más polinucleosomas sintéticos como se describe anteriormente. Cada miembro de un banco de este tipo tiene uno o más códigos de barras únicos, con una secuencia y posición en el ADN polinucleosómico que son indicativas de un patrón único de modificaciones polinucleosómicas.

55 En un banco de la invención (un banco de mononucleosomas o un banco de polinucleosomas), las histonas se pueden modificar en cualquiera de una variedad de formas. Estas modificaciones pueden comprender, por ejemplo, isoformas de histonas, PTM, y/o aminoácidos no naturales.

60 Las isoformas o variantes de histonas pueden ser naturales o artificiales. Se caracterizan por sustituciones aminoacídicas (por ejemplo, las variantes de histona H3 más comunes son H3.1, H3.2, H3.3) o inserciones aminoacídicas dentro de las secuencias de proteínas o extensiones en el extremo de las secuencias de proteínas (por ejemplo, macro-H2A). Una lista parcial de isoformas de histonas en seres humanos incluye:

65 (a) Histona H2A:

H2AF, H2AFB1, H2AFB2, H2AFB3, H2AFJ, H2AFV, H2AFX, H2AFY, H2AFY2, H2AFZ, H2A1, HIST1H2AA, HIST1H2AB, HIST1H2AC, HIST1H2AD, macro-H2A, HIST1H2AE, HIST1H2AG, HIST1H2AI, HIST1H2AJ, HIST1H2AK, HIST1H2AL, HIST1H2AM, H2A2 HIST2H2AA3, HIST2H2AC

5

(b) Histona H2B:

H2BF, H2BFM, H2BFS, H2BFWT, H2B1, HIST1H2BA, HIST1H2BB, HIST1H2BC, HIST1H2BD, HIST1H2BE, HIST1H2BF, HIST1H2BG, HIST1H2BH, HIST1H2BI, HIST1H2BJ, HIST1H2BK, HIST1H2BL, HIST1H2BM, HIST1H2BN, HIST1H2BO, H2B2, HIST2H2BE

10

(c) Histona H3:

H3A1, HIST1H3A, HIST1H3B, HIST1H3C, HIST1H3D, HIST1H3E, HIST1H3F, HIST1H3G, HIST1H3H, HIST1H3I, HIST1H3J, H3A2, HIST2H3C, H3A, HIST3H3, CENP-A

15

(d) Histona H4:

H41, HIST1H4A, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E, HIST1H4F, HIST1H4G, HIST1H4H, HIST1H4I, HIST1H4J, HIST1H4K, HIST1H4L, H44, HIST4H4

20

(e) Histona fijadora H1:

H1F, H1F0, H1FNT, H1FOO, H1FX, H1H1, HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, HIST1H1T

25

Otras isoformas de histonas serán evidentes para los expertos.

Adicionalmente, se han observado mutaciones en histonas en cánceres (por ejemplo, Lys27Met en la cola de H3.3 se produce con frecuencia en tumores del tronco encefálico infantiles); y dichos mutantes se pueden incluir también en bancos de la invención.

30

Una variedad de PTM de histonas será evidente para el experto. Estas incluyen cualquier modificación de histona natural, por ejemplo, metilación, acetilación, fosforilación, ubiquitinación, ADP-ribosilación, SUMOilación, glucosilación, alquilación, acilación, isomerización cis/trans de propilo, nitrosilación oxidación. Las PTM que aún no se han descubierto o caracterizado se incluyen en la invención.

35

Los aminoácidos no naturales incluyen análogos sintéticos de PTM, que pueden ser química y/o bioquímicamente inertes, agentes de fotoreticulación, marcadores fluorescentes, marcadores de isótopos u otros que serán evidentes para el experto.

40

Las modificaciones se pueden producir en un sitio, o en más de un sitio, en un nucleosoma.

Un "código de barras" como se usa en el presente documento es una secuencia de ácido nucleico que, conjuntamente con su localización en una molécula de ADN, se puede usar para identificar inequívocamente esa molécula de ADN, por ejemplo en el contexto de un banco de nucleosomas. El número de códigos de barras se dicta por la complejidad del banco que se va a usar, que a su vez es dependiente del número y combinaciones de las variantes de histonas (en los ejemplos mostrados en el presente documento, estas histonas difieren en su estado de PTM de histona), secuencias de ADN, proteínas asociadas a cromatina adicionales usadas para formar una única variante de matriz de cromatina o nucleosoma. Por ejemplo, un código de barras de 1 nucleótido (nt) puede codificar 4 miembros del banco, un código de barras de 2 nt 16 variantes, un código de barras de 3 nt 64 variantes, 4 nt 256 variantes, 5 nt 1.024 variantes y así sucesivamente. La longitud del/de los código(s) de barras de ADN se determina por el tamaño del banco. Dependiendo del tamaño del banco, el código de barras de ADN tiene un número de bases suficiente para proporcionar un número suficiente de variaciones para codificar únicamente cada miembro del banco. El/Los código(s) de barras puede(n) ser ADN monocatenario (ss) o ADN bicatenario (ds) (como se muestra en los ejemplos en el presente documento) o una combinación de los mismos.

50

55

En los ejemplos descritos en el presente documento, se usa un código de barras de 6 nucleótidos, que en principio codifica 4.096 variantes de matrices de cromatina o nucleosoma diferentes. En general, los códigos de barras de longitudes de 4-12 nucleótidos cubren la mayoría de las aplicaciones con tamaños de banco que son realistas, pero el código de barras puede ser más largo, si se necesita un mayor poder combinatorio.

60

Una "secuencia de posicionamiento de nucleosoma (NPS)" es una secuencia de ADN bicatenario natural o sintético de al menos 146 pares de bases que interactúa fuertemente con histonas y complejos de histona, en particular octámeros de histona (que consisten en 2 copias de las histonas H2A, H2B, H3 y H4). una NPS forma el nucleosoma con una posición y orientación específicas del octámero de histona con respecto al ADN. Este complejo

65

- 5 histona-ADN debe ser estable durante un periodo de tiempo prolongado para su almacenamiento (meses a 4 °C), y durante manipulaciones bioquímicas estándar (a concentraciones de pocas decenas de nanomolares (nM) en tampones comunes y horas a 30 °C). La NPS usada en los ejemplos en el presente documento es una secuencia artificial denominada 601, que se une a octámeros de histona aproximadamente 100 veces más fuertemente que el ADN bruto (Lowary y Widom, 1998). Se puede usar cualquier secuencia de ADN artificial o natural alternativa que cumple los criterios descritos anteriormente, de las que muchas serán evidentes para el experto. Por ejemplo, una NPS se puede unir a octámeros de histona aproximadamente 10 veces, 20 veces, 30 veces, 40 veces, 50 veces, 60 veces, 70 veces, 80 veces, 90 veces, 100 veces, 125 veces, 150 veces, 200 veces, 250 veces, o más, más fuertemente que el ADN bruto.
- 10 Las extensiones de ADN de los ADN de mononucleosoma o polinucleosoma de la invención pueden tomar cualquiera de una variedad de formas. Por ejemplo, pueden ser códigos de barras de ADN, sitios de cebado de ADN (por ejemplo para la lectura de secuenciación posterior o para amplificaciones de PCR), fijadores de ADN (como se explica en el siguiente párrafo), secuencias de posicionamiento alternativas, sitios de unión a proteínas (para proteínas histonas o distintas de histonas adicionales), sustratos de ADN enzimáticos, ADN con bases modificadas, o cualquier otra molécula distinta de ADN artificial, tal como asas con afinidad (por ejemplo, biotina) o sondas fluorescentes.
- 15 Los fijadores de ADN en ADN de mononucleosoma o polinucleosoma de la invención pueden ser de varias longitudes y composiciones. En la naturaleza, los nucleosomas se separan normalmente por ~10-90 pb del ADN fijador. Estos fijadores varían entre diferentes tejidos, especies o incluso dentro de un único genoma celular, y en la composición de bases. Los fijadores artificiales se caracterizan por su resistencia a envolverse alrededor de octámeros de histona. Prácticamente, un límite superior de los apéndices de ADN está en el intervalo de 100-1000 pb ya que secuencias más largas podrían perturbar el posicionamiento de las histonas en la hebra de ADN.
- 20 Los ADN de mononucleosoma o polinucleosoma de la invención pueden comprender una o bases de ADN no modificadas, bases con modificaciones naturales, tales como metilación, alquilación u oxidación, o bases con modificaciones artificiales. Una variedad de modificaciones adecuadas será evidente para el experto.
- 25 Una variedad de proteínas asociadas a cromatina distintas de histonas será evidente para el experto e incluyen factores de transcripción, interactuantes y modificadores de histonas, y proteínas remodeladoras de cromatina.
- Otro aspecto de la invención es un kit, por ejemplo para llevar a cabo uno de los procedimientos descritos en el presente documento como se define por las reivindicaciones adjuntas. El kit comprende nucleosomas (mononucleosomas o polinucleosomas) o un banco de mononucleosomas o polinucleosomas banco de la invención. El kit puede comprender una lista (compendio, algoritmo, resumen, medio legible por ordenador, o similares) que indica la correlación (relación, asociación, relación predeterminada) entre cada código de barras único y patrón único de modificaciones de nucleosomas. Una lista ejemplar de este tipo se muestra en la figura 12B. En modos de realización de la invención, los nucleosomas o bancos están en tubos de ensayo, los pocillos de una placa multipocillo, o en una cámara de reacción de un dispositivo microfluídico. Otros elementos opcionales de un kit de la invención incluye tampones, componentes de medios, sustratos, cofactores, inhibidores adecuados, y similares; un ordenador o me legible por ordenador para almacenar y/o evaluar los resultados de las matrices; o materiales de envasado.
- 35 Otro aspecto de la invención es un procedimiento para determinar la especificidad de afinidades y patrones de reconocimiento de lectores de cromatina como se define por las reivindicaciones adjuntas, el procedimiento comprende incubar (poner en contacto) un banco de la invención con uno o varios interactuantes y/o modificadores de cromatina de origen recombinante, o incubar un banco de la invención con interactuantes y/o modificadores de cromatina derivados de un extracto celular nuclear de una línea celular que se va a investigar (por ejemplo, incluyendo células derivadas de pacientes con cáncer humanos), aislar los miembros del banco unidos y/o modificados, e identificar y/o cuantificar los miembros del banco unidos o modificados y cualquier marca añadida o marca retirada. El procedimiento puede comprender analizar complejos remodeladores de cromatina grandes. El procedimiento puede comprender analizar una línea celular, incluyendo células derivadas de pacientes con cáncer humanos. Otro aspecto de la invención es un procedimiento para identificar las modificaciones asociadas con un interactuante o modificador, que comprende multiplexar un número mayor que uno de interactuantes y/o modificadores de cromatina con un banco de la invención, y dividir el banco en el mismo número de subbancos de acuerdo con las modificaciones, e identificar las modificaciones asociadas con cada interactuante o modificador.
- 45 Otro aspecto de la invención es un procedimiento para identificar y crear perfiles de la especificidad de fármacos epigenéticos, que comprende combinar una molécula candidata con un banco de la invención, y detectar la modulación de una modificación de nucleosoma (por ejemplo, inhibir o extinguir fármacos de proteínas o enzimas que interactúan con cromatina), identificando de este modo los fármacos epigenéticos candidatos que modulan modificaciones de nucleosomas.
- 60 El banco de nucleosomas puede comprender además una lista de códigos de barras de ADN y las modificaciones de nucleosomas asociadas y la composición de cada nucleosoma con código de barras.
- 65

El mononucleosoma o polinucleosoma sintético puede comprender un código de barras de ADN en el extremo 5' y/o 3' de o en cualquier lugar dentro de la molécula de ADN.

5 Otro aspecto de la invención es un mononucleosoma sintético como se define en la reivindicación 13, que comprende un complejo de

10 (a) un octámero de proteína, que contiene 2 copias cada una de las histonas H2A, H2B, H3 y H4, y opcionalmente, la histona fijadora H1, en el que al menos una de las histonas no está modificada y/o en el que al menos una de las histonas se modifica para formar un patrón de modificaciones de histona, (por ejemplo, isoformas de histona, PTM y/o aminoácidos no naturales) y

(b) una molécula de ADN nucleosómico que comprende

15 (i) secuencia de posicionamiento de nucleosoma (NPS) fuerte,

(ii) uno o más códigos de barras de ADN situados en las posiciones definidas en el ADN nucleosómico (por ejemplo, situados dentro de, en o cerca de un extremo del ADN nucleosómico, tal como a una distancia específica de la NPS u otro punto fijo en el ADN),

20 (iii) extensiones de ADN, incluyendo fijadores de ADN, en el extremo 5' y/o 3' de la NPS y/o dentro de la NPS (por ejemplo, moléculas distintas de ADN artificiales y secuencias de ADN unidas covalentemente), y opcionalmente

25 (c) una o más proteínas asociadas a cromatina distintas de histona,

en el que la secuencia y posición del/de los código(s) de barras en el ADN nucleosómico son indicativas del patrón de las modificaciones de nucleosomas en el mononucleosoma.

30 Otro aspecto de la invención es un procedimiento para ensamblar un mononucleosoma de la invención, que comprende combinar proteínas histonas y ADN de nucleosoma con código de barras con un ADN con tampón con MMTV marcado con biotina, en una proporción predeterminada.

35 En un banco de la invención, la NPS en los mononucleosomas o polinucleosomas en general es suficientemente fuerte para que el banco sea estable, y no se produce ningún reordenamiento de ADN entre los miembros del banco de mono y polinucleosomas después de un almacenamiento prolongado, por ejemplo durante al menos un mes a 4 °C. En aspectos de la invención, las modificaciones de histonas y/o ADN comprenden un conjunto representativo de estados de cromatina biológicamente relevantes. En aspectos de la invención, la proporción de los miembros del banco de mono y polinucleosoma es equimolar (1:1 para cada miembro del banco) o no equimolar, variando de 1 a 1000 (por ejemplo, 1:10, 1:50, 1:100, 1:150, 1:200, 1:300, 1:400, 1:500, 1:600, 1:700, 1:800, 1:900) para uno o para un subconjunto de los miembros del banco (por ejemplo, variando de 1 a 1000, tal como 1, 10, 50, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900 o 1000) en una proporción predeterminada fija.

40 Los modos de realización de la invención incluyen lo siguiente.

45 **(A) Aplicación de códigos de barras de ADN de la historia sintética de cada miembro del banco de CA o MN**

50 Este modo de realización se refiere a la fabricación de bancos de matrices de cromatinas (CA) diseñadoras y mononucleosomas (MN) diseñadores, donde cada miembro del banco porta (a) código(s) de barras de ADN que codifica(s) la historia sintética específica de cada variante de CA o MN.

Un MN es un complejo que consiste en

55 (1) un octámero de proteína, que contiene 2 copias de las histonas canónicas H2A, H2B, H3 y H4 (o versiones modificadas de las mismas), y en algunos casos, la histona fijadora H1,

(2) una molécula de ADN nucleosómico que comprende

60 (i) secuencia de posicionamiento de nucleosoma (NPS) fuerte,

(ii) uno o más códigos de barras de ADN situados en las posiciones definidas en el ADN nucleosómico (por ejemplo, situados dentro de, en o cerca de un extremo del ADN nucleosómico, tal como a una distancia específica de la NPS u otro punto fijo en el ADN),

(iii) extensiones de ADN, incluyendo fijadores de ADN, en el extremo 5' y/o 3' de la NPS y/o dentro de la NPS (por ejemplo, moléculas distintas de ADN artificiales y secuencias de ADN unidas covalentemente), y opcionalmente

5 (c) una o más proteínas asociadas a cromatina distintas de histona.

Específicamente, cada miembro del banco de MN porta

10 (a) una combinación única de variantes de histonas, incluyendo isoformas de histonas, patrones de PTM de histonas, histonas con aminoácidos no naturales.

y

15 (b) una variante de ADN nucleosómico única, que contiene una NPS, código(s) de barras de ADN y/o extensiones de ADN. El ADN puede contener las bases de ADN canónicas, bases con modificaciones naturales (tales como metilación u oxidación), o bien bases con modificaciones artificiales, y opcionalmente,

(c) una o más proteínas asociadas a cromatina distintas de histona.

20 La composición de cada MN único, con diferentes patrones de PTM de histonas y/o modificación de ADN y/u otras proteínas histonas y distintas de histonas, se codifica en (a) secuencia(s) de ADN (en el presente documento denominada como código de barras de MN) en cualquier lugar de, en, o cerca del extremo del ADN nucleosómico, tal como el extremo 5' (fig. 1). El límite superior del tamaño de banco se define por la combinatoria de variantes de histonas y ADN. Por motivos prácticos, el tamaño de banco se define por el experimento posterior y típicamente

25 varía de decenas a cientos a miles de miembros de banco. Por ejemplo, el tamaño puede ser, o puede ser menor que aproximadamente, o mayor que aproximadamente 10, 20, 30, 50, 75, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1250, 1500, 2000, 2500, 3000, 3500, 4000 o más miembros de banco.

30 Una CA es un complejo que consiste en unidades de MN (véase la definición de MN anterior), que pueden estar (a) modificadas de forma uniforme o (b) modificadas de forma única. La longitud de la matriz es variable y típicamente varía de 2-12 MN, tal como un dímero, trímero, pentámero, y así sucesivamente. Los MN se conectan entre sí en una secuencia definida. Una matriz de cromatina sintética también se puede denominar aquí polinucleosoma, en contraste con un mononucleosoma sintético.

35 El número puede ser cualquier número hasta 8, 10, 12, 14, 15, 16, 18 o 20.

Cada miembro del banco de CA está compuesto de MN modificados individualmente con conectividad definida y porta:

40 (a) MN, con

45 (i) una combinación única de variantes de histonas, incluyendo isoformas de histonas, patrones de PTM de histonas, histonas con aminoácidos no naturales,

(ii) una variante de ADN nucleosómico única, que contiene una o varias NPS diferentes, código(s) de barras de ADN, y/o fijadores de ADN. El ADN puede contener las bases de ADN canónicas, bases con modificaciones naturales (tales como metilación u oxidación), o bien bases con modificaciones artificiales,

50 y/o

(c) histonas fijadoras y/u otras proteínas distintas de histonas.

55 El patrón de PTM de proteínas y/o modificación de ADN, conectividad de MN, longitud e identidad del ADN, presencia y patrón de modificación de histonas fijadoras, y/o proteínas distintas de histonas se codifican en (a) secuencia(s) de ADN (en el presente documento denominada código de barras de CA) en cualquier lugar dentro de, en, o cerca del extremo del ADN de matriz, tal como el extremo 5' (fig. 2). El límite superior del tamaño de banco se define por la combinatoria de las PTM individuales y/o modificaciones de ADN, conectividad de MN, variante de ADN, histonas fijadoras y/o proteínas distintas de histonas. Por motivos prácticos, el tamaño de banco se define por

60 el experimento posterior y típicamente varía de cientos a miles de miembros de banco.

65 Los códigos de barras de MN o CA marcan de forma única e inequívoca la composición química del MN o CA dentro del banco. Estos bancos con código de barras se pueden usar tanto para someter a prueba como para generar diversas hipótesis bioquímicas y biofísicas, tales como creación de perfiles de especificidades de sustrato de interactuantes o modificadores de cromatina (fig. 3) a través del procedimiento de descodificación del código de

barras (de forma análoga a los procedimientos existentes basados en bancos químicos codificados por ADN (Buller, Mannocci y Scheuermann, 2010; Clark, 2010)). Los procedimientos adecuados incluyen

- 5 (1) restricción-digestión
- (2) reacción en cadena de la polimerasa (PCR)
- (3) hibridación de micromatrices de ADN
- 10 (4) secuenciación de ADN, tal como las tecnologías de secuenciación de próxima generación (Mardis, 2008) IonTorrent (Rothberg *et al.*, 2011) o Illumina.

En el caso de secuenciación de ADN, los sitios de cebado de secuenciación directo (FW) e inverso (RV) requeridos se pueden añadir al ADN de matriz o nucleosómico en cualquier fase del procedimiento, tal como por clonación molecular, PCR o fijación de ADN.

Además de la identificación de los MN o CA con las propiedades bioquímicas o biofísicas deseadas, la cuantificación de los miembros de banco aislados es factible, en particular cuando se usan tecnologías de secuenciación de próxima generación (Mardis, 2008) como lectura. En combinación con una etapa de cuantificación de ADN absoluta antes de la secuenciación, por ejemplo por PCR cuantitativa (qPCR), se pueden obtener afinidades de unión relativas de sustratos de MN o CA en un experimento multiplexado individual, como se explica a continuación.

Por ejemplo, una proteína que interactúa con cromatina (tal como un lector de cromatina) o que modifica la cromatina (tal como un escritor o borrador de cromatina) de origen recombinante o derivado de extractos celulares nucleares se incuban con el banco de MN o CA con código de barras (fig. 3). En el caso particular descrito en este ejemplo, es un banco de MN con un código de barras ds unido antes de la formación de MN en el extremo 5' del ADN Widom 601 nucleosómico, un tramo artificial de 147 pb de ADN bicatenario (ds) al que los octámeros de histona se unen con alta afinidad (Lowary y Widom, 1998). El tamaño y composición de los bancos se pueden adaptar al experimento posterior de forma modular. En la siguiente etapa, los sustratos de MN o CA con las propiedades bioquímicas y biofísicas deseadas se aíslan por procedimientos apropiados, incluyendo:

- (1) experimentos de precipitación (afinidad o inmunoprecipitación)
- 35 (2) separación por propiedades físicas o químicas diferenciales después de unión o modificación de MN o CA, tal como movilidad electroforética (matrices de desplazamiento de movilidad electroforética), hidrofobicidad, carga (cromatografía de intercambio iónico) o tamaño (cromatografía de exclusión por tamaño, SEC)
- (3) clasificación celular de moléculas activadas por fluorescencia (FAMS)
- 40 usando
- (1) marcaje del interactuante (lector de cromatina)
- 45 (2) marcadores (afinidad, asa química, sonda de fluorescencia) en el interactuante para aislamiento directo (por ejemplo, usando clasificación de moléculas basada en fluorescencia) u otra funcionalización bioquímica o química (lector de cromatina)
- (3) proteína secundaria que reconoce el interactuante (lector de cromatina)
- 50 (4) anticuerpos frente a marca unida o retirada (modificador de cromatina)
- (5) proteína secundaria (lector) que reconoce la marca unida o retirada (modificador de cromatina)
- 55 (6) sustratos enzimáticos modificados con marcadores artificiales (afinidad, asa química, sonda de fluorescencia) para aislamiento directo (por ejemplo usando técnicas de clasificación de moléculas basada en fluorescencia) u otra funcionalización bioquímica o química (modificador de cromatina)

Por ejemplo, se usa un anticuerpo frente a la proteína o la modificación unida o retirada para (a) precipitar el interactuante de cromatina, tal como un lector, en complejo con sus fijadores de MN o CA más fuertes, o (b) aislar los sustratos de MN o CA preferentes del modificador de cromatina, tal como un escritor o borrador. Después del aislamiento de ADN, los sustratos o interactuantes de cromatina se identifican y cuantifican a través de la descodificación del código de barras de MN o CA usando metodologías tales como hibridación de micromatrices de ADN o secuenciación de ADN (Mardis, 2008), por ejemplo IonTorrent (Rothberg *et al.*, 2011) o Illumina.

65 **(B) Aplicación de códigos de barras de ADN de la historia experimental de cada miembro de banco de MN o CA.**

Este modo de realización implica manipulaciones por un procedimiento bioquímico o biofísico, tales como manipulaciones bioquímicas o biofísicas en presencia de moléculas adicionales que modulan el procedimiento, incluyendo inhibidores y/o activadores de proteínas interactuantes con cromatina y enzimas modificadoras de cromatina. Es conocida la aplicación de códigos de barras de ADN en el contexto diferente de bancos de moléculas pequeñas de marcado (Buller *et al.*, 2010; Clark, 2010; Kleiner, Dumelin, y Liu, 2011) y anticuerpos (Agasti, Liong, Peterson, Lee, y Weissleder, 2012; Krutzik y Nolan, 2006).

Por ejemplo, un banco de MN o CA con código de barras se manipula por diversos procedimientos bioquímicos o biofísicos (fig. 4). En el caso particular descrito en este ejemplo, es un banco de MN con un código de barras unido antes de la formación de MN en el extremo 5' del ADN 601 nucleosómico. Las etapas de manipulación se codifican posteriormente en códigos de barras separados (en el presente documento denominados código de barras del experimento).

Variante A: Si el experimento se somete solo a una etapa de manipulación experimental, el código de barras del experimento (múltiple) se puede unir después del aislamiento de sustrato y ADN por PCR. En esta etapa de PCR, la identidad de la variante de MN o CA se acopla al experimento específico a través de la generación de una secuencia de ADN que contiene ambos códigos de barras de MN o CA así como el código de barras del experimento (múltiple). La longitud de la secuencia de ADN con doble código de barras, que también comprende los sitios de cebado FW y RV para la posterior secuenciación, está limitada por la longitud de la lectura fiable por el procedimiento de secuenciación de ADN elegido.

Variante B: Si se realizan múltiples manipulaciones experimentales, los códigos de barras de los experimentos se pueden unir al extremo 5' o 3' del ADN de matriz o nucleosómico de cada miembro de banco.

Los subbancos que se someten a manipulación por el procedimiento bioquímico o biofísico específico se aíslan por procedimientos adecuados como se explica anteriormente, y un código de barras que codifica la manipulación específica realizada se fija a todos los miembros de subbanco. Los bancos con códigos de barras diferenciales se agrupan y se dividen de nuevo para la posterior manipulación por un segundo procedimiento bioquímico o biofísico y se manejan como se describe para la primera etapa uniendo un segundo código de barras de experimento al extremo 5' o 3' del ADN. Esto se puede repetir según se desee. En la última etapa de aplicación de códigos de barras, el sitio de cebado de secuenciación FW también se une, por ejemplo por fijación de ADN. Después del aislamiento de ADN, el sitio de cebado RV para la secuenciación de ADN se añade por PCR. Los MN o CA respectivos se identifican y cuantifican a través de la descodificación del código de barras de manipulación y el código de barras de MN o CA usando secuenciación de ADN. Este procedimiento se puede adaptar a experimentos en presencia de moléculas (por ejemplo moléculas pequeñas o biomoléculas más grandes, tales como péptidos o proteínas) que modulan la actividad o función de interactuantes y/o modificadores de cromatina, tales como lectores, escritores o borradores de cromatina o de origen recombinante o derivados de extractos celulares nucleares.

El código de barras de experimento codifica inequívocamente cada procedimiento bioquímico o biofísico que la variante de MN o CA ha sufrido durante el transcurso del experimento.

Los bancos de mononucleosomas y matrices de cromatina sintéticas difieren de los nucleosomas y matrices de cromatina previamente aislados de la naturaleza de muchas formas, por ejemplo en que están sintetizados, químicamente puros, y contienen modificaciones de histonas y ADN predeterminadas en patrones controlados, y que tienen uno o más códigos de barras únicos que especifican una variante de mono o polinucleosoma dada. También comprenden secuencias NPS fuertes y definidas, como se analiza en otra parte en el presente documento. La cromatina obtenida de organismos con sustratos naturales tiene modificaciones desconocidas, sin códigos de barras, y no es adecuada para un banco de la invención, o para su manipulación o marcaje como se logra usando los mononucleosomas y matrices sintéticos de la invención.

Las composiciones y procedimientos de aplicación de códigos de barras de ADN descritos se pueden utilizar (1) para cribar y/o crear perfiles de moléculas que interactúan con y/o modifican MN y CA; (2) para el descubrimiento y/o creación de perfiles de interactuantes y/o modificadores de cromatina y su especificidad por sustrato de MN o CA preferente con respecto a la variante de MN o CA; (3) para el descubrimiento y creación de perfiles de moléculas que modulan los interactuantes y/o modificadores de cromatina; y (4) para la creación de perfiles de variantes de MN y CA con respecto a sus propiedades bioquímicas y biofísicas. Los procedimientos de acuerdo con la invención incluyen los siguiente:

(1) creación de perfiles de interactuantes de cromatina, tales como lectores de histonas y su unión preferente por patrones de PTM y/o modificación de ADN específicos.

