

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 764 504**

51 Int. Cl.:

C12Q 1/68 (2008.01)

G01N 33/52 (2006.01)

G01N 33/53 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **16.04.2015 PCT/IL2015/050409**

87 Fecha y número de publicación internacional: **22.10.2015 WO15159295**

96 Fecha de presentación y número de la solicitud europea: **16.04.2015 E 15723307 (3)**

97 Fecha y número de publicación de la concesión europea: **16.10.2019 EP 3132057**

54 Título: **Métodos y equipos para analizar restos de unión al ADN fijados al ADN**

30 Prioridad:

17.04.2014 US 201461980630 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
03.06.2020

73 Titular/es:

**YEDA RESEARCH AND DEVELOPMENT CO., LTD. (100.0%)
at the Weizmann Institute of Science, P.O. Box 95
7610002 Rehovot, IL**

72 Inventor/es:

**AMIT, IDO;
LARA-ASTIASO, DAVID y
GURY, MEITAL**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 764 504 T3

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

DESCRIPCIÓN

Métodos y equipos para analizar restos de unión al ADN fijados al ADN

La presente descripción, en algunos aspectos de la misma, se refiere a métodos y equipos para analizar restos de unión a ADN fijados al ADN y, más especialmente, pero no exclusivamente, para analizar la unión a
5 histonas en una célula.

Las funciones y la respuesta específicas del tipo de célula se basan en una regulación génica estricta, orquestada por el panorama de la cromatina. El dogma actual de diferenciación celular establecido hace más de medio siglo por Conrad Waddington presenta un modelo de cierre progresivo del genoma. Según este
10 modelo, la diferenciación es una transición gradual desde un estado de cromatina abierta en células madre multipotentes a un estado de cromatina compactada en células diferenciadas. Este modelo está respaldado por la modificación reciente de histonas de todo el genoma que analiza la presencia o el nivel de células madre embrionarias en comparación con las células diferenciadas terminales. La hematopoyesis es un proceso de diferenciación paradigmático en el que una única célula madre hematopoyética da lugar a una gran cantidad de tipos de células (esencialmente todo el sistema sanguíneo) a través de una serie de células
15 madre intermedias caracterizadas. La regulación de la cromatina tiene una función principal en la hematopoyesis y las mutaciones o la pérdida de los factores de cromatina alteran críticamente el resultado hematopoyético. Además, la presencia de análisis de cromatina en todo el genoma o el nivel de estudios han puesto de manifiesto grandes diferencias en las modificaciones de histonas y las cartografías de unión de TF en diferentes inmunocitos maduros. Todos estos resultados apuntan a reajustes dinámicos de cromatina en
20 algún momento durante la hematopoyesis.

El estudio exhaustivo de los episodios de cromatina durante la hematopoyesis se ha visto obstaculizado por la baja sensibilidad y reproducibilidad para pequeños números de células con los protocolos actuales de inmunoprecipitación de cromatina (ChIP). Estos protocolos requieren varias etapas enzimáticas con un rendimiento limitado cuando el ADN de entrada está por debajo de la banda de nanogramos. Mientras que
25 una célula de mamífero diploide tiene por término medio aproximadamente 4-8 pg de ADN, las pérdidas después de ChIP reducen el ADN disponible para el análisis en 2 a 3 órdenes de magnitud, estableciendo el límite inferior para el análisis de cromatina en todo el genoma en 50.000 células (TS Furey, *Nat. Rev. Genet.* 13, 840-852 (2012)).

La amplificación del material de ChIP alivia parcialmente este problema a costa de introducir sesgos de
30 amplificación (P. Shankaranarayanan *et al.*, *Nat. Methods* 8, 565-567 (2011); M. Garber *et al.*, *Mol. Cell* 47, 810-822 (2012)). Además, dichos procesos de amplificación hacen que los protocolos de ChIP sean laboriosos y propensos a sesgos.

Los antecedentes adicionales incluyen Blecher Gonen *et al.*, *Nature Protocols*, 8, 539-554 (2013), Solicitud de Patente de los EE.UU. nº 20140024052, documentos WO 2013134261 y WO 2002014550.

35 El documento WO 2012047726 describe que se proporcionan métodos y equipos para la inmunoprecipitación de cromatina.

El documento WO 2013134261 se refiere a secuenciación microfluídica y epigenética. En una serie de realizaciones, las células contenidas dentro de un gran número de gotitas microfluídicas se lisan y se marca el ADN (p. ej., de nucleosomas) dentro de las gotas, p. ej., con adaptadores que contienen una secuencia de
40 identificación. Los adaptadores también pueden contener otras secuencias, p. ej., secuencias de restricción, secuencias principales, etc., para ayudar en el análisis posterior. Después de marcar con adaptadores, el ADN de las diferentes células puede combinarse y analizarse, p. ej., para determinar información epigenética sobre las células. Por ejemplo, el ADN puede separarse basándose en determinadas modificaciones (p. ej., metilación), y el ADN de los nucleosomas separados puede secuenciarse utilizando técnicas como la
45 inmunoprecipitación de cromatina ("ChIP"). En algunos casos, las secuencias de ADN también puede alinearse con genomas, p. ej., para determinar qué partes del genoma se modificaron epigenéticamente, p. ej., por metilación.

El documento WO 2012050963 describe agentes de control para ensayos de inmunoprecipitación, métodos para usar los agentes de control y equipos que comprenden los agentes de control.

50 Ronnie Blecher-Gonen *et al.*, se refieren a "High-throughput chromatin immunoprecipitation for genomewide mapping of in vivo protein-DNA interactions and epigenomic states", *Nature protocols*, (20130221), vol. 8, nº 3, doi: 10.1038/nprot.2013.023, ISSN 1754-2189, páginas 539 - 554.

David R. Lara-Astiaso *et al.*, se refieren a "Chromatin state dynamics during blood formation", *Science*, (20140807), vol. 345, nº 6199, doi: 10.1126/science.1255083, ISSN 0036-8075, páginas 943 - 949.

55 Dahl John Arne *et al.* se refieren a "Q²ChIP, a quick and quantitative chromatin immunoprecipitation assay, unravels epigenetic dynamics of developmentally regulated genes in human carcinoma cells" células madre,

alphamed press, dyton, OH, EE. UU., vol.25, nº 4, 1 de abril de 2007, páginas 1037-1046, XP002604155.

El documento WO 2011/096926 se refiere a mejoras en las técnicas de secuenciación de alto rendimiento de inmunoprecipitación de cromatina que han permitido la creación de cartografías de cromatina a partir de tamaños de muestras biológicas limitados que no pueden evaluarse utilizando inmunoprecipitación convencional de cromatina- protocolos de secuenciación. Por ejemplo, se utiliza un cebador universal modificado que incorpora enzimas de restricción en fragmentos de inmunoprecipitación de cromatina antes de la amplificación. El método mejorado permite que los tamaños de muestra sean varios órdenes de magnitud menores que los requeridos para las técnicas ChIPSeq normales.

John Arne Dahl *et al.* se refieren a "microChiP-a rapid micro chromatin immunoprecipitation assay for small cell samples and biopsies", *Nucleic Acids Research*, vol. 36, nº 3, 1 de febrero de 2008, página e15, XP055465549.

Acevedo Luis G. *et al.* se refieren a "Genome-scale ChIP-chip analysis using 10,000 human cells", *Biotechniques*

rapid dispatches, informa Healthcare, EE. UU., vol.43, nº 6, 1 de diciembre de 2007, páginas 791-797, XP002604156.

Compendio de la invención

La presente invención se refiere a un método de caracterizar la dinámica de la cromatina, comprendiendo el método:

- (a) obtener al menos dos muestras de poblaciones celulares, comprendiendo cada muestra células de un tipo de células diferentes o procedentes de un determinado sujeto;
- (b) reticular la cromatina celular total de al menos dichas dos muestras;
- (c) aislar dicha cromatina celular total en un soporte sólido;
- (d) marcar el ADN de dicha cromatina celular total, en donde el marcaje distingue entre dicha cromatina celular total de la primera de al menos dichas dos muestras y dicha cromatina celular total de la segunda de al menos dichas dos muestras;
- (e) liberar dicha cromatina celular total de dicho soporte sólido;
- (f) agrupar al menos dos muestras de cromatina celular total;
- (g) aislar el ADN que forma un complejo con una proteína de unión al ADN de dicha cromatina celular total usando un agente que se une específicamente a dicha proteína de unión al ADN; y
- (h) analizar dicho ADN de dicha cromatina celular total.

Preferiblemente, dicho análisis comprende secuenciar dicho ADN.

Preferiblemente, comprende además amplificar dicho ADN después de la etapa (g) y antes de la etapa (h).

Preferiblemente, comprende además analizar dicha proteína de unión a ADN de dicha cromatina.

Preferiblemente, comprende además someter a ultrasonidos dicho ADN después de la etapa (b) y antes de la etapa (c) de tal manera que dicho ADN no sea más largo de 500 bases.

Preferiblemente, dicha proteína de unión a ADN es una histona.

Preferiblemente, dicho agente es un anticuerpo.

Preferiblemente, no se está llevando a cabo utilizando un dispositivo microfluídico.

Preferiblemente, el número de células de la muestra está comprendido entre 100 y 10.000

Breve descripción de las diversas vistas de los dibujos

Algunos aspectos de la descripción se describen en la presente memoria, a modo de ejemplo solamente, con referencia a los dibujos e imágenes que se acompañan. Con referencia específica ahora a los dibujos en detalle, se enfatiza que los detalles mostrados son a modo de ejemplo y para exposición ilustrativa de la descripción. A este respecto, la descripción considerada con los dibujos pone en evidencia para los expertos en la técnica cómo se pueden poner en práctica aspectos de la invención.

En los dibujos:

Figs. 1A-D. Inmunoprecipitación de cromatina con primera indización para describir modificaciones de histonas y unión a TF. (A) Diagrama esquemático del enfoque de iChIP para inmunoprecipitación de cromatina que implica una etapa inicial de código de barras de cromatina antes de ChIP con el anticuerpo deseado. (B) Perfiles normalizados H3K4me3 de picos encontrados en una región de 100 Kb en el locus TNF (los genes se indican a continuación) obtenidos con iChIP de cantidades decrecientes de células dendríticas procedentes de la médula ósea (BMDC). Arriba, en rojo, el perfil de H3K4me3 obtenido usando ChIP convencional con 20 millones de células (15). A continuación, en tonos azules, los perfiles de H3K4me3 obtenidos con iChIP. (C) Perfiles normalizados de PU.1, H3K4me1, H3K4me2, H3K4me3 y H3K27ac encontrados en una región de 250 Kb en el locus TNF obtenido con iChIP con 10⁴ BMDC. (D) Gráficos de dispersión que muestran la correlación entre las reproducciones representativas de HK4me3 iChIP. De izquierda a derecha: correlación entre 10⁴ reproducciones de células, 500 reproducciones de células y 500 en comparación con 10⁴ células.

Figs. 2A-D. Dinámica de la cromatina en hematopoyesis. (A) Diagrama esquemático de las fases de diferenciación hematopoyética incluidas en este estudio (se muestran las plaquetas, pero no se incluyeron). La flecha punteada indica progenitores intermedios no medidos en este estudio. Código de color: verde para progenitores multipotentes, naranja para linaje mielocítico (incluidos los progenitores oligopotentes CMP y GMP), azul para linaje linfoide, rojo para linaje eritrocítico. (B) Dendrograma de agrupamiento de tipos de linfocitos basado en los perfiles H3K4me1 (izquierda) y niveles de RNA-seq (derecha) que muestran la asociación diferencial de progenitores de linaje (CMP, GMP, MEP). (C) Ejemplos representativos de la señal H3K4me1 (tipos de células marcadas a la izquierda) en varios locus (de izquierda a derecha): Gata 2 para progenitores, F7-F10 y S100a8 para linaje mielocítico, Gypa y CpoX para linaje eritrocítico, Bcl1 1b para linfocitos T. Se muestran coberturas de lecturas normalizadas en una región de 100 Kb alrededor del cuerpo del gen. (D) Perfiles de modificaciones H3K4me1, H3K4me2, H3K4me3, H3K27ac y niveles de expresión de ARN en dos locus de genes específicos de linaje: S 100a8 en (mielocítico) y Gypa (eritrocítico). Se muestran características para los tipos de células específicas del linaje y las células madre multipotentes en una región de 100 Kb alrededor del cuerpo del gen. Los potenciadores específicos del supuesto linaje están sombreados en azul. Los genes específicos del linaje se indican a continuación (negro), así como otros genes en los locus (gris claro).

Figs. 3A-F. La hematopoyesis evoluciona mediante la ganancia y pérdida de potenciadores específicos de linaje (A) Cartografía térmica que muestra 48.415 potenciadores de hematopoyesis agrupados con medios de K (K = 9) del recuento medio de lecturas dentro de las regiones H3K4me1 (16) anotadas con la conservación de la secuencia genómica de los potenciadores, derecha (escala azul; (16)). Las anotaciones para locus de genes específicos de linaje importantes se muestran a la izquierda. (B) Vista esquemática en árbol de tres grupos de potenciadores representativos; progenitores (V), mielocíticos (VI) y eritrocíticos (IX). El relleno de color representa fases con la marca H3K4me1 en el diagrama de barras del potenciador (C) que muestra el número de potenciadores ganados (arriba, gris claro) y perdidos (abajo, gris oscuro) durante el desarrollo de HSC a células hematopoyéticas maduras a lo largo de cada linaje. (D) Gráfico de barras que muestra el porcentaje de potenciadores de nueva procedencia (a partir de C) creados en cada etapa de desarrollo desde HSC hasta células hematopoyéticas maduras. (E) Gráficos de dispersión que muestran la señal H3K27ac frente a la señal H3K4me1 tanto en las células CMP como en las EryA. Los potenciadores activos (H3K4me1 positivo, H3K27ac positivo) están coloreados en rojo, los potenciadores preparados (H3K4me1 positivo, H3K27ac negativo) en gris claro, potenciadores inactivos (H3K4me1 negativo, H3K27ac negativo) en azul. (F) Proporciones de potenciadores activos (rojo) y preparados (gris claro) en cada uno de los tipos de células estudiados. H3K27Ac negativo) en azul. (F) Proporciones de potenciadores activos (rojo) y preparados (gris claro) en cada uno de los tipos de células estudiados. H3K27Ac negativo) en azul. (F) Proporciones de potenciadores activos (rojo) y preparados (gris claro) en cada uno de los tipos de células estudiados.