Un lector de cromatina, o versiones del mismo, por ejemplo que contienen (a) uno o bien (b, c) múltiples módulos lectores (que residen (b) dentro de una cadena polipeptídica o (c) en diferentes cadenas polipeptídicas dentro de un complejo proteico más grande, fig. 5), se incuban con el banco con código de barras (en este ejemplo, que contiene un único código de barras en el extremo 5' del ADN de mayor o nucleosómico) en solución. Como enfoque

alternativo, se puede preparar un lisado celular nuclear del organismo que se va a estudiar e incubar con el banco con código de barras. Esto está seguido del aislamiento de los fijadores de MN o CA, tal como por precipitación de anticuerpo del lector de cromatina o una marca de afinidad unida al lector. Después del aislamiento de ADN, un segundo código de barras, o múltiple, que codifica el experimento específico (de precipitación) ('MP'), así como los sitios de cebado FW y RV para la secuenciación de ADN se unen por PCR. (Nota: opcionalmente, el sitio de cebado FW y RV también se puede incluir en el ADN nucleosómico antes de la formación de MN). Los fijadores preferentes se identifican a través de la descodificación de los códigos de barras múltiples o de MN o CA usando secuenciación de ADN. Se pueden realizar múltiples experimentos, por ejemplo usando concentraciones variables de proteína/MN, lectores variables, truncamientos o mutantes de los mismos, en paralelo y leerse en una única etapa de secuenciación debido a la etapa de PCR de multiplexación.

(2) creación de perfiles de modificadores de cromatina, tales como escritores o borradores de histona, y sus patrones de reconocimiento de modificación de sustrato preferentes

Un escritor o borrador de cromatina, o versiones de los mismo, tales como (a) el dominio catalítico, (b) la enzima de longitud completa, o (c) un complejo multisubunitario grande (fig. 6), se incuba con el banco (en este caso particular, que contiene un único código de barras en el extremo 5' del ADN de matriz o nucleosómico) en presencia de cualquier sustrato requerido, tal como S-adenosina-metionina (SAM), Acetil-Coenzima A (AcCoA), y trifosfato de adenosina (ATP). Como enfoque alternativo, se puede preparar un lisado celular nuclear del organismo que se va a estudiar e incubar con el banco con código de barras en presencia de cualquier sustrato requerido, tal como SAM, AcCoA y ATP. Esta etapa está seguida del aislamiento de los sustratos de MN o CA que se han modificado con éxito por el escritor o borrador de cromatina, tal como por precipitación de anticuerpo de la marca unida o retirada. En el caso de un escritor de cromatina, se usa un anticuerpo frente a la marca unida para la precipitación de los MN/CA que se han modificado por el escritor de cromatina. En el caso de un borrador de cromatina, se usa un anticuerpo frente a la marca retirada para sustraer los MN/CA que no se han seleccionado por los borradores de cromatina, dejando atrás los sustratos preferentes del borrador de histona. Después del aislamiento de ADN, un segundo código de barras, o múltiple, que codifica el experimento específico, así como los sitios de cebado FW y RV para la secuenciación de ADN se añaden por PCR (nota: opcionalmente, también se puede incluir el sitio de cebado FW y RV en el ADN nucleosómico antes de la formación de MN). Los sustratos se identifican a través de la descodificación de los códigos de barras múltiples o de MN o CA usando secuenciación de ADN. Se pueden realizar múltiples experimentos, por ejemplo usando concentraciones variables de proteína/MN, enzimas variables, truncamientos o mutantes de los mismos, en paralelo y leerse en una única etapa de secuenciación debido a la etapa de PCR de multiplexación.

(3) creación de perfiles de la firma epigenética de una línea celular

Usando la estrategia explicada anteriormente, se pueden someter a ensayo convenientemente las actividades de los modificadores de histonas en extractos nucleares además. Este ajuste permite la identificación de actividades de modificación de cromatina e interferencias específicas, de las que algunas son características de un tipo celular dado. En particular, las células cancerosas poseen tendencias de modificación de cromatina distintas. Por ejemplo, EZH2 es un marcador de cáncer de mama invasor (Kleer *et al.*, 2003), y la capacidad para medir actividades enzimáticas en lugar de la abundancia de muestras tisulares es de alto valor de diagnóstico (Spacil *et al.*, 2013). Se incuban bancos de nucleosomas con código de barras con extractos nucleares de biopsias tisulares para catalogar las firmas de las modificaciones de cromatina (tales como modificaciones de histonas y/o ADN) para distintos tipos celulares y estados de enfermedad, permitiendo así el diagnóstico de disfunción de la bioquímica nuclear.

(4) creación de perfiles de estabilidad de MN

Se expone el banco de MN con código de barras (en este caso particular, que contiene un único código de barras en el extremo 5' del ADN nucleosómico) a diversas condiciones experimentales que desestabilizan los MN, tales como incremento de concentraciones salinas (fig. 7). Como enfoque alternativo, se puede preparar un lisado celular nuclear del organismo que se va a estudiar e incubar con el banco con código de barras. Se usa la liberación de ADN nucleosómico después de cada incremento salino para controlar la estabilidad de los MN respectivos. Después del aislamiento de ADN, un segundo código de barras, o múltiple, que codifica el experimento específico, tal como la concentración salina usada, así como los sitios de cebado FW y RV para la secuenciación de ADN se añaden por PCR. (Nota: opcionalmente, el sitio de cebado FW y RV también se puede incluir en el ADN nucleosómico antes de la formación de MN). Los MN diferencialmente estables se identifican a través de la descodificación de los códigos de barras múltiples y de MN usando secuenciación de ADN. Estas pruebas de estabilidad se pueden extender a los experimentos en presencia de proteínas que modulan la estabilidad de MN, tales como chaperonas de histonas o factores remodeladores de cromatina, realizar en paralelo, y leerse en una única etapa de secuenciación debido a la etapa de PCR de multiplexación.

(5) creación de perfiles de estabilidad de CA

Se expone un banco de CA (en este caso particular, que contiene un único código de barras en el extremo 5' del ADN de matriz) a condiciones experimentales que desestabilizan las matrices de cromatina, tales como incremento

de concentraciones salinas (de forma análoga a la fig. 7). Como enfoque alternativo, se puede preparar un lisado celular nuclear del organismo que se va a estudiar e incubar con el banco con código de barras. Se usa la liberación de ADN de cromatina después de cada incremento salino para controlar la estabilidad de las CA respectivas. Después del aislamiento de ADN, un segundo código de barras, o múltiple, que codifica el experimento específico, tal como la concentración salina usada, así como los sitios de cebado FW y RV para la secuenciación de ADN se añaden por PCR (nota: opcionalmente, también se puede incluir el sitio de cebado FW y RV en el ADN nucleosómico antes de la formación de MN). Las CA diferencialmente estables se identifican a través de la descodificación de los códigos de barras múltiples o de CA usando secuenciación de ADN. Estas pruebas de estabilidad se pueden extender a los experimentos en presencia de proteínas que modulan la estabilidad de la matriz, tales como chaperonas de histonas o factores remodeladores de cromatina, que se pueden realizar en paralelo, y leerse en una única etapa de secuenciación debido a la etapa de PCR de multiplexación.

(6) creación de perfiles de accesibilidad de CA

El banco de CA se expone a condiciones experimentales que modulan el plegamiento de los MN y/o CA, tales como incremento de concentraciones salinas. Como enfoque alternativo, se puede preparar un lisado celular nuclear del organismo que se va a estudiar e incubar con el banco con código de barras. La accesibilidad de los miembros de banco de CA se puede investigar por diversas formas, tales como por reconocimiento de un patrón de PTM (por ejemplo por un lector de histona) que se fija dentro de la CA o por reconocimiento de un sitio de unión a ADN tal como para un factor de transcripción incluido dentro del ADN de CA. Después del aislamiento de ADN, un segundo código de barras, o múltiple, que codifica el experimento específico, así como los sitios de cebado FW y RV para la secuenciación de ADN se añaden por PCR (nota: opcionalmente, también se puede incluir el sitio de cebado FW y RV en el ADN nucleosómico antes de la formación de MN). Las CA respectivas se identifican a través de la descodificación de los códigos de barras múltiples o de CA usando secuenciación de ADN. Estas pruebas de accesibilidad se pueden extender a experimentos en presencia de proteínas que modulan la compactación/descompactación de matrices de cromatina, que se pueden realizar en paralelo y leerse en una única etapa de secuenciación debido a la etapa de PCR de multiplexación.

(7) cribado de moléculas que modulan la actividad de interactuantes y modificadores de cromatina, tales como lectores/escriutores/borradores de histonas

Se incuban uno o varios MN o CA con código de barras con el lector, escritor o borrador de cromatina de interés en presencia de miembros de un banco molecular (por ejemplo moléculas pequeñas, péptidos, ácidos nucleicos, ácidos peptidonucleicos, foldámeros) que contiene un inhibidor(es) putativo(s) (fig. 8). La elección de los sustratos de MN o CA se determina por el lector, escritor o borrador que se use en el experimento específico. Como enfoque alternativo, se puede preparar un lisado celular nuclear del organismo que se va a estudiar e incubar con el banco con código de barras en presencia del banco molecular. Después de cada etapa de incubación, se añade un código de barras que codifica la respectiva molécula candidata usada (en el presente documento denominado código de barras inhibidor) así como el sitio de cebado FW al extremo 5' del ADN de cromatina o nucleosómico. Los MN/CA se agrupan y se realiza el aislamiento de los subbancos que mostraron interacción o modificación de cromatina alterada como resultado de la(s) molécula(s) específica(s) usada(s), tal como por precipitación (frente a (a) la marca unida o (b) retirada o (c) el lector). Después del aislamiento de ADN, el sitio de cebado RV para secuenciación de ADN se añade por PCR (nota: opcionalmente, también se puede incluir el sitio de cebado FW y RV en el ADN nucleosómico antes de la formación de MN). Los resultados positivos se identifican a través de la descodificación del código de barras inhibidor usando secuenciación de ADN.

Los siguientes son ejemplos de modificaciones de histonas, modificadores de histonas (incluyendo escriutores y borradores de histonas), lectores de histonas, modificaciones de ADN y modificadores de ADN y lectores de ADN, así como tipos celulares que se pueden usar de acuerdo con la invención.

Lectores de histonas, incluyendo proteínas que contienen el/los siguiente(s) dominio(s):

Bromodominio (BD)

Homeodominio vegetal (PHD)

PHD en tándem

Cromodominio

WD40

Tudor

Tudor doble/tándem

MBT

Repeticiones de anquirina

5 zf-CF

Dominio PWWP ("PWWP" divulgado como SEQ ID NO: 1)

10 14-3-3

BRCT

UBA

15 **Escritores de histonas, incluyendo:**

Histona acetiltransferasas (HAT)

20 Histona aciltransferasas,

Histona metiltransferasas (HMT)

Cinasas

25 Ubiquitinasas (UB)

ADP-ribosiltransferasas

Glucosiltransferasas

30 Prolina isomerasas

Complejos remodeladores de histonas

35 **Borradores de histonas, incluyendo:**

Histona desacetilasas (HDAC)

40 Histona desmetilasas (HDM)

Desubiquitinasas (DUB)

Fosfatasas

45 Arginina desiminasas

Modificadores de ADN, incluyendo:

50 ADN metiltransferasas (DNMT)

Metilcitosina hidroxilasas/oxidasas (enzimas de la familia TET)

Lectores de modificación de ADN, incluyendo:

55 Dominio de unión metil-CpG (MBD)

Dominio asociado a los dedos de SET y Ring (SRA)

Modificaciones de ADN, incluyendo:

60 [Citosina] metilación / metilcitosina

[Citosina] hidroximetilación / hidroximetilcitosina

65 [Citosina] formilación / formilcitosina

[Citosina] carboxilación / carboxicitosina

[Adenosina] metilación / metiladenosina

5 [Guanidina] oxidación / oxoguanidina

Dimerización de timidina

10 Sitios abásicos

Mellas monocatenarias

Lisados celulares nucleares, incluyendo los que se originan de:

15 células humanas (tales como células 293-T, COR-L23, HEK293, HeLa, Jurkat, NIH-3T3)

células cancerosas humanas (tales como 721, U937, BCP1, A2780, A-549, A431, CML-T1, DU145, H1299, KYO1, MCF-7, Raji, THP1)

20 así como células de cualquier otro organismo de origen sano o con enfermedad

Los enfoques que se han usado por otros investigadores para investigar las proteínas que interactúan con y/o modifican la cromatina se describen a continuación.

25 **(1) Bancos de CA pequeñas o MN después de la digestión de cromatina, usando típicamente tratamiento con nucleasa microcócica (MNasa)**

Este enfoque es una versión de una técnica experimental de inmunoprecipitación llamada inmunoprecipitación de cromatina (ChIP)(Schones y Zhao, 2008), que se usa para investigar;

30 (a) la interacción entre proteínas y ADN en la célula

(b) la abundancia y localización de proteínas, incluyendo histonas, en regiones genómicas específicas

35 (c) la abundancia y localización de PTM de histonas específicas

Típicamente, la cromatina endógena y sus proteínas asociadas se reticulan en un lisado celular

40 la cromatina se digiere por MNasa para dar un banco de CA y/o MN

las proteínas o marcas de PTM de interés se inmunoprecipitan selectivamente, y los fragmentos de ADN asociados se purifican y se determina su secuencia.

45 Aunque estos tipos de muestras representan biológica grandes y biológicamente relevantes y proporcionan información valiosa sobre sus locus genómicos, existen desventajas en su uso para la creación de perfiles de interactuantes y modificadores de cromatina, tales como:

50 (a) la impureza/heterogeneidad de las muestras, tanto dentro de la muestra (contaminación con proteínas interactuantes con cromatina endógena) así como entre diferentes experimentos (composición de banco difícil de reproducir de experimento a experimento, en parte debido al reordenamiento *in vitro*)

55 (b) la lectura de la composición de MN o CA después de la precipitación, puesto que su historia sintética no está codificada. La lectura es dependiente de cualquier anticuerpo (frente a una proteína o marca específica), a veces en combinación con espectrometría de masas (EM)(Britton, Gonzales-Cope, Zee, y Garcia, 2011). La EM, en particular, es un procedimiento no sesgado muy sensible que no requiere ningún marcaje del banco. Sin embargo, la EM solo puede crear perfiles de patrones de reconocimiento de lectores/escritores/borradores de histonas dentro de una cadena polipeptídica, es decir, los patrones de PTM dentro de una histona, pero no puede acoplar las PTM que residen en diferentes histonas en un contexto nucleosómico inalterado. Además, aunque es sensible, la EM no tiene un límite de detección con determinada modificación tal como la fosforilación, lo que es especialmente problemático. La EM no puede discriminar entre determinadas modificaciones, por ejemplo dimetilarginina simétrica frente a asimétrica. Por último, las señales iónicas detectadas en EM no son adaptables a cualquier tipo de amplificación (a diferencia de la información basada en ADN), situando de nuevo restricciones prácticas del mundo real sobre la sensibilidad.

65 **(2) Bancos de péptidos de histonas con extremo N químicamente definido con marcas de PTM específicas**

- Se han sintetizado bancos grandes (que contienen hasta miles de miembros) de péptidos de colas de histonas con extremo N modificado usando química en fase sólida (Garske *et al.*, 2010) y usado para crear perfiles de diversos dominios de lectores de histonas conocidos para determinar la unión de los patrones de PTM. La construcción de bancos de péptidos es, en comparación con MN o CA inalterados, más simple, más rápida (se puede automatizar) y modular (por ejemplo usando tecnología de división y agrupación). Además, todos los miembros de banco se pueden codificar, por ejemplo por separación física sobre soporte sólido. Se puede determinar la identidad de los péptidos modificados diferencialmente, tal como por EM. Se pueden cribar las interacciones de proteínas con PTM de histonas que residen en las colas de histona dentro de una cadena polipeptídica, mientras que las interacciones con (i) patrones de PTM dentro de los dominios globulares de las histonas, en diferentes histonas, o en diferentes nucleosomas en un ajuste mono y polinucleosómico fisiológico; y (ii) patrones de modificaciones de ADN o (iii) una combinación de los mismos, no se pueden investigar, tal como la unión multivalente de lectores de cromatina para patrones de PTM específicos en el contexto nucleosómico.

(3) MN individuales químicamente definidos que contienen patrones de PTM específicos

- Se han usado sustratos de MN químicamente definidos inalterados con un patrón de PTM específico en experimentos de precipitación individual tradicionales para investigar el concepto de multivalencia de un lector de histona, *factor de transcripción de los dedos de PHD del bromodominio* (BPTF) (Ruthenburg *et al.*, 2011). Este trabajo mostró la importancia del contexto nucleosómico natural para la lectura de patrones de PTM de histonas, pero experimenta un rendimiento extremadamente bajo y un requisito de diseño experimental basado en hipótesis. Cada experimento de precipitación investigó el evento de unión de un único par de lector de histona-MN y no se pudo realizar con variantes de MN múltiples al mismo tiempo, como se propone en la invención descrita, puesto que su historia sintética no se codificó para permitir la identificación y cuantificación del fijador preferente. Se construyó un banco de mononucleosomas de dos nucleosomas por Kingston y colaboradores (con el propósito de someterlos a un factor remodelador de cromatina) uniendo dos fluoróforos diferentes, Cy3 y Cy5, al extremo 5' de una NPS (Goldman, Garlick y Kingston, 2010). Las desventajas de este enfoque, en comparación con la invención descrita, incluyen:

(1) tamaño de banco de nucleosomas muy pequeño, que está limitado por la disponibilidad y la idoneidad de moléculas fluorescentes ortogonales (el tamaño de banco es de 2 en la publicación descrita, pero se podría escalar a aproximadamente 4)

(2) sensibilidad baja de la lectura (lectura basada en ADN frente a fluorescencia), por lo tanto poco rentable y con ineficacia material

(3) ninguna multiplexación de experimento posible debido a la lectura basada en fluorescencia

(4) normalización difícil de los datos

Hasta la fecha, no se ha desarrollado con éxito ningún procedimiento para la bioquímica de cromatina de alto rendimiento cuantitativa, lo que requiere la construcción de bancos de matrices de cromatina o mononucleosomas grandes y diversos pero químicamente definidos para crear perfiles de proteínas interactuantes/modificadoras de cromatina con una lectura apropiada de forma no sesgada (Allis y Muir, 2011; Fierz y Muir, 2012).

Las composiciones y procedimientos de acuerdo con la invención superan las desventajas de los tres enfoques explicados anteriormente y proporcionan una solución para la generación de dichos bancos de cromatinas diseñadoras, el aislamiento, identificación, y cuantificación de moléculas interactuantes con propiedades deseables. Se pueden incluir los siguientes rasgos característicos:

- recapitulación de estados de cromatina natural presentes *in vivo* en forma de mono y polinucleosomas modificados
- preparaciones homogéneas, que dan como resultado sustratos mono y polinucleosómicos químicamente definidos aunque naturales
- estabilidad hacia reordenamiento de ADN durante la preparación de nucleosomas, procesamiento y posterior ensayos bioquímicos usando una NPS artificial
- naturaleza de alto rendimiento de las matrices bioquímicas realizadas en bancos mono y polinucleosómicos gracias a la estrategia de la aplicación de códigos de barras (codificación inequívoca tanto de (a) variante de MN o CA como de (b) experimento)
- lectura sensible (es decir, amplificable) y cuantitativa gracias a la etapa de PCR y secuenciación de próxima generación

Es ventajoso el uso de códigos de barras de ADN únicos para codificar las

propiedades bioquímicas y/o biofísicas individuales de cada miembro del banco de MN o CA. Esos códigos de barras se unen a o se incluyen dentro de las respectivas secuencias polinucleosómicas de ADN.

Tradicionalmente, la aplicación de códigos de barras de ADN se realiza a nivel genómico, donde los genes de interés se marcan con códigos de barras moleculares únicos para facilitar la identificación de los respectivos grupos de proteínas a través de amplificación, marcado e hibridación de micromatrices con código de barras, como se muestra en el ejemplo de sondas de nucleosomas con mutantes de histonas sintéticas H3 y H4 (Dai, Hyland, Yuan, Huang, y Bader, 2008)

Se ha descrito el uso de códigos de barras para la identificación distinta del nivel genómico para bancos químicos con códigos de barras de ADN (Buller *et al.*, 2010; Clark, 2010; Kleiner *et al.*, 2011), donde se introducen tramos de ADN como asas artificiales para marcar de forma única cada miembro de banco de molécula pequeña, y bancos de anticuerpos codificados con ADN (Agasti *et al.*, 2012; Krutzik y Nolan, 2006).

El uso de matrices de cromatina y nucleosomas con código de barras de acuerdo con la invención proporciona varias ventajas distintas. Por ejemplo, los códigos de barras codifican inequívocamente la variante de MN o CA individual en el banco y se puede codificar por hibridación de micromatrices o secuenciación de ADN para obtener información cuantitativa sobre los sustratos o fijadores preferentes. Adicionalmente, esta estrategia de aplicación de códigos de barras se puede emplear para codificar cada manipulación bioquímica o biofísica realizada en un banco dado de forma modular.

Los procedimientos de preparación de bancos de mononucleosomas y bancos de matrices de cromatina incluyen lo siguiente:

(1) Síntesis de histonas

(a) Histonas naturales (wt) y modificadas postraduccionalmente naturales y versiones de las mismas usando procedimientos descritos.

- La síntesis de proteína recombinante, síntesis de péptidos en fase sólida, o una combinación de las mismas (usando tecnologías tales como Native Chemical Ligation (NCL) (Dawson y Kent, 2000), Expressed Protein Ligation (Muir, 2003) (EPL) o procedimientos de supresión Amber (Wang, Xie y Schultz, 2006)) se usan para sintetizar las variantes de histonas que se incorporan posteriormente en los MN o CA, incluyendo histonas naturales, modificadas postraduccionalmente, marcadas artificialmente o truncadas (fig. 9a).

(b) Histonas modificadas postraduccionalmente que portan análogos de lisina metilada (Kme) o lisina acetilada (Kac) usando los procedimientos descritos.

- Se alquilan mutantes de cisteína para dar metilaminoetilcisteínas (análogos de Kme)(Simon *et al.*, 2007) o se someten a reacción tiol-eno para dar acetamidoetilcisteínas (análogos de Kac)(Cao, Korolev y Nordenskiöld, 2011) usando la estrategia descrita por Shokat y colaboradores y Nordenskiöld y colaboradores, respectivamente.

La síntesis de una única histona modificadas postraduccionalmente consume mucho tiempo pero se puede acelerar a través de automatización y paralelización, tal como a través del desarrollo de protocolos y uniones de NCL modulares, así como la inclusión de histonas que contienen análogos de aminoácidos modificados.

(2) Formación de octámeros

Las histonas se ensamblan por medio de la adición de histonas wt y/o modificadas en proporciones iguales, diálisis de GdmHCl 6 M a NaCl 2 M, seguido de purificación de SEC (Dyer *et al.*, 2004; Luger, Rechsteiner y Richmond, 1999). De forma alternativa, se puede realizar la formación de octámeros a una escala de tan solo 1 nmol de histona (aproximadamente 50 µg de histona total, dependiendo de la variante de histona). Sin embargo, la escala se puede incrementar si se necesita más material, o disminuir, siempre que se usen dispositivos de diálisis apropiados para alojar los volúmenes usados. Las concentraciones se pueden medir por espectrometría UV a 280 nm y sustracción del fondo a 300 nm, y se pueden obtener los coeficientes de extinción calculados usando páginas web comunes tales como el sitio web expasy.org/protparam. Las histonas se mezclan en proporciones equimolares en hielo en aproximadamente 55 µl de tampón de desplegado (cloruro de guanidinio 6 M, Tris-HCl 20 mM, pH 7,5 a 4 °C, Na-EDTA 1 mM, DTT 1 mM) para proporcionar una concentración de proteína total de aproximadamente 1 mg/ml a 4 °C. Las mezclas se disponen en botones de diálisis MINI (valor de corte 3.500 Da) y se dializa frente a 3x 600 ml de tampón de replegado (NaCl 2 M, Tris-HCl 10 mM, pH 7,5 a 4 °C, Na-EDTA 1 mM, DTT 1 mM) durante al menos 4 h cada uno a 4 °C, con una etapa de diálisis durante la noche. Al día siguiente, las mezclas se transfieren a tubos

Eppendorf y se hacen girar a 17.000 g durante al menos 5 min a 4 °C para retirar precipitados. Se transfieren los sobrenadantes a un tubo limpio. Se añade glicerol al 50 % (v/v), y se miden las concentraciones de octámeros usando su absorción UV y fueron típicamente 2-5 µM. Los octámeros se pueden procesar directamente para ensamblajes de MN, y/o se almacenan a -20 °C.

5

(3) Preparación de ADN de matrices y nucleosómico modificado y no modificado con código de barras

(a) Andamio nucleosómico y de matriz:

Se puede obtener ADN no modificado por procedimientos conocidos tales como clonación molecular, PCR o fijación de ADN de piezas del mismo, o síntesis química. Se puede usar cualquier secuencia de ADN, siempre que pueda dirigir el posicionamiento nucleosómico, tal como la secuencia 601 Widom (Lowary y Widom, 1998) usada en los presentes ejemplos. Se pueden utilizar diversas longitudes y tipos de NPS y fijadores de ADN, dependiendo de la aplicación necesitada. En general, se requiere una NPS suficientemente fuerte para una aplicación de códigos de barras a los miembros del banco con éxito y una identificación para evitar el reordenamiento de ADN entre miembros del banco a una concentración de decenas de nM (por ejemplo, 5, 10, 20, 30, 40, 50 o más). Es necesario que se sintetice el ADN modificado usando procedimientos apropiados y, dependiendo de la(s) modificación/modificaciones introducida(s), incluyendo procedimientos enzimáticos o químicos.

15

(b) Código(s) de barras

Se incorporan uno o varios códigos de barras en cualquier parte dentro del ADN de matriz o nucleosómico, en, o cerca de sus extremos 5' o 3', o dentro de regiones fijadoras. La elección y la longitud del/de los código(s) de barras se pueden ajustar al experimento específico y a la extensión del poder combinatorio necesario para codificar todo el banco. Los códigos de barras codifican cualquier propiedad bioquímica o biofísica de los miembros de banco, tales como la(s) variante(s) de histona, variante(s) de ADN, conectividad de MN, ADN fijador, histonas fijadoras, proteínas distintas de histonas, y/o tipo de manipulación. Por ejemplo, como se muestra en el experimento demostrativo preliminar descrito en la sección 9, un tramo nucleotídico de 6 pb se puede unir al extremo 5' del ADN de matriz o nucleosómico, que codifica hasta 4096 variantes de MN o CA. Estos códigos de barras se pueden introducir por procedimientos tales como:

25

30

(i) clonación molecular en ADN plasmídico seguido de restricción enzimática para liberar el ADN de matriz o nucleosómico

35

(ii) PCR

(iii) Fijaciones de ADN enzimáticas en el extremo 5' o 3' del ADN de matriz o nucleosómico antes de la formación de MN o CA

40

(iv) Fijaciones de ADN enzimáticas en el extremo 5' o 3' del ADN de matriz o nucleosómico después de la formación de MN o CA

(v) combinaciones de los mismas.

(c) Sitios de cebado de secuenciación FW y RV.

Si se usa la secuenciación de ADN como lectura, se pueden introducir sitios de cebado de secuenciación FW y RV antes de o después de la formación de MN o CA o después de que se haya realizado el experimento, o una combinación de los mismos, usando tecnologías tales como clonación molecular, PCR o fijación de ADN.

50

En una versión de bancos de MN con código de barras (fig. 10a), un tramo de ADN bicatenario (ds) de 190 pb que contiene un sitio de cebado FW de 30 pb, compatible con una lectura de secuenciación de próxima generación posterior usando el secuenciador IonTorrent® (Rothberg *et al.*, 2011), un código de barras de 6 pb que codifica la variante de MN, un fijador de 4 pb, 147 pb de ADN 601 nucleosómico (Lowary y Widom, 1998), y un apéndice de 3 pb corto en el extremo 3' se prepara por PCR usando la secuencia 601 nucleosómica como molde.

55

En otra versión de bancos de MN con código de barras, un tramo de ADN ds de 177 pb que contiene el ADN nucleosómico 601 de Widom (Lowary y Widom, 1998) con un saliente 5'-AA-3' en el extremo 5' de la hebra inferior y un saliente 5'-CAC-3' en el extremo 3' de la hebra superior ('BC-601'; fig. 10b) se prepara por (1) liberación de la secuencia 601 nucleosómica, clonada como una repetición 12-mera en un plásmido circular o bien producida por PCR, usando BsaI y DralI; (2) hibridación de ADN monocatenario complementario que abarca nt 10-30 del sitio de cebado IonTorrent directo ('FW-iT₁₀₋₃₀') (Rothberg *et al.*, 2011) y los respectivos códigos de barras de MN de 6 pb ('BC-MN'; hebra inferior que contiene un saliente 5'-AA-3' en el extremo 3' y un saliente 5'-CATC-3' en el extremo 5'); (3) combinación de estas secuencias de ADN hibridadas con el ADN 601 nucleosómico; (4) fosforilación *in situ* usando ADN T4 polinucleótido cinasa (PNK); (5) fijación usando ADN ligasa T4 para proporcionar 'BC-601'; y (6) purificación del producto de ADN final (fig. 10b). La concentración del producto de ADN final se determina por

60

65

absorción UV del ADN ds a 260 nm con una sustracción de fondo a 340 nm, y se pueden obtener los coeficientes de extinción calculados usando sitios web comunes tales como el sitio web biophysics.idtdna.com/UVSpectrum.html.

(4) Formación de MN

Se pueden ensamblar mononucleosomas por adición de ADN nucleosómico con código de barras a la respectiva variante de octámero purificado con SEC, seguido de diálisis con tampón salino de alto a bajo usando procedimientos descritos (Dyer *et al.*, 2004; Luger *et al.*, 1999). Esto puede estar seguido de una etapa de purificación, tal como electroforesis en gel preparativa o cromatografía de intercambio iónico.

Se puede determinar una proporción de ADN a octámero correcta experimentalmente para un ensamblaje de MN o CA con éxito (Dyer *et al.*, 2004; Luger *et al.*, 1999). Para la producción a gran escala donde se predetermina la proporción, el procedimiento se puede acelerar por automatización.

De forma alternativa, se pueden realizar ensamblajes de MN en una forma de alto rendimiento tan solo en tres días con un nuevo protocolo basado en el uso de ADN de tampón, tal como la secuencia de ADN MMTV (Flaus y Richmond, 1998). Se instala un asa con afinidad de biotina en el extremo 5' del ADN MMTV por PCR (fig. 9b), lo que facilita la retirada de complejos histona-MMTV ensamblados incorrectamente y ADN MMTV potencialmente libre, y elimina por tanto las siguientes etapas limitantes del tiempo y el material: (1) requisito de purificación del octámero de histona antes del ensamblaje de nucleosoma, tal como por SEC; y (2) pruebas de proporción ADN a octámero empíricas. Este nuevo protocolo permite la preparación de material en una escala de ensamblaje de nucleosoma y octámero de histona muy pequeña (1 nmol de cada respectiva histona, pero se puede incrementar o disminuir además, en caso necesario) de forma altamente paralelizada de decenas, cientos y potencialmente miles de variantes de nucleosomas. Típicamente, se realizan ensamblajes de nucleosomas en una escala de decenas de pmol (por ejemplo, 49 pmol), pero esto se puede aumentar o reducir a escala, dependiendo de la aplicación necesaria. Típicamente, se combina 1 equivalente (eq) (por ejemplo, 49 pmol) de los respectivos octámeros (por ejemplo, de un banco de octámeros 39-mero) individualmente con 0,6 eq (con respecto al octámero del ADN 601 nucleosómico) (por ejemplo, 29 pmol) de ADN con código de barras ('BC-601') y 0,4 eq (con respecto a octámeros) (por ejemplo, 20 pmol) de ADN BIO-MMTV a una concentración de octámero de aproximadamente 0,7 μ M (por ejemplo, volumen total de aproximadamente 70 μ M pmol) en tampón de replegado (Tris-HCl 10 mM, pH 7,5 a 4 °C, NaCl 2 M, EDTA 1 mM, DTT 1 mM) a 4 °C. Los ensamblajes de MN se pueden realizar usando (1) la dilución, (2) la diálisis en etapas, o (3) el procedimiento de diálisis continua como se describe (Dyer *et al.*, 2004; Luger *et al.*, 1999). En este caso particular, se empleó una diálisis continua usando una bomba peristáltica. Por ejemplo, se pueden disponer las mezclas del ensamblaje de MN en botones de diálisis MINI y dializar frente a 200 ml de tampón de inicio MN (KCl 1,4 M, Tris-HCl 10 mM, pH 7,8, EDTA 0,1 mM, DTT 1 mM) durante 1 h a 4 °C en agitación. Durante un transcurso de aproximadamente 6 h, se añaden 320 ml de tampón terminador de MN (KCl 10 mM, Tris-HCl 10 mM, pH 7,8, EDTA 0,1 mM, DTT 1 mM) a una tasa de 1,0 ml/min en agitación, y se dializan las mezclas frente a ese tampón resultante durante otra hora en agitación a 4 °C. Posteriormente, se dializan las mezclas frente a 2x 200 ml de tampón terminador de MN (una etapa durante 1 h, la otra durante al menos 4 h en agitación a 4 °C). Se transfieren las mezclas a tubos Eppendorf, se hacen girar a 17.000 g durante al menos 5 min a 4 °C para retirar los precipitados, los sobrenadantes se transfieren a tubos limpios y se complementan con inhibidores de proteasas, tales como fluoruro de fenilmetilsulfonilo 0,5 mM (PMSF). Se eliminan los MN ensamblados del ADN MMTV (en forma libre y/o unidos a histonas) incubando las soluciones con cantidades suficientes de perlas recubiertas de estreptavidina durante 1 h a temperatura ambiente (TA) en un nutador de extremo a extremo. Se eliminan las soluciones de las perlas y se hace girar a 17.000 g durante al menos 5 min a 4 °C para retirar los precipitados. Se transfieren los sobrenadantes en tubos limpios. Se realiza la cuantificación de los MN finales usando la absorbancia del ADN ds nucleosómico a 260 nm con una sustracción de fondo a 340 nm. Se pueden obtener los coeficientes de extinción calculados usando sitios web tales como el sitio web biophysics.idtdna.com/UVSpectrum.html. Para analizar la calidad de los MN resultantes, típicamente se complementan 0,5 pmol de MN resultantes con aproximadamente un 15 % v/v de sacarosa y se analiza por electroforesis en gel de poliacrilamida natural y tinción por EtBr. Se realiza la visualización usando el ImageQuant LAS4000 (GE Healthcare). Para el almacenamiento, se añade un 20 % v/v de glicerol, y los MN o bien se dividen en alícuotas de 3 pmol, se someten a ultracongelado rápido y se almacenan a -80 °C de forma individual o combinada en proporciones equimolares, se concentra usando un concentrador de proteínas (Vivaspin 500, valor de corte 10.000 Da) a aproximadamente 1,0 μ M (concentración de nucleosoma total), se dividen en alícuotas de 3 pmol, se someten a ultracongelado rápido y se almacenan a -80 °C. En general, se pueden preparar nucleosomas en tan solo tres días, comenzando a partir de histonas recombinantes y/o sintéticas, de forma altamente paralelizada, en la escala de decenas, cientos y potencialmente miles de variantes en el margen temporal descrito con robótica disponible.

(5) Formación de CA

Una matriz uniforme con un solo tipo de subunidad de MN (o variante) se puede ensamblar de forma similar al (4) anterior. En una matriz no uniforme, está presente más de un tipo de subunidad de MN, y los MN se ensamblan individualmente y se fijan entre sí en una secuencia definida usando fijación de ADN con salientes de ADN únicos (Blacketer, Feely y Shogren-Knaak, 2010). En algunas matrices, cada MN puede ser único en sus modificaciones.

(6) Formación de bancos de MN o CA: agrupamiento de MN modificados deseados MN en concentraciones deseadas

5 Se puede formar un banco por adición de MN/CA con códigos de barras de ADN de forma única para dar una composición deseada de los bancos. Las proporciones de los miembros de banco pueden ser equimolares o bien no equimolares, por ejemplo para recapitular las diferentes distribuciones de tipos de MN/CA, es decir, estados de cromatina, *in vivo*.

10 Se puede usar una etapa de purificación para purificar los miembros de banco, en caso necesario, por ejemplo por electroforesis en gel preparativa, intercambio iónico o filtración en gel (Bao, Chakravarthy, Muthurajan y Luger, 2003). Puesto que a los MN y CA se les aplica códigos de barras, el banco agrupado resultante se puede purificar en una etapa.

15 Un banco puede incluir recipientes separados para cada miembro de MN o CA, o puede incluir miembros de MN o CA múltiples en un único recipiente, o puede incluir todos los miembros de un banco o subbanco en un único recipiente. Un banco puede incluir tanto MN como CA.

(7) Aislamiento, identificación y cuantificación de miembros de banco con propiedades bioquímicas o biofísicas deseadas**(a) Ensayo bioquímico o biofísico sobre el banco de MN/CA codificado.**

25 Los procedimientos para los ensayos bioquímicos o biofísicos a continuación se describen para bancos de MN, pero se pueden aplicar a bancos de CA de forma idéntica.

(i) Perfil de unión de un lector de cromatina

30 Se inmoviliza un lector de cromatina (recombinante o derivado de un extracto celular nuclear) sobre un soporte sólido, por ejemplo a través de una marca de afinidad o inmunoprecipitación, y se incuba con el banco de nucleosomas en tampones proteicos estándar. Estos incluyen tampón 2-amino-2-hidroximetil-1,3-propanodiol (Tris), tampón fosfato y tampón ácido (4-(2-hidroxietil)-1-piperacinetanosulfónico (Hepes), a un pH próximo a neutro (tal como 6,5-8), y deben contener todos los reactivos y cofactores adicionales requeridos para (a) la estabilidad e integridad de nucleosomas (tales como reactivos reductores, inhibidores de proteasas), estabilidad del lector de cromatina (tales como glicerol, sales), y (c) la especificidad del evento de unión (tales como sales o detergentes). La incubación se realiza típicamente a una concentración de nucleosoma total de decenas de nM en un volumen de pocos cientos de μ l, pero ambas cantidades se pueden incrementar o disminuir. La temperatura de incubación es típicamente de 4 °C, pero puede ser cualquier entre 4 °C y una temperatura que todavía se tolere por el lector de nucleosoma y cromatina, tal como 37 °C. El tiempo de incubación es típicamente de 4 h, cuando el experimento se realiza a 4 °C, pero se puede ajustar al experimento específico. De forma alternativa, los eventos de unión entre el lector de cromatina y el/los nucleosoma(s) se pueden realizar primero en solución, y la inmovilización de los complejos lector de cromatina-nucleosoma se puede realizar después de esto.

(ii) Patrón de modificación enzimática de un escritor y borrador de cromatina

45 Se incuba un escritor de cromatina (recombinante o derivado de un extracto celular nuclear) con el banco de nucleosomas en tampones proteicos estándar. Estos incluyen tampón 2-amino-2-hidroximetil-1,3-propanodiol (Tris), tampón fosfato, y tampón ácido (4-(2-hidroxietil)-1-piperacinetanosulfónico (Hepes), a un pH próximo a neutro (tal como 6,5-8), y debe contener todos los reactivos y cofactores adicionales requeridos para (a) la estabilidad e integridad de nucleosomas (tales como reactivos reductores, inhibidores de proteasas), (b) la reacción enzimática (tales como sustratos, por ejemplo ATP, SAM y/o AcCoA), (c) la estabilidad de las enzimas (tales como sales o glicerol), (d) la estabilidad de los productos de reacción (por ejemplo inhibidores de la reacción inversa, por ejemplo inhibidores HDAC cuando sigue una reducir de histona acetiltransferasas), y (e) la especificidad de la etapa de inmunoprecipitación posterior (tales como sales, glicerol o detergentes). La reacción se realiza típicamente a una concentración de nucleosoma total de decenas de nM en un volumen de pocas decenas de μ l a concentraciones enzimáticas apropiadas, pero se puede incrementar o disminuir como se desee. La temperatura de incubación es típicamente de 25-37 °C, pero puede ser cualquiera entre 4 °C y una temperatura que todavía se tolere por el escritor de nucleosoma y cromatina. El tiempo de incubación es típicamente de 10-60 min, cuando el experimento se realiza a 30 °C, pero se puede ajustar al experimento específico. Posteriormente, se añade un anticuerpo que se une a un aminoácido enzimáticamente modificado dentro de la secuencia de histona, típicamente a una concentración de pocas decenas de μ g/ml durante 1 h a TA. Posteriormente, se capturan los complejos anticuerpo-nucleosoma por perlas de proteína G o A por incubación durante típicamente 1,5 horas a TA. De forma alternativa, la etapa de unión entre el anticuerpo y el/los nucleosoma(s) enzimáticamente modificado(s) se puede realizar en solución, y la inmovilización de los complejos anticuerpo-nucleosoma se puede realizar después de esto. Para un borrador, los sustratos de nucleosomas que no se modificaron se reducen por inmunoprecipitación con un anticuerpo frente a la marca retirada en consecuencia.

(b) Aislamiento de ADN

La separación del ADN de matriz o nucleosómico de las proteínas asociadas se puede realizar por procedimientos estándar, tales como por digestión proteica, tal como tratamiento con proteinasa K, que está seguido de purificación de ADN (por ejemplo, kit de purificación de PCR de Qiagen). La cantidad de ADN absoluta se puede determinar usando técnicas de cuantificación de ADN, tales como por espectrometría UV, o hibridación de sondas fluorescentes, tales como Qubit o qPCR.

(c) Descodificación de códigos de barras usando hibridación de micromatrices

La decodificación se puede lograr usando un chip de micromatriz con las secuencias de ADN inmovilizadas usando procedimientos estándar (Heller, 2002).

(d) Multiplexación y decodificación de códigos de barras usando secuenciación de ADN

Los sitios de cebado de secuenciación directo (FW) e inverso (RV) se pueden incluir durante la preparación de ADN de matriz o nucleosómico usando metodologías tales como PCR, clonación molecular o fijación de ADN. De forma alternativa, los sitios de cebado se pueden añadir después de la formación de MN o CA y la unión o el experimento enzimático usando PCR o fijación de ADN, lo que permite la inserción simultánea de un código de barras de multiplexación que codifica un experimento específico. Una combinación de unión pre y posexperimental de sitios de cebado es factible. El banco de ADN resultante, que contiene por tanto códigos de barras de nucleosomas con una propiedad bioquímica o biofísica dada, se puede agrupar, someter a un secuenciador de próxima generación, y se puede lograr la decodificación de las lecturas de secuenciación (como se ejemplifica por la secuenciación IonTorrent® en los ejemplos específicos mostrados) clasificando los datos de acuerdo con la lista de códigos de barras múltiples experimentales en primer lugar, seguido de la clasificación de acuerdo con la lista de códigos de barras de nucleosoma y la normalización de la entrada del banco, tal como por la herramienta divisora de código de barras Fastx Toolkit, sitio web hannonlab.cshl.edu/fastx_toolkit/ind3ex.html (fig. 10d).

En el modo de realización mostrado en la fig. 10a y 10b, el sitio de cebado FW (compatible con la posterior etapa de secuenciación de próxima generación por IonTorrent® (Rothberg *et al.*, 2011)) se incluye durante la etapa de PCR para generar el ADN 601 nucleosómico con código de barras con un sitio de cebado FW de 30 pb (fig. 10a) o se fija junto con el código de barras de nucleosoma al bloque constitutivo de 601 nucleosómico 601 por fijación de ADN T4 ('BC-601', fig. 10b). Después del aislamiento de ADN después del experimento bioquímico apropiado, el sitio de cebado del adaptador RV de 23 pb compatible con el posterior secuenciador de próxima generación IonTorrent® (Rothberg *et al.*, 2011) así como el código de barras múltiple (experimento) de 6 pb así como las primeras 10 pb del sitio de cebado FW IonTorrent® se añaden por una etapa de PCR posterior (fig. 10c).

El uso de secuenciación de ADN para una lectura requiere material mínimo. Sin embargo, se requiere una determinada cantidad para producir los octámeros de histona y MN, y para los experimentos de precipitación, por ejemplo usando perlas, y otras técnicas que se pueden usar para proporcionar resultados de alto rendimiento. Estos incluyen implementar dispositivos de microfluidos (Weibel y Whitesides, 2006; Whitesides, 2006) para paralelizar y miniaturizar la formación de MN y octámeros, y también para proporcionar funcionamientos de alto rendimiento usando los bancos para pruebas de creación de perfiles y cribado.

Los modos de realización de la invención incluyen un kit que contiene bancos de MN o CA definidos que se pueden producir y distribuir y usar para crear perfiles interactuantes y/o modificadores de cromatina y para cribar moléculas que modulan su actividad. Los kits, composiciones y procedimientos de la invención se pueden usar para el descubrimiento y creación de perfiles de interactuantes y/o modificadores de cromatina existentes o nuevos; como herramienta de diagnóstico para el análisis de firmas epigenéticas de una línea celular dada, incluyendo las derivadas de pacientes con cáncer humanos; y para el descubrimiento y creación de perfiles de fármacos epigenéticos existentes o nuevos.

Con el uso de las técnicas descritas aquí, se pueden preparar mononucleosomas y matrices de cromatina que tienen cualquier modificación deseada en tipo y número, ya sea a la histona y proteína distinta de histona, el ADN de matriz o nucleosómico, o combinaciones de los mismos. Con hasta aproximadamente 100 modificaciones posibles en las proteínas histonas solas, la combinatoria es extremadamente alta. Sin embargo, solo un número limitado de las modificaciones y combinaciones son biológicamente relevantes. Por tanto, un banco seleccionado puede tener solo cientos de combinaciones de modificaciones de histonas, correspondientes al conjunto limitado de modificaciones biológicamente relevantes típicamente encontrado en eucariotas. Es decir, para la eficacia y relevancia biológica, los bancos pueden excluir las modificaciones de histonas postraduccionales no naturales y las modificaciones no naturales de ADN en mononucleosomas. Por ejemplo, las modificaciones se pueden basar en las encontradas en seres humanos y/o levadura.

Los mononucleosomas y matrices de cromatina pueden diferir de la cromatina natural en varias formas, incluyendo el uso de un código de barras sintético correlacionado con la síntesis y modificación del nucleosoma o matriz, una NPS sintética fuerte, sitios de reconocimiento de ADN y/u otras secuencias de ADN sintéticas. Tienen normalmente

5 alta estabilidad y homogeneidad, lo que permite experimentos controlados sin variables de confusión tales como diferentes secuencias de ADN en cada nucleosoma, o desacoplamiento de ADN de la proteína y "reordenamiento" de los mononucleosomas sintéticos. Las matrices de cromatina pueden tener un número N preciso y predeterminado de unidades de nucleosoma, donde N es 2-96. Todos estos rasgos característicos son importantes para el análisis de alto rendimiento.

Se pueden usar ordenadores para analizar las grandes cantidades de datos que se pueden requerir para algunos análisis de alto rendimiento, tal como el análisis de resultados de grandes cantidades de secuencias de ADN.

10 Por ejemplo, se puede usar un medio legible por ordenador no transitorio, que comprende instrucciones que, cuando se ejecutan por un ordenador, hacen que el ordenador

15 a) identifique la presencia y localización de un código de barras de interés que codifica el experimento particular (múltiple, o código de barras de experimento),

b) compare el código de barras obtenido en a) con una base de datos de códigos de barras indexados con respecto a un experimento particular,

20 c) identifique la presencia y localización de un código de barras de interés en un ADN de nucleosoma de un nucleosoma sintético que ha interactuado con un interactuante o modificador de histona de interés (por ejemplo, en una secuencia de ADN obtenida por secuenciación el ADN de un punto de partida fijo en el ADN de nucleosoma),

25 d) compare el código de barras obtenido en c) con una base de datos de códigos de barras indexados con respecto a patrones particulares de modificaciones de nucleosomas en nucleosomas, y

e) identifique el patrón de modificaciones de nucleosomas asociadas con el código de barras, determinando de este modo la modificación asociada con el interactuante o modificador.

30 Un experto reconocerá las etapas adicionales que se realizan por el ordenador, u otra serie de etapas para llevar a cabo otros procedimientos de la invención.

También se describe un procedimiento para establecer la modificación de nucleosomas asociada con el interactuante o modificador de interés, que comprende

35 a) identificar por (en) un ordenador la presencia y localización de un código de barras de interés que codifica el experimento particular (múltiple, o código de barras de experimento),

40 b) comparar el código de barras obtenido en a) con una base de datos de códigos de barras indexados con respecto a un experimento particular,

45 c) identificar la presencia y localización de un código de barras de interés en un ADN de nucleosoma de un nucleosoma sintético que ha interactuado con un interactuante o modificador de histona de interés (por ejemplo, en una secuencia de ADN obtenida por secuenciación el ADN de un punto de partida fijo en el ADN de nucleosoma),

d) comparar el código de barras obtenido en c) con una base de datos de códigos de barras indexados con respecto a patrones particulares de modificaciones de nucleosomas en nucleosomas; y

50 e) identificar el patrón de modificaciones de nucleosomas asociadas con el código de barras, determinando de este modo la modificación asociada con el interactuante o modificador.

También se describe un sistema para establecer la modificación de nucleosomas asociada con el interactuante o modificador de interés, que comprende

55 memoria y

un procesador configurado para

60 a) identificar la presencia y localización de un código de barras de interés que codifica el experimento particular (múltiple, o código de barras de experimento),

b) comparar el código de barras obtenido en a) con una base de datos de códigos de barras indexados con respecto a un experimento particular,

65

c) identificar la presencia y localización de un código de barras de interés en un ADN de nucleosoma de un nucleosoma sintético que ha interactuado con un interactuante o modificador de histona de interés (por ejemplo, en una secuencia de ADN obtenida por secuenciación el ADN de un punto de partida fijo en el ADN de nucleosoma),

5 d) comparar el código de barras obtenido en c) con una base de datos de códigos de barras indexados con respecto a patrones particulares de modificaciones de nucleosomas en nucleosomas (por ejemplo, con respecto al experimento particular, tal como un múltiple o un código de barras de experimento), y

10 e) identificar el patrón de modificaciones de nucleosomas asociadas con el código de barras, determinando de este modo la modificación asociada con el interactuante o modificador.

La figura 17 representa una arquitectura ejemplar para implementar un dispositivo informático 1700 de acuerdo con uno o más modos de realización, que se puede usar para implementar cualquiera de los dispositivos, o cualquier otro sistema de ordenador o componente de dispositivo informático de los mismos. Se apreciará que otros dispositivos que se pueden usar con el dispositivo informático 1700, tal como un cliente o un servidor, se pueden configurar de forma similar. Como se ilustra en la fig. 17, el dispositivo informático 1700 puede incluir un bus 1710, un procesador 1720, una memoria 1730, una memoria de solo lectura (ROM) 1740, un dispositivo de almacenamiento 1750, un dispositivo de entrada 1760, un dispositivo de salida 1770 y una interfaz de comunicación 1780.

El bus 1710 puede incluir una o más interconexiones que permiten la comunicación entre los componentes del dispositivo informático 1700. El procesador 1720 puede incluir cualquier tipo de procesador, microprocesador o lógica de procesamiento que puede interpretar y ejecutar instrucciones (por ejemplo, una matriz de puertas programable por campo (FPGA)). El procesador 1720 puede incluir un único dispositivo (por ejemplo, un único núcleo) y/o un grupo de dispositivos (por ejemplo, multinúcleo). La memoria 1730 puede incluir una memoria de acceso aleatorio (RAM) u otro tipo de dispositivo de almacenamiento dinámico que puede almacenar información e instrucciones para la ejecución por el procesador 1720. La memoria 1730 también se puede usar para almacenar variables temporales u otra información intermedia durante la ejecución de las instrucciones por el procesador 1720.

La ROM 1740 puede incluir un dispositivo ROM y/u otro tipo de dispositivo de almacenamiento estático que puede almacenar información estática e instrucciones para el procesador 1720. El dispositivo de almacenamiento 1750 puede incluir un disco magnético y/o disco óptico y su correspondiente unidad de disco para almacenar información y/o instrucciones. El dispositivo de almacenamiento 1750 puede incluir un único dispositivo de almacenamiento o múltiples dispositivos de almacenamiento, tales como múltiples dispositivos de almacenamiento que funcionan en paralelo. Además, el dispositivo de almacenamiento 1750 puede residir localmente en el dispositivo informático 1700 y/o puede ser remoto con respecto a un servidor y conectarse el mismo por medio de red y/u otro tipo de conexión, tal como un enlace o canal dedicado.

El dispositivo de entrada 1760 puede incluir cualquier mecanismo o combinación de mecanismos que permiten a un operador introducir información al dispositivo informático 1700, tal como un teclado, un ratón, un dispositivo de pantalla táctil, un micrófono, un dispositivo señalador de tipo lápiz y/o un dispositivo de entrada biométrico, tal como un dispositivo de reconocimiento de voz y/o un dispositivo de escaneo de huellas. El dispositivo de salida 1770 puede incluir cualquier mecanismo o combinación de mecanismos que extrae información para el operador, incluyendo una pantalla, una impresora, un altavoz, etc.

La interfaz de comunicación 1780 puede incluir cualquier mecanismo de tipo transceptor que permite que el dispositivo informático 1700 se comunique con otros dispositivos y/o sistemas, tal como un cliente, un servidor, un administrador de licencias, un proveedor, etc. Por ejemplo, la interfaz de comunicación 1780 puede incluir una o más interfaces, tales como una primera interfaz acoplada a una red y/o una segunda interfaz acoplada a un administrador de licencias. De forma alternativa, la interfaz de comunicación 1780 puede incluir otros mecanismos (por ejemplo, una interfaz inalámbrica) para la comunicación por medio de una red, tal como una red inalámbrica. En un equipamiento, la interfaz de comunicación 1780 puede incluir lógica para enviar código a un dispositivo de destino, tal como un dispositivo de destino que puede incluir equipo físico de uso general (por ejemplo, un factor de forma de ordenador personal), equipo físico dedicado (por ejemplo, un dispositivo de procesamiento de señal digital (DSP) adaptado para ejecutar una versión compilada de un modelo o una parte de un modelo), etc.

El dispositivo informático 1700 puede realizar determinadas funciones en respuesta al procesador 1720 ejecutar instrucciones de programas informáticos contenidas en un medio legible por ordenador, tal como la memoria 1730. En modos de realización alternativos, se pueden usar circuitos cableados en lugar de o en combinación con instrucciones de programas informáticos para implementar rasgos característicos consecuentes con los principios de la divulgación. Por tanto, las implementaciones consecuentes con los principios de la divulgación no se limitan a cualquier combinación específica de circuitos de equipos físicos y programas informáticos.

Los modos de realización ejemplares se pueden realizar de muchas formas diferentes como un componente de programa informático. Por ejemplo, puede ser un paquete informático independiente, una combinación de paquetes

informáticos, o puede ser un paquete informático incorporado como una "herramienta" en un producto de programa informático más grande. Se puede descargar de una red, por ejemplo, un sitio web, como un producto independiente o como un paquete de complemento para su instalación en una aplicación informática existente. También puede estar disponible como aplicación informática cliente-servidor, o como una aplicación informática habilitada para web. También se puede realizar como un paquete informático en un dispositivo físico.

Se han expuesto numerosos detalles específicos para proporcionar una comprensión completa de los modos de realización. Sin embargo, se entenderá que se pueden practicar los modos de realización sin estos detalles específicos. En otros casos, no se han descrito en detalle funcionamientos, componentes y circuitos bien conocidos para no confundir los modos de realización. Se puede apreciar que los detalles estructurales y funcionales específicos son representativos y no limitan necesariamente el alcance de los modos de realización.

Aunque se pueden ilustrar y describir algunos modos de realización que comprenden módulos o componentes funcionales ejemplares que realizan diversos funcionamientos, se puede apreciar que se pueden implementar dichos módulos o componentes por uno o más componentes de equipos físicos, componentes de programas informáticos y/o combinación de los mismos. Los módulos y/o componentes funcionales se pueden implementar, por ejemplo, por lógica (por ejemplo, instrucciones, datos y/o código) que se va a ejecutar por un dispositivo lógico (por ejemplo, procesador). Dicha lógica se puede almacenar de forma interna o externa a un dispositivo lógico en uno o más tipos de medios de almacenamiento legible por ordenador.