Figs. 4A-F. Los potenciadores específicos de linaje están asociados con cohortes de factores de transcripción (A) La señal ATAC-seq coexiste temporalmente con H3K4me1 en potenciadores específicos de linaje. Se muestran perfiles normalizados en 9 células hematopoyéticas para ATAC-seq (negro) y H3K4me1 en locus específico eritrocítico, Gypa (rojo) y mielocítico específico, F7-10 (naranja). Se muestran picos encontrados en una región de 100 Kb alrededor del cuerpo del gen. Supuestos potenciadores (K4me1) sombreados en azul, secuencia de inicio de la transcripción sombreado en rojo. El zoom en el pico ATAC-seq se muestra a continuación. (B) Gráfico de dispersión que muestra la correlación entre la señal H3K4me1 y la señal ATAC-seq restringida de potenciador en cuatro tipos de células: linfocitos B, citolíticos, células MPP y CMP (los colores representan densidad o puntos). (C) Mapa térmico que muestra los valores p del motivo del factor de transcripción (prueba de Kolmogorov-Smirnov) para los potenciadores específicos del tipo de célula indicados (16). El rojo indica enriquecimiento significativo ($p < 10^{-5}$) del motivo asociado al TF marcado, el verde indica agotamiento del motivo ($p < 10^{-5}$). El blanco indica que no hay enriquecimiento significativo o ninguna expresión de ARN. (D) Gráficos de barras que muestran perfiles de expresión génica a través de los tipos de células hematopoyéticas para factores de transcripción representativos (de C). (E) Esquema del árbol hematopoyético que muestra los factores de transcripción representativos que regulan los potenciadores de linaje en cada tipo de célula según el modelo logístico. (F) Esquema del establecimiento de potenciadores

específicos de linaje en el linaje mielocítico mediante un mecanismo convencional (derecha), así como mediante el establecimiento mediado por TF de potenciadores de nueva procedencia (izquierda).

Descripción de aspectos específicos de la descripción

La presente descripción, en algunos aspectos de la misma, se refiere a métodos y equipos para analizar restos de unión a ADN unidos al ADN y, más particularmente, pero no exclusivamente, para analizar la unión a histonas en una célula.

Antes de explicar al menos un aspecto de la descripción en detalle, debe entenderse que la descripción no está necesariamente limitada en su aplicación a los detalles establecidos en la siguiente descripción o ejemplificados por los ejemplos. La descripción es capaz de otros aspectos o de ser practicada o llevada a cabo de varias maneras.

Los protocolos actuales de inmunoprecipitación de cromatina (ChIP) requieren varias etapas enzimáticas con un rendimiento limitado cuando el ADN de entrada está por debajo de la banda de los nanogramos. Mientras que una célula de mamífero diploide media tiene aproximadamente 4-8 pg de ADN, las pérdidas después de ChIP reducen el ADN disponible para el análisis en 2-3 órdenes de magnitud, estableciendo el límite inferior para el análisis de cromatina en todo el genoma en 50.000 células. La amplificación del material de ChIP alivia parcialmente este problema a costa de introducir sesgos de amplificación.

Los presentes inventores han ideado un método novedoso para analizar la cromatina procedente de un pequeño número de células. En este protocolo, el código de barras se realiza directamente en la cromatina celular total, evitando así las reacciones enzimáticas de bajo aporte que se producen en ChIP convencional. Sobre todo, esto permite que múltiples muestras con códigos de barras de cromatina se agrupen para ChIP en el mismo pozo, reduciendo aún más los requisitos iniciales de entrada y aumentando la reproducibilidad de muestras cruzadas. Para minimizar las etapas de centrifugación, las células se reticulan antes de la clasificación y el fraccionamiento del ADN. Después, la cromatina cortada se inmoviliza e indiza. La cromatina indizada se libera y se combina con cromatina de otras muestras. Por último, se realiza la ChIP con el anticuerpo deseado, y un solo grupo de código de barras de cromatina puede dividirse en múltiples ChIP para caracterizar varias modificaciones de cromatina.

Los presentes inventores demostraron que este protocolo es muy reproducible para bajos números de células (unos pocos cientos de células) al tiempo que aumenta la sensibilidad y el rendimiento (Figura 1). Mientras se reduce la presente descripción a la práctica, los presentes inventores codificaron cantidades decrecientes (10.000 a 500 células) de cromatina aislada de células dendríticas procedentes de médula ósea (BMDC), por triplicado, y realizaron ChIP en la cromatina codificada con un anticuerpo para histona H3 lisina 4 mono- y trimetilada (H3K4me1 y H3K4me3) (Figuras 1B-D). Ellos confirmaron la reproducibilidad de los picos de H3K4me3 entre reproducciones de 10.000 células ($r = 0,95$) por debajo de 500 células ($r = 0,85$). Las características de ChIP que obtuvieron se correlacionaron con ChIP-seq convencional en 10 millones de BMDC ($r = 0,92$).

Por consiguiente, los presentes inventores proponen que su protocolo ChIP pueda usarse para caracterizar otras modificaciones de histonas y secuencias de unión de factores de transcripción para números de células bajos (Figuras 1C-D).

Por lo tanto, según un aspecto de la presente descripción, se proporciona un método de análisis de ADN que, en una célula, se une a un resto de unión a ADN, comprendiendo el método:

- (a) obtener al menos dos muestras de complejos de ADN unidos a un resto de unión a ADN;
- (b) marcar el ADN de los complejos, en donde el marcaje distingue entre los complejos de al menos la primera de las dos muestras y los complejos de al menos la segunda de las dos muestras;
- (c) agrupar al menos dos muestras de complejos; y
- (d) aislar los complejos usando un agente que se une específicamente al resto de unión al ADN; y
- (e) analizar el ADN de los complejos.

El ADN que se analiza puede proceder de una célula eucariota o procariota. Según una realización, el ADN procede una célula de mamífero (p. ej., humano). En la célula (es decir, en el entorno *in vivo*) de donde procede, el ADN puede unirse de forma permanente o temporal al resto de unión al ADN.

Según un aspecto concreto, el método no se lleva a cabo utilizando un dispositivo microfluídico.

Como se emplea en la presente memoria, la frase "resto de unión al ADN" se refiere a un agente que se une al ADN (en una secuencia específica o de manera no específica. El resto de unión al ADN puede unirse al ADN mediante intercalado, unión al surco y/o enlace covalente.

ES 2 764 504 T3

En un aspecto, el resto de unión a ADN es un polipéptido o péptido de unión a ADN.

En otro aspecto, el resto de unión a ADN es un fármaco (p. ej., un agente de molécula pequeña).

5 Los polipéptidos de unión al ADN incluyen factores de transcripción que modulan el proceso de transcripción, varias polimerasas, nucleasas que escinden moléculas de ADN e histonas que están involucradas en la encapsulación y la transcripción de cromosomas en el núcleo celular. Las proteínas de unión al ADN pueden incorporar dominios como el dedo de cinc, la hélice-giro-hélice y la cremallera de leucina (entre muchos otros) que facilitan la unión al ácido nucleico.

Según un aspecto concreto, la proteína de unión al ADN es una histona.

10 La siguiente es una lista de proteínas de histonas humanas que pueden analizarse según los métodos descritos en la presente memoria.

Tabla 1

Superfamilia	Familia	Subfamilia	Miembros
Enlazador	H1	H1F	H1F0, H1FNT, H1FOO, H1FX
		H1H1	HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, HIST1H1T
Núcleo	H2A	H2AF	H2AFB1, H2AFB2, H2AFB3, H2AFJ, H2AFV, H2AFX, H2AFY, H2AFY2, H2AFZ
		H2A1	HIST1H2AA, HIST1H2AB, HIST1H2AC, HIST1H2AD, HIST1H2AE, HIST1H2AG, HIST1H2AI, HIST1H2AJ, HIST1H2AK, HIST1H2AL, HIST1H2AM
		H2A2	HIST2H2AA3, HIST2H2AC
	H2B	H2BF	H2BFM, H2BFS, H2BFWT
		H2B1	HIST1H2BA, HIST1H2BB, HIST1H2BC, HIST1H2BD, HIST1H2BE, HIST1H2BF, HIST1H2BG, HIST1H2BH, HIST1H2BI, HIST1H2BJ, HIST1H2BK, HIST1H2BL, HIST1H2BM, HIST1H2BN, HIST1H2BO
		H2B2	HIST2H2BE
	H3	H3A1	HIST1H3A, HIST1H3B, HIST1H3C, HIST1H3D, HIST1H3E, HIST1H3F, HIST1H3G, HIST1H3H, HIST1H3I, HIST1H3J
		H3A2	HIST2H3C
		H3A3	HIST3H3
	H4	H41	HIST1H4A, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E, HIST1H4F, HIST1H4G, HIST1H4H, HIST1H4I, HIST1H4J, HIST1H4K, HIST1H4L
		H44	HIST4H4

Los ejemplos de modificaciones de histonas en la regulación de la transcripción se resumen en la tabla 2 a continuación.

Tabla 2

Tipo de modificación	Histona									
	H3K4	H3K9	H3K14	H3K27	H3K79	H3K36	H4K20	H2BK5		
monometilación	activación	activación		activación	activación		activación	activación		
dimetilación		represión		represión	activación					
trimetilación	activación	represión		represión	activación, represión	activación		represión		
acetilación		activación	activación	activación						

Etapa 1: Obtención de al menos dos muestras de complejos de ADN unidos al resto de unión al ADN.

Las muestras son muestras celulares. La obtención de muestras celulares se lleva a cabo según métodos conocidos en la técnica. Las células pueden obtenerse de un fluido corporal (p. ej., sangre) o un tejido corporal. Las células pueden obtenerse de un sujeto (p. ej., de un mamífero) o pueden formar parte de un cultivo celular. Las células pueden surgir de un organismo sano, o de uno que está enfermo o se sospecha que está enfermo. Según un aspecto concreto, cada muestra comprende células de un tipo de célula concreta (es decir, una población homogénea de células). Así, por ejemplo, una de las muestras puede comprender granulocitos, otra de las muestras puede comprender macrófagos, otra de las muestras puede comprender monocitos, etc. Las muestras de un solo tipo de célula pueden obtenerse utilizando métodos conocidos en la técnica, por ejemplo, por clasificación FAC. Según aún otro aspecto, cada muestra comprende células de un origen determinado (p. ej. de un sujeto determinado).

Según otros aspectos, cada muestra comprende 100-10.000 células, 100-5.000 células, 100-2.500 células, 100-1.000 células, 100-7.500 células, 100-5.000 células, 100-2.500 células, 100-1.000 células, 100-750 células, 200-750 células (p. ej., aproximadamente 500 células).

- 15 Según un aspecto, las células se reticulan de forma reversible para garantizar que los restos de unión al ADN que se unen al ADN en el entorno *in vivo* (es decir, en la célula) permanecen unidos durante el procedimiento de inmunoprecipitación. Los agentes que pueden usarse para la reticulación reversible incluyen, entre otros, formaldehído o luz ultravioleta. Otros agentes incluyen, entre otros, compuestos homobifuncionales difluoro-2,4-dinitrobenzono (DFDNB), dimetil pimelimidato (DMP), disuccinimidil suberato (DSS), el reactivo carbodiimida EDC, psoralenos incluidos 4,5',8-trimetilpsoraleno, azidas fotoactivables como ¹²⁵I(S-[2-(4-azidosalicilamido) etil]tio]-2-tiopiridina), conocida también como AET, (N-[4-(p-azidosalicilamido)butil]-3'[2'-pridilditio]propionamida) también conocido como APDP, el reactivo químico de reticulación Ni(II)-NH₂-Gly-Gly-His-COOH conocido también como Ni-GGH, sulfosuccinimidil 2-[(4-azidosalicil)amino]etil]-1,3-ditiopropionato) también conocido como SASD, (N-14-(2-hidroxibenzoil)-N-1 1(4-azidobenzoil) -9-oxo-8,1 1,14-triaza-4,5-diteratodecanoato).

- 30 Los complejos de este aspecto de la descripción están aislados de las células. Por lo tanto, la presente descripción contempla la lisis de las células para liberar los complejos desde el interior. La lisis celular se puede realizar utilizando protocolos habituales que pueden llevar a cabo con éxito los expertos en la técnica, incluida la destrucción mecánica de las membranas celulares, como por congelación y descongelación repetidas, homogeneización, exposición a ultrasonidos, presión o filtración y el uso de enzimas y/o detergentes (p. ej. SDS). A los efectos de la inmunoprecipitación cromosómica, es importante que los quelantes de metales como EDTA y EGTA, así como los inhibidores de la proteasa, se agreguen a la reacción para evitar la degradación de los complejos de ADN de proteínas.

- 35 Según un aspecto concreto, el ADN de los complejos no tiene más de 1.000 pares de bases, y más preferiblemente no más de 500 pares de bases. Si el ADN en la muestra es más largo, la presente descripción contempla una etapa de corte o escisión del ADN. Esto puede efectuarse por exposición a ultrasonidos durante varios periodos de tiempo. El tiempo preciso para la exposición a ultrasonidos depende de las células en la muestra y la determinación del tiempo está dentro de la experiencia de un experto en la técnica. Ejemplos de baños de ultrasonidos que se pueden utilizar incluyen el NGS Bioruptor Sonicator (Diagenode) o el sonifier/sonicator Branson modelo 250, así como la digestión por enzimas de restricción por enzimas de corte frecuentes y raras que incluidas, entre otras, Ace I, Aci I, Acl I, Afe I, Afl It, Afl El, Age I, Ahd I, Alu I, Alw I, AlwN I, Apa I, ApaL I, Apo I, Asc I, Ase I, Ava I, Ava II, Avr II, Bae I, BamH I, Ban I, Ban π, Bbs I, Bbv I, BbvC I, BceA I, Beg I, BciV I, Bel I, Bfa I, BfrB I, Bgl I, Bgl II, Blp I, Bmr I, Bpm I, BsaA I, BsaB I, BsaH I, Bsa I, BsaJ I, BsaW I, BsaX I, BseR I, Bsg I, BsiE I, BsiHKA I, BsiW I, Bsl I, BsmA I, Bs B I, BsmF I, Bsm I, BsoB I, Bsp12861, BspD I, BspE I, BspH I, BspM I, BsrB I, BsrD I, BsrF I, BsrG I, Bsr I, BssH II, BssK I, BssS I, BstAP I, BstB I, BstE II, BstF5 I, BstN I, BstU I, BstX I, BstY I, BstZ171, Bsu361, Btg I, Btr I, Bts I, Cac8 I, Cla I, Dde I, Dpn I, Dpn II, Dra I, Dra HI, Drd I, Eae I, Eag I, Ear I, Eci I, EcoN I, EcoO109 I, EcoR I, EcoR V, Fau I, Fnu4H I, Fok I, Fse I, Fsp I, Hae π, Hae lü, Hga I, Hha I, Hinc II, Hind m, Hinf I, HinPI I, Hpa I, Hpa II, HpyI88 I, HpyI88 IE, Hpy99 1, HpyCH4iπ, HpyCH4IV, HpyCH4V, Hph I, Kas I, Kpn I, Mbo I, Mbo II, Mfe I, Mlu I, Mly I, Mnl I, Msc I, Mse I, Msl I, MspAI I, Msp I, Mwo I, Nae I, Nar I, Nci I, Nco I, Nde I, NgoM IV, Nhe I, Nla in, Nla IV, Not I, Nru I, Nsi I, Nsp I, Pac I, PaeR7 1, Pci I, PflF I, PflM I, Pie I, Pme I, Pml I, PpuM I, PshA I, Psi I, PspG I, PspOM I, Pst I, Pvu I, Pvu H, Rsa I, Rsr II, Sac I, Sac π, Sal I, Sap I, Sau3A I, Sau96 1, Sbf I, Sea I, ScrF I, SexA I, SfaN I, Sfc I, Sfo I, SgrA I, Sma I, Sml I, SnaB I, Spe, Sph I, Ssp I, Stu I, Sty I, Swa I, Taq I, Tfi I, Tli I, Tse I, Tsp45 I, Tsp509 I, TspR I, Tth1 11 1, Xba I, Xcm I, Xho I, Xma I y Xmn I.

- 55 Según un aspecto concreto, la enzima no es MNasa.

Otras enzimas que pueden usarse se describen además en la presente memoria a continuación.

Etapa 2: Marcaje del ADN de los complejos.

Según un aspecto, el marcaje se realiza después de la inmovilización de los complejos de ADN aislados. Los

inventores actuales conciben cualquier forma de inmovilización siempre que no interfiera con el marcaje del ADN.

Según un aspecto particular, los complejos no están inmovilizados en gotitas microfluídicas.

Según un aspecto, los complejos se inmovilizan sobre un soporte sólido. Los ejemplos de soportes sólidos contemplados por la presente descripción incluyen, entre otros, sefarosa, quitina, proteína A reticulada con agarosa, proteína G reticulada con agarosa, agarosa reticulada con otras proteínas, ubiquitina reticulada con agarosa, resina tiofílica, proteína G reticulada con agarosa, proteína L reticulada con agarosa y cualquier material de soporte que permita un aumento en la eficiencia de la purificación de complejos de proteína/ADN.

Según otro aspecto, los complejos se inmovilizan en un soporte sólido usando un anticuerpo que se une al menos al 30%, más preferiblemente al 40%, más preferiblemente al 50% de los complejos en la muestra. El anticuerpo de este aspecto de la presente descripción puede ser policlonal o monoclonal. Los anticuerpos pueden unirse a las proteínas completas, así como contra subconjuntos de aminoácidos epítopos particulares presentes en esas proteínas. Los anticuerpos pueden ser de cualquier origen (p. ej., conejo, cabra, humanizados).

Así, por ejemplo, cuando la muestra comprende cromatina, el anticuerpo puede unirse específicamente a una histona H2, H3 o H4. Según un aspecto particular, el anticuerpo se une específicamente a H3. Los anticuerpos que reconocen histonas están disponibles en el mercado de varios proveedores, incluidos por ejemplo Abeam y Pierce. Para la inmovilización, los anticuerpos están unidos a un soporte sólido incluido, entre otros, a perlas magnéticas. Otros soportes en fase sólida contemplados por la presente descripción incluyen, entre otros, sefarosa, quitina, proteína A reticulada con agarosa, proteína G reticulada con agarosa, agarosa reticulada con otras proteínas, ubiquitina reticulada con agarosa, resina tiofílica, proteína G reticulada con agarosa, proteína L reticulada con agarosa y cualquier material de soporte que permita un aumento en la eficiencia de la purificación de complejos de proteína/ADN.

Los métodos para unir anticuerpos a soportes sólidos son conocidos en la técnica. Por ejemplo, el enlace de anticuerpos con perlas magnéticas de soporte en fase sólida se puede lograr mediante el protocolo habitual (información y especificaciones del producto de Dynal Corporation) y los conocidos y expertos en la técnica son capaces de lograr crear este enlace. Las perlas se lavan brevemente en un amortiguador apropiado (p. ej., solución salina tamponada con fosfato (PBS), pH 7,4). Se añaden aproximadamente 0,1-1,5 µg de anticuerpo por ml de perlas, se ajusta el volumen y la mezcla se incuba durante un período de tiempo adecuado (p. ej., 12-24 horas a 4°C). Las perlas se recogen posteriormente con un imán y se elimina el sobrenadante. Las perlas se pueden lavar al menos una vez más (p. ej., en Tris-HCl 10 mM, pH 7,6) durante 16 a 24 horas más el complejo perla/anticuerpo está listo para la inmunoprecipitación de complejos proteína/ADN.

Las perlas magnéticas contempladas por la presente descripción incluyen las creadas por Dynal Corporation, como por ejemplo Dynabeads M-450 Tosylactivated (Dynal Corporation). Otros Dynabeads M-450 sin recubrimiento, Dynabeads M-280 Tosylactivated, Dynabeads M-450 Sheep anti-Mouse IgG, Dynabeads M-450 Goat anti-Mouse IgG, Dynabeads M-450 Sheep anti-Rat IgG, Dynabeads M-450 Rat anti-Mouse IgM, Dynabeads M-280 Sheep anti-Mouse IgG, Dynabeads M-280 Sheep anti-Rabbit IgG, Dynabeads M-450 Sheep anti-Mouse IgG1, Dynabeads M-450 Rat anti-Mouse IgG1, Dynabeads M-450 Rat anti-Mouse IgG2a, Dynabeads M-450 Rat anti-Mouse IgG2b, Dynabeads M-450 Rat anti-Mouse IgG3. Otras perlas magnéticas que también se contemplan en la presente descripción como útiles para los fines de inmunoprecipitación incluyen Dynabeads recubiertas con estreptavidina.

Un método alternativo para unir anticuerpos a perlas magnéticas u otro material de soporte en fase sólida contemplado por la presente descripción es el procedimiento de reticulación química. La reticulación de anticuerpos con perlas puede realizarse por una variedad de métodos, pero puede implicar la utilización de un reactivo químico que facilita la unión del anticuerpo a la perla seguido de varios pasos de neutralización y lavado para preparar aún más las perlas recubiertas de anticuerpo para inmunoprecipitación. Otro método más para unir anticuerpos a perlas magnéticas contemplados por la presente descripción es el procedimiento de reticulación por UV. Un tercer método para unir anticuerpos a perlas magnéticas contemplados por la presente descripción es el procedimiento de reticulación enzimática.

Se puede emplear con éxito un elemento de soporte de columna en lugar de perlas para fines de fase sólida. Además, elementos de soporte tales como placas de Petri, filtros, tubos de ensayo recubiertos químicamente o tubos eppendorff que pueden tener capacidad para unir perlas recubiertas de anticuerpos u otros materiales de soporte de fase sólida recubiertos de anticuerpos también pueden ser empleados por la presente descripción.

Marcaje de ADN:

El marcaje puede realizarse en el extremo 3' del ADN, el extremo 5' del ADN o en ambos extremos 3' y 5' del ADN. Según un aspecto concreto, se usa una etiqueta idéntica para marcar los extremos 3' y 5' del ADN. Los

5 marcadores incluyen tintes fluorescentes, puntos cuánticos, partículas magnéticas, partículas metálicas y tintes coloreados. Los marcadores genéricos incluyen marcadores que se unen de forma no específica a ácidos nucleicos (p. ej., colorantes intercaladores, colorantes de unión al surco de ácido nucleico y aglutinantes del surco menor) o proteínas. Los ejemplos de tintes intercaladores incluyen YOYO-1, TOTO-3, Syber Green y bromuro de etidio.

10 Según un aspecto concreto, el ADN se marca mediante una reacción de ligadura a un adaptador que contiene una secuencia de código de barras. En un aspecto, el adaptador comprende un adaptador Solexa. En un aspecto, el adaptador comprende un adaptador Illumina. El ADN también puede marcarse usando una enzima (p. ej., Tn5 transposasa, enzima de Tagment ADN) que interviene en la fragmentación del ADN bicatenario y liga oligonucleótidos sintéticos.

15 La ligadura puede ser una ligadura de extremo romo o usar una secuencia monocatenaria sobresaliente (p. ej., la secuencia puede ser primero con cola A). El adaptador puede comprender más secuencias, p. ej., una secuencia reconocible por un cebador de PCR, secuencias que son necesarias para unirse a una superficie de la celda de flujo (secuencias P5 y P7), una secuencia que codifica un activador para una ARN polimerasa (como se describe con más detalle más adelante en la presente memoria) y/o una secuencia de restricción. En un aspecto, el adaptador no comprende secuencias que codifican una secuencia de enzima de restricción. La secuencia de código de barras se puede usar para identificar una molécula, muestra o biblioteca concreta. La secuencia de código de barras puede estar comprendida entre 3 y 400 nucleótidos, más preferiblemente entre 3 y 200 e incluso más preferiblemente entre 3 y 100 nucleótidos. Por lo tanto, la secuencia del código de barras puede ser de 6 nucleótidos, 7 nucleótidos, 8, nucleótidos, nueve nucleótidos o diez nucleótidos. El código de barras es generalmente de 4 a 15 nucleótidos.

20 Las secuencias activadoras de ARN polimerasa son conocidas en la técnica e incluyen, por ejemplo, la secuencia activadora de ARN polimerasa T7, p. ej., SEQ ID n° 10 (CGATTGAGGCCGTAATACGACTCACTATAGGGGC).

25 Para utilizar una población de adaptadores para identificar una población de células, la secuencia de identificación del adaptador difiere según el tipo de célula, mientras que el resto del adaptador es idéntico. Como cada tipo de célula está marcado con un adaptador que contiene una secuencia de identificación diferente, se pueden distinguir los ácidos nucleicos que generan estas células.

30 La eliminación de los adaptadores no ligados puede realizarse utilizando cualquier método conocido en la técnica (p. ej., TrisCl 10 mM). El amortiguador para la eliminación de adaptadores no ligados también puede comprender inhibidores de proteasa.

35 Si los complejos se inmovilizaron antes de la etapa de marcaje, la siguiente etapa comprende la liberación de los complejos inmovilizados. Los presentes inventores contemplan cualquier método para liberar los complejos siempre que el ADN de los complejos continúe estando marcado. Los métodos incluyen el uso de detergentes (p. ej., DTT, desoxicolato de sodio), alto contenido de sal (p. ej., 200-700 mM, p. ej., sal de 500 mM, NaCl) y/o calor (p. ej., aproximadamente 37°C). Los inhibidores de proteasa también pueden incluirse en el amortiguador. Preferiblemente, el método usado libera más del 40% de los complejos, más preferiblemente más del 50% de los complejos, e incluso más preferiblemente más del 60% de los complejos.

Etapa 3: Agrupar los complejos

40 Después del marcaje (y la liberación del agente inmovilizador), se agrupan las muestras. Los complejos pueden purificarse y/o concentrarse en esta etapa. Esto puede efectuarse utilizando cualquier método conocido en la técnica, incluida la ultracentrifugación (p. ej., utilizando un centrifugador con un límite de 50 kDa). Además, o alternativamente, los complejos se pueden lavar antes de la siguiente etapa para asegurar que los complejos sean capaces de unirse al agente. Por ejemplo, la concentración final de sal y detergente debe ser compatible con la integridad del anticuerpo. Así, por ejemplo, la concentración de detergente debería ser inferior a aproximadamente 1 mM y la concentración de sal (p. ej., NaCl) debería ser inferior a aproximadamente 150 mM. Un ejemplo de amortiguador que puede usarse para incubar los complejos es el siguiente: Tris-HCl 10 mM, pH 8,0, NaCl 140 mM, Triton X-100 al 1%, SDS al 0,1%, DOC al 0,1%, EDTA 1mM, 1X inhibidores de proteasa.

50 *Etapa 4: Aislar los complejos utilizando un agente que se une específicamente al resto de unión al ADN*

Los agentes que se unen a restos de unión a ADN son generalmente anticuerpos como se describe en la presente memoria anteriormente.

Según un aspecto, el anticuerpo se une específicamente a un factor de transcripción. Los factores de transcripción ejemplares incluyen, entre otros, los descritos en el documento WO 2002014550.

55 Según otro aspecto, el anticuerpo se une a una histona modificada. Los ejemplos de modificaciones de histonas que pueden estudiarse incluyen la acetilación, la metilación, la ubiquitilación, la fosforilación y la

sumoilación. Así, por ejemplo, el anticuerpo puede unirse específicamente a H3K4me1, H3K4me2, H3K4me3 o H3K27Ac. Dichos anticuerpos están disponibles en el mercado en varios proveedores, por ejemplo Abcam.

Tras un aislamiento suficiente de los complejos proteína/ADN utilizando las tecnologías de inmunoprecipitación descritas anteriormente, los complejos se reticulan a la inversa para liberar los fragmentos de ADN para su posterior análisis. Los conocidos y expertos en la técnica pueden lograr invertir las reticulaciones mediante protocolos convencionales de inmunoprecipitación de cromosomas. La inversión de las reticulaciones se logra mediante una incubación de los complejos proteína/ADN aislados a altas temperaturas, preferiblemente por encima de 50°C durante al menos 6 horas (p. ej., 65°C durante aproximadamente 8 horas). La presente descripción contempla que la inversión de las reticulaciones por métodos químicos como el tratamiento con álcali, así como UV o manipulación enzimática pueden realizarse con éxito y están cubiertos por la descripción ahora descrita para los propósitos de la presente descripción, siempre que el ADN del complejo no se altere de ninguna manera de modo que no puede someterse a análisis de secuencias.

Para aumentar la sensibilidad, el ADN liberado (reticulado a la inversa) puede experimentar una etapa de transcripción *in vitro* (según este aspecto, la secuencia del adaptador en la etapa de marcaje debe comprender un punto de fijación a ARN polimerasa, como se describe más adelante en la presente memoria). El ADN se incuba con una ARN polimerasa (p. ej., T7), trifosfatos de ribonucleótidos, preferiblemente en un sistema amortiguador que incluye DTT e iones de magnesio. La muestra se incuba con una DNAsa para eliminar el ADN de la muestra.