Algunos modos de realización pueden comprender un artículo de fabricación. Un artículo de fabricación puede comprender un medio de almacenamiento para almacenar lógica. Los ejemplos de un medio de almacenamiento pueden incluir uno o más tipos de medios de almacenamiento legibles por ordenador que pueden almacenar datos electrónicos, incluyendo memoria volátil o memoria no volátil, memoria extraíble o no extraíble, memoria borrable o no borrable, memoria grabable o regrabable, etcétera. Los ejemplos de medios de almacenamiento incluyen discos duros, unidades de discos, unidades de estado sólido y cualquier otro medio de almacenamiento tangible.

También se apreciará que los modos de realización descritos ilustran implementaciones ejemplares, y que los módulos y/o componentes funcionales se pueden implementar de otras diversas formas que sean consecuentes con los modos de realización descritos. Además, los funcionamientos realizados por dichos módulos o componentes se pueden combinar y/o separar para una implementación dada y se pueden realizar por un número mayor o número menor de módulos o componentes.

Algunas de las figuras pueden incluir un diagrama de flujo. Aunque dichas figuras pueden incluir un flujo lógico particular, se puede apreciar que el flujo lógico proporciona simplemente una implementación ejemplar de la funcionalidad general. Además, el flujo lógico no se tiene que ejecutar necesariamente en el orden presentado a menos que se indique de otro modo. Además, el flujo lógico se puede implementar por un elemento físico, un elemento informático ejecutado por un procesador, o cualquier combinación de los mismos.

EJEMPLOS

Los siguientes experimentos demuestran la viabilidad de la estrategia de aplicación de códigos de barras para codificar la composición de variantes de MN para la identificación del patrón de PTM preferente en el contexto nucleosómico de un lector de histona, un escritor de histona y la actividad de lectura, escritura y borrado de histona combinadas de un lisado celular nuclear derivado de células 293T humanas, que se puede lograr en una semana, comenzando a partir de histonas recombinantes y/o sintéticas disponibles (11a y b).

Se generó un banco de 39 miembros de variantes de MN con código de barras individual, que contiene combinaciones de histonas wt y/o las histonas modificadas H2A, H2B, H3 y H4 (fig. 12a).

Se expresaron las histonas humanas wt en *E.coli* y se purificaron usando metodologías descritas (Dyer *et al.*, 2004; Luger *et al.*, 1999). Se prepararon las histonas H2A, H2B, H3 y H4 humanas modificadas usando NCL de péptidos N terminales preparados por síntesis peptídica en fase sólida e histonas truncadas de forma N terminal recombinantes con una cisteína N terminal (fig. 9a) (Dawson y Kent, 2000; Fierz *et al.*, 2011; Fierz, Kilic, Hieb, Luger y Muir, 2012) con las siguientes PTM: ubiquitinación de lisina (ub) en K119 y K120 para H2A y H2B, respectivamente, trimetilación de lisina (me3) en K4, K9, K27 así como pentaacetilación de lisina (acpoli, K9/14/18/23/27ac) para H3, y monoacetilación de lisina (ac) en K5, K8, K12, K16, K20, así como pentaacetilación de lisina (H4Kac₅, K5/8/12/16/20ac) para H4. Se purificaron las proteínas por HPLC de fase inversa C18 hasta una pureza de un > 95 %, como se estima por ESI-EM. Se combinaron las histonas para formar 39 nucleosomas con diferentes combinaciones de PTM (para una lista completa, véase la fig. 12b).

Se usó fijación de ADN T4 para unir los códigos de barras de MN, junto con los nucleótidos 10-30 del sitio de cebado FW compatible con el secuenciador IonTorrent® (Rothberg *et al.*, 2011), a las secuencias de ADN 601 nucleosómico en el extremo 5' para codificar cada variante de MN única, dando como resultado moléculas de ADN que contenían un sitio de cebado FW de 20 pb (con un saliente 5'-AA-3' monocatenario (ss) en el extremo 3' de la hebra de ADN inferior), un código de barras de 6 pb (que codifica la variante de MN respectiva), un fijador de 4 pb (que se origina

del sitio de fijación de ADN Bsal no palindrómico), la secuencia de ADN nucleosómico 601 de 147 pb, y un apéndice 5'-CAC-3' de 3 nt en el extremo 3' de la hebra de ADN superior como resultado de la digestión de restricción DraIII del ADN nucleosómico 601 (fig. 10b). En particular, se preparó el bloque constitutivo de 601 de 147 pb que contiene un saliente 5'-Bsal y un 3'-DraIII por digestión de un fragmento de ADN producido por (a) Phusion PCR usando un

molde de ADN 601 (Lowary y Widom, 1998) (cebadores purificados por HPLC; cebador FW: 5'-ACCCTAGTCTCTGATGCTGGAGAATCCCGGTGCCGAGG-3' (SEQ ID NO: 2), cebador RV: 5'-CTACCACATCGTGGGATGTATATATCTGACACGTGCCTGG-3') (SEQ ID NO: 3) o de (b) un plásmido que contiene 12 copias de la secuencia deseada (flanqueada por sitios EcoRV en cada lado; secuencia completa de 1 unidad repetitiva:

5'-GATATCACCCCTAGGTCTCTGATGCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTG

GTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTTA

ACCGCCAAGGGGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCCTGT

CACGCGGTGAACAGCGATATC -3') (SEQ ID NO: 4). Se produjo el producto de PCR a una escala de 500 mg y se purificó usando un kit de purificación de PCR de Qiagen. Se digirió el producto de PCR a 20 °C durante 20 h con

Bsal y DraIII (0,5 mg de ADN en un volumen total de 0,5 ml que contenía 100 U/ml de cada enzima de restricción) para liberar la secuencia de 601 de 147 pb, se purificó usando un kit de purificación de PCR de Qiagen, se precipitó

con etanol (70 % vol), se redisolvió en tampón de elución (Tris-HCl 10 mM, pH 8,5, EB), se cuantificó por espectroscopía UV a 260 nm con sustracción de fondo a 340 nm ($\epsilon = 2407002 \text{ l} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$), y se almacenó en

alícuotas a -80 °C. Se produjo el plásmido en células competentes DH5 α en una escala ca 1 l y se purificó usando un kit Qiafilter Plasmid Giga de Qiagen. Se digirió el plásmido a 20 °C durante 20 h con Bsal y DraIII para liberar la

secuencia 601 de 147 pb (5 mg de ADN en un volumen total de 5 ml que contenía 100 U/ml de cada enzima de restricción). Se purificaron las reacciones de digestión por electroforesis con acrilamida y precipitación del producto

purificado usando un 70 % vol. de etanol. Se redisolvió el sedimento en EB, se cuantificó por espectroscopía UV a 260 nm con sustracción de fondo a 340 nm ($\epsilon = 2407002 \text{ l} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$) y se almacenó en alícuotas a -80 °C. Se

produjo el ADN ds FW-iT₁₀₋₃₀-BC-MN hibridando cantidades iguales de 5'-CCTGCGTGTCTCCGACTCAGHXXXXH-3' (SEQ ID NO: 5) (hebra superior) y 5'-CATCDXXXXDCTGAGTCCGAGACACGCAGGAA-3' (SEQ ID NO: 6) (hebra inferior) por incubación a 95 °C durante 5 min y se enfrió lentamente hasta TA durante 1 h (fig. 10b). En una reacción

de fijación típica, se combinaron 1,25 μM del bloque constitutivo de 601 con 1,1 eq FW-iT₁₀₋₃₀ ds y se incubó con 0,1 U/ μl polinucleótido cinasa (PNK) en un volumen de 800 μl de tampón ligasa T4 durante 1 h a 37 °C. Posteriormente, se añadieron 10 U/ μl de ADN ligasa T4 y se incubó durante 1 h a TA (fig. 10b). Se controló la

reacción de fijación por electroforesis en gel natural seguido de tinción de ADN con bromuro de etidio (gel de acrilamida al 5 %, 200 V, 40 min; fig. 12c). Se purificó el producto final usando un kit de purificación de PCR de Qiagen, se eluyó con 50 μl de EB, y se cuantificó por espectroscopía UV a 260 nm con sustracción de fondo a 340 nm ($\epsilon = 2886629 \text{ l} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$). Se preparó el ADN BIO-MMTV por Phusion PCR usando un molde de ADN

MMTV (Flaus y Richmond, 1998) con los siguientes cebadores: 5'-biotina-

TATCACTTGCAACAGTCCCTAACATTCACCTC-3' (SEQ ID NO: 7) (cebador FW) y 5'-

ATCCAAAAAAGTGTGCCGAGTCCGG-3' (SEQ ID NO: 8) (cebador RV). Se purificó el producto de PCR usando un kit de purificación de PCR de QIAGEN seguido de precipitación usando 70 % vol. de etanol, redisolviendo el

sedimento en EB, se cuantificó por espectroscopía UV a 260 nm con sustracción de fondo a 340 nm ($\epsilon = 2414925 \text{ l} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$), y se almacenó en alícuotas a -80 °C.

Se ensamblaron los octámeros con las respectivas variantes de histonas replegando cantidades estequiométricas de histonas individuales de GdmHCl a una escala de 1 nmol (con respecto a cada variante de histona, aproximadamente 50 μg de histona total, dependiendo de la variante de histona) sin purificación adicional (fig. 9b). Se midieron las concentraciones por espectrometría UV a 280 nm y sustracción de fondo a 300 nm (coeficientes de

extinción: $\epsilon_{\text{H2A}} = 4470 \text{ l} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$; $\epsilon_{\text{H2B}} = 7450 \text{ l} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$; $\epsilon_{\text{H3}} = 4470 \text{ l} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$; $\epsilon_{\text{H4}} = 5960 \text{ l} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$). Se mezclaron las histonas en proporciones equimolares en hielo en aproximadamente 55 μl de tampón de desplegado (cloruro de guanidinio 6 M, Tris-HCl 20 mM, pH 7,5 a 4 °C, Na-EDTA 1 mM, DTT 1 mM) para proporcionar una

concentración de proteína total de aproximadamente 1 mg/ml. Se dispusieron las mezclas en botones de diálisis MINI (valor de corte 3.500 Da) y se dializó frente a 3x 600 ml de tampón de replegado (NaCl 2 M, Tris-HCl 10 mM, pH 7,5 a 4 °C, Na-EDTA 1 mM, DTT 1 mM) durante al menos 4 h cada una a 4 °C, con una etapa de diálisis durante la noche. Al día siguiente, las mezclas se transfirieron a tubos Eppendorf y se hacen girar a 17.000 g durante al

menos 5 min a 4 °C para retirar precipitados. Se transfirieron los sobrenadantes a tubos limpios. Se añadió glicerol al 50 % (v/v), y se midieron las concentraciones de octámeros usando su absorción UV ($\epsilon_{\text{Octámero}} = 35760 \text{ l} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$) y estuvieron típicamente en el intervalo de 2-5 μM . Se procesó directamente una fracción de los octámeros para los

ensamblajes de MN, y se almacenó el resto a -20 °C. Los rendimientos típicos fueron de un 60-80 % para dar aproximadamente 300-400 pmol de octámeros de histona.

En este caso particular, solo se usó una fracción de los octámeros de histona disponibles para el ensamblaje de nucleosoma adicional. Se combinaron octámeros de histona apropiados a una concentración de aproximadamente 1,0 μM en 70 μl de tampón de reconstitución con 0,6 μM de ADN 'BC-601' con código de barras (fig. 9b) y 0,4 μM de ADN BIO-MMTV tampón a 4 °C. Se transfirió la mezcla a una unidad de diálisis Slide-A-Lyzer MINI y se dializó a 4 °C frente a tampón de reconstitución que contenía 200 ml de KCl 1,4 M durante 1 h. Posteriormente, se añadieron 330 ml de tampón que contenía KCl 10 mM a una tasa de 1 ml/min, seguido de dos etapas de diálisis finales frente a un tampón de reconstitución que contenía KCl 10 mM (1 h y durante la noche). Se logró la purificación de complejos

de histona ensamblados no productivamente y BIO-MMTV potencialmente libre por purificación por afinidad con estreptavidina usando 50 μl de MyOne Dynabeads (Invitrogen) durante 1 h a TA, seguido de la adición de un 20 %

vol. de glicerol y PMSF 0,5 mM. Se evaluó la calidad de los nucleosomas por separación en un gel de acrilamida al 5 % pasado en 0,5 x tampón Tris/borato/EDTA (TBE) (200 V, 40 min), seguido de tinción de ADN con bromuro de etidio (fig. 13, izquierda). Se cuantificaron los MN por espectroscopía UV a 260 nm con sustracción de fondo a 340 nm ($\epsilon = 2886629 \text{ l} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$). Se combinaron los nucleosomas con código de barras en proporciones equimolares para formar el banco y se concentró usando unidades de filtro centrífugo Vivaspin 500 (valor de corte de peso molecular 10.000 Da). Se ultracongeló el banco de nucleosomas en alícuotas a una concentración de aproximadamente 1 μM y se almacenó a - 80 °C. Los rendimientos típicos estuvieron próximos a un 60 %, dando como resultado un rendimiento global de aproximadamente un 40 %, comenzando a partir de las histonas individuales. Por ejemplo, 1 nmol de cada histona proporcionaría aproximadamente 200 pmoles de nucleosomas finales, lo que es suficiente para aproximadamente 1000 experimentos de unión a lectores de histonas (véase el ejemplo 1A y 1B), o > 10.000 experimentos de escritores de histonas enzimáticos (véase el ejemplo 2 y 3).

Se evaluó la integridad del banco de MN agrupado después del almacenamiento prolongado durante > 1 mes a 4 °C por PAGE natural, seguido de tinción de ADN por bromuro de etidio (fig. 13, derecha). Adicionalmente, se sondaron los nucleosomas individuales para determinar su estabilidad en solución, en particular con respecto al reordenamiento de ADN. Por lo tanto, se inmunoprecipitó el banco con anticuerpos específicos dirigidos frente a una marca preinstalada, α -H3K4me3 (abcam, ab8580). Se complementaron 15 μl de una mezcla de nucleosomas de banco (total) 30 nM (es decir, 12 fmoles de cada variante de MN por precipitación de anticuerpo) en tampón (Tris 50 mM, pH 8,0, EDTA 0,1 mM, PMSF 1 mM, Na-butilato 100 mM, glicerol al 10 %, DTT 1 mM) con el anticuerpo α -H3K4me3 en un volumen total de 100 μl de tampón de anticuerpo (AB) (Tris-HCl 20 mM, pH 7,5, NaCl 50 mM, EDTA 5 mM) para dar una concentración de anticuerpo final de 15 $\mu\text{g}/\text{ml}$, y se incubó a TA durante 1 h en un rotor de extremo a extremo. Posteriormente, se añadieron 100 μl de tampón AB así como 10 μl de una suspensión de perlas de proteína G (Invitrogen) y se incubó la mezcla en un rotor de extremo a extremo durante 1 h a TA. Se lavaron las perlas 4 veces con 200 μl de tampón AB, y se eluyó el ADN usando 100 μl de tampón de elución de ADN (Tris 100 mM, pH 7,8, EDTA 10 mM, SDS al 1 % SDS, β -mercaptoetanol 10 mM (β ME), 200 $\mu\text{g}/\text{ml}$ de proteinasa K, NEB) durante 1,5 h a 50 °C y se purificó usando el kit de purificación de PCR de Qiagen. Se eluyó el ADN de precipitación resultante con 50 μl de tampón EB y se cuantificó por el kit de cuantificación de ADN de alta sensibilidad Qubit (Invitrogen). Se diluyó el ADN con H₂O hasta una concentración final de 2 $\text{pg}/\mu\text{l}$ (los factores de dilución del respectivo experimento se consideraron después durante el análisis de los datos de secuenciación, véase a continuación). Las muestras de entrada (10-50 % de la entrada de banco inicial, dependiendo del experimento) se procesaron de forma idéntica. Se produjo una mezcla estándar interna para la etapa de PCR de multiplexación por Phusion PCR usando un molde de 601 y los siguientes cebadores: FW: 5'-CCTGCGTGTCTCCGACTCAGXXXXXXXXGATGCTGGAGAATCCCGTGCCGAGG-3' (SEQ ID NO: 9) (A estándar: CTCAGT, B estándar: CATGCT, C estándar: TGAGTC, D estándar: ACTGCA); RV: 5'-GTGACAGGATGTATATATCTGACACGTGCCTGG-3' (SEQ ID NO: 10). Se purificaron los productos de PCR usando un kit de purificación Qiagen, se eluyó con tampón de ADN EB, se cuantificó usando el kit de cuantificación de ADN de alta sensibilidad Qubit (Invitrogen), y se mezcló en EB hasta una concentración de ADN total de 2 $\text{pg}/\mu\text{l}$ con la siguiente distribución: 1.000 eq de A estándar, 100 eq de B estándar, 10 eq de C estándar y 1 eq de D estándar.

En una reacción de PCR múltiple típica, se combinaron 9 pg de cada ADN de precipitación con 1 pg de la mezcla estándar interna en presencia de 0,01 U/ μl de Phusion, dNTPs 0,2 mM, 0,5 μM de cada uno del cebador FW (FW-iT: 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3') (SEQ ID NO: 11) y el respectivo cebador RV con código de barras (RV-601-BC-EXP-RV-iT: 5'-CCTCTCTATGGGCAGTCCGGTATBXXXXDGGTGCTAGAGCTGTCTACGACCAATTG AGC-3' (SEQ ID NO: 12); programa de ciclos de PCR: desnaturalización inicial, 30 s/98 °C; desnaturalización, 10 s/98 °C; hibridación, 15 s/62 °C; extensión, 5 s/72 °C; total 15 ciclos; extensión final, 7 min/72 °C; fig. 10c). Se purificaron los productos de PCR usando un kit de purificación de PCR de Qiagen y se eluyó con 50 μl de EB. Se agruparon las secuencias de ADN multiplexadas en volúmenes iguales y se secuenció usando un aparato IonTorrent Personal Genome Machine de acuerdo con las instrucciones del fabricante (Merriman, R D Team, y Rothberg, 2012; Rothberg *et al.*, 2011). Para el análisis de datos, en primer lugar se clasificaron las lecturas de secuenciación brutas de acuerdo con los códigos de barras experimentales en 3'. Posteriormente, se clasificaron las lecturas de acuerdo con los códigos de barras de MN en 5' (fig. 10d). Se multiplicaron las lecturas individuales de cada variante de MN por el factor de dilución de ADN de la muestra de ADN del experimento de precipitación respectivo. Finalmente, se normalizaron estas lecturas corregidas frente a las lecturas de los MN individuales de la entrada de banco (promediadas a partir de al menos 2 muestras de entrada). Los valores resultantes representaron las eficacias de precipitación de los MN individuales expresadas como % entrada en el experimento respectivo. En algunos casos, se normalizaron además las eficacias de precipitación frente a la eficacia de precipitación de una variante.

Se aislaron solo los nucleosomas que portaban un código de barras de H3K4me3 cuando se usó un anticuerpo α -H3K4me3 en la precipitación (eficacia de precipitación de un 50-60 % en comparación con la entrada); los equipados con diferentes PTM se unieron a niveles de fondo únicamente (eficacia de precipitación menor de un 0,5 % en comparación con la entrada), lo que sugiere que no se observó intercambio de ADN entre los miembros de banco, incluso cuando se expuso el banco a ciclos de congelación-descongelación o almacenamiento prolongado durante meses a 4 °C (fig. 14a).

EJEMPLO 1A: creación de perfiles de dos módulos de lectores de histonas adyacentes de un factor de transcripción de los dedos de homeodominio vegetal (PHD) del bromodominio BPTF expresado de forma recombinante: con un banco de nucleosomas diversamente modificado y estable en cuestión, se trató de perfilar las propiedades de unión del BPTF. Puesto que esta colección de MN cubrió variantes que no se han investigado previamente en el contexto de BPTF (Ruthenburg *et al.*, 2011), tal como las que portan Kac₅ en H3 o H4, Kme3 en las posiciones 9 y 27 en H3, o Kub en H2A y H2B (fig. 12a y b) (Ruthenburg *et al.*, 2011), el banco permitió interrogar sistemáticamente el comportamiento de unión nucleosómica multivalente de BPTF con un tamaño de muestreo de sustrato mayor. Adicionalmente, se minimizaron las variaciones experimentales ya que todos los nucleosomas se sondaron para la unión a BPTF simultáneamente en un único experimento, una ventaja clave única de los nucleosomas con código de barras.

Un exceso de construcciones de BPTF de glutatión-S-transferasa (GST) fusionadas de forma N terminal que contenía el módulo PHD-BD acoplado o bien los respectivos dominios individuales (aproximadamente 200 pmoles, fig. 13b) se inmovilizaron en 3 µl de una suspensión de resina de glutatión (1 h, TA) y se incubó con el banco de nucleosomas codificado (0,125 pmoles de cada variante de MN) en un volumen de 200 µl en tampón GDP300 (HEPES 20 mM, pH 7,9, KCl 30 mM, 20 % p/v de glicerol, EDTA 0,2 mM, DTT 1 mM, NP-40 al 0,01 %; concentración de nucleosoma final: 24 nM) durante 4 h a 4 °C en un nutador de extremo a extremo. Se lavaron las perlas 4 veces con 200 µl de tampón GDP300. Se realizaron el aislamiento, purificación, multiplexación, descodificación y normalización de ADN como se describe anteriormente. BPTF-BD con marca GST no interactuó con nucleosomas acetilados en las condiciones sometidas a prueba, lo que es consecuente con los resultados del trabajo previo (fig. 14c). Por el contrario, BPTF-PHD con marca GST reconoció los nucleosomas trimetilados en H3K4, independientemente de si estaban presentes otras marcas en el mismo nucleosoma o no. No se observó unión cuando se instaló cualquier Kme3 en otras posiciones dentro de la cola H3 o se unió la marca Kub voluminosa a H2A o H2B (fig. 13d, rojo, derecha). Cuando se usó el dominio PHD-BD doble como señuelo, se aislaron todos los nucleosomas con marcas H3K4me3, mientras que los que solo llevaban lisina(s) acetilada(s) no se aislaron (fig. 13d, rojo, izquierda). Sin embargo, la coexistencia de H3K4me3 y H4Kac dentro del mismo nucleosoma potenció la eficacia de unión en 1,5-3,5 veces (marcas individuales en K5, K8, K12, K16 o K20) en comparación con H3K4me3 solo (fig. 13d, rojo). Se había observado un incremento de 2-3 veces en la unión nucleosómica previamente con nucleosomas modificados con H3K4me3-H4K16ac, y la contribución de Kac a la multivalencia fue menos pronunciada en sitios adyacentes en la cola H4 (Ruthenburg *et al.*, 2011). En este caso, esta dependencia posicional se nivela, aunque se encontró una potenciación fuerte de aproximadamente 7 veces con nucleosomas que contenían H4Kac₅, una variante que no se había caracterizado previamente (fig.13d, rojo). Puesto que el bolsillo de unión a Kac del BD de BPTF solo puede alojar una lisina acetilada, es probable que la afinidad potenciada sea una consecuencia de la aidez de unión resultante de múltiples lisinas acetiladas en la cola H4. Se verificaron estos hallazgos usando experimentos de precipitación aislados tradicionales y se encontraron comportamientos de unión similares.

EJEMPLO 1B: creación de perfiles de dos módulos de lectores de histonas adyacentes (potencialmente) de una histona acetiltransferasa p300 expresada de forma recombinante: Habiendo demostrado la capacidad de esta tecnología para examinar rápidamente el comportamiento de unión multivalente de BPTF, se decidió extender la investigación al coactivador transcripcional p300. Se incubó p300-BD-PHD con marca GST recombinante (fig. 15a) con el banco de nucleosomas como se describe en el ejemplo 1B. Se realizaron el aislamiento de fijadores de nucleosomas, aislamiento, purificación, multiplexación, descodificación y normalización de ADN como se describe anteriormente. Aunque la unión a nucleosoma fue más débil que para BPTF en las condiciones usadas, se detectó una asociación fuerte del módulo p300-BD-PHD con marca GST con nucleosomas que contenían H4Kac₅ (fig. 15b, rojo). Puesto que las marcas Kme3 adicionales no corroboraron ni debilitaron la fuerza de asociación, este nuevo hallazgo propone que el dedo de PHD no se une a H3K4 trimetilado o bien no modificado. Para desglosar las contribuciones individuales del dedo de PHD del BD para la asociación con nucleosomas acetilados, se repitieron los experimentos con cada dominio solo. Aunque el dedo de PHD no mostró una interacción detectable para cualquier variante de nucleosoma (fig. 15c), el BD presentó un comportamiento de unión similar al módulo BD-PHD (fig. 14d), lo que sugiere que el dedo de PHD es prescindible para esta asociación. Conjuntamente con informes anteriores, estos datos de precipitación de banco permitieron generar la hipótesis de que p300 puede funcionar por medio de un bucle de retroalimentación positiva por unión a sus propias marcas y/o bien por una regulación alostérica mediada por Kac, potencialmente a través de su BD. Esta tecnología desarrollada es altamente adecuada para someter a prueba directamente dicha teoría debido a su modularidad y capacidad para combinar la unión con ensayos enzimáticos usando el mismo banco de nucleosomas.

EJEMPLO 2: creación de perfiles de un escritor de histona expresado de forma recombinante, la histona acetiltransferasa p300: Se incubaron los nucleosomas codificados (a una concentración de 30 nM) con 3 nM de p300 humana de longitud completa recombinante, se preparó en células de insecto Sf9, en ausencia o presencia de AcCoA 10 µM durante 1 h a 30 °C en tampón de histona acetiltransferasa (HAT) (Tris 50 mM, pH 8,0, EDTA 0,1 mM, PMSF 1 mM, Na-butilato 10 mM, glicerol al 10 %, DTT 1 mM). Se aislaron los productos de reacción, o subconjuntos de los mismos, por inmunoprecipitación usando anticuerpos específicos para H4K5ac y H3K18ac (fig.16 a y b). En resumen, se complementaron 15 µl de la reacción con el respectivo anticuerpo (es decir, 12 fmoles de cada variante de MN por reacción y precipitación de anticuerpo), en este caso los anticuerpos α-H3K18ac o α-H4K5ac, se ajustó a un volumen total de 100 µl de tampón AB para dar una concentración de anticuerpo final de 15 µg/ml, y se incubó

durante 1 h a TA en un rotor de extremo a extremo. Posteriormente, se añadieron 100 µl de tampón AB así como 10 µl de una suspensión de perlas de proteína G y se incubó en un rotor de extremo a extremo durante 1,5 h a TA. Se lavaron las perlas 4 veces con 200 µl de tampón AB. Se realizaron la precipitación de anticuerpos, aislamiento, purificación, multiplexación, descodificación y normalización de ADN como se describe anteriormente. Cuando se trató el banco con AcCoA y se inmunoprecipitó con α-H3K18ac, se observó un incremento en la precipitación de entrada de un 0,3 % a aproximadamente un 4 % para nucleosomas wt, lo que reflejó un incremento en la acetilación de lisina en ese sitio provocada por p300 (fig. 16a). La extensión de la acetilación no se vio influenciada por la existencia de las marcas Kme3 o Kub. Sin embargo, se encontró una acetilación de H3 significativamente más fuerte (fig. 16a, 13 veces en comparación con wt) en los nucleosomas poliacetilados en la cola H4, que está en consonancia con nuestros datos de unión de banco (fig. 15a). De forma interesante, tan solo una marca Kac en H4 puede recapitular esta interferencia de H4Kac a H3Kac, aunque con menor eficacia. Este efecto fue dependiente de la posición exacta de la marca dentro de la cola y fue más pronunciada para H4K12ac (fig. 16a). Para someter a prueba si existió interferencia en el sentido inverso, se inmunoprecipitó la mezcla de reacción con un anticuerpo α-H4K5ac (fig. 16b). Se observó un pequeño incremento en la acetilación para nucleosomas wt en presencia de AcCoA (entrada de un 0,5 a un 0,9 %), lo que no se vio influenciado por las modificaciones de Kub o Kme3 (fig. 16b, derecha). Por el contrario, los nucleosomas que contenían H3Kac₅ se modificaron más fácilmente (fig. 16b, incremento de 6 veces en comparación con wt), lo que sugiere una interferencia entre colas de H3Kac a H4Kac además. Adicionalmente, los nucleosomas con marcas de K12ac preexistentes dieron lugar a una eficacia de precipitación elevada (entrada de un 4,4 %, fig. 16b, derecha), lo que propone una interferencia entre colas adicional dentro de la cola H4. Sin embargo, este incremento también pudo ser el resultado de una afinidad de anticuerpo mayor por nucleosomas poliacetilados, en comparación con monoacetilados. Estos experimentos muestran que p300 funciona por medio de un mecanismo de retroalimentación positiva a través de interferencias entre colas, y potencialmente intracoladas, mediadas por Kac dependientes de la posición. Por lo tanto, se probaron estos nuevos hallazgos con nucleosomas modificados individualmente tradicionales que contienen marcas Kac preinstaladas dentro de la cola H3 o H4 con procedimientos independientes de anticuerpos, tales como ensayos enzimáticos basados en geles radioactivos y espectrometría de masas, lo que confirma los resultados de cribado de banco.