Para una mayor mejora de la sensibilidad, se puede llevar a cabo una etapa adicional para asegurar que ambos extremos de la molécula estén codificados con barras. Por lo tanto, la presente descripción contempla ligar otro adaptador de secuenciación a las moléculas de ARN sintetizado *in vitro* usando una enzima ARN ligasa (p. ej., ARN de T4 ligasa). Un amortiguador ejemplar para realizar esta reacción es el siguiente: DMSO al 9,5%, ATP 1 mM, PEG8000 al 20% y 1 U/μl de T4 ligasa en Tris HCl 50 mM pH 7,5, MgCl₂ 10 mM y DTT 1 mM.

La transcripción inversa se puede llevar a cabo para convertir el ARN sintetizado en ADN. Un ejemplo de enzima transcriptasa inversa es la enzima Affinity Script RT (disponible en el mercado en Agilent). Una mezcla de reacción ejemplar puede contener un DTT, dNTP, la enzima RT enriquecido con amortiguador adecuado, y un cebador complementario al adaptador ligado.

El ADN puede secuenciarse usando cualquier método conocido en la técnica, p. ej., secuenciación de ADN masivamente en paralelo, secuenciación por síntesis, secuenciación por ligadura, pirosecuenciación 454, amplificación en grupo, amplificación en puente y amplificación por PCR, aunque preferiblemente, el método comprende un método de secuenciación de alto rendimiento. Los métodos típicos incluyen la tecnología de secuenciación y la instrumentación analítica ofrecida por Roche 454 Life Sciences™, Branford, Connecticut, a la que a veces se hace referencia en este documento como "tecnología 454" o "secuenciación 454"; la tecnología de secuenciación y la instrumentación analítica ofrecida por Illumina, Inc., San Diego, California, (su tecnología de secuenciación Solexa a veces se denomina en la presente memoria "método Solexa" o "tecnología Solexa"); o la tecnología de secuenciación y la instrumentación analítica ofrecida por ABI, Applied Biosystems, Indianapolis, Inc., que a veces en la presente memoria se denomina plataforma o metodología ABI-Solid™.

Otros métodos conocidos para la secuenciación incluyen, por ejemplo, los descritos en: Sanger, F. *et al.*, *Proc. Natl. Acad. Sci. USA* 75, 5463-5467 (1977); Maxam, A. M. y Gilbert, W. *Proc. Natl. Acad. Sci. USA* 74, 560-564 (1977); Ronaghi, M. *et al.*, *Science* 281, 363, 365 (1998); Lysov, L. *et al.*, *Dokl Akad. Nauk SSSR* 303, 1508-1511 (1988); Bains W. y Smith G. C. J. *Theor. Biol.* 135, 303-307 (1988); Drnanac, R. *et al.*, *Genomics* 4, 114-128 (1989); Khrapko, KR *et al.*, *FEBS Lett.* 256.118-122 (1989); Pevzner P. A. *J. Biomol. Struct. Dyn.* 7, 63-73 (1989); y Southern, E. M. *et al.*, *Genomics* 13, 1008-1017 (1992). También puede usarse la reacción de secuenciación basada en pirofosfato como se describe, por ejemplo, en las patentes de Estados Unidos n° 6.274.320, n° 6.258.568 y n° 6.210.891.

Después de la secuenciación, el ADN puede alinearse con genomas, p. ej., para determinar qué partes del genoma se modificaron epigenéticamente, p. ej., mediante metilación. El análisis de las secuencias puede proporcionar información relacionada con los posibles sitios de unión del factor de transcripción y/o el perfil epigenético, como se describe con más detalle en el apartado Ejemplos a continuación.

Equipos

Cualquiera de las composiciones descritas en la presente memoria puede estar comprendida en un equipo. En un ejemplo no restrictivo, el equipo comprende los siguientes componentes, estando cada componente en un recipiente adecuado:

- (i) un agente para inmovilizar al menos el 50% de la cromatina de una célula (p. ej. un filtro o un anticuerpo que se une específicamente a una histona seleccionada del grupo que consiste en H2, H3 y H4).

(ii) al menos un anticuerpo que se une específicamente a un subgrupo de la cromatina; y

(iii) un gran número de secuencias de ADN de códigos de barras.

5 El equipo puede comprender más componentes incluidos, entre otros, una ARN polimerasa, una DNAsa y/o una transcriptasa inversa. Los componentes adicionales incluyen un reticulante, una enzima proteasa, trifosfatos de nucleótidos y/o una ligasa. El equipo también puede comprender los amortiguadores apropiados para llevar a cabo el procedimiento de inmunoprecipitación descrito en la presente memoria. Ejemplos de amortiguadores se describen en la presente memoria anteriormente y en el apartado Ejemplos más adelante en la presente memoria.

10 En otro ejemplo no restrictivo, el equipo comprende los siguientes componentes, estando cada componente en un recipiente adecuado:

(i) Una enzima de ADN de Tagment; y al menos uno de los siguientes

(ii) al menos un anticuerpo que se une específicamente a la proteína del complejo ADN-proteína; y

(iii) un agente de marcaje de ADN (p. ej. los adaptadores que comprenden las secuencias de código de barras como las descritas en la presente memoria anteriormente).

15 El equipo puede comprender más componentes incluidos, entre otros, una ARN polimerasa, una DNAsa, trifosfatos de nucleótidos y/o una transcriptasa inversa. Otros componentes incluyen un reticulante, una enzima proteasa y/o una ligasa. El equipo también puede comprender los amortiguadores apropiados para llevar a cabo el procedimiento de inmunoprecipitación descrito en la presente memoria. Ejemplos de amortiguadores se describen en la presente memoria anteriormente y en el apartado Ejemplos más adelante en la presente memoria.

20 En otro ejemplo no restrictivo, el equipo comprende los siguientes componentes, estando cada componente en un recipiente adecuado:

(i) al menos un agente seleccionado del grupo que consiste en una ARN polimerasa, una DNAsa y una transcriptasa inversa; y

25 (ii) al menos un anticuerpo que se une específicamente a la proteína del complejo de proteínas del ADN.

Los componentes adicionales incluyen un agente reticulante, trifosfatos de nucleótidos, una enzima proteasa y/o una ligasa. El equipo también puede comprender los amortiguadores apropiados llevar a cabo el procedimiento de inmunoprecipitación descrito anteriormente en la presente memoria. Ejemplos de amortiguadores se han descrito anteriormente en la presente memoria y en el apartado Ejemplos más adelante.

Según un aspecto concreto, los equipos de este aspecto de la presente descripción no comprenden MNAsa.

Los recipientes de los equipos generalmente incluirán al menos un vial, tubo de ensayo, matraz, botella, jeringuilla u otros recipientes, en los que se puede colocar un componente, y preferiblemente, adecuadamente en alícuotas. Cuando hay más de un componente en el equipo, el equipo generalmente contendrá un segundo, tercer u otro recipiente más en el que se pueden colocar por separado otros componentes. Sin embargo, varias combinaciones de componentes pueden estar comprendidas en un recipiente.

40 Cuando los componentes del equipo se proporcionan en una o más soluciones líquidas, la solución líquida puede ser una solución acuosa. Sin embargo, los componentes del equipo pueden proporcionarse en polvo(s) seco(s). Cuando reactivos y/o los componentes se proporcionan en polvo seco, el polvo puede redisolverse añadiendo un disolvente adecuado.

Un equipo incluirá preferiblemente instrucciones para el empleo, los componentes del equipo, así como el uso de cualquier otro reactivo no incluido en el equipo. Las instrucciones pueden incluir las variaciones que se pueden realizar.

45 Como se usa en la presente memoria, el término "aproximadamente" se refiere a $\pm 10\%$.

Los términos y expresiones "comprende", "que comprende", "incluye", "Incluido", "que tiene" y sus conjugados significan "incluido pero no limitado a".

La expresión "que consiste en" significa "incluido y limitado a".

50 La expresión "que consiste esencialmente en" significa que la composición, método o estructura puede incluir más ingredientes, etapas y/o partes, pero solo si los ingredientes, etapas y/o partes adicionales no alteran

materialmente las características básicas y novedosas de la composición, método o estructura reivindicada.

En toda esta aplicación, varios aspectos de esta descripción pueden presentarse en formato de intervalo. Esto debe entenderse que la descripción en formato de intervalo es meramente por conveniencia y brevedad y no debe interpretarse como una limitación inflexible en el alcance de la descripción. Por consiguiente, la

5 descripción de un intervalo debe considerarse que ha descrito específicamente todo lo posible todos los subintervalos posibles, así como cada uno de los valores numéricos dentro de ese intervalo. Por ejemplo, la descripción de un intervalo, como de 1 a 6 debe considerarse que ha descrito específicamente subintervalos tal como de 1 a 3, de 1 a 4, de 1 a 5, de 2 a 4, de 2 a 6, de 3 a 6, etc., así como cada uno de los números dentro de ese intervalo, por ejemplo, 1, 2, 3, 4, 5 y 6. Esto se aplica independientemente de la amplitud del

10 intervalo.

Como se emplea en la presente memoria, el término "método" se refiere a maneras, medios, técnicas y procedimientos para llevar a cabo una tarea dada que incluidos, entre otros, esas maneras, medios, técnicas y procedimientos conocidos o desarrollados fácilmente a partir de maneras, medios, técnicas y procedimientos conocidos por profesionales de las técnicas química, farmacológica, biológica, bioquímica y

15 médica.

Se aprecia que determinadas características de la descripción, que, para mayor claridad, se describen en el contexto de aspectos separados, también se pueden proporcionar en combinación en un solo aspecto. Por el contrario, varias características de la descripción, que, por brevedad, se describen en el contexto de un solo aspecto, también se pueden proporcionar por separado o en cualquier subcombinación adecuada o como

20 adecuada en cualquier otro aspecto descrito de la descripción. Algunas características descritas en el contexto de varios aspectos no deben considerarse características esenciales de esas realizaciones, a menos que el aspecto no funcione sin esos elementos.

Varios aspectos y aspectos de la presente descripción como se ha descrito anteriormente en la presente memoria y como se reivindica en el apartado reivindicaciones más adelante hallan apoyo experimental en los

25 ejemplos siguientes.

Ejemplos

A continuación, se hace referencia a los siguientes ejemplos, que junto con las descripciones anteriores ilustran algunos aspectos de la descripción de manera no restrictiva.

30 Generalmente, la nomenclatura utilizada en la presente memoria y los procedimientos de laboratorio utilizados en la presente incluyen técnicas moleculares, bioquímicas, microbiológicas y de ADN recombinado. Dichas técnicas se explican completamente en la bibliografía. Véanse, por ejemplo, "Molecular Cloning: A laboratory Manual" Sambrook *et al.*, (1989); "Current Protocols in Molecular Biology" volúmenes I-III Ausubel, R. M., ed. (1994); Ausubel *et al.*, "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Maryland (1989); Perbal, "A Practical Guide to Molecular Cloning, John Wiley & Sons, Nueva York (1988);

35 Watson *et al.*, "Recombinant DNA, Scientific American Books, Nueva York; Birren *et al.* (eds.) "Genome Analysis: A Laboratory Manual Series", Vols. 1-4, Cold Spring Harbor Laboratory Press, Nueva York (1998); metodologías expuestas en las Pat. De EE.UU. n° 4.666.828; n° 4.683.202; n° 4.801.531; n° 5.192.659 y n° 5.272.057; "Cell Biology: A Laboratory Handbook", volúmenes I-III Cellis, J. E., ed. (1994); "Culture of Animal Cells - A Manual of Basic Technique" por Freshney, Wiley Liss, N. Y. (1994), tercera edición: "Current

40 Protocols in Immunology" volúmenes I-III Coligan J. E., ed. (1994); Stites *et al.* (eds), "Basic and Clinical Immunology" (8ª Edition), Appleton & Lange, Norwalk, CT. (1994); Mishell and Shigi (eds), "Selected Methods in Cellular Immunology". W. H. Freeman and Co., Nueva York (1980); inmunoanálisis disponibles se describen extensamente en la patente y en la bibliografía científica, véanse, por ejemplo, las Pat. de EE.UU. n° 3.791.932; n° 3.839.153; n° 3.850.752; n° 3.850.578; n° 3.853.987; n° 3.867.517; n° 3.879.262; n°

45 3.901.654; n° 3.935.074; n° 3.984.533; n° 3.996.345; n° 4.034.074; n° 4.098.876; n° 4.879.219; n° 5.011.771 y n° 5.281.521; "Oligonucleotide Synthesis" Gait, M. J., ed. (1984); "Nucleic Acid Hybridization" Hames, B. D., and Higgins S. J., eds. (1985); "Transcription and Translation" Hames, B. D., and Higgins S. J., eds. (1984); "Animal Cell Culture" Freshney, R. I., ed. (1986); "Immobilized Cells and Enzymes" IRL Press, (1986); "A Practical Guide to Molecular Cloning Perbal, B., (1984) y "Methods in Enzymology" vol. 1-317,

50 Academic Press: "PCR Protocols: A Guide To Methods And Applications", Academic Press, San Diego, CA. (1990); Marshak *et al.*, "Strategies for Protein Purification and Characterization - A Laboratory Course Manual" CSHL Press (1996). A lo largo de este documento se proporcionan otras referencias generales. Se cree que los procedimientos en la presente memoria son bien conocidos en la técnica y se proporcionan para la conveniencia del lector. Toda la información contenida en los mismos se incorpora en la presente memoria

55 por referencia.

Materiales y métodos

Recogida de células: Se recogen células reticuladas clasificadas por FACS en tubos de FACS de 5ml que contienen 500 µl de suero de ternera fetal, diluido en 4 ml de amortiguador de recogida iChIP (12 mM TrisCl,

0,1x PBS, EDTA 6 mM, 1,2x inhibidores de proteasa (Roche) y granulados por centrifugación durante 15 min a 1.000xG usando un rotor giratorio con baja aceleración y opciones de freno. Se retira el sobrenadante dejando una para evitar perturbar el sedimento celular. Las células se vuelven a poner en suspensión en la cobertura de amortiguador de 200 µl, se transfieren a tubos de 0,2 ml en alícuotas de 10.000 a 20.000 5 células, a continuación se sedimentan las células a 1000xG durante 15 min utilizando la configuración anterior. El sobrenadante se elimina dejando 10 µl de amortiguador que cubre el sedimento celular. Las células se congelan almacenadas a -80°C.

Exposición a ultrasonidos: Alícuotas celulares se descongelan en hielo y 2 µl de SDS al 3% se agrega para lograr una concentración final de SDS al 0,5%. Las células se vuelven a poner en suspensión completamente 10 y se lisan en hielo durante 10 min. Los lisados se transfirieron a microtubos Bioruptor de 0,1 ml (Diagenode C30010015, Lieja, Bélgica) y la cromatina se cortó con un NGS Bioruptor Sonicator (Diagenode) a alta intensidad y ciclos de 30" ON/30" OFF. El tiempo de exposición a ultrasonidos se calibró para los diferentes tipos de células: 20 min para las MEP, EryA y EryB; 25 min para las CMP; 30 min para LT-HSC, ST-HSC, MPP y GMP; 40 min para granulocitos, monocitos y macrófagos y 45 min para linfocitos NK, T y B.

15 *Inmovilización de cromatina:* Después de la exposición a ultrasonidos, los extractos de cromatina se transfirieron a una placa de 96 pocillos y se diluyeron 1 a 5 con amortiguador de equilibrio para exposición a ultrasonidos (TrisCl 10 mM, NaCl 140 mM, desoxicolato de sodio al 0,1%, TX-100 al 1%, EDTA 1 mM, 1X inhibidores de la proteasa (Roche) para lograr una concentración de SDS de 0,1%. Para inmovilizar la cromatina en perlas magnéticas, se añadieron 15µl de proteína Dynabeads (Life Tech) y 1,3 ug de anticuerpo anti-H3 (ab1791) a los extractos de cromatina diluidos y se incubaron durante 20 horas a 4°C. 20

Indización de cromatina: Se usó la captura de perlas con imanes para agregar, lavar y eliminar eficientemente las diferentes mezclas madre utilizadas en el proceso de indización. Todas las reacciones se hicieron mientras la cromatina se unía a las perlas magnéticas recubiertas con H3. Después de la inmovilización, los inmunocomplejos de cromatina unidos por perlas IP se magnetizaron y se lavaron 3 veces con 150 µl de 25 inhibidores de la proteasa Tris Cl 10 mM + 1X inhibidores de proteasa sin EDTA (Roche). Después de los lavados la cromatina unida a las perlas se volvió a poner en suspensión en 20 µl del mismo amortiguador. La reparación final de cromatina se realizó agregando 30 µl de una mezcla madre: 25 µl 2x mezcla ER (Tris-Cl 50 mM pH 7,5, MgCl₂ 20 mM, DTT 20 mM, ATP 2 mM, dNTP 1 mM), 2 µl de enzima T4 PNK (10 U/µl NEB) y 2 µl de T4 polimerasa (3 U/µl NEB) a cada pocillo. Las muestras se incubaron en un termociclador a 12°C 30 durante 25 min, 25°C durante 25 min y finalmente se enfrió a 4°C. Después de la reparación final, la cromatina unida a las perlas se lavó una vez con 150 µl de TrisCl 10 mM + inhibidores de proteasa y se volvió a poner en suspensión en 40 µl del mismo amortiguador. La cromatina se acabó en A agregando 20 µl de mezcla madre (añadir 17 µl de base A a la mezcla, 3 µl de Klenow (3'→5' exonucleasa, 3 U/µl, NEB) a cada pocillo y se incubó a 37°C durante 30 min en un termociclador. Después del final de la reparación la 35 cromatina unida a las perlas se lavó una vez con 150 µl de TrisCl 10 mM + inhibidores de proteasa y se volvió a poner en suspensión en 19 µl del mismo amortiguador. La cromatina se indizó agregando 5 µl de adaptadores indizados en forma de Y 0,75 µM (que contienen secuencias P5 y P7) a cada pocillo que se ligaron a los extremos del ADN de la cromatina agregando 34 µl de mezcla maestra AL (29 µl 2x de amortiguador Quick Ligation y 5 µl Quick DNA ligasa (NEB) a cada pocillo. Las muestras se mezclaron 40 completamente y se incubaron a 25°C durante 40 min en un termociclador. La cromatina indizada unida a las perlas se lavó una vez con 150 µl de TrisCl 10 mM + inhibidores de la proteasa para eliminar los adaptadores no ligados.

Liberación de cromatina: Se usaron condiciones desnaturalizantes (DTT, alto contenido de sal y detergente) y calor para liberar la cromatina indizada de las perlas magnéticas recubiertas de anticuerpo. Inmediatamente 45 después del lavado posterior a la indización, se tomaron muestras del imán, se volvieron a poner en suspensión las perlas en 12,5 µl de 100mM de DTT e se incubó durante 5 min a temperatura ambiente. Después, 12,5 µl de amortiguador de liberación de cromatina (NaCl 500 mM, SDS al 2%, desoxicolato de sodio al 2%, 2x inhibidores de proteasa) se añadieron a cada pocillo, las muestras se mezclaron completamente y se incubaron a 37°C durante 30 minutos. Después de la incubación de liberación, las perlas 50 magnéticas se volvieron a suspender completamente y se agruparon en grupos de n muestras, dando como resultado un volumen por grupo de 200-250 µl. Las muestras de cromatina indizadas agrupadas se concentraron usando un umbral Centricon (Amicon) de 50 KDa.

Inmunoprecipitación de cromatina: Se añadió anticuerpo de referencia y se incubó a 4°C. durante 3 h, después se añadieron 50 µl con perlas magnéticas de proteína G y se incubó IP durante 1 hora más. Para 55 cada CHIP, se usaron 1,5 µg de anti H3K4me1 (ab8895) y 2,5 µg de anti H3K4me2 (ab32356), anti H3K4me3 (Millipore, 07-473) y anti H3K27ac (ab4729).

TF iChIP: Se añadieron 50 µl de proteína G magnética Dynabeads (Invitrogen) acopladas al anticuerpo de referencia y se incubaron 4 h a 4°C. Para el acoplamiento, las perlas se lavaron una vez (200 µl) en un 60 tampón de unión/bloqueo (PBS, Tween 20 al 0,5%, BSA al 0,5%), se incubaron con 10 µg de anticuerpo en tampón de unión/bloqueo durante 1 hora a temperatura ambiente, y a continuación se lavaron para eliminar el exceso de anticuerpo. Por cada CHIP, se usaron 10 µg de anti PU.1 (Santa Cruz, sc352-x)

Lavados y elución de ADN con ChIP: Se utilizó una placa de 96 pocillos e imán (Invitrogen) en todas las etapas adicionales. Se eliminó el amortiguador CHIP y las muestras se lavaron 5 veces con RIPA frío (200 µl por lavado), dos veces con amortiguador RIPA enriquecido con NaCl 500 mM (200 µl por lavado), dos veces con amortiguador LiCl (TE 10 mM, LiCl 250 mM, NP-40 al 0,5%, 0,5% DOC), una vez con TE (Tris-HCl 10 mM pH 8,0, EDTA 1 mM), y después se eluyó en 50 µl de SDS al 0,5%, NaCl 300 mM, EDTA 5 mM, Tris HCl 10mM a pH 8,0. El eluido se trató sucesivamente con 2 µl de RNasaA (Roche, 11119915001) durante 30 min a t 37°C, 2,5 µl de Proteinasa K (NEB, P8102S) durante dos horas a 37°C y 8 horas a 65°C para invertir la reticulación de formaldehído.

10 *Aislamiento de ADN con ChIP:* Las etapas de limpieza de SPRI se realizaron usando placas de 96 pocillos e imanes. Se añadieron 90 µl de SPRI a las muestras reticuladas inversamente, se mezclaron con pipeta 15 veces y se incubaron durante 6 minutos. El sobrenadante se separó de las perlas usando una placa e imán de 96 pocillos durante 5 minutos. Las perlas se lavaron en el imán con etanol al 70% y después se secaron al aire durante 5 minutos. El ADN se eluyó en 23 µl de amortiguador EB (Tris-HCl 10 mM, pH 8,0) con pipeta mezclando 25 veces.

15 *Amplificación de la biblioteca, control de calidad y secuenciación:* La biblioteca se completó y amplificó mediante una reacción de PCR con 0,5 µM de cebadores directo de PCR e inverso de PCR y una mezcla lista para PCR (Kapa Biosystems). El cebador directo contiene las secuencias Illumina P5-Read1 y el cebador inverso contiene las secuencias P7-Read2. La biblioteca unicelular agrupada amplificada se purificó con 1x volúmenes de perlas SPRI. La concentración de la biblioteca se midió con un fluorómetro Qubit (Life Technologies) y el tamaño medio de la molécula se determinó con un instrumento 2200 TapeStation (Agilent Technologies). La calidad de la biblioteca se determinó además por qPCR, midiendo el enriquecimiento de activadores activos ubicuos (Actina-B y GAPDH) frente al fondo (Cryaa). Las bibliotecas se seleccionaron por tamaño con 0,6x volúmenes de perlas SPRI eliminando grandes fragmentos de ADN (> 600 pb). Las bibliotecas iChIP se secuenciaron usando un Illumina HiSeq 1500.

25 *Aislamiento de ARN:* Un mínimo de 5.000 células se clasificaron en 200 µl de amortiguador de lisis/unión (Life Technologies), se lisaron durante 5 minutos y se congelaron a -80°C. Los lisados celulares se descongelaron y el ARN mensajero se capturó con 12 µl de oligo Dynabeads (dT) (Life Technologies) y se lavaron según las instrucciones de fabricación. El ARN mensajero purificado se eluyó a 70°C con 10 µl de Tris-Cl 10 mM, pH 7,5 y se almacenó a -80°C.

30 *Construcción de la biblioteca RNAseq:* Para todas las poblaciones hematopoyéticas se utilizó el protocolo MARS-seq (Science 343, 776-779 (2014)) desarrollado para RNA-seq monocelular para producir bibliotecas de RNA-seq. En resumen, el protocolo consiste en cebadores especialmente diseñados con identificadores moleculares únicos para el recuento exacto de moléculas y una etapa de amplificación lineal del grupo de ARNm inicial, seguido de una etapa de construcción de la biblioteca. De esta manera, la diversidad del grupo original de ARN mensajeros se conserva incluso si la cantidad de ARN de entrada es baja. Se preparó un mínimo de 2 bibliotecas replicadas para cada una de las diferentes poblaciones hematopoyéticas. La Tabla 1 resume los cebadores utilizados en los experimentos descritos a continuación.

Tabla 1

NOMBRE DEL CEBADOR	SECUENCIA Y MODIFICACIONES
Adaptador universal	5'-ACACTCTTTCCCTACACGAGCGCTCTCCGATC*T-3', donde * indica modificación de fosfonato (SEQ. ID. n°: 1)
Adaptadores indizados	5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCCAGTCACXXXXXXATCTCGTA TGCCGCTCTCTGT-3', (SEQ ID n°: 2) donde XXXXXX es el código de barras para multiplexar la muestra
Adaptadores indizados en forma de Y	- Hibridar el Adaptador ssUniversal y Adaptadores ssIndizados para obtener Adaptadores Indizados en forma de Y
Directo de PCR	5'-AATGATACGGGACCACCGAGATCTACACTCTTTCCCTACACGAC-3' (SEQ ID n°: 3)
Inverso de PCR	5'-CAAGCAGAAGACGGCATACGAGAT-3' (SEQ ID n°: 4)
Cebador RT codificado por barras	5'-CGATTGAGGCCGGTAATACGACTCACTATAGGGGCGAGGTGTGCTCTCCGATCTXXXXXXNNNNNTTTTTTTTTTTT TTTTTTN-3', donde XXXXXX es el código de barras y NNNN es el RMT (SEQ. ID. n°: 5)
Adaptador de ligadura	5'-AGATCGGAAGAGCGTGTAG-3', modificada con un grupo fosfato en 5' y un espaciador C3 (bloqueador) en 3' (SEQ. ID. n°: 6)
Segundo cebador RT	5'-TCTAGCCTTCTCGACACATC-3' (SEQ ID n°: 7)
P5_Rd1 directo de PCR	5'-AATGATACGGGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTCCGATCT-3' (SEQ ID n°: 8)
P7_Rd2 inverso de PCR	5'-CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACCGTGTGC TCTCCGATCT-3' (SEQ ID n°: 9)

Se describe a continuación el protocolo RNAseq utilizado para las células hematopoyéticas:

1. Amplificación lineal del grupo de ARNm.

Se colocaron 4 µl de ARNm purificado en placas de 384 pocillos. En primer lugar, para abrir estructuras secundarias de ARN y permitir la hibridación del cebador RT, la placa de 384 pocillos se incubó a 72°C durante 3 minutos y se transfirió inmediatamente a un bloque térmico Inheco de 384 pocillos integrado en Bravo y se ajustó a 4°C. Después, 2 µl de una mezcla de reacción de RT (DTT 10 mM, dNTP 4 mM, 2,5 U/µl de enzima superíndice III RT en Tris-HCl 50 mM (pH 8,3), KCl 75 mM MgCl₂, 3 mM) se añadieron en cada pocillo de la placa de 384 pocillos y se mezcló una vez la reacción. Se reemplazaron las boquillas y el proceso se repitió en todos los pozos. La placa de 384 pocillos se centrifugó y se trasladó a un ciclador 384 (Eppendorf) para la incubación siguiente: 2 min a 42°C, 50 min a 50°C, 5 min a 85°C. Se agruparon las muestras indizadas con una cantidad equivalente de ADNc. El ADNc agrupado se convirtió en ADN bicatenario con un segundo equipo de síntesis de cadena (NEB) en una reacción con 20 µl, incubando durante 2,5 h a 16°C. El producto se purificó con 1,4 volúmenes de perlas SPRI, se eluyó en 8 µl y se transcribió *in vitro* (con las perlas) a 37°C durante la noche para amplificación lineal usando el equipo T7 High Yield RNA polymerase IVT (NEB). Después de IVT, la plantilla de ADN se eliminó con Turbo DNase I (Ambion) 15 minutos a 37°C y el ARN amplificado (ARNa) se purificó con 1,2x volúmenes de perlas SPRI.