EJEMPLO 3: creación de perfiles de lectores, escritores y borradores de histonas de un extracto celular nuclear derivado de células 293T humanas: Para investigar la firma epigenética de una línea celular dada, se creó un perfil de banco de nucleosomas con código de barras con un extracto nuclear derivado de células 293T humanas preparadas como se describe anteriormente (Dignam, Lebovitz y Roeder, 1983). Se incubaron 15 µl del banco de nucleosomas codificados (a una concentración de 30 nM, es decir, 12 fmoles de cada variante de MN por reacción y precipitación de anticuerpo) con 7,5 µl del extracto nuclear en presencia de AcCoA 20 µM, SAM 10 µM y ATP 10 µM durante 1 h a 30 °C en tampón HAT. Se aislaron los nucleosomas que se acetilaron en H3K14 como resultado de las actividades de acetiltransferasa de las HAT endógenas en el extracto nuclear usando un anticuerpo α-H3K14ac. Se realizaron la precipitación de anticuerpos, aislamiento, purificación, multiplexación, descodificación y normalización de ADN como se describe anteriormente. Se observó un incremento en la acetilación de H3K14 de una HAT no caracterizada, potencialmente p300, nucleosomas wt. Se incrementó significativamente la extensión de la acetilación para las variantes de MN que portaban marcas H3K4me3 preinstaladas y/o una sola de múltiples Kac dentro de la cola H4 (fig. 16c), lo que sugiere interferencias enzimáticas escritores, borradores y/o lectores de histonas. Estos resultados son intrigantes puesto que demuestran que este procedimiento experimental se puede explotar como herramienta de diagnóstico rápido y eficaz para investigar la firma epigenética de una línea celular dada, incluyendo la derivada de pacientes con cáncer humanos.

De la descripción anterior, un experto en la técnica puede averiguar fácilmente las características esenciales de la presente invención, como se define por las reivindicaciones adjuntas.

Referencias.

- Agasti, S. S., Liong, M., Peterson, V. M., Lee, H., & Weissleder, R. (2012). Photocleavable DNA Barcode-Antibody Conjugates Allow Sensitive and Multiplexed Protein Analysis in Single Cells. *Journal of the American Chemical Society*, 134(45), 18499-18502. doi:10.1021/ja307689w
- Allis, C. D., & Muir, T. W. (2011). Spreading chromatin into chemical biology. *Chembiochem : a European journal of chemical biology*, 12(2), 264-279. doi:10.1002/cbic.201000761
- Bao, Y., Chakravarthy, S., Muthurajan, U. M., & Luger, K. (2003). Reconstitution of nucleosome core particles from recombinant histones and DNA.... in *enzymology*.
- Blacketer, M. J., Feely, S. J., & Shogren-Knaak, M. A. (2010). Nucleosome interactions and stability in an ordered nucleosome array model system. *The Journal of biological chemistry*, 285(45), 34597-34607. doi:10.1074/jbc.M110.140061
- Britton, L.-M. P., Gonzales-Cope, M., Zee, B. M., & Garcia, B. A. (2011). Breaking the histone code with quantitative mass spectrometry. *Expert review of proteomics*, 8(5), 631-643. doi:10.1586/epr.11.47

- Buller, F., Mannocci, L., & Scheuermann, J. (2010). Drug discovery with DNA-encoded chemical libraries. *Bioconjugate*
- 5 Cao, Y., Korolev, N., & Nordenskiöld, L. (2011). A Direct Method for Site-Specific Protein Acetylation. *Angewandte Chemie*
- Clark, M. A. (2010). Selecting chemicals: the emerging utility of DNA-encoded libraries. *Current Opinion in Chemical Biology*, 14(3), 396-403. doi:10.1016/j.cbpa.2010.02.017
- 10 Dai, J., Hyland, E. M., Yuan, D. S., Huang, H., & Bader, J. S. (2008). Probing nucleosome function: a highly versatile library of synthetic histone H3 and H4 mutants. *Cell*.
- Dawson, P. E., & Kent, S. B. (2000). Synthesis of native proteins by chemical ligation. *Annual Review of Biochemistry*, 69, 923-960. doi:10.1146/annurev.biochem.69.1.923
- 15 Dignam, J. D., Lebovitz, R. M., & Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Research*.
- Dyer, P. N., Edayathumangalam, R. S., White, C. L., Bao, Y., Chakravarthy, S., Muthurajan, U. M., & Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods in enzymology*, 375, 23-44.
- 20 ENCODE Project Consortium, Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57-74. doi:10.1038/nature11247
- 25 Fierz, B., & Muir, T. W. (2012). Chromatin as an expansive canvas for chemical biology. *Nature chemical biology*, 8(5), 417-427. doi:10.1038/nchembio.938
- 30 Fierz, B., Chatterjee, C., McGinty, R. K., Bar-Dagan, M., Raleigh, D. P., & Muir, T. W. (2011). Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nature chemical biology*, 7(2), 113-119. doi:10.1038/nchembio.501
- 35 Fierz, B., Kilic, S., Hieb, A. R., Luger, K., & Muir, T. W. (2012). Stability of Nucleosomes Containing Homogenously Ubiquitylated H2A and H2B Prepared Using Semisynthesis. *Journal of the American Chemical Society*, 134(48), 19548-19551. doi:10.1021/ja308908p
- Flaus, A., & Richmond, T. J. (1998). Positioning and stability of nucleosomes on MMTV 3'LTR sequences. *Journal of Molecular Biology*, 275(3), 427-441. doi:10.1006/jmbi.1997.1464
- 40 Garske, A. L., Oliver, S. S., Wagner, E. K., Musselman, C. A., LeRoy, G., Garcia, B. A., et al. (2010). Combinatorial profiling of chromatin binding modules reveals multisite discrimination. *Nature Methods*, 6(4), 283-290. doi:10.1038/nchembio.319
- 45 Goldman, J. A., Garlick, J. D., & Kingston, R. E. (2010). Chromatin remodeling by imitation switch (ISWI) class ATP-dependent remodelers is stimulated by histone variant H2A.Z. *The Journal of biological chemistry*, 285(7), 4645-4651. doi:10.1074/jbc.M109.072348
- Heller, M. J. (2002). DNA microarray technology: devices, systems, and applications. *Annual review of biomedical engineering*, 4, 129-153. doi:10.1146/annurev.bioeng.4.020702.153438
- 50 Kleer, C. G., Cao, Q., Varambally, S., Shen, R., Ota, I., Tomlins, S. A., et al. (2003). EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proceedings of the National Academy of Sciences*, 100(20), 11606-11611. doi:10.1073/pnas.1933744100
- 55 Kleiner, R. E., Dumelin, C. E., & Liu, D. R. (2011). Small-molecule discovery from DNA-encoded chemical libraries. *Chemical Society Reviews*.
- Krutzik, P. O., & Nolan, G. P. (2006). Fluorescent cell barcoding in flow cytometry allows high-throughput drug screening and signaling profiling. *Nature Methods*, 3(5), 361-368. doi:10.1038/nmeth872
- 60 Lowary, P. T., & Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology*, 276(1), 19-42. doi:10.1006/jmbi.1997.1494
- 65

Luger, K., Rechsteiner, T. J., & Richmond, T. J. (1999). Preparation of nucleosome core particle from recombinant histones. *Methods in enzymology*, 304, 3-19.

5 Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annual review of genomics and human genetics*, 9, 387-402. doi:10.1146/annurev.genom.9.081307.164359

Merriman, B., R D Team, I. T., & Rothberg, J. M. (2012). Progress in Ion Torrent semiconductor chip based sequencing. *Electrophoresis*, 33(23), 3397-3417. doi: 10.1002/elps.201200424

10 Muir, T. W. (2003). Semisynthesis of proteins by expressed protein ligation. *Annual Review of Biochemistry*, 72, 249-289. doi:10.1146/annurev.biochem.72.121801.161900

15 Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356), 348-352. doi:10.1038/nature10242

20 Ruthenburg, A. J., Li, H., Milne, T. A., Dewell, S., McGinty, R. K., Yuen, M., et al. (2011). Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions. *Cell*, 145(5), 692-706. doi:10.1016/j.cell.2011.03.053

Schones, D. E., & Zhao, K. (2008). Genome-wide approaches to studying chromatin modifications. *Nature reviews. Genetics*, 9(3), 179-191. doi:10.1038/nrg2270

25 Simon, M. D., Chu, F., Racki, L. R., la Cruz, de, C. C., Burlingame, A. L., Panning, B., et al. (2007). The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell*, 128(5), 1003-1012. doi:10.1016/j.cell.2006.12.041

30 Spacil, Z., Tatipaka, H., Barcenas, M., Scott, C. R., Turecek, F., & Gelb, M. H. (2013). High-throughput assay of 9 lysosomal enzymes for newborn screening. *Clinical chemistry*, 59(3), 502-511. doi:10.1373/clinchem.2012.189936

Wang, L., Xie, J., & Schultz, P. G. (2006). Expanding the genetic code. *Annual review of biophysics and biomolecular structure*.

35 Weibel, D. B., & Whitesides, G. M. (2006). Applications of microfluidics in chemical biology. *Current Opinion in Chemical Biology*.

Whitesides, G. M. (2006). The origins and the future of microfluidics. *Nature*, 442(7101), 368-373. doi:10.1038/nature05058

40 JPeterson et al., US 5,972,608, "Assays and Reagents for Chromatin Remodeling Enzymes and their Modulators."

45 Steinman et al., US 2009/0062130, "Nucleosome-Based Biosensor."

LISTADO DE SECUENCIAS

<110> CONSEJO DE ADMINISTRACIÓN DE LA UNIVERSIDAD DE PRINCETON

50 <120> APLICACIÓN DE CÓDIGOS DE BARRAS DE ADN DE BANCOS DE MATRICES DE CROMATINAS Y MONONUCLEOSOMAS DISEÑADORES PARA LA CREACIÓN DE PERFILES DE LECTORES, ESCRITORES, BORRADORES Y MODULADORES DE CROMATINA DE LOS MISMOS

<130> 32108-348453

55 140
<141>

60 <150> 61/712 148
<151> 10-10-2012

<150> 61/656.233
<151> 06-06-2012
<160> 57

65 <170> Patentin versión 3.5

ES 2 693 349 T3

<210> 1
 <211> 4
 <212> FRT
 5 Desconocido

<220>
 <223> Descripción de desconocido: motivo peptídico "PWWP"

10 <400>
Pro Trp Trp Pro
1

15 <210> 2
 <211> 39
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Descripción de secuencia artificial: cebador sintético

<400> 2
 acctaggct tctgatgctg gagaatcccg gtgccgagg 39

25 <210> 3
 <211> 40
 <212> ADN
 Secuencia artificial

30 <220>
 <223> Descripción de secuencia artificial: cebador sintético

<400> 3
 ctaccacatc gtgggatgta tatactgac acgtgcctgg 40

35 <211> 191
 <212> ADN
 <213> Secuencia artificial

40 <220>
 <223> Descripción de secuencia artificial: polinucleótido sintético

<400> 4
gatatcacc taggtctctg atgctggaga atcccgggtgc cgaggccgct caattggtcg 60
tagacagctc tagcaccgct taaacgcacg tacgcgctgt cccccgcgtt ttaaccgcca 120
aggggattac tccctagtct ccaggcacgt gtcagatata tacatcctgt cacgcggtga 180
acagcgatat c 191

45 <210> 5
 <211> 26
 <212> ADN
 <213> Secuencia artificial

50 <220>
 <223> Descripción de secuencia artificial: oligonucleótido sintético

55 <220>
 <221> modified_base
 <222> (22)..(25)
 <223> a, c, t, g, desconocida u otra

<400> 5
 cctgcgtgtc tccgactcag hnnnh 26

5 <210> 6
 <211> 32
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Descripción de secuencia artificial: oligonucleótido sintético

10 <220>
 <221> modified_base
 <222> (6)..(9)
 <223> a, c, t, g, desconocida u otra

15 <400> 6
 catcdnnnd ctgagtcgga gacacgcagg aa 32

20 <210> 7
 <211> 31
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Descripción de secuencia artificial: cebador sintético

25 <220>
 <223> 5'-biotina

30 <400> 7
 tatcactgac aacagtccta acattcacct c 31

35 <210> 8
 <211> 25
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Descripción de secuencia artificial: cebador sintético

40 <400> 8
 atcctcaaaaa ctgtgccgca gtcgg 25

45 <210> 9
 <211> 52
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Descripción de secuencia artificial: oligonucleótido sintético

50 <220>
 <221> modified_base
 <222> (21)..(26)
 <223> a, c, t, g, desconocida u otra

55 <400> 9
 cctgcgtgac tccgactcag nnnnnngatg ctggagaatc ccggtgccga gg 52

60 <210> 10
 <211> 33
 <212> ADN
 <213> Secuencia artificial

65 <220>
 <223> Descripción de secuencia artificial: cebador sintético

ES 2 693 349 T3

<400> 10
 gtgacaggat gtatatatct gacacgtgcc tgg 33

5 <210> 11
 <211> 30
 <212> ADN
 <213> Secuencia artificial

10 <220>
 <223> Descripción de secuencia artificial: cebador sintético

<400> 11
 ccatctcatc cctgcggtgc tccgactcag 30

15 <210> 12
 <211> 59
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> Descripción de secuencia artificial: cebador sintético

<220>
 <221> modified_base
 <222> (25)..(28)
 25 <223> a, c, t, g, desconocida u otra

<400> 12
 cctctctatg ggcagtcggt gatbnnndg gtgctagagc tgtctacgac caattgagc 59

30 <210> 13
 <211> 152
 <212> ADN
 <213> Secuencia artificial

35 <220>
 <223> Descripción de secuencia artificial: polinucleótido sintético

<220>
 <223> 5'-biotina

40 <400> 13
 tatcacttgc aacagtccta acattcacct cttgtgtggt tgtgtctggt cgccatccc 60
 tctccgctog tcaacttatcc ttcactttcc agagggtccc cccgagacc ccggcgacc 120
 tggtcggccg actgcggcac agttttttgg at 152

45 <210> 14
 <211> 62
 <212> ADN
 <213> Secuencia artificial

50 <220>
 <223> Descripción de secuencia artificial: cebador sintético

<400> 14
 55 ccatctcatc cctgcggtgc tccgactcag atgcaggatg ctggagaatc ccggtgccga 60
 gg 62
 <210> 15
 <211> 147
 <212> ADN
 60 <213> Secuencia artificial

ES 2 693 349 T3

	<220>		
	<223>	Descripción de secuencia artificial: polinucleótido sintético	
5	<400>	15	
		ctggagaatc ccggtgccga ggccgctcaa ttggtcgtag acagctctag caccgcttaa	60
		acgcacgtac gcgctgtccc ccgcgtttta accgccaagg ggattactcc ctagtctcca	120
		ggcacgtgtc agatatatac atcctgt	147
	<210>	16	
	<211>	20	
10	<212>	ADN	
	<213>	Secuencia artificial	
	<220>		
	<223>	Descripción de secuencia artificial: oligonucleótido sintético	
15	<400>	16	
		cctgcgtgtc tccgactcag	20
	<210>	17	
20	<211>	23	
	<212>	ADN	
	<213>	Secuencia artificial	
	<220>		
25	<223>	Descripción de secuencia artificial: oligonucleótido sintético	
	<400>	17	
		cctctctatg ggcagtcggt gat	23
30	<210>	18	
	<211>	30	
	<212>	ADN	
	<213>	Secuencia artificial	
35	<220>		
	<223>	Descripción de secuencia artificial: oligonucleótido sintético	
	<400>	18	
40		ggtgctagag ctgtctacga ccaattgagc	30
	<210>	19	
	<211>	6	
	<212>	ADN	
45	<213>	Secuencia artificial	
	<220>		
	<223>	Descripción de secuencia artificial: oligonucleótido sintético	
50	<400>	19	
		atgca	6
	<210>	20	
	<211>	6	
55	<212>	ADN	
	<213>	Secuencia artificial	
	<220>		
	<223>	Descripción de secuencia artificial: oligonucleótido sintético	
60	<400>	20	
		atcata	6

ES 2 693 349 T3

5
<210> 21
<211> 6
<212> ADN
<213> Secuencia artificial

<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

10
<400> 21
agatac 6

<210> 22
<211> 6
<212> ADN
<213> Secuencia artificial

15
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

20
<400> 22
cgatgc 6

<210> 23
<211> 6
<212> ADN
<213> Secuencia artificial

25
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

30
<400> 23
ctgtat 6

<210> 24
<211> 6
<212> ADN
<213> Secuencia artificial

35
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

40
<400> 24
agatgt 6

45
<210> 25
<211> 6
<212> ADN
<213> Secuencia artificial

50
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

<400> 25
catac 6

55
<210> 26
<211> 6
<212> ADN
<213> Secuencia artificial

60
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

65
<400> 26
cgtgac 6

ES 2 693 349 T3

5
<210> 27
<211> 6
<212> ADN
<213> Secuencia artificial

<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

10
<400> 27
atctga 6

<210> 28
<211> 6
<212> ADN
<213> Secuencia artificial

15
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

20
<400> 28
tcactc 6

<210> 29
<211> 6
<212> ADN
<213> Secuencia artificial

25
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

30
<400> 29
atgtgc 6

<210> 30
<211> 6
<212> ADN
<213> Secuencia artificial

35
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

40
<400> 30
tctgta 6

<210> 31
<211> 6
<212> ADN
<213> Secuencia artificial

45
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

50
<400> 31
atgaca 6

<210> 32
<211> 6
<212> ADN
<213> Secuencia artificial

55
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

60
<400> 32
cgatgt 6

ES 2 693 349 T3

5
<210> 33
<211> 6
<212> ADN
<213> Secuencia artificial

<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

10
<400> 33
acgcac 6

<210> 34
<211> 6
<212> ADN
<213> Secuencia artificial

15
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

20
<400> 34
cgatga 6

<210> 35
<211> 6
<212> ADN
<213> Secuencia artificial

25
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

30
<400> 35
tacact 6

<210> 36
<211> 6
<212> ADN
<213> Secuencia artificial

35
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

40
<400> 36
cataca 6

<210> 37
<211> 6
<212> ADN
<213> Secuencia artificial

45
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

50
<400> 37
catcgc 6

<210> 38
<211> 6
<212> ADN
<213> Secuencia artificial

55
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

60
<400> 38
tcgtga 6

65

ES 2 693 349 T3

5
<210> 39
<211> 6
<212> ADN
<213> Secuencia artificial

<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

10
<400> 39
atatct 6

<210> 40
<211> 6
<212> ADN
<213> Secuencia artificial

15
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

20
<400> 40
atcgac 6

<210> 41
<211> 6
<212> ADN
<213> Secuencia artificial

25
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

30
<400> 41
cgtatc 6

<210> 42
<211> 6
<212> ADN
<213> Secuencia artificial

35
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

40
<400> 42
ctctca 6

45
<210> 43
<211> 6
<212> ADN
<213> Secuencia artificial

50
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

55
<400> 43
acacgt 6

<210> 44
<211> 6
<212> ADN
<213> Secuencia artificial

60
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

65
<400> 44
cagctc 6

ES 2 693 349 T3

5
<210> 45
<211> 6
<212> ADN
<213> Secuencia artificial

<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

10
<400> 45
tgcgct 6

<210> 46
<211> 6
<212> ADN
<213> Secuencia artificial

15
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

20
<400> 46
cgtagc 6

<210> 47
<211> 6
<212> ADN
<213> Secuencia artificial

25
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

30
<400> 47
tgtatc 6

<210> 48
<211> 6
<212> ADN
<213> Secuencia artificial

35
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

40
<400> 48
acatct 6

45
<210> 49
<211> 6
<212> ADN
<213> Secuencia artificial

50
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

55
<400> 49
agatga 6

<210> 50
<211> 6
<212> ADN
<213> Secuencia artificial

60
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

65
<400> 50
cacatc 6

ES 2 693 349 T3

5
<210> 51
<211> 6
<212> ADN
<213> Secuencia artificial

<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

10
<400> 51
cgcgca 6

<210> 52
<211> 6
<212> ADN
<213> Secuencia artificial

15
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

20
<400> 52
acagac 6

<210> 53
<211> 6
<212> ADN
<213> Secuencia artificial

25
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

30
<400> 53
agtact 6

<210> 54
<211> 6
<212> ADN
<213> Secuencia artificial

35
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

40
<400> 54
tagcgc 6

<210> 55
<211> 6
<212> ADN
<213> Secuencia artificial

45
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

50
<400> 55
ctatca 6

<210> 56
<211> 6
<212> ADN
<213> Secuencia artificial

55
<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

60
<400> 56
atagac 6

65

ES 2 693 349 T3

<210> 57
<211> 6
<212> ADN
<213> Secuencia artificial

5

<220>
<223> Descripción de secuencia artificial: oligonucleótido sintético

10

<400> 57
tatgtc 6

REIVINDICACIONES

1. Un banco de mononucleosomas sintéticos que comprende dos o más tipos de mononucleosomas,

5 en el que cada mononucleosoma comprende un complejo de

(a) un octámero de proteína, que contiene 2 copias cada una de las histonas H2A, H2B, H3 y H4, y opcionalmente, la histona fijadora H1, en el que al menos una de las histonas no está modificada y/o en el que al menos una de las histonas se modifica para formar un patrón de modificaciones de histona, y

10

(b) una molécula de ADN nucleosómico, que comprende

(i) una secuencia de posicionamiento de nucleosoma (NPS) fuerte,

15

(ii) uno o más códigos de barras de ADN situados en una posición/posiciones definida(s) en el ADN nucleosómico, y, opcionalmente,

(iii) extensiones de ADN, en el extremo 5' y/o 3' de la NPS y/o dentro de la NPS,

20

en el que el ADN no está modificado y/o en el que al menos uno de los nucleótidos en el ADN se modifica para formar un patrón único de modificaciones de ADN,

y opcionalmente

25

(c) una o más proteínas asociadas a cromatina distintas de histona,

en el que cada mononucleosoma en el banco tiene (i) un patrón único de modificaciones de histonas y/o un patrón único de modificaciones de ADN, formando de este modo un patrón único de modificaciones de nucleosomas, y (ii) uno o más códigos de barras únicos con una secuencia y posición en el ADN nucleosómico que son indicativas del patrón único de modificaciones de nucleosomas en el mononucleosoma.

30

2. Un polinucleosoma sintético, que comprende dos o más mononucleosomas sintéticos unidos entre sí por una molécula de ADN definida, teniendo los mononucleosomas una conectividad definida,

35

en el que cada mononucleosoma comprende un complejo de

(a) un octámero de proteína, que contiene 2 copias cada una de las histonas H2A, H2B, H3 y H4, y opcionalmente, la histona fijadora H1, en el que al menos una de las histonas no está modificada, y/o al menos una de las histonas se modifica para formar un patrón de modificaciones de histonas mononucleosómicas; y

40

(b) una molécula de ADN mononucleosómico, que comprende

(i) una secuencia de posicionamiento de nucleosoma (NPS) fuerte,

45

y, opcionalmente,

(ii) extensiones de ADN, en el extremo 5' y/o 3' de la NPS y/o dentro de la NPS,

50

en el que el ADN no está modificado y/o en el que al menos uno de los nucleótidos en el ADN se modifica para formar un patrón único de modificaciones de ADN mononucleosómico,

y opcionalmente

55

(c) una o más proteínas asociadas a cromatina distintas de histona,

en el que el patrón de modificaciones de histonas mononucleosómicas y/o el patrón de modificaciones de ADN mononucleosómico en el polinucleosoma puede ser uniforme o puede ser diferente, lo que da como resultado un patrón único de modificaciones polinucleosómicas; y

60

en el que el polinucleosoma comprende uno o más códigos de barras situados en una posición definida en el ADN polinucleosómico con una secuencia y posición en el ADN polinucleosómico que son indicativas del patrón único de modificaciones polinucleosómicas.

65

3. Un banco de polinucleosomas sintéticos, que comprende dos o más polinucleosomas sintéticos de la reivindicación 2,

en el que cada miembro del banco tiene uno o más códigos de barras únicos, lo que es indicativo de un patrón único de modificaciones polinucleosómicas.

5 **4.** El banco de la reivindicación 1 o la reivindicación 3, en el que las modificaciones de las histonas comprenden isoformas de histonas, modificaciones postraduccionales (PTM) y/o aminoácidos no naturales.

5. El banco de la reivindicación 4, en el que

10 (a) la isoforma de histona comprende una o más sustituciones aminoacídicas o inserciones aminoacídicas dentro de las secuencias de proteínas o extensiones en el extremo de las secuencias de proteínas; y/o

15 (b) la PTM se selecciona de cualquier modificación de histona natural, incluyendo metilación, acetilación, fosforilación, ubiquitilación, SUMOilación, ADP-ribosilación, glucosilación, alquilación, acilación, isomerización cis/trans de propilo, nitrosilación y oxidación, con modificación en un sitio, o modificación en más de un sitio por nucleosoma, y/o

20 (c) el aminoácido no natural se selecciona de análogos sintéticos de PTM, que pueden ser química y bioquímicamente inertes, fotorreticuladores, marcadores fluorescentes o marcadores de isótopos.

6. El banco de la reivindicación 1 o la reivindicación 3, en el que las modificaciones de la molécula de ADN comprenden una o más bases de ADN con modificaciones naturales, o bases con modificaciones artificiales.

25 **7.** Un kit que comprende el banco de la reivindicación 1 o la reivindicación 3, en el que los mononucleosomas o polinucleosomas están opcionalmente en uno o más recipientes y en el que los recipientes son tubos de ensayo, pocillos de una placa multipocillo, o cámaras de reacción de un dispositivo microfluidico, que comprende además una lista que indica la correlación entre cada código de barras único y patrón único de modificaciones de histonas.

30 **8.** Un procedimiento para determinar la especificidad de los patrones de reconocimiento y afinidades de uno o más lectores de cromatina que son de origen recombinante y/o se extraen de extractos celulares nucleares, o para determinar la especificidad e interferencias de modificación de cromatina de modificadores de cromatina que son de origen recombinante y/o se extraen de extractos celulares nucleares de una o más líneas celulares, que comprende

35 (i) poner en contacto (por ejemplo, incubar) un banco de la reivindicación 1 o la reivindicación 3 con el lector de cromatina o el modificador de cromatina,

(ii) aislar miembros de banco unidos y/o modificados, y

40 (iii) identificar y/o cuantificar los miembros de banco unidos o modificados y cualquier marca añadida o marca retirada.

9. Un procedimiento para identificar las modificaciones asociadas con un interactuante o modificador, que comprende

45 multiplexar un número mayor que uno de interactuantes y/o modificadores de cromatina con un banco de mononucleosomas o polinucleosomas de las reivindicaciones 1 o 3, y dividir el banco en el mismo número de subbancos de acuerdo con las modificaciones, e identificar las modificaciones asociadas con cada interactuante o modificador.

50 **10.** Un procedimiento para identificar y crear perfiles de la especificidad de fármacos epigenéticos, que comprende

combinar una molécula candidata con un banco de la reivindicación 1 o la reivindicación 3, y

detectar la modulación de una modificación de nucleosoma,

55 identificando de este modo fármacos epigenéticos candidatos que modulan modificaciones de nucleosomas.