2. Preparación de la biblioteca para la secuenciación de alto rendimiento.

El ARNa se fragmentó químicamente en moléculas cortas (tamaño medio ~200 nucleótidos) incubando 3 minutos a 70°C en solución de fragmentación de ARN Zn²⁺ (Ambion) y se purificó con dos volúmenes de perlas SPRI. El ARNa (5 µl) se preincubó 3 min a 70°C con 1 µl de adaptador de ligadura 100 µM; a continuación, 14 µl de una mezcla que contenía 9,5% de DMSO, ATP 1 mM, PEG8000 al 20% y se añadió 1 U/µl de T4 ligasa en Tris HCl 50 mM, pH 7,5, MgCl₂ 10 mM y 1 mM DTT. La reacción se incubó a 22°C durante 2 h. El producto ligado se transcribió de forma inversa usando la enzima Affinity Script RT (Agilent; la mezcla de reacción contiene amortiguador Affinity Script RT, DTT 10 mM, dNTP 4 mM, 2,5 U/µl de enzima RT) y un cebador complementario al adaptador ligado. La reacción se incubó durante 2 minutos a 42°C, 45 minutos a 50°C y 5 minutos a 85°C. El ADNc se purificó con 1,5x volúmenes de perlas SPRI. La biblioteca se completó y amplificó mediante una reacción de PCR anidada con 0,5 µM de cebadores P5_Rd1 y P7_Rd2 y una mezcla lista para PCR (Kapa Biosystems). El cebador directo contiene las secuencias Illumina P5-Read1 y el cebador inverso contiene las secuencias P7-Read2. La biblioteca agrupada amplificada se purificó con 0,7x volúmenes de perlas SPRI para eliminar las sobras del cebador. La concentración de la biblioteca se midió con un fluorómetro Qubit (Life Technologies) y el tamaño medio de la molécula se determinó con un instrumento 2200 TapeStation (Agilent). Las bibliotecas MARS-Seq se secuenciaron usando un Illumina HiSeq 1500.

ATAC-seq: Para caracterizar la cromatina abierta, se utilizó el análisis para el protocolo de cromatina accesible a transposasa (ATAC-seq) desarrollado por Buenrosto *et al.* (*Nat. Methods* 10, 1213-1218 (2013)) con los siguientes cambios: se clasificaron diferentes poblaciones de células hematopoyéticas en 400 µl de amortiguador MACS (1x PBS, BSA al 0,5%, EDTA 2 mM) y se sedimentaron por centrifugación durante 15 minutos a 500 g y 4°C utilizando un rotor giratorio con baja aceleración y ajustes de freno. Los sedimentos celulares se lavaron una vez con 1x PBS y las células se sedimentaron por centrifugación usando los ajustes anteriores. Los sedimentos celulares se volvieron a poner en suspensión en 25 µl de amortiguador de lisis (Tris-HCl 10 mM pH 7,4, NaCl 10 mM, MgCl₂ 3 mM), Igepal CA-630 al 0,1%) y los núcleos se sedimentaron por centrifugación durante 30 minutos a 500 g, 4°C usando un rotor oscilante con baja aceleración y ajustes de freno. El sobrenadante se descartó y los núcleos se volvieron a poner en suspensión en 25 µl de amortiguador de reacción que contenía 2 µl de Tn5 transposasa y 12,5 µl de amortiguador TD (equipo de preparación de muestras Nextera de Illumina). La reacción se incubó a 37°C durante una hora. Después se añadieron 5 µl de amortiguador de limpieza (NaCl 900 mM, EDTA 300 mM), 2 µl de SDS al 5% y 2 µl de Proteínasa K (NEB) y se incubaron durante 30 minutos a 40°C. El ADN tagmentado se aisló usando limpieza de perlas 2x SPRI. Para la amplificación de la biblioteca, se realizaron dos PCR de 9 ciclos sucesivas para enriquecer pequeños fragmentos de ADN tagmentado. Se usaron 2 µl de cebadores de indización incluidos en el equipo Nextera Index y la mezcla preparada KAPA HiFi HotStart. Después de la primera PCR, las bibliotecas se seleccionaron para fragmentos pequeños (menos de 600 pb) utilizando la limpieza SPRI. Después se realizó una segunda PCR en las mismas condiciones para obtener la biblioteca final. La concentración de ADN se midió con un fluorómetro Qubit (Life Technologies) y los tamaños de la biblioteca se determinaron usando TapeStation (Agilent Technologies). Las bibliotecas se secuenciaron en un HiSeq 1500 para un promedio de 20 millones de lecturas por muestra.

Diferenciación ex vivo de las BMDC: Para obtener las BMDC, las células de médula ósea se colocaron en placas a una densidad de 200.000 células/ml en platos de plástico tratados con cultivo sin tejido (10 ml de medio por placa). El 2º día, las células se alimentaron con otros 10 ml de medio por plato. El 5º día, las células se recogieron de 15 ml del sobrenadante centrifugando a 1.400 rpm durante 5 minutos; los sedimentos se volvieron a poner en suspensión con 5 ml de medio y se añadieron por primera vez al plato original. Las células se alimentaron con otros 5 ml de medio el 7º día. El medio BMDC contiene: RPMI

(Gibco) enriquecido con FBS al 10% inactivado por calor (Gibco), β -mercaptoetanol (50 μ M, Gibco), L-glutamina (2 mM, Biological Industries) penicilina/estreptomicina (100 U/ml, Biological Industries), aminoácidos no esenciales MEM (IX, Biological Industries), HEPES (10 mM, Biological Industries), piruvato de sodio (1 mM, Biological Industries) y GM-CSF (20 ng/ml; Peprotech).

- 5 *Recolección de las BMDC para la valoración de iChIP:* Las células se fijaron durante 8 minutos con formaldehído al 1%, se enfriaron con glicina y se lavaron dos veces con PBS enfriado con hielo. Las células se volvieron a poner en suspensión en iChIP Harvesting Buffer, se hizo el recuento y se transfirió a tubos de 0,2 ml en alícuotas de 10 μ l que contenían 10.000, 3.000, 1.000 y 500 células.

- Aislamiento de células madre hematopoyéticas:* Se extirparon fémures, pelvis y tibias de 6 ratones hembra C57BL/6J (de 8 a 12 semanas de edad) y se enjuagaron células de médula ósea con amortiguador MACS. Las células se enriquecieron con microperlas del equipo CD117/c según las instrucciones del fabricante (AutoMACS, Miltenyi Biotec; número de pedido: 130-091-224). Las células enriquecidas con equipo C se tiñeron con Lin (Ter1199, Gr1, CD11b, B220, CD3, CD4, CD8), equipo c, Sca1, CD34, FcgR-II, F1k2 y anticuerpos IL7R (eBioscience) durante 30 minutos y se clasificaron con el clasificador celular FACSAria III (BD Biosciences). La Tabla 2 a continuación resume los anticuerpos utilizados para este experimento.

ES 2 764 504 T3

Tabla 2

Marcador	Fluoróforo	Clon	
B220	e450	RA3-6B2	eBioscience
B220	PE	RA-6B2	eBioscience
equipo c	APC	2B8	eBioscience
CD11b	e450	M1/70	eBioscience
CD11b	APC	M1/70	eBioscience
CD19	PE-Cy7	Bio1D3	eBioscience
CD3	E450	17A2	eBioscience
CD34	FITC	RAM34	eBioscience
CD19	PE-Cy7	Bio1D3	eBioscience
CD3	e450	17A2	eBioscience
CD34	FITC	RAM34	eBioscience
CD4	e450	GK1.5	eBioscience
CD4	Alexa700	GK1.5	eBioscience
CD71	APC	R17217	eBioscience
CD8	FITC	53-6.7	eBioscience
CD8a	e450	53-6.7	eBioscience
F4/80	FITC	BM8	eBioscience
FcgR-II	PE-Cy7	93	eBioscience
Flk2	PE	A2F10	eBioscience
Gr1	e450	RB6-8C5	eBioscience
Gr1	PcrCP-Cy5.5	RB6-8C5	Bioleyenda
I-Ab	PacBlue	AF6-120.1	eBioscience
IL7-R	FITC	A7R34	eBioscience
NK1.1	e450	PK136	eBioscience
NK1.1l	APC	PK136	eBioscience
Sca-1	PerCP-Cy5.5	D7	eBioscience
TCR-β	FITC	H57-597	eBioscience
Ter119	e450	TER-119	eBioscience

Las poblaciones celulares se identificaron como:

- LT-HSC: Lin-, c-Kit+, Sca-1+, Flk2-, CD34-
- 5 ST-HSC: Lin-, c-Kit+, Sca-1+, Flk2-, CD34+
- MPP: Lin-, c-Kit+, Sca-1+, Flk2+, CD34+
- CMP: Lin-, c-Kit+, Sca-1+, FcgRII bajo, CD34+

GMP: Lin-, c-Kit+, Sca-1+, FcγRII alto, CD34+

MEP: Lin-, c-Kit+, Sca-1+, FcγRII-, CD34-

CLPs: Lin-, Flk2+, I17R+

5 *Aislamiento de células maduras procedentes de médula ósea:* Se lavaron células de médula ósea de fémures y tibias de ratones hembra C57BL/6J (de 8 a 12 semanas de edad), se pusieron en suspensión con amortiguador MACS y se incubaron 3 minutos en solución de lisis de glóbulos rojos (Sigma). Las células se tiñeron con anticuerpos CD3, B220, NK1.1, Gr1, CD11b, CD115, F4/80 (eBioscience), y después de la filtración a través de un filtro de 70 μm se clasificaron con un clasificador de células FACSAria III (BD Biosciences). La Tabla 2 anterior resume los anticuerpos utilizados para este experimento. Las poblaciones celulares se identificaron como:

Granulocitos: CD3-, B220-, NK1.1-, CD11b+, Gr1+, SSC alto

Macrófagos: CD3-, B220-, F4/80+, CD115-, SSC bajo

Monocitos: CD3-, B220-, NK1.1-, F4/80-, CD115+ SSC bajo

15 *Aislamiento de células linfoides esplénicas:* Se extrajeron bazos de ratones hembra C57BL/6J (de 8 a 12 semanas de edad), se disociaron con un disociador MACS suave (Miltenyi Biotec, Alemania) en suspensión única y, después de lavar con amortiguador MACS, se incubaron durante 5 minutos en solución de lisis de glóbulos rojos (Sigma). Después se lavaron las células, se volvieron a poner en suspensión en amortiguador MACS y se tiñeron con: CD3, CD4, CD8, B220, CD19, TCR-β, I-Ab, Ter119 y NK1.1 (eBioscience). Después del lavado, las células se filtraron a través de un filtro de 70 μm y se clasificaron con un clasificador de células
20 FACSAria III (BD Biosciences). La Tabla 2 anterior resume los anticuerpos utilizados para este experimento. Las poblaciones celulares se identificaron como:

Linfocitos B: CD3-, B220+, CD19+.

Linfocitos T CD4+: CD3+, B220-, CD4+, CD8-.

Linfocitos T CD8+: CD3+, B220-, CD4-, CD8+.

25 Linfocitos NK: CD4-, CD8-, B220-, I-Ab-, Ter119-, TCR-β-, NK1.1+.

Aislamiento de células eritrocíticas esplénicas: Se extrajeron bazos de ratones hembra C57BL/6J (de 8 a 12 semanas de edad), se disociaron en esplenocitos aislados en amortiguador MACS, con un disociador MACS suave (Miltenyi Biotec, Alemania) y se tiñeron con: B220, Ter119 y CD71 (eBioscience). Después del lavado, las células se filtraron a través de un filtro de 70 μm y se clasificaron con un clasificador de células FACSAria
30 III (BD Biosciences). La tabla 2, anterior resume los anticuerpos utilizados para este experimento. Las poblaciones celulares se identificaron como:

EryA: Ter119+, CD71+, FSC alto.

EryB: Ter119+, CD71+, FSC bajo.

35 *Secuenciación de ChIP y descubrimiento de picos:* La secuenciación de ChIP se realizó en Illumina HiSeq-1500 y las bibliotecas agrupadas se secuenciaron a una profundidad de secuenciación de ~10-15 millones de lecturas alineadas por muestra. Se prepararon bibliotecas por triplicado o duplicado. Las lecturas se cartografiaron en el conjunto mm9 de ratón utilizando el programa 'bowtie2' con los parámetros predeterminados, solo se usaron etiquetas que cartografiaban de forma exclusiva al genoma para un análisis posterior. Debido al bajo número inicial de células por muestra (~10.000) y para evitar artefactos clonales
40 introducidos en la etapa de amplificación por PCR, el número de lecturas idénticas en cada posición genómica se restringió a un máximo de 3.

La identificación de regiones enriquecidas con ChIP-seq (picos) en cada muestra se realizó utilizando HOMER. Para las modificaciones de histonas, se identificaron picos buscando lugares de alta densidad de lectura usando un intervalo deslizante de 1000 pb. Los presentes inventores necesitaban que los picos
45 adyacentes estén al menos a 1.000 pb de distancia para evitar la detección redundante. El umbral para el número de etiquetas que determinaron un pico válido se seleccionó a una tasa de descubrimiento falso de 0,001. Se utilizó el siguiente comando HOMER: cmd = findPeaks <simple tag directory> -L () -C 3 -size 1000 -minDist 1000 -tbp 3 -o <output file>

50 *Construcción del catálogo de potenciadores:* Los picos de todas las muestras (16 tipos de células con 2-3 reproducciones) se combinaron en un catálogo unificado para cada modificación por separado. Se descartaron los picos que no se superponían entre al menos dos reproducciones (distancia entre centros > 500 pb); el aspecto en un tipo de celda fue suficiente para ser incluida en el catálogo final. La redundancia de

los picos superpuestos (distancia entre centros < 500 pb) se eliminó seleccionando un pico representativo como el que tenía la señal más fuerte, extendido por 1000 pb en cada dirección desde su centro. Los gráficos de dispersión y las correlaciones entre las reproducciones se calcularon en el vector de recuentos de lecturas transformadas logarítmicamente para la recolección de picos (figura 1D).

- 5 En total, el catálogo de H3K4me1 y H3K4me2 tenía 110.844 y 69.970 picos, respectivamente, con 66.338 picos comunes. Para generar el catálogo final de potenciadores, en primer lugar se contaron el número de lecturas de H3K4me3 en cada supuesto pico. Se observó una distribución bimodal de los contenidos de H3K4me3 y se ajustó a un modelo de mezcla gaussiana 2 para seleccionar un umbral. Se eliminaron las regiones con altos contenidos de H3K4me3, lo que dio como resultado 48.415 regiones en el conjunto final de potenciadores.

A partir de este momento en adelante, este catálogo de potenciadores se utilizó para el resto del análisis en este trabajo. Para cada potenciador en este catálogo, el número de lecturas se contó dentro de 2 kb alrededor de su centro para cada modificación. Se utilizó el texto `annotatePeaks.pl` del paquete HOMER, normalizando cada biblioteca a 10,000,000 de lecturas.

- 15 *Análisis de agrupación y correlación:* El catálogo de potenciadores se agrupó en base a los recuentos transformados logarítmicamente H3K4me1 utilizando medios de K (matlab R2012b `k` significa realización) con $K = 9$. Los presentes inventores han probado otros k (5-20) y observaron que aumentar el número de grupos conduce a ajustes cuantitativos de los patrones, pero no introduce nuevos patrones de cambio potenciador. Las matrices de correlación para las modificaciones de histonas entre tipos de células se calcularon en el catálogo de potenciadores transformados logarítmicamente.

- 20 *Dinámica y estado del potenciador:* Para definir el número de potenciadores dinámicos, primero dividimos los potenciadores en 3 categorías para cada tipo de celda: "desactivado" (número de lecturas < 25), "activado" (número de lecturas > 50) e "intermedio" de otra manera. Los potenciadores que se clasificaron al menos una vez tanto en estado "activado" como "desactivado" se clasificaron como dinámicos. Para cuantificar el número de potenciadores que se abren o cierran durante el proceso de diferenciación, se contaron el número de potenciadores "desactivado"/"activado" en la raíz del árbol hematopoyético (LT-HSC) que invirtió el estado en el tipo de célula madura relevante. Para determinar aún más la etapa precisa de ganancia/pérdida (figura 3D), los presentes inventores examinaron la ruta de HSC a las células maduras y seleccionaron la etapa en la que ocurrió el cambio máximo en los niveles de H3K4me1 ocurridos.

- 30 La actividad potenciadora se determinó en base a la relación de los niveles de H3K4me1 y H3K27ac. Los presentes inventores se centraron solo en las regiones "activado" a base de H3K4me1 al definir los potenciadores "activos" / "preparados". A diferencia de la distribución bimodal observada en los niveles de H3K4me1 en el conjunto de regiones, H3K27ac mostró una distribución normal. Por lo tanto, seleccionaron un umbral para las regiones "activas" de lecturas de H3K27ac > 25. Cambiar el umbral influyó en la relación "activo"/"preparado" pero no cambió el orden relativo de los tipos de células. Por ejemplo, los eritrocitos utilizan más potenciadores que CMP, independientemente del umbral específico seleccionado.

- 35 *Análisis de ATAC-seq:* Se cartografiaron y analizaron los datos de ATAC-seq utilizando el mismo conducto que los datos de ChIP-seq. Se extrajo un perfil de cobertura de lectura para cada región en el catálogo de mejoradores. El valor de accesibilidad de cromatina para cada potenciador en un tipo de célula dado se definió en base a la cobertura máxima en la región. El catálogo se volvió a centrar en base a la posición del valor máximo en el perfil de cobertura ATAC.

- 40 *Análisis de enriquecimiento de motivos:* Se buscaron motivos en el intervalo de 1000 pb centrado en el centro ATAC usando el comando `homer2` del paquete HOMER. Para cada región, los presentes inventores extrajeron la mejor puntuación de motivo ('`homer2 find -mscore`') para cada motivo en el catálogo de motivos JASPAR CORE (descargado en marzo de 2014). Las puntuaciones de enriquecimiento se calcularon usando la prueba de Kolmogorov-Smirnov de dos muestras, comparando las puntuaciones de motivos en el conjunto de potenciadores "desactivados" con los de los potenciadores "activado" para cada tipo de célula. Los motivos correspondientes a TF no expresados en un determinado tipo de célula se excluyeron del análisis.

- 45 *Modelo logístico:* Para predecir el estado potenciador ("activado" o "desactivado") a partir de su secuencia, se practicó con un modelo logístico.

En primer lugar, los presentes inventores estimaron si cada factor de transcripción se expresaba en cada tipo de célula. Se usaron mediciones de RNA-seq, definiendo "expresado" como que tiene más de 5 lecturas por millón de etiquetas, y más del 20% de sus niveles máximos de expresión en todas las células.

- 55 Las etiquetas de preparación fueron el estado de cada potenciador en cada tipo de célula, eliminando las de estado intermedio. El indicador recibió las puntuaciones de los 216 motivos del potenciador enmascarados por el estado de expresión del factor de transcripción correspondiente. Por lo tanto, la predicción podría utilizar las puntuaciones de motivos de factores que se expresan en el tipo de célula relevante. Los parámetros se estimaron utilizando la función matlab R2012b `glmfit`. Para evaluar la exactitud, la sensibilidad

y la precisión del modelo, se utilizó 5 veces la validación cruzada, cada vez practicando en el 80% de los tipos de células X potenciadoras y probando en el 20% restante. Los resultados publicados son el promedio de 1000 series de validación cruzada, cada una con un reparto aleatorio de 5 veces.

Resultados

- 5 Para caracterizar la dinámica de la cromatina de poblaciones celulares *in vivo* escasas, se desarrolló un primer enfoque de indización IP de cromatina (iChIP). En este protocolo, el código de barras se realiza directamente en la cromatina celular total (figura 1A), evitando así las reacciones enzimáticas de poca entrada que ocurren en ChIP convencional. Sobre todo, esto permite que múltiples muestras con códigos de barras de cromatina se agrupen para ChIP en el mismo pozo, reduciendo aún más los requisitos iniciales de entrada y aumentando la reproducibilidad de muestras cruzadas. Para minimizar las etapas de centrifugación, las células se reticulan antes de la clasificación y la exposición a ultrasonidos. Después, la cromatina cortada se inmoviliza en perlas magnéticas recubiertas con anticuerpo anti-H3 y se indiza. La cromatina indizada se libera de las perlas H3 y se combina con cromatina de otras muestras. Por último, se realiza ChIP con el anticuerpo deseado, y un solo grupo de cromatina con código de barras puede dividirse en múltiples ChIP para caracterizar varias modificaciones de cromatina.

15 El protocolo iChIP es muy reproducible para pocos números de células (unos pocos cientos de células) al tiempo que aumenta la sensibilidad y el rendimiento (figuras 1A-D). Para comparar iChIP, cantidades decrecientes (10.000 a 500 células) de cromatina aisladas de células dendríticas procedentes de médula ósea (BMDC) se codificaron con un código de barras, por triplicado, y se realizó ChIP en la cromatina con código de barras con un anticuerpo para histona mono- y tri-metilada H3 lisina 4 (H3K4me1 y H3K4me3) (figuras 1B-D). La reproducibilidad de los picos de H3K4me3 se confirmó entre reproducciones de 10.000 células ($r = 0,95$) hasta 500 células ($r = 0,85$). Los perfiles de iChIP se correlacionaron con ChIP-seq convencional en 10 millones de BMDC ($r = 0,92$) (16). iChIP es aplicable para caracterizar otras modificaciones de histonas y factores de transcripción para números de células bajos (figuras 1C-D). La reproducibilidad y sensibilidad para pequeño número de células de iChIP abre la vía para la caracterización *in vivo* de la dinámica de cromatina durante la hematopoyesis.

20 Se caracterizaron cuatro modificaciones de histonas (H3K4me1, H3K4me2, H3K4me3 y H3K27ac) en tres reproducciones *in vivo* en 16 etapas de desarrollo del compromiso hematopoyético utilizando 5.000 células por marca. Las poblaciones celulares elegidas comprenden todas las etapas progenitoras multipotentes y los progenitores oligopotenciales de linaje común de cada uno de los principales linajes de sangre (mielocítico, linfocítico y eritrocítico), así como varios tipos de células diferenciadas terminalmente de cada uno de los linajes (16). Las reproducciones de la marca histona ChIP y RNAseq mostraron gran reproducibilidad (promedio $r = 0,950$). Las marcas de histona se usaron para definir las regiones activadora (H3K4me3 alta) y potenciadora (H3K4me1/2 alta y H3K4me3 baja) dando como resultado el análisis de 48.415 potenciadores y 17.923 activadores. La actividad potenciadora se definió a partir de los niveles de H3K27ac y los valores de expresión de ARN del gen más cercano (16-18). El análisis del genoma completo de la expresión de ARN y la actividad activadora (señal H3K4me3) identificó cuatro patrones principales para las células madre, linfocíticas, mielocíticas y eritrocíticas (figura 2B). En contraste, la señal de H3K4me1 puso de manifiesto que los progenitores de linaje eran más similares a las células diferenciadas dentro de su linaje que a los progenitores de otros linajes celulares (figura 2B). Esta distinción sugiere que el establecimiento del potenciador se inicia en un compromiso de linaje inicialmente y puede poner de manifiesto el potencial de diferenciación de la descendencia antes de la ejecución del programa de expresión de ARN.

35 Los cambios en las marcas de cromatina durante la hematopoyesis (especialmente H3K4me1 y H3K4me2) sugieren actividad específica de linaje de elementos reguladores. De los 48.415 potenciadores hematopoyéticos, el 90% (43.428) cambió de estado durante la hematopoyesis. Se encontró que el 60% (26.393) de estos potenciadores dinámicos muestran el comportamiento estereotípico en el que inicialmente están marcados en células madre hematopoyéticas (HSC), pero se mantienen solo en el linaje relevante (figura 2C (Gata2)). Sorprendentemente, se descubrió que una proporción significativa de potenciadores dinámicos (40%, 17,035) se crean por primera vez durante el proceso de diferenciación (16). Por ejemplo, los potenciadores presentes en los locus de genes mielocíticos IL-1b, CD14, S 100a8 y F7 (factor de coagulación VII) se crean por primera vez o se vuelven más prominentes en el linaje mielocítico (figura 2C). Igualmente, los potenciadores de Ebfl y Cr2 se crean en linfocitos B; potenciadores de Bc11 lb y CD3g en linfocitos T; potenciadores de granzima A y Ncr1 en linfocitos citolíticos; y potenciadores de Gata 1, Gypa (Glycophorin A) y CpoX (Coproporphyrinogen Oxidase) en el linaje eritrocítico (figura 2).

55 La creación de potenciadores específicos del linaje por primera vez se produce principalmente en la raíz del punto de compromiso en el primer progenitor del linaje, mientras que el cierre de los potenciadores se produce de forma más gradual (figura 2). Por ejemplo, los locus S100a8 y F7 - F10 muestran la señal H3K4me1 específicamente en el linaje mielocítico con creación por primera vez de los potenciadores en la raíz del punto de compromiso mielocítico en el progenitor mielocítico común (CMP) (figura 2C). Igualmente, los locus CpoX y Gypa muestran la señal H3K4me1 específicamente en el linaje de eritrocitos con la creación por primera vez de los potenciadores en la raíz del punto de compromiso eritrocítico en el progenitor

eritrocítico de megacariocitos (MEP) (figura 2C). Sobre todo, se observó una adquisición gradual de las diferentes modificaciones de cromatina en los potenciadores específicos del linaje durante la hematopoyesis (figura 2D). Por ejemplo, en ambos locus S100a8 (linaje mielocítico) y Gypa (linaje eritrocítico), H3K4me1/2 aparece en primer lugar en el progenitor de linaje raíz (CMP y MEP, respectivamente), mientras que H3K27ac (activación de locus) se adquiere junto con la transcripción activa de ARN solo una vez que las células

5 (granulocitos y eritrocitos B) se diferencian terminalmente (figura 2D). Globalmente, el 32% de los potenciadores activados (H3K27ac) en células diferenciadas terminalmente están inicialmente preparados (solo H3Kme1) en los progenitores de linaje.

La agrupación de todos los 48.415 picos de H3K4me1 por sus perfiles dinámicos durante la hematopoyesis puso de manifiesto nueve grupos principales, según la biología subyacente del sistema (figura 3A). La señal H3K4me2 muestra patrones similares en los nueve grupos. El grupo I comprende potenciadores compartidos a lo largo de la hematopoyesis. Los grupos II-IV agrupan potenciadores específicos del linaje ya marcados en HSC y compartidos con progenitores hematopoyéticos. Por último, los grupos VI (figura 3B en el medio), VII VIII, IX (figura 3B en la parte inferior) agrupan potenciadores por primera vez del grupo que son específicos

10 para un linaje concreto y no se marcaron en HSC; con 6.382 potenciadores mielocíticos, 5.834 linfocíticos y 4.819 eritrocíticos. También se encontró un grupo de 6.612 potenciadores (grupo V) que compartía exclusivamente entre los progenitores (figura 3B arriba). Sobre todo, los grupos de potenciadores de eritrocitos y progenitores presentan una conservación de secuencia relativamente alta en el clado de mamíferos, con los potenciadores mielocíticos y linfocitos que muestran menor conservación. Estos datos

15 sugieren tasas diferenciales de dinámica evolutiva en estas regiones *cis* reguladoras (figura 3A).

Los potenciadores recién formados podían crearse en un punto de ramificación específico o crearse gradualmente durante el proceso de desarrollo. Para examinar más a fondo esta dinámica, los presentes inventores generaron un catálogo de potenciadores que son dinámicos (ganados o perdidos) durante el proceso de diferenciación de las células madre hematopoyéticas a largo plazo (LT-HSC) a células maduras, terminalmente diferenciadas (figura 3C). Se determinó, para cada potenciador, la etapa de ganancia/pérdida

20 a lo largo de la ruta de diferenciación. En la diferenciación eritrocítica, el 65% del repertorio potenciador por primera vez se obtiene en la etapa MEP (figura 3D). Del mismo modo, el 40-50% de los potenciadores mielocíticos por primera vez se obtienen en el primer paso del compromiso mielocítico, durante la transición de progenitor multipotente (MPP) a CMP, mientras que la transición de CMP a GMP implica menos ganancias (15-30%). Juntas, las fases CMP y GMP son responsables del 63-80% de los potenciadores obtenidos en las células mielocíticas diferenciadas terminalmente (figura 3D), coherente con la similitud potenciadora global de CMP y GMP con el linaje mielocítico (figura 2B). Por el contrario, la pérdida del potenciador es un proceso más gradual que se inicia en la fase CMP o MEP (para el desarrollo mielocítico y eritrocítico, respectivamente) con una gran proporción (40-50%) de potenciadores perdidos en la última y definitiva etapa

25 de diferenciación a las células maduras.