11. El banco de la reivindicación 1 que comprende además una lista de códigos de barras de ADN y las modificaciones de nucleosomas asociadas y la composición de cada nucleosoma con código de barras.

60 **12.** Un mononucleosoma sintético, que comprende un complejo de

(a) un octámero de proteína, que contiene 2 copias cada una de las histonas H2A, H2B, H3 y H4, y opcionalmente, la histona fijadora H1, en el que al menos una de las histonas no está modificada y/o en el que al menos una de las histonas se modifica para formar un patrón de modificaciones de histona, y

65

(b) una molécula de ADN nucleosómico, que comprende

(i) una secuencia de posicionamiento de nucleosoma (NPS) fuerte,

5 (ii) uno o más códigos de barras de ADN situados en una posición/posiciones definida(s) en el ADN nucleosómico, y, opcionalmente,

(iii) extensiones de ADN, en el extremo 5' y/o 3' de la NPS y/o dentro de la NPS,

10 en el que el ADN no está modificado y/o en el que al menos uno de los nucleótidos en el ADN se modifica para formar un patrón único de modificaciones de ADN,

y opcionalmente

15 (c) una o más proteínas asociadas a cromatina distintas de histona,

en el que el mononucleosoma tiene un patrón único de modificaciones de histonas y/o un patrón único de modificaciones de ADN, formando de este modo un patrón único de modificaciones de nucleosomas, y

20 en el que la secuencia y posición del/de los código(s) de barras en el ADN nucleosómico son indicativas del patrón único de las modificaciones de nucleosomas en el mononucleosoma.

25 **13.** Un mononucleosoma sintético de la reivindicación 12 que comprende un código de barras de ADN en el extremo 5' y/o 3', o dentro de, el ADN nucleosómico, o un polinucleosoma sintético de la reivindicación 2 que comprende un código de barras en el extremo 5' y/o 3', o dentro de, el ADN polinucleosómico.

30 **14.** Un procedimiento para ensamblar un mononucleosoma de la reivindicación 12, que comprende combinar proteínas histonas y ADN nucleosómico con código de barras con un ADN con tampón con MMTV marcado con biotina, en una proporción predeterminada.

15. El banco de la reivindicación 1 o la reivindicación 3, en el que la NPS es suficientemente fuerte de modo que el banco es estable, y no se produce reordenamiento de ADN significativo entre miembros de banco de mono y polinucleosomas después del almacenamiento durante al menos un mes a 4 °C.

35 **16.** El banco de la reivindicación 1 o la reivindicación 3, en el que

40 (a) las modificaciones de histonas y/o ADN comprenden un conjunto representativo de estados de cromatina biológicamente relevantes; y/o (b) la proporción de los miembros de banco de mono y polinucleosomas es equimolar (1:1 para cada miembro de banco) o no equimolar, variando de 1 a 1000 para uno o un subconjunto de los miembros de banco, en una proporción predeterminada fija.

17. El banco de la reivindicación 3 que comprende además una lista que indica la correlación entre cada código de barras único y el patrón único de modificaciones polinucleosómicas.

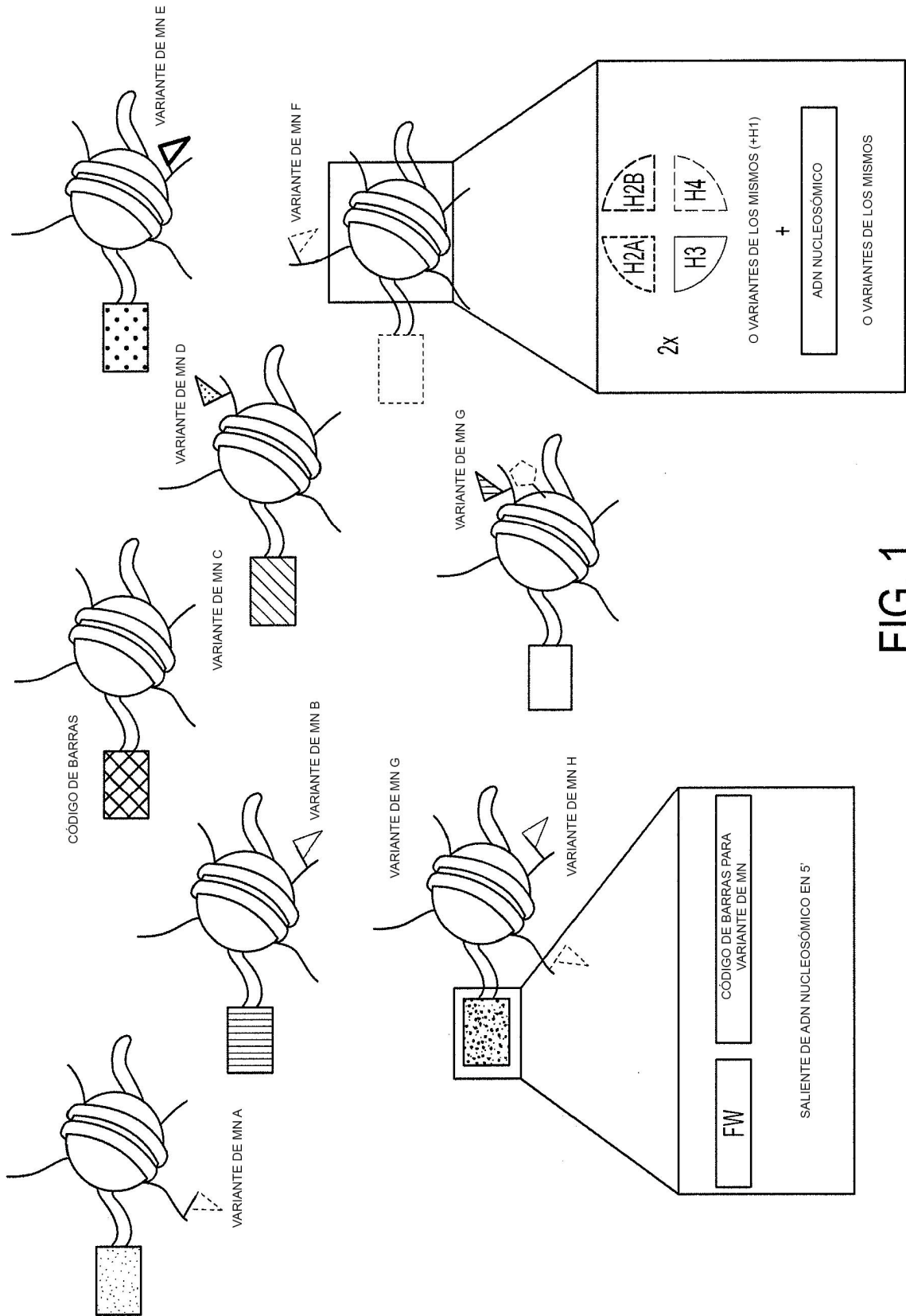


FIG. 1

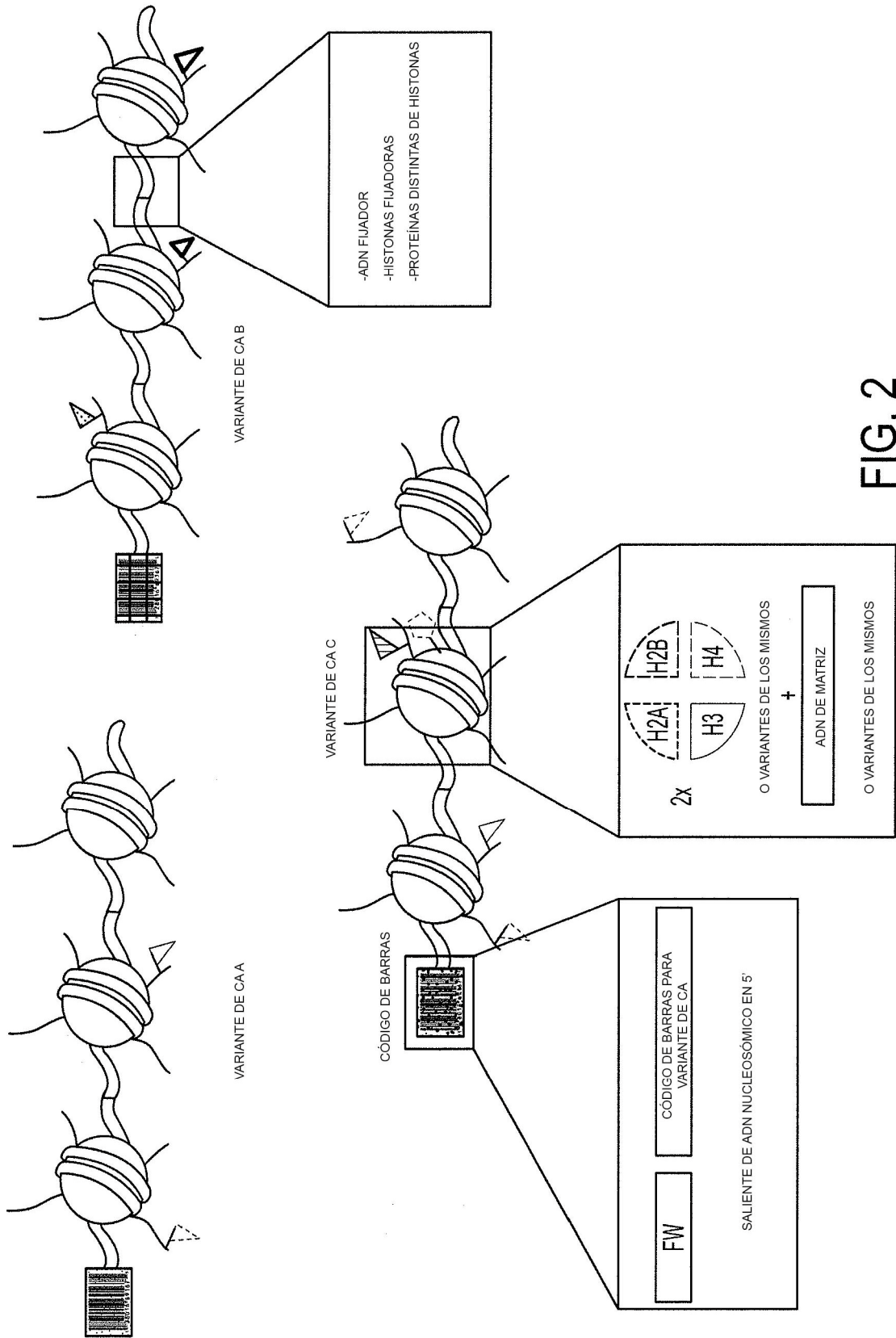


FIG. 2

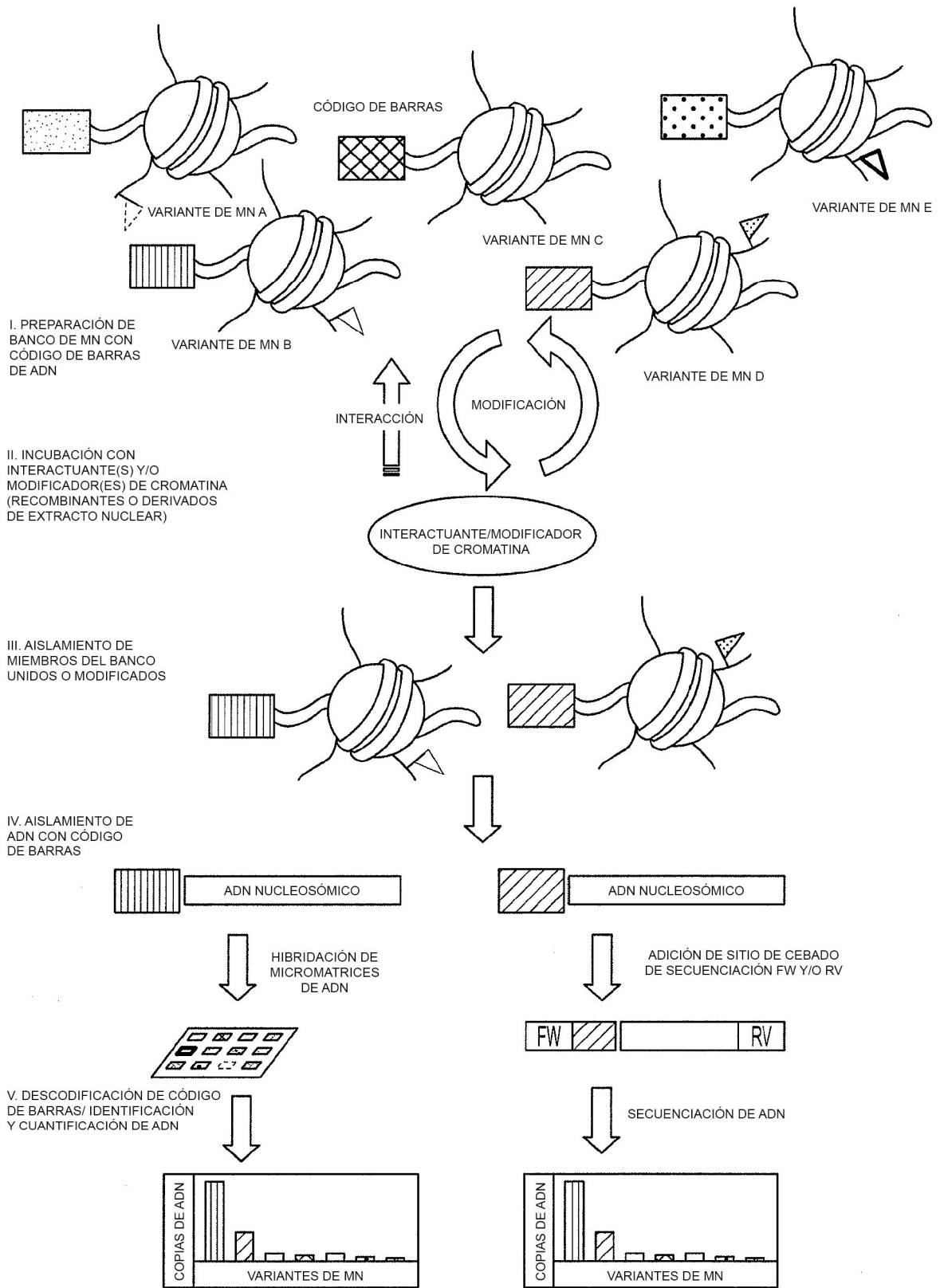


FIG. 3

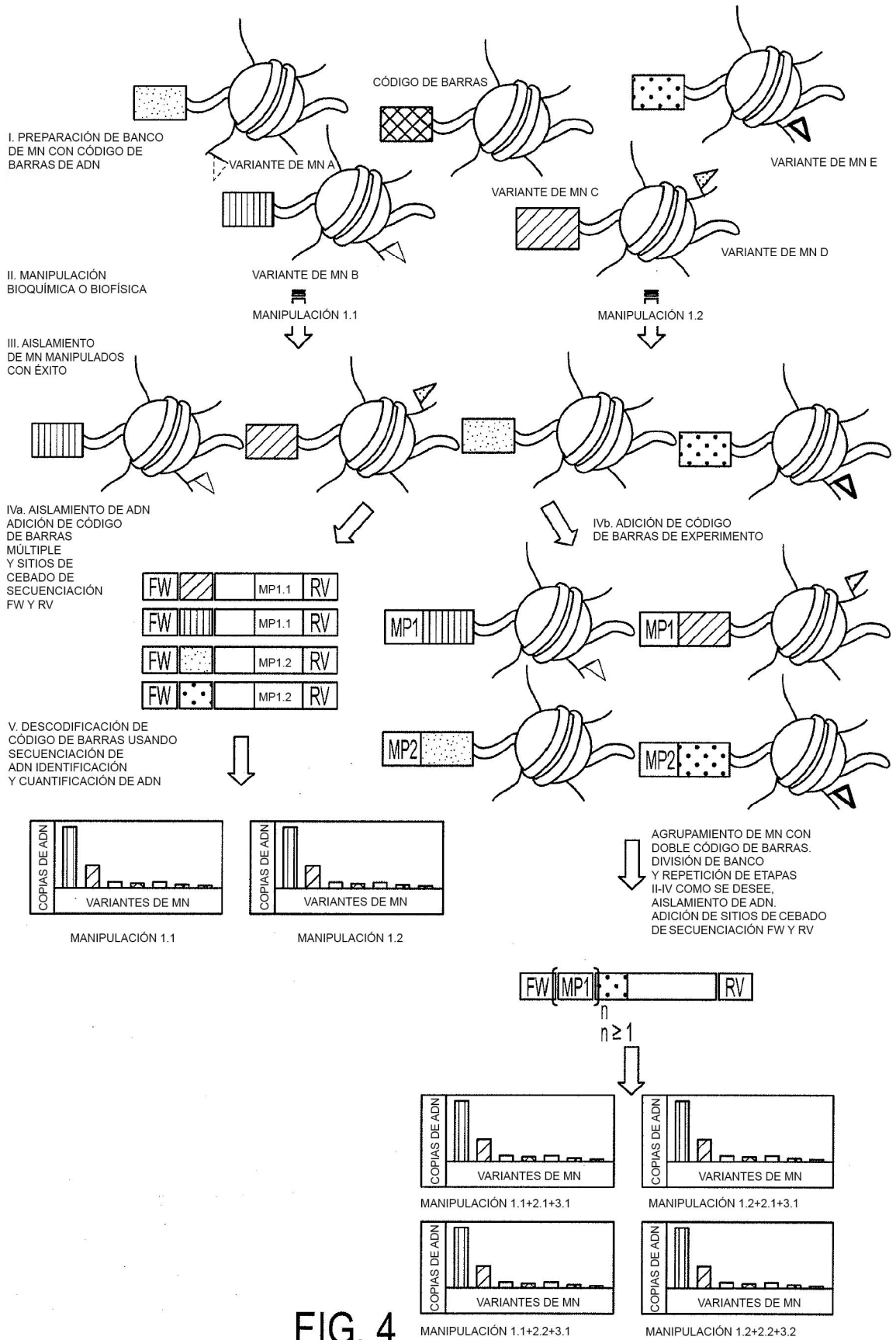


FIG. 4

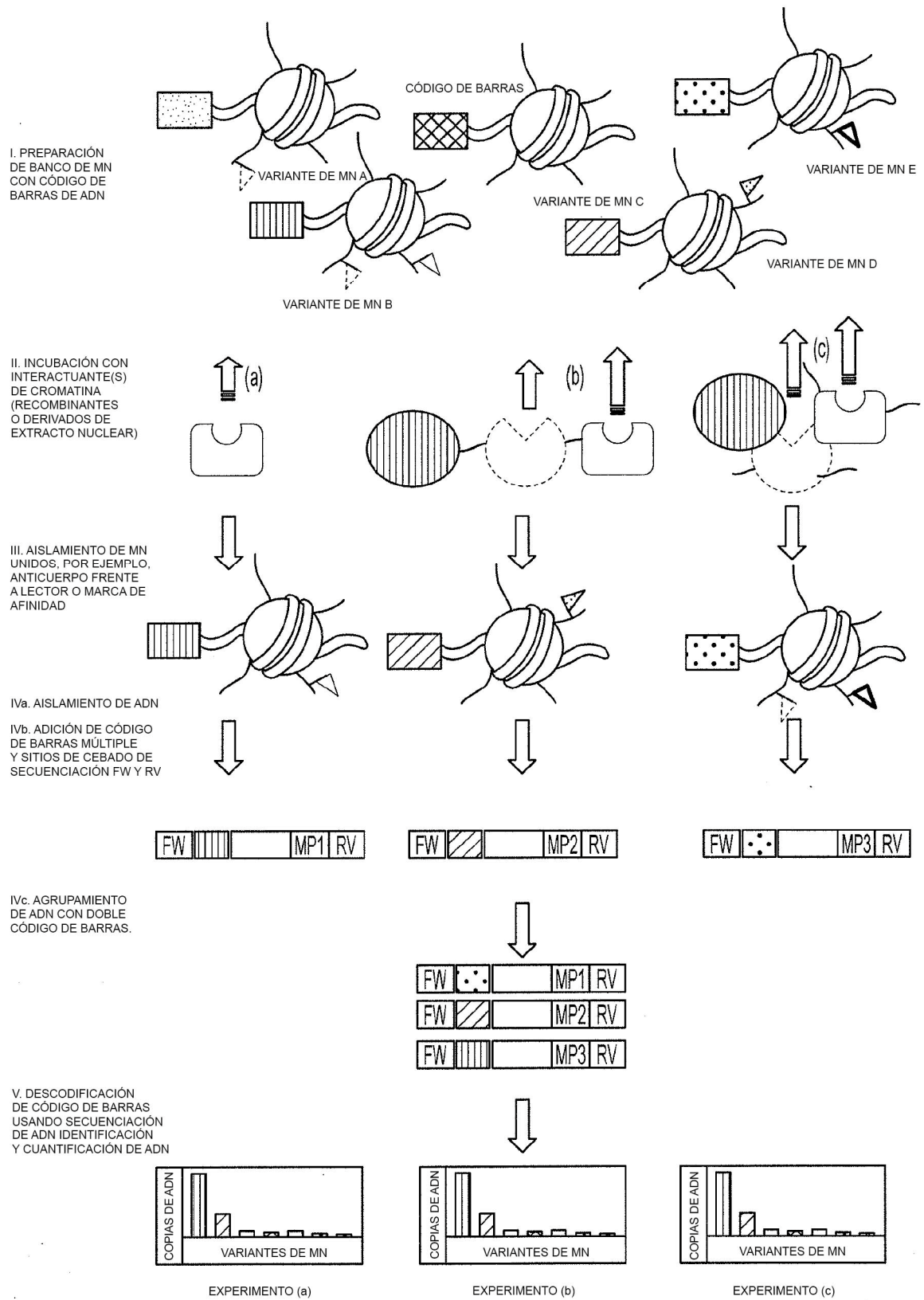
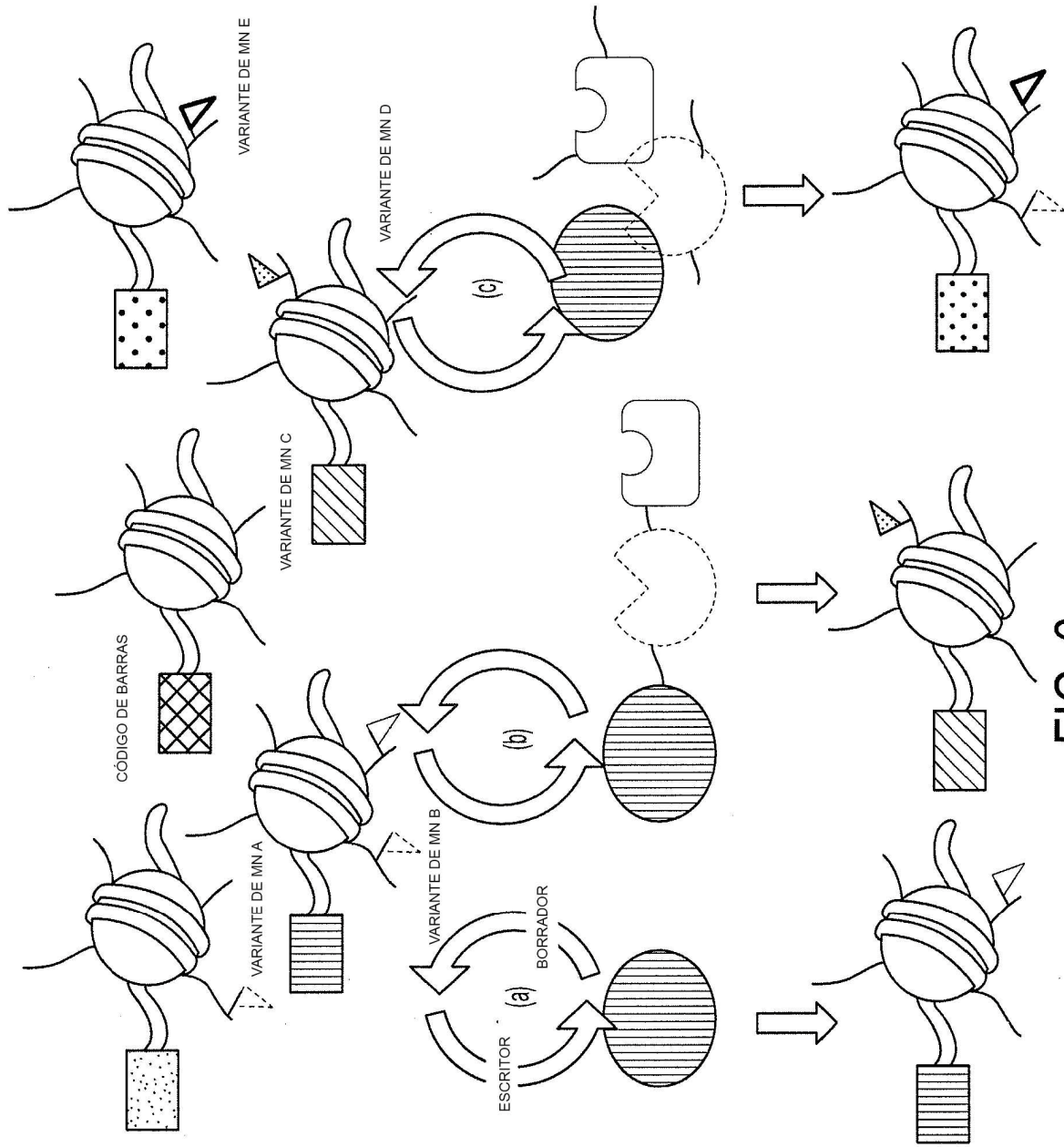


FIG. 5



I. PREPARACIÓN DE BANCO DE MN CON CÓDIGO DE BARRAS DE ADN

II. INCUBACIÓN CON MODIFICADOR(ES) DE CROMATINA, TALES COMO ESCRITOR(ES) O BORRADOR(ES) DE CROMATINA (RECOMBINANTES O DERIVADOS DE EXTRACTO NUCLEAR)

III. AISLAMIENTO DE MN MODIFICADOS CON ÉXITO, POR EJEMPLO, ANTICUERPO FRENTE A PARA AISLAR SUSTRATOS PARA AISLAR SUSTRATOS (BORRADOR) PARA SUSTRATER LOS NO SUSTRATOS

FIG. 6

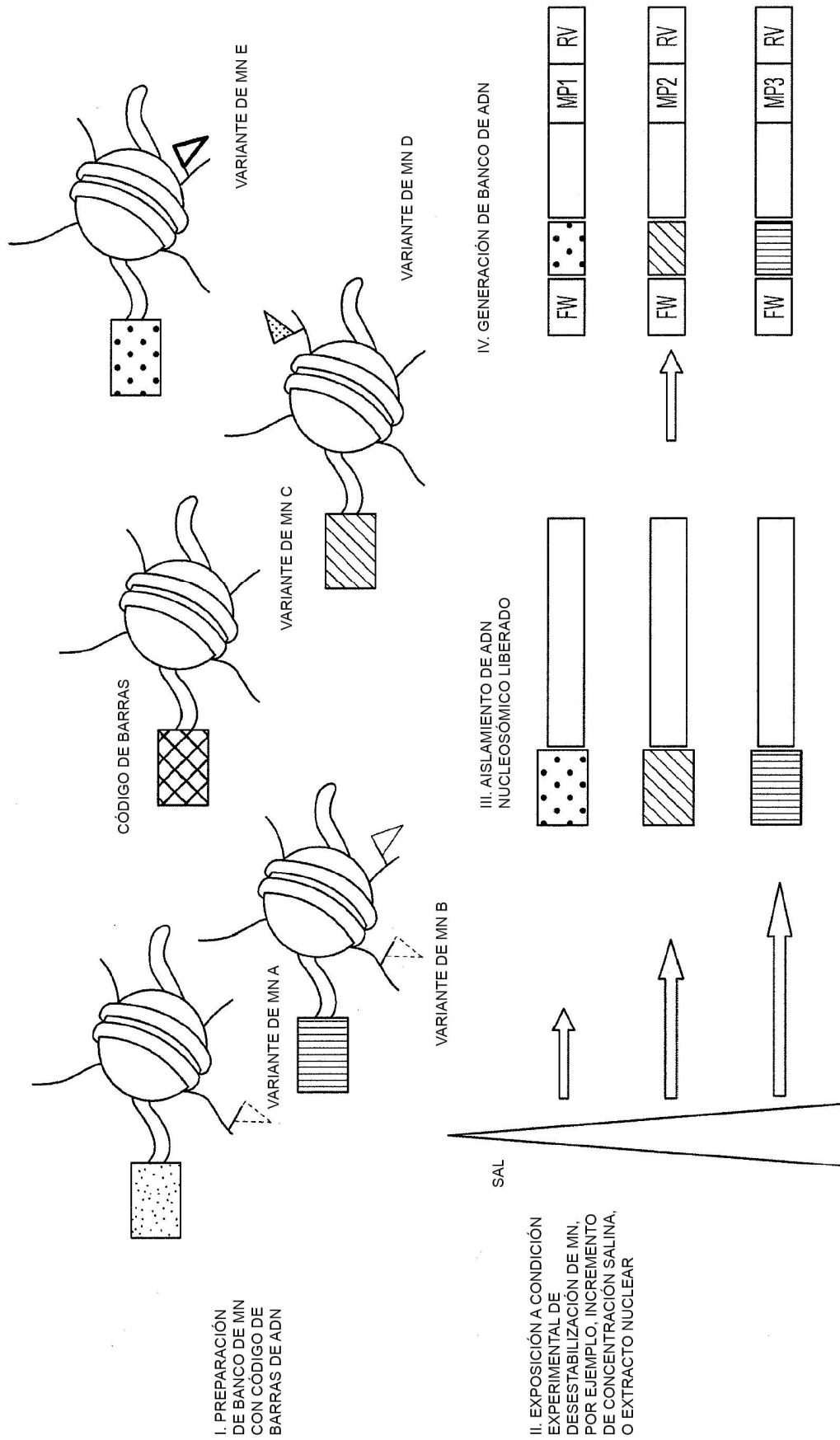


FIG. 7

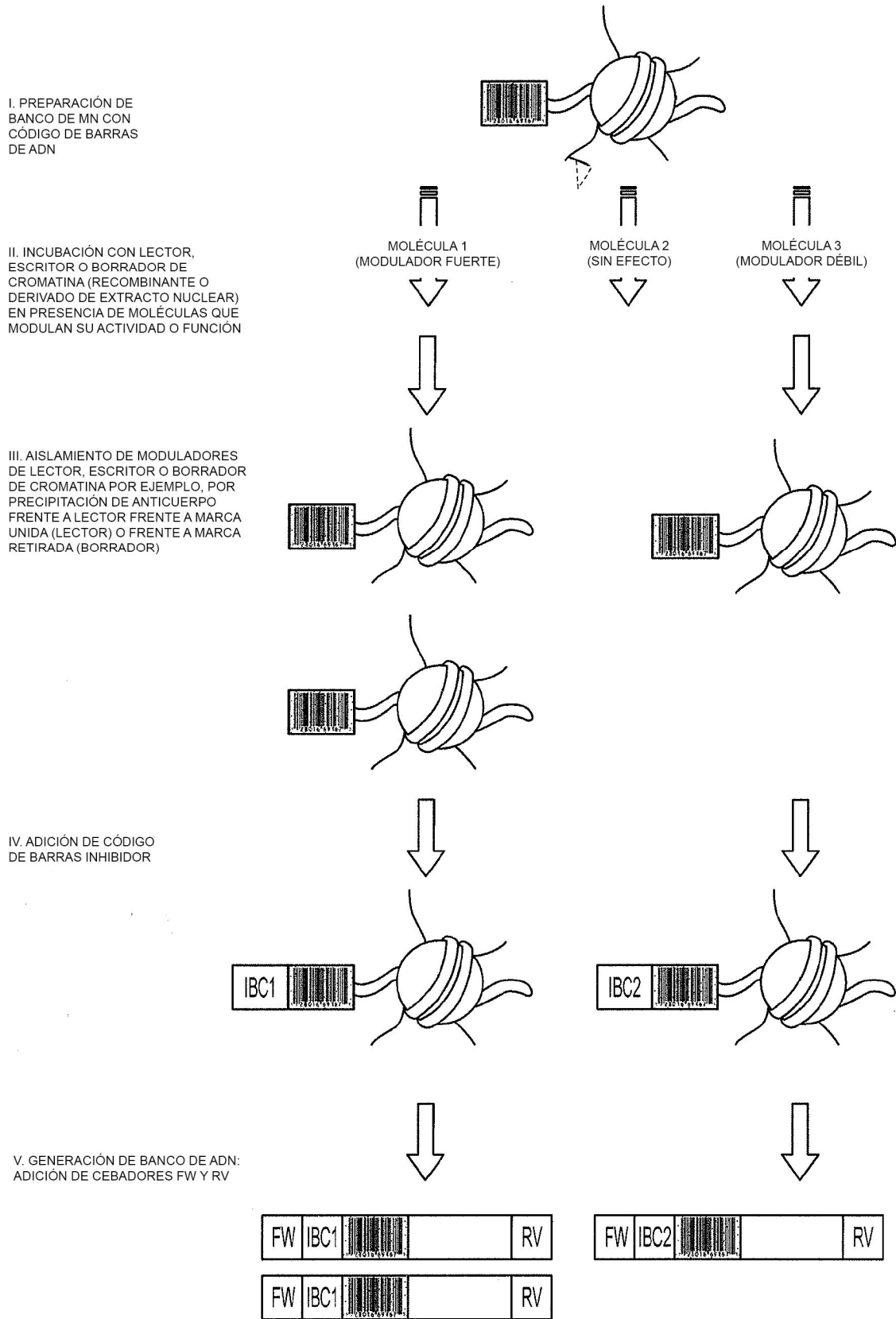


FIG. 8

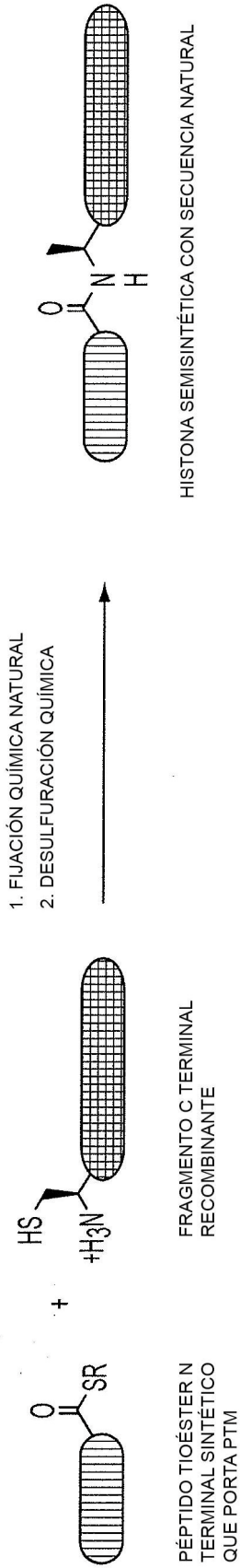
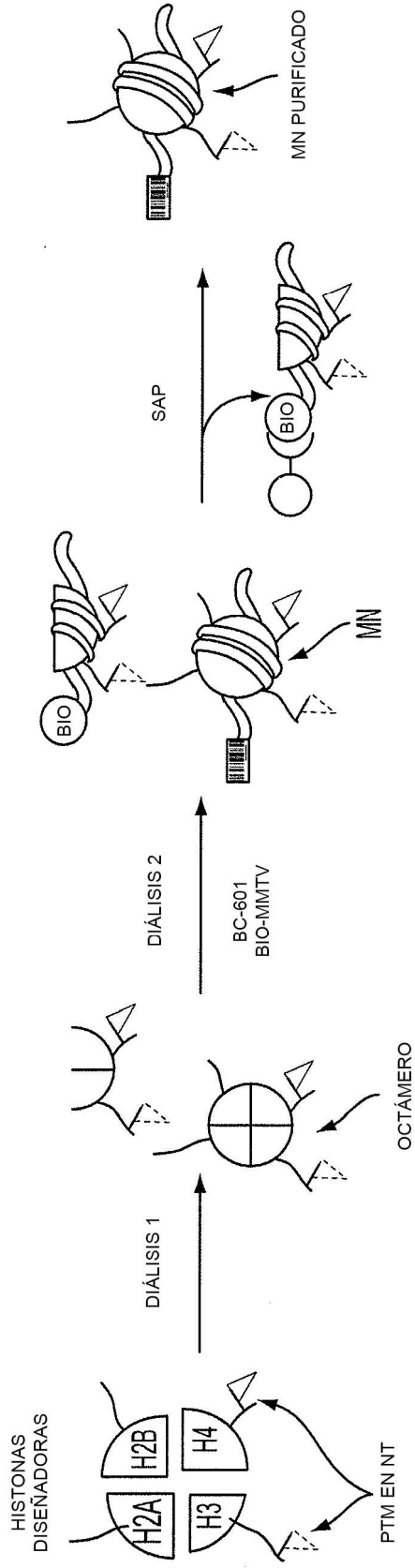


FIG. 9A



BIO-MMTV: 5'-biotina-TATCACTTGCAACAGTCCTAACATTACCTCTTGTGTGTTTGTGTTCCGCAATCCCGTCTCCGCTCGTCACCTTATCCCTTC
ACTTTCAGAGGGTCCCGCCGAGACCCCGGCGACCTGGTGGCGGACTGCGGCACAGTTTTTGGAT

FIG. 9B

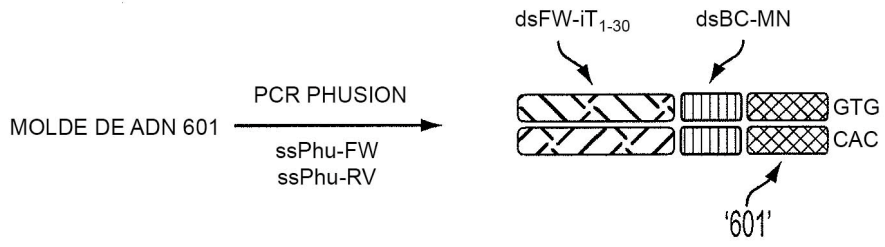


FIG. 10A

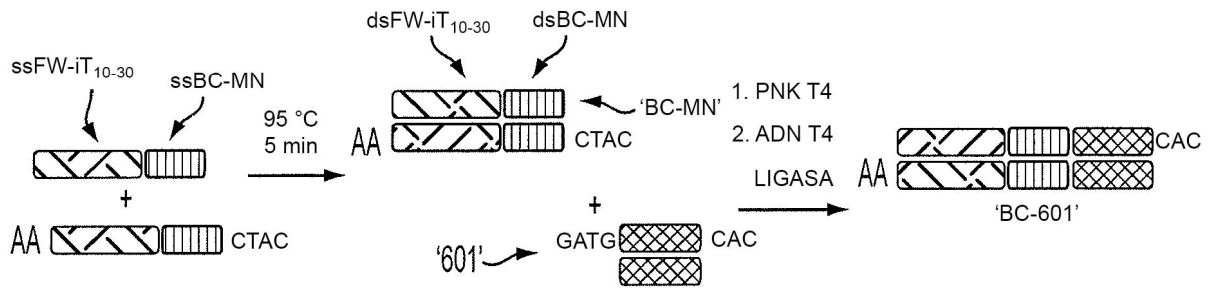


FIG. 10B

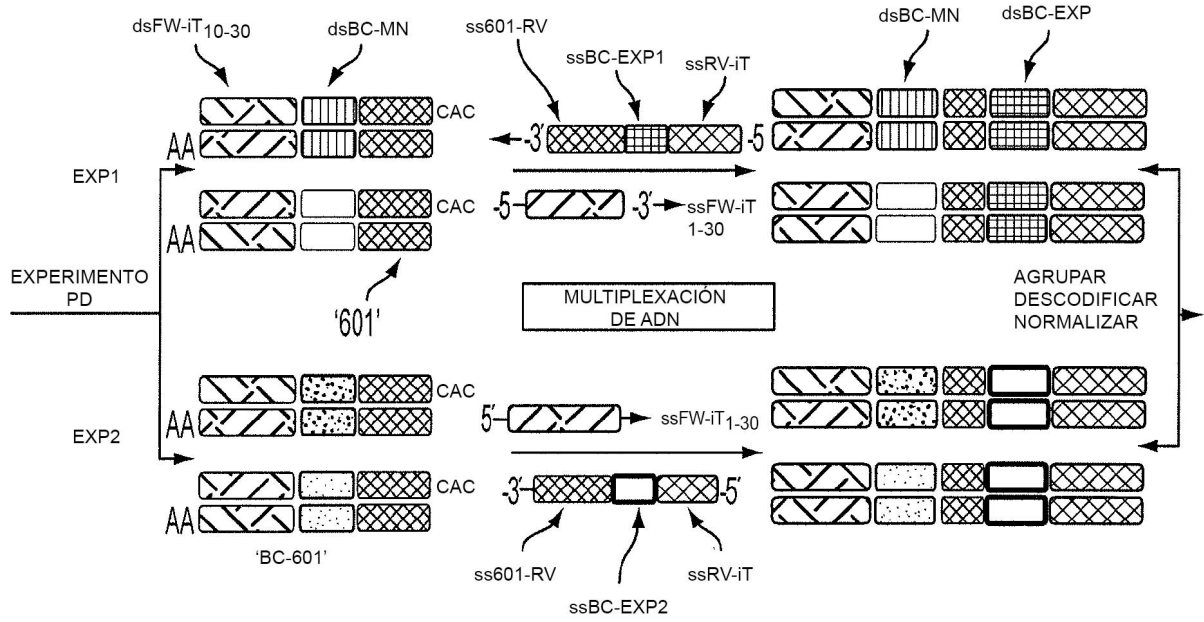
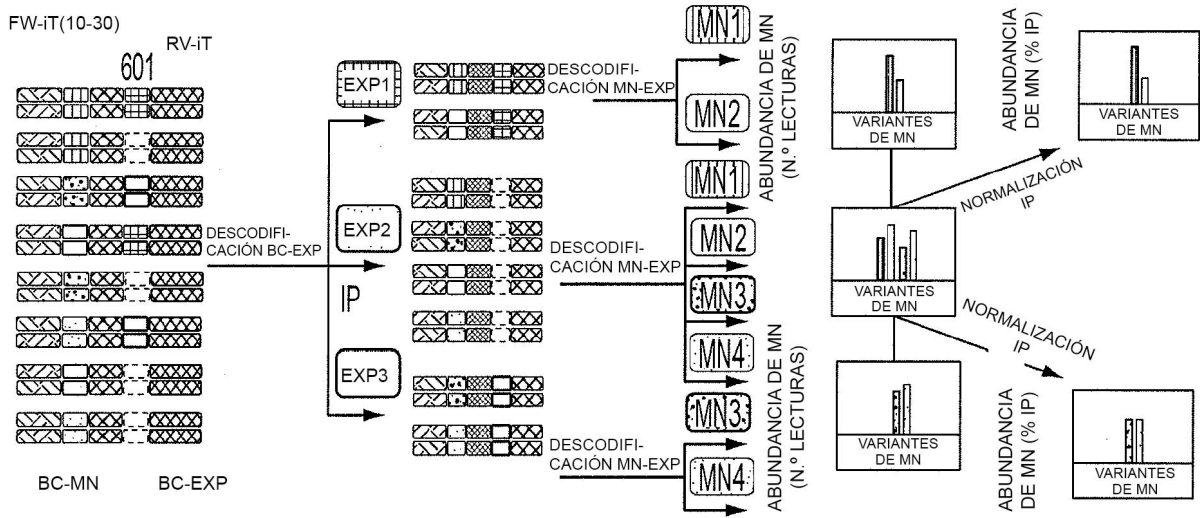


FIG. 10C



ssPhu-cebador FW:	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGATGCAGGATGCTGGAGAATCCCGGTGCCGAGG-3'
ssPhu-cebador RV:	5'-GTGACAGGATGTATATATCTGACACGTGCCTGG-3'
ds601:	5'-CTGGAGAATCCCGGTGCCGAGGCCGTCAATTGGTCTAGACAGCTCTAGCACCGCTTAAACGCACGTACGGCTGTCCCCCGCGTTTAAACCGCCAGGGGATTACTCCCTAGTCTCCAGGCAGGTGCAGATATATACATCCTGT-3' y hebra inferior complementaria
ss/ds BC-MN:	5'-HXXXXH-3' (ds: y hebra inferior complementaria)
ss/ds FW-iT10-30:	5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3' (ds: y hebra inferior complementaria)
ss/ds FW-iT1-30:	5'-CCTGCGTGTCTCCGACTCAG-3' (y hebra inferior complementaria)
ssRV-iT:	5'-CCTCTCTATGGGCAGTCCGGTAT-3'
ss601-RV:	5'-GGTGTAGAGCTGTCTACGACCAATTGAGC-3'
ss/dsBC-EXP:	5'DXXXXB-3 (ds: y hebra superior complementaria)

FIG. 10D

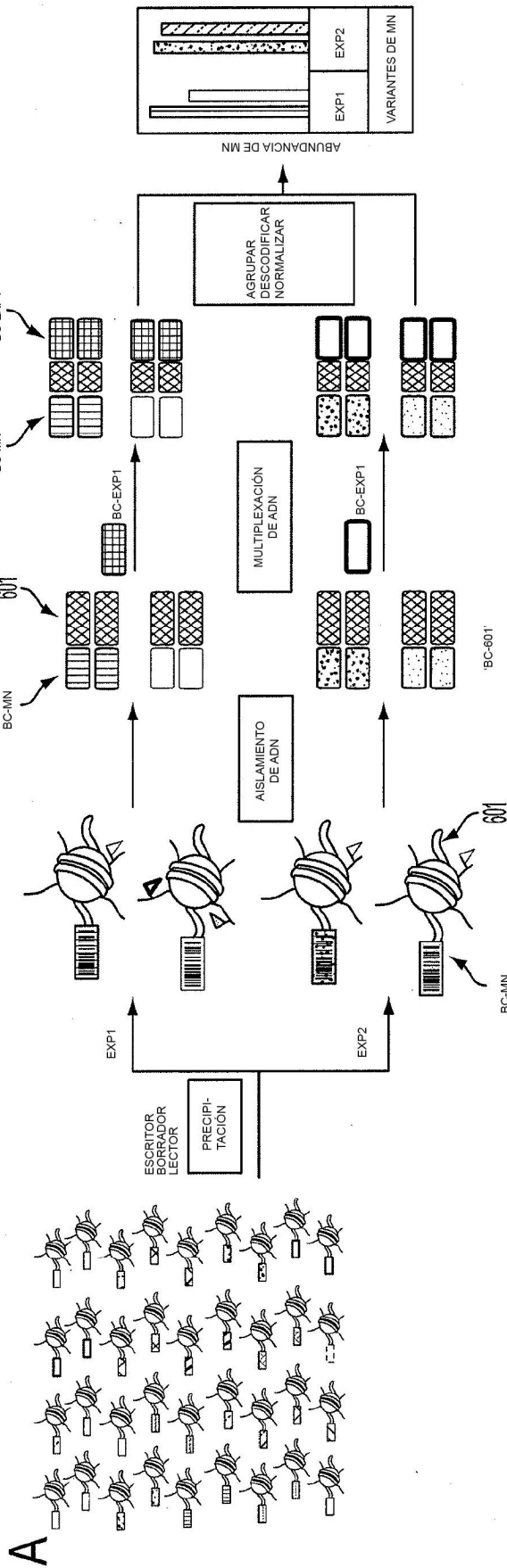


FIG. 11A

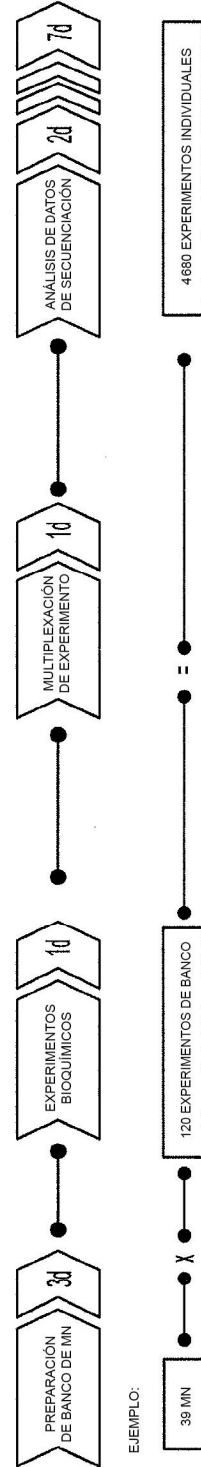


FIG. 11B

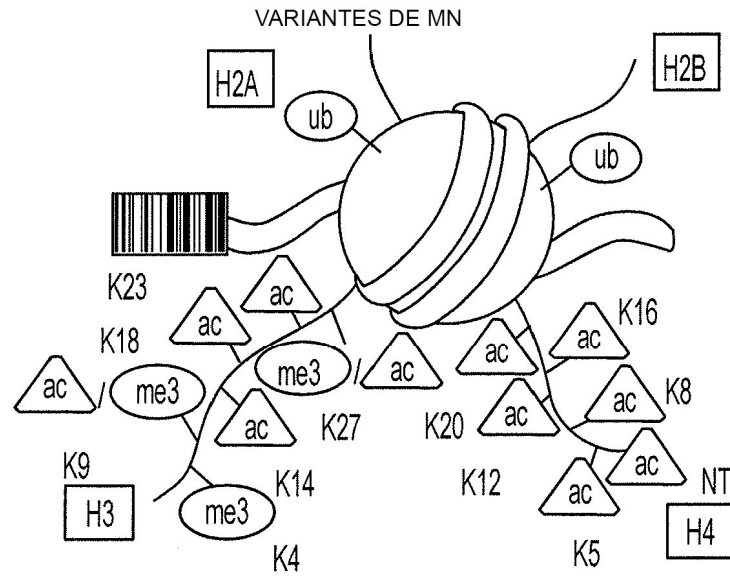


FIG. 12A

ID MN	H2A	H2B	H3	H4	código de barras de MN
1	wt	wt	wt	wt	AGTGCA
2	wt	wt	wt	K5ac	ATCATA
3	wt	wt	wt	K8ac	AGATAC
4	wt	wt	wt	K12ac	CGATGC
5	wt	wt	wt	K16ac	CTGTAT
6	wt	wt	wt	K20ac	AGATGT
7	wt	wt	wt	Kac 5	CATATC
8	wt	wt	K4me3	wt	CGTGTC
9	wt	wt	K4me3	K5ac	ATCTGA
10	wt	wt	K4me3	K8ac	TCACTC
11	wt	wt	K4me3	K12ac	ATGTGC
12	wt	wt	K4me3	K16ac	TCTGTA
13	wt	wt	K4me3	K20ac	ATGACA
14	wt	wt	K4me3	Kac 5	CGATGT
15	wt	wt	K9me3	wt	ACGCAC
16	wt	wt	K9me3	K5ac	CGATGA
17	wt	wt	K9me3	K8ac	TACACT
18	wt	wt	K9me3	K12ac	CATACA
19	wt	wt	K9me3	K16ac	CATCGC
20	wt	wt	K9me3	K20ac	TCGTGA
21	wt	wt	K9me3	Kac 5	ATATCT
22	wt	wt	K27me3	wt	ATCGAC
23	wt	wt	K27me3	K5ac	CGTATC
24	wt	wt	K27me3	K8ac	CTCTCA
25	wt	wt	K27me3	K12ac	ACACGT
26	wt	wt	K27me3	K16ac	CAGCTC
27	wt	wt	K27me3	K20ac	TGCGCT
28	wt	wt	K27me3	Kac 5	CGTAGC
29	wt	wt	Kac 5	wt	TGTATC
30	wt	ubH2B	wt	wt	ACATCT
31	wt	ubH2B	K4me3	wt	AGATGA
32	wt	ubH2B	K9me3	wt	CACATC
33	wt	ubH2B	K27me3	wt	CGCGCA
34	wt	ubH2B	Kac 5	wt	ACAGAC
35	wt	ubH2B	wt	Kac 5	AGTACT
36	ubH2A	wt	wt	wt	TAGCGC
37	ubH2A	wt	K4me3	wt	CTATCA
38	ubH2A	wt	K9me3	wt	ATAGAC
39	ubH2A	wt	K27me3	wt	TATGTC

FIG. 12B

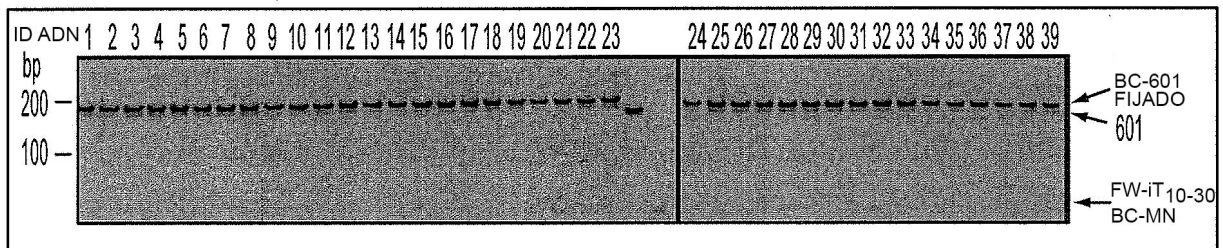


FIG. 12C

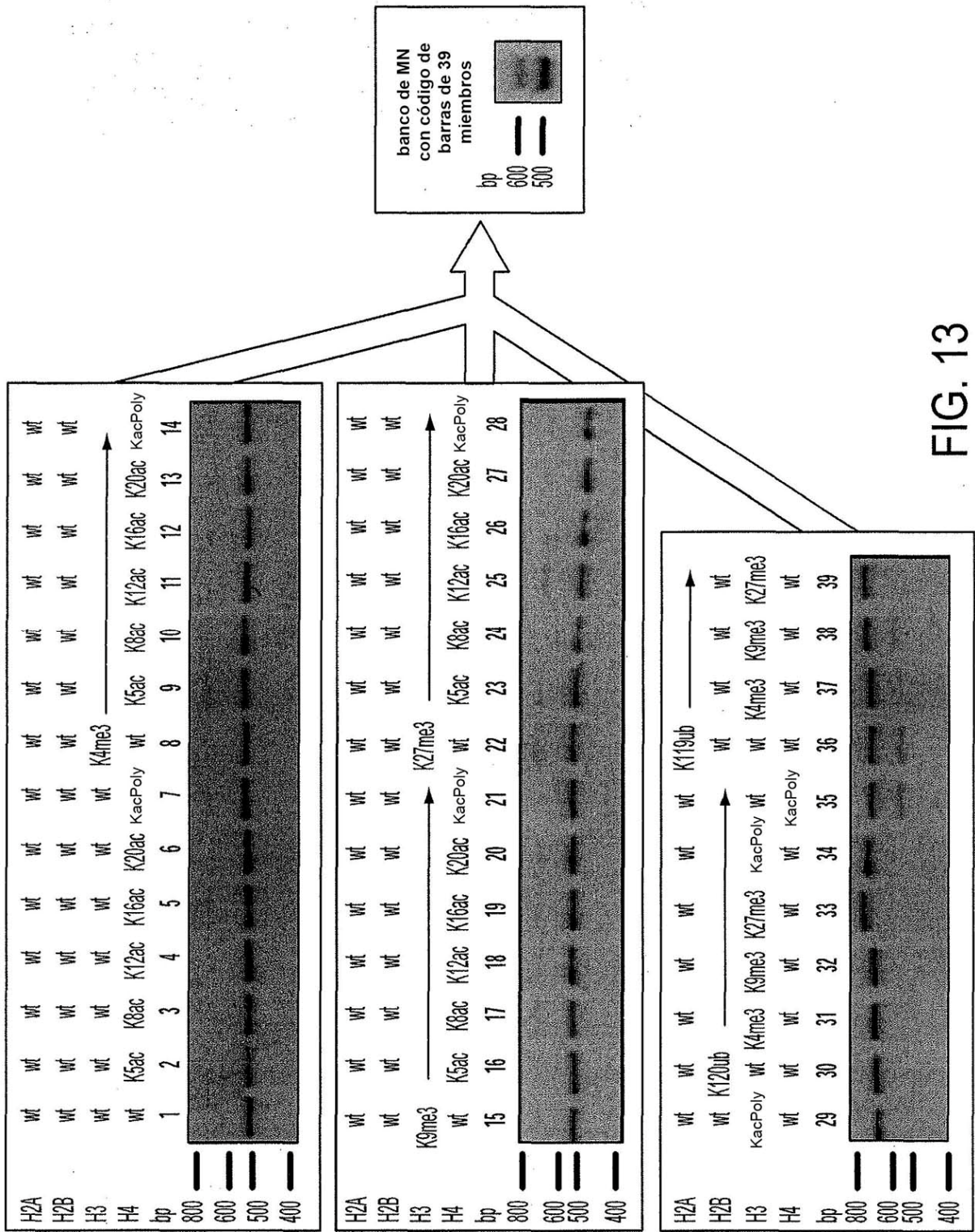


FIG. 13

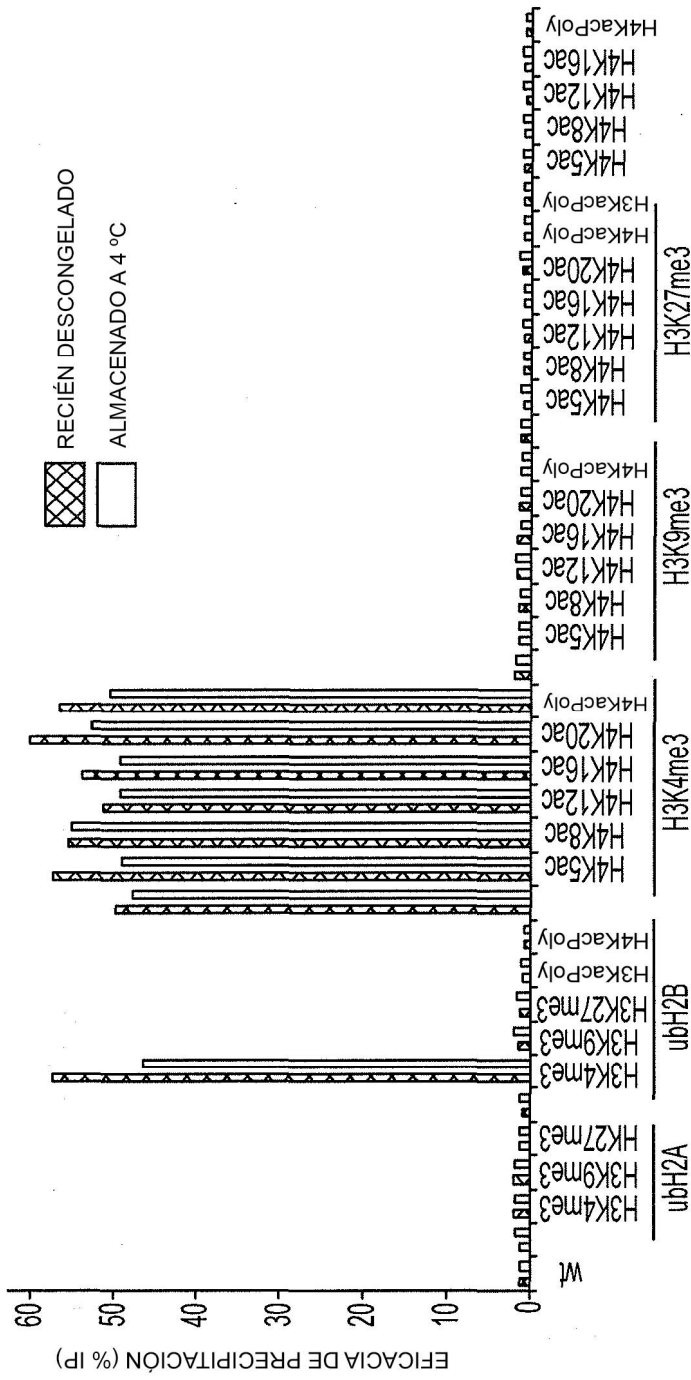


FIG. 14A

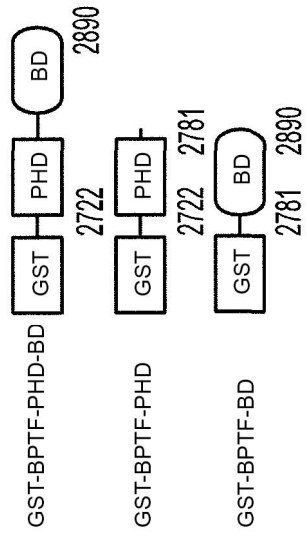


FIG. 14B

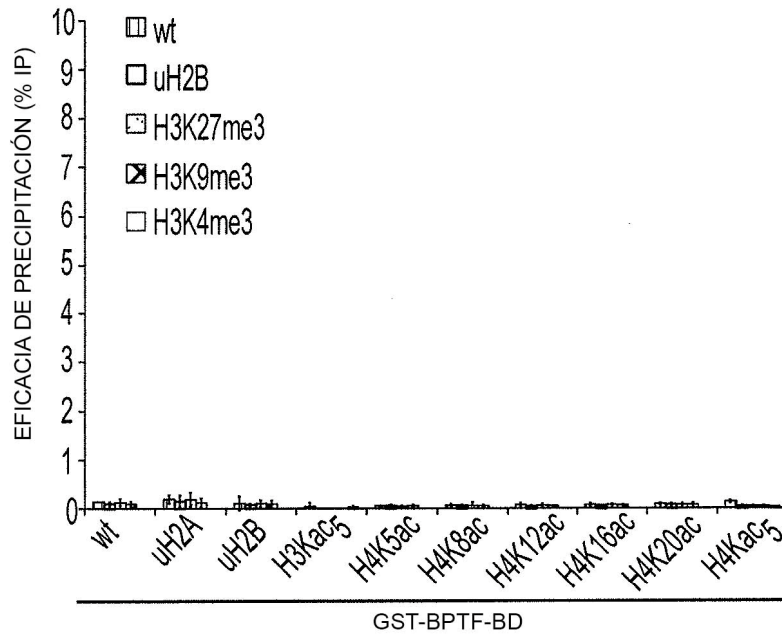


FIG. 14C

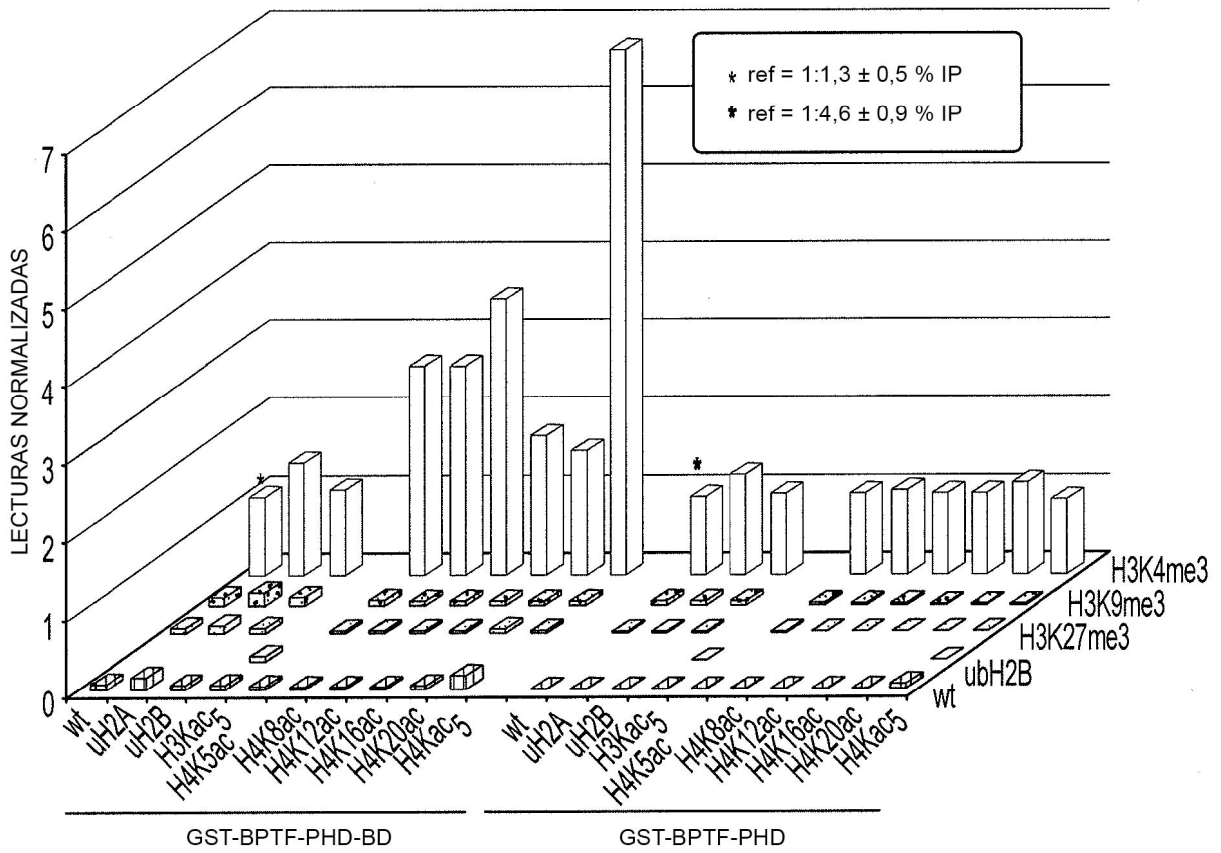


FIG. 14D

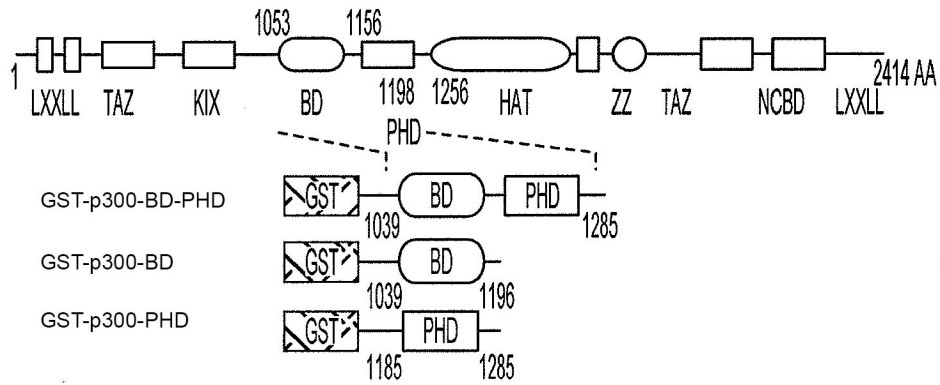


FIG. 15A

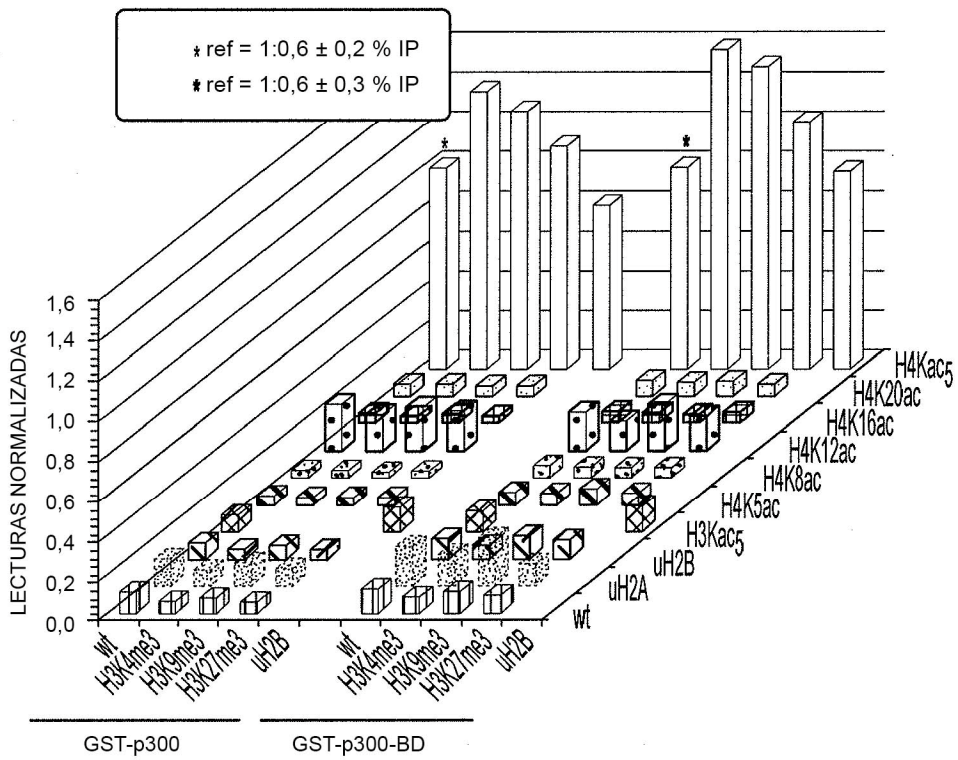


FIG. 15B

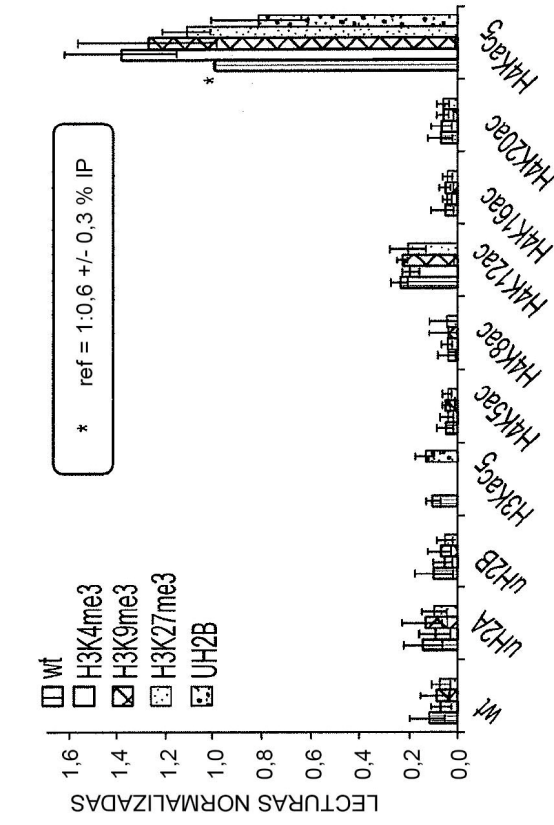


FIG. 15D

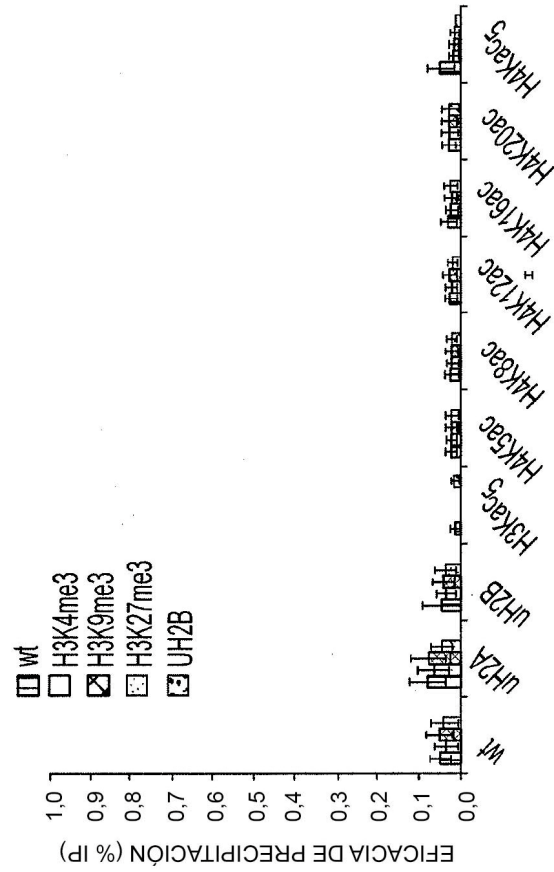


FIG. 15C

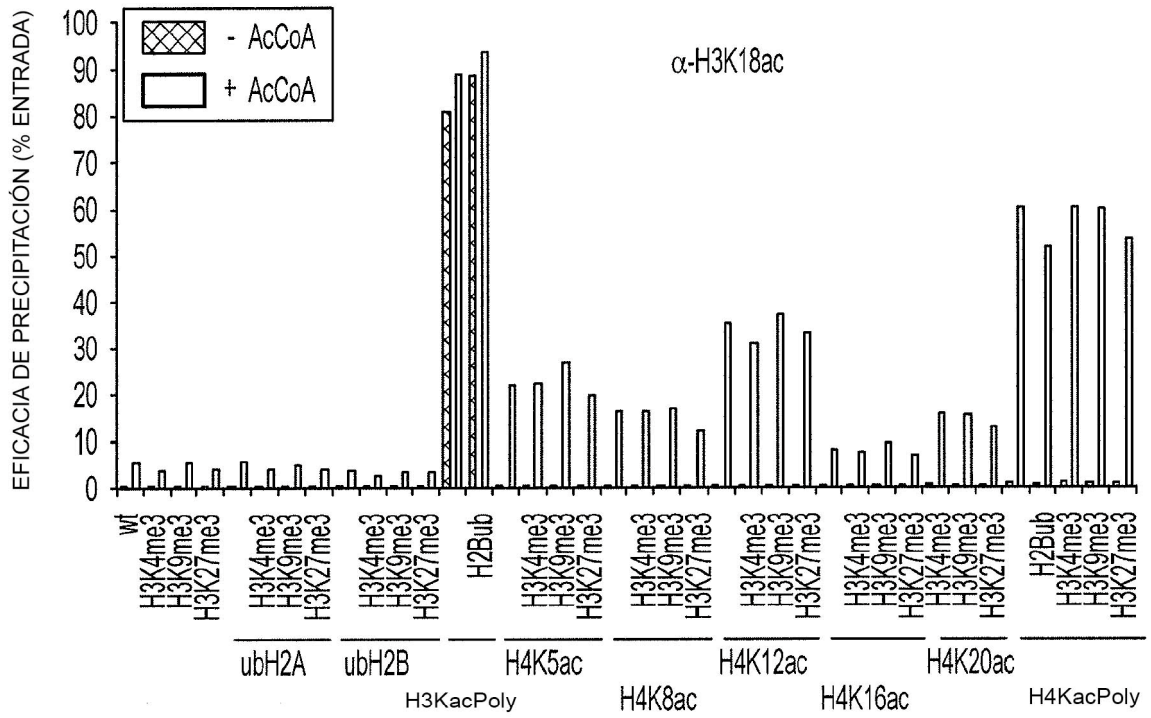


FIG. 16A

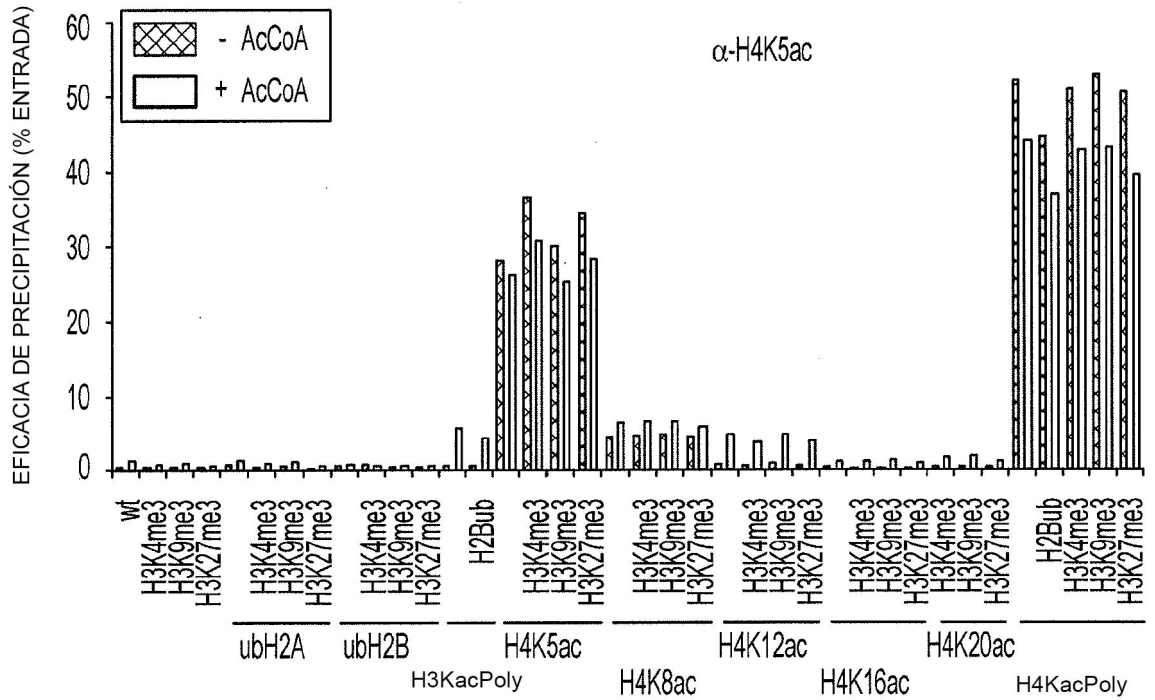


FIG. 16B

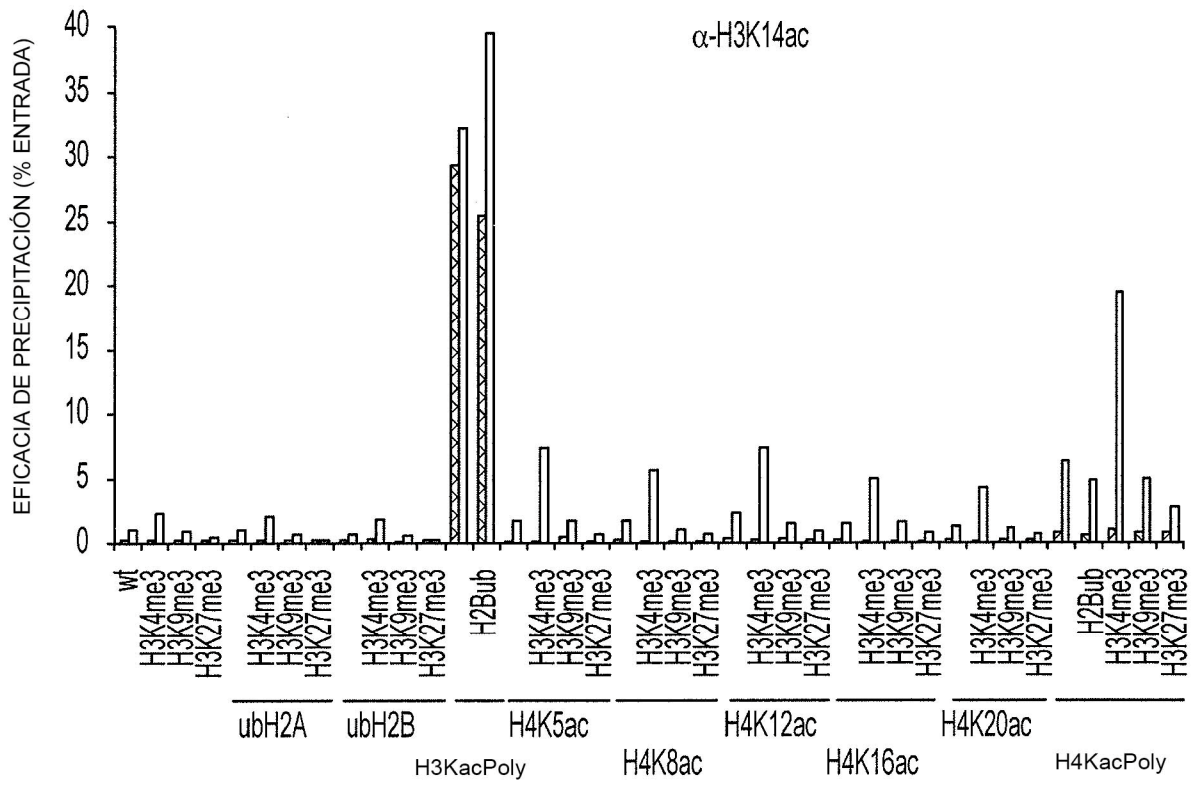


FIG. 16C

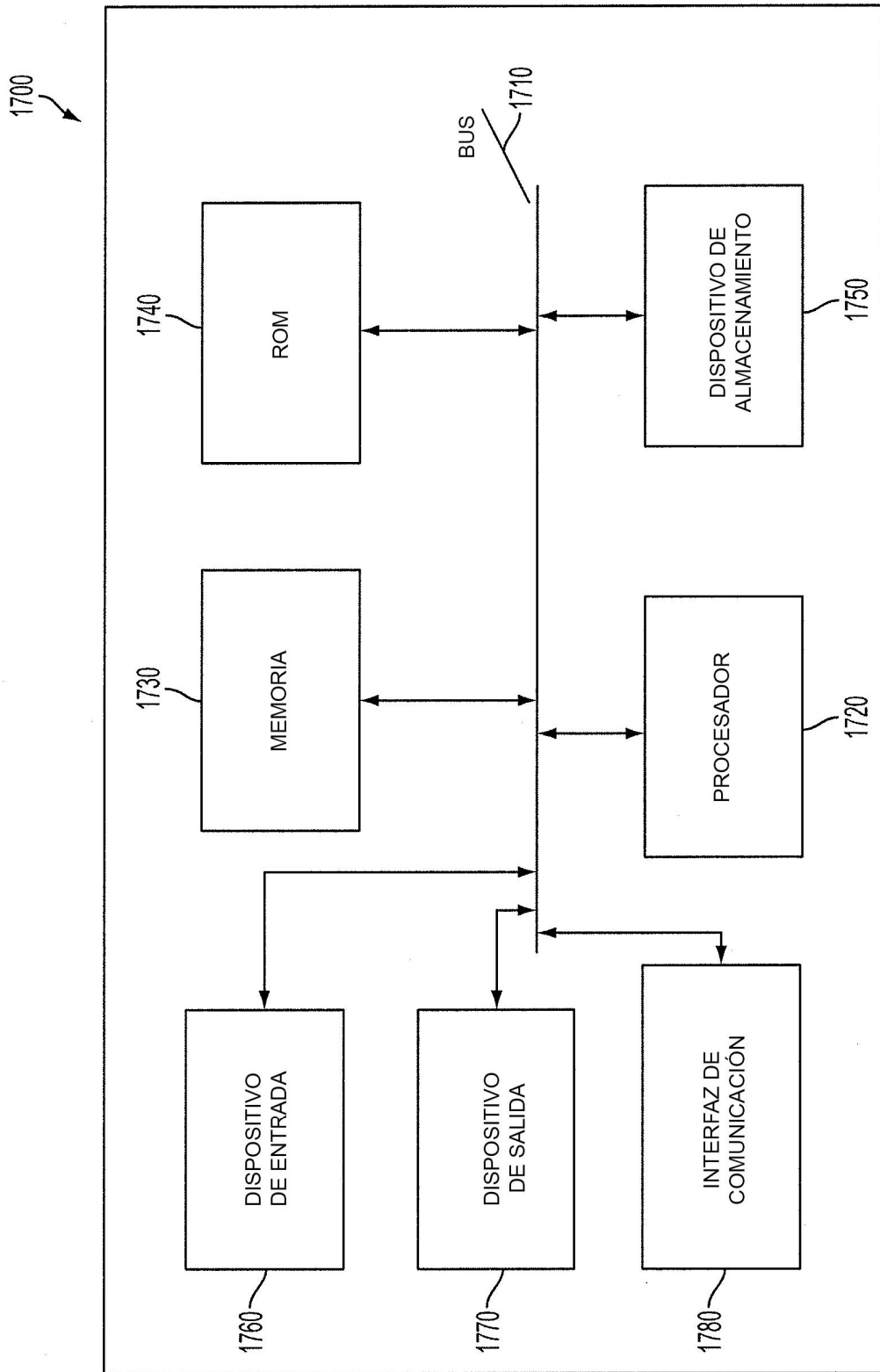


FIG. 17