Los estudios de todo el genoma demuestran que, mientras que H3K4me1 marca los potenciadores tanto preparados como activos, H3K27ac solo marca los potenciadores activos (19, 20). Dado que los potenciadores preparados representan posibles programas de expresión génica, la proporción de estos potenciadores en un tipo celular dado se aproxima al potencial regulador actual de la célula (20). El análisis

30 del potencial regulador en la hematopoyesis demuestra que las células madre son más plásticas que los tipos de células diferenciadas, con eritrocitos que utilizan la mayoría (78%) de sus potenciadores, mientras que las CMP usan solo el 33% de su potencial potenciador (figura 3E-F). Dentro del grupo progenitor, CMP y GMP son más plásticos que las células madre multipotentes (33-37% frente al 62-65% de uso de potenciadores); probablemente debido a la expansión por primera vez de los potenciadores mielocíticos (figuras 2B, 3A y F).

35 En especial, existe un amplio espectro de plasticidad a través de las células diferenciadas terminalmente, con el linaje mielocítico (macrófagos, monocitos y granulocitos) que presentan grados de plasticidad más altos que los eritrocitos, linfocitos B y linfocitos plasmáticos; esto es coherente con la mayor versatilidad funcional de las células mielocíticas en comparación con otros tipos de células hematopoyéticas (21).

La creación por primera vez de H3K4me1 en la hematopoyesis es simultánea con el aumento de la accesibilidad a la cromatina (figura 4A, B). La accesibilidad a la cromatina, "cromatina abierta" (22), se midió durante el proceso de desarrollo en diez tipos de células hematopoyéticas utilizando el ensayo de cromatina accesible a la transposasa seguido de secuenciación (ATAC-seq) (23). En locus de genes eritrocíticos (Gypa) y mielocíticos (F7 y F10), la accesibilidad a la cromatina sigue el patrón temporal observado para H3K4me1 (figura 4A). Como es de esperar, la señal de ATAC se enriquece tanto en activadores activos como en

40 potenciadores positivos a H3K4me1, así como en regiones aislantes y otras reguladoras (23). Para evaluar la proporción de potenciadores por primera vez que también presentan la creación por primera vez de cromatina abierta, los presentes inventores trazaron la señal ATAC-seq en las regiones del catálogo de potenciadores H3K4me1. Compararon la señal ATAC-seq con las intensidades de H3K4me1 y H3K27ac. Se observó un patrón similar entre la señal ATAC y H3K4me1 ($r = 0,75$) y un acuerdo debilitado con H3K27ac ($r = 0,62$) (figura 4B) que sugiere que el proceso de ganancia o pérdida de la marca H3K4me1 en potenciadores ocurre

45 simultáneamente a la formación de sitios de cromatina abierta.

La creación de potenciadores específicos de linaje está regulado por la actividad de los factores de

transcripción específicos del linaje (24) (figura 4C-E). Utilizando los picos de ATAC (16) y el catálogo de potenciadores descrito en este documento, buscaron motivos de unión de factor de transcripción enriquecidos en cada tipo de célula (figura 4C) para identificar los factores determinantes del linaje. Se encontró que, en línea con sus funciones identificadas, PU. 1, Gata1 y Foxo1 pueden clasificarse como reguladores potenciales de potenciadores mielocíticos, eritrocíticos y linfocíticos, respectivamente (3, 11, 25, 26).

Para identificar sistemáticamente los posibles reguladores, los presentes inventores generaron un modelo de regresión logística para predecir la actividad potenciadora en cada etapa a partir de las puntuaciones de los motivos de unión al ADN y la expresión de TF (figura 4D, E). El modelo logístico predice con precisión la actividad específica del tipo de célula potenciadora con un 75% de precisión (con un 70% de sensibilidad y un 80% de especificidad). Sobre todo, el modelo les permitió dilucidar los factores de transcripción que controlan la dinámica de la cromatina y la especificación de linaje en la hematopoyesis (25, 26). El modelo identificó los factores determinantes del linaje mielocítico conocido, PU.1, Cebpb y Cebpa como reguladores de potenciadores mielocíticos; además, el modelo sugiere una jerarquía entre los factores Cebp con Cebpa activa en los progenitores (CMP/GMP) y Cebpb reemplazando a Cebpa en los tipos de células diferenciadas (figura 4E). Del mismo modo, Meis1, Hoxa9 y Erg se identificaron como reguladores potenciales de potenciadores de células madre, Pax5 en linfocitos B, Klf1 en células eritrocíticas y Ets1 en células linfocíticas (figuras 4C-E). Se identificaron muchos factores de transcripción que han estado implicados en el desarrollo del linaje pero que no se han asociado previamente con la regulación de la cromatina de la determinación del linaje, como Irf1 e Irf2 en linfocitos B y Cebpe en los granulocitos (27, 28). Se destacan nuevos reguladores potenciales de linajes hematopoyéticos: ATF3 en monocitos y Tcf712, Mef2a y Runx2 en linfocitos citolíticos (figura 4E). Todos juntos, los descubrimientos actuales demuestran que la cromatina es muy dinámica durante la hematopoyesis, orquestada por un conjunto definido de factores de transcripción.

En conclusión, iChIP permite la ejecución de ChIP reproducible y sensible en solo unos pocos cientos de células de una manera ampliamente aplicable en organismos y tejidos. Se demostró que los potenciadores preparados se crean en progenitores de linaje antes de su activación y preceden a la expresión de ARN en la diferenciación de linaje posterior. Estos potenciadores se crean simultáneamente con la formación de puntos de cromatina abiertos. Los presentes inventores demostraron entonces que la mayor parte de la dinámica potenciadora puede explicarse por la actividad de conocidos factores específicos de linaje, así como por los nuevos reguladores candidatos. Estos resultados sugieren un nuevo modelo para la dinámica de la cromatina durante la diferenciación (figura 4F) y demuestran que el desarrollo implica una reorganización dinámica masiva del paisaje de cromatina. Si bien algunos potenciadores están preestablecidos en células madre hematopoyéticas, como sugirió el modelo de desarrollo convencional, un número comparable de potenciadores parece haberse creado por primera vez durante la hematopoyesis. Se puede suponer que la creación de potenciadores recién preparados en las etapas iniciales del compromiso de linaje inicia programas reguladores que posteriormente se aplican en células diferenciadas, mientras que el cierre de potenciadores ocurre durante las fases de diferenciación posteriores. Esto sugiere que el potencial potenciador celular alcanza su máximo no en la fase HSC, sino durante las etapas progenitoras oligopotentes. En conjunto, estas observaciones remodelan la comprensión actual del papel de la cromatina y los factores pioneros durante la diferenciación.

Listado de secuencias

<110> Yeda Research And Development Co. Ltd.
AMIT, Ido
45 LARA-ASTIASO, David

<120> MÉTODOS Y KITS PARA ANALIZAR CELULAS ESPECÍFICAS DE UNIÓN DE ADN CÉLULAS ADJUNTAS

50 <130> 61930

<150> 61/980,630
<151> 17-04-2014

55 <160> 10
<170> PatentIn versión 3.5

<210> 1
<211> 33
60 <212> ADN
<213> Secuencia artificial

<220>

ES 2 764 504 T3

<223> oligonucleótido de ADN de cadena sencilla

<220>
 <221> característica miscelánea

5 <222> (32)..(32)
 <223> modificación de fosfotionato

<400> 1
 acactctttc cctacacgac gctcttccga tct 33

10 <210> 2
 <211> 61
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> oligonucleótido de ADN de cadena sencilla

<220>
 <221> característica miscelánea

20 <222> (34)..(39)
 <223> código de barras para multiplexación de muestras

<220>
 <221> misc_feature

25 <222> (34)..(39)
 <223> a o g o c o t / u, desconocido u otro

<400> 2
 gatcggaaga gcacacgtct gaactccagt cacnnnnna tctcgtatgc cgttttctgt 60

t 61

30 <210> 3
 <211> 45
 <212> ADN
 <213> Secuencia artificial

35 <220>
 <223> oligonucleótido de ADN de cadena sencilla

<400> 3

40 aatgatacgg cgaccaccga gatctacact ctttcctac acgac 45

<210> 4
 <211> 24
 <212> ADN

45 <213> Secuencia artificial

<220>
 <223> oligonucleótido de ADN de cadena sencilla

50 <400> 4
 caagcagaag acggcatacg agat 24

<210> 5
 <211> 86

55 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador oligonucleotídico de ADN de cadena sencilla RT con código de barras

60 <220>
 <221> característica miscelánea
 <222> (56)..(65)
 <223> a o g o c o t / u, desconocido u otro

ES 2 764 504 T3

- <220>
 <221> característica miscelánea
 <222> (56)..(61)
 5 <223> código de barras para multiplexación de muestras
- <220>
 <221> característica miscelánea
 <222> (62)..(65)
 10 <223> RMT
- <220>
 <221> característica miscelánea
 <222> (86)..(86)
 15 <223> a o g o c
 o t / u, desconocido u otro
 <400> 5
 cgattgaggc cggtaatacg actcactata ggggacgacgt gtgctcttcc gatctnnnnn 60
 nnnnnnttttt tttttttttt tttttt 86
- 20 <210> 6
 <211> 22
 <212> ADN
 <213> Secuencia artificial
- 25 <220>
 <223> Adaptador de ligadura de oligonucleótidos de ADN de cadena sencilla
- <220>
 <221> característica miscelánea
 30 <222> (1)..(1)
 <223> 5 'fosforilada
- <220>
 <221> característica miscelánea
 35 <222> (22)..(22)
 <223> Espaciador 3 'C3 (bloqueador)
- <400> 6
 agatcggaag agcgtcgtgt ag 22
 40
- <210> 7
 <211> 22
 <212> ADN
 <213> Secuencia artificial
 45
- <220>
 <223> oligonucleótido de ADN de cadena sencilla
- <400> 7
 50 tctagccttc tcgcagcaca tc 22
- <210> 8
 <211> 58
 <212> ADN
 55 <213> Secuencia artificial
- <220>
 <223> oligonucleótido de ADN de cadena sencilla
- 60 <400> 8
 aatgatacgg cgaccaccga gatctacact ctttcctac acgacgctct tccgatct 58
- <210> 9
 <211> 58

ES 2 764 504 T3

<212> ADN
<213> Secuencia artificial

<220>
5 <223> oligonucleótido de ADN de cadena sencilla
<400> 9
caagcagaag acggcatacg agatgtgact ggagttcaga cgtgtgctct tccgatct 58

<210> 10
10 <211> 35
<212> ADN
<213> Secuencia artificial

<220>
15 <223> Secuencia promotora de ARN polimerasa T7

<400> 10
cgattgaggc cggtaatacg actcactata ggggc 35

REIVINDICACIONES

1. Un método para caracterizar la dinámica de la cromatina de las poblaciones celulares obtenidas, comprendiendo el método:
 - 5 (a) obtener al menos dos muestras de poblaciones celulares, comprendiendo cada muestra células de un tipo celular diferente o procedentes de un sujeto concreto;
 - (b) reticular la cromatina celular total de al menos dichas dos muestras;
 - (c) aislar dicha cromatina celular total en un soporte sólido;
 - (d) marcar el ADN de dicha cromatina celular total, en donde dicho marcaje distingue entre dicha cromatina celular total
- 10 de la primera de al menos dichas dos muestras y dicha cromatina celular total de la segunda de al menos dichas dos muestras;
 - (e) liberar dicha cromatina celular total de dicho soporte sólido;
 - (f) agrupar al menos dichas dos muestras de cromatina celular total; y
 - (g) aislar el ADN que forma un complejo con una proteína de unión al ADN de dicha cromatina celular total
- 15 usando un agente que se une específicamente a dicha proteína de unión a ADN; y
 - (h) analizar dicho ADN de dicha cromatina celular total.
2. El método de la reivindicación 1, en donde dicho análisis comprende secuenciar dicho ADN.
3. El método de las reivindicaciones 1 o 2, que comprende además amplificar dicho ADN después de la etapa (g) y antes de la etapa (h).
- 20 4. El método de una cualquiera de las reivindicaciones 1-3, que comprende además analizar dicha proteína de unión al ADN de dicha cromatina.
 5. El método de una cualquiera de las reivindicaciones 1-4 que comprende además someter a ultrasonidos dicho ADN después de la etapa (b) y antes de la etapa (c) de tal manera que dicho ADN no tenga más de 500 bases.
- 25 6. El método de la reivindicación 1, en donde dicha proteína de unión a ADN es una histona.
 7. El método de una cualquiera de las reivindicaciones 1-6, en donde dicho agente es un anticuerpo.
 8. El método de cualquiera de las reivindicaciones 1-7, que no se lleva a cabo utilizando un dispositivo microfluídico.
 9. El método de la reivindicación 1, en donde el número de células de la muestra está comprendido entre 100
- 30 y 10.000.

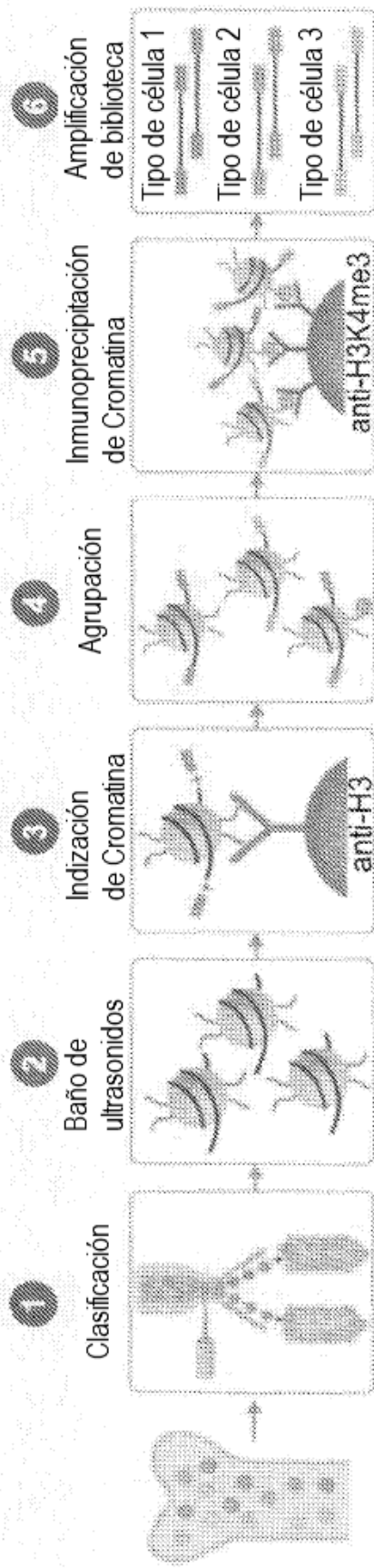


FIG. 1A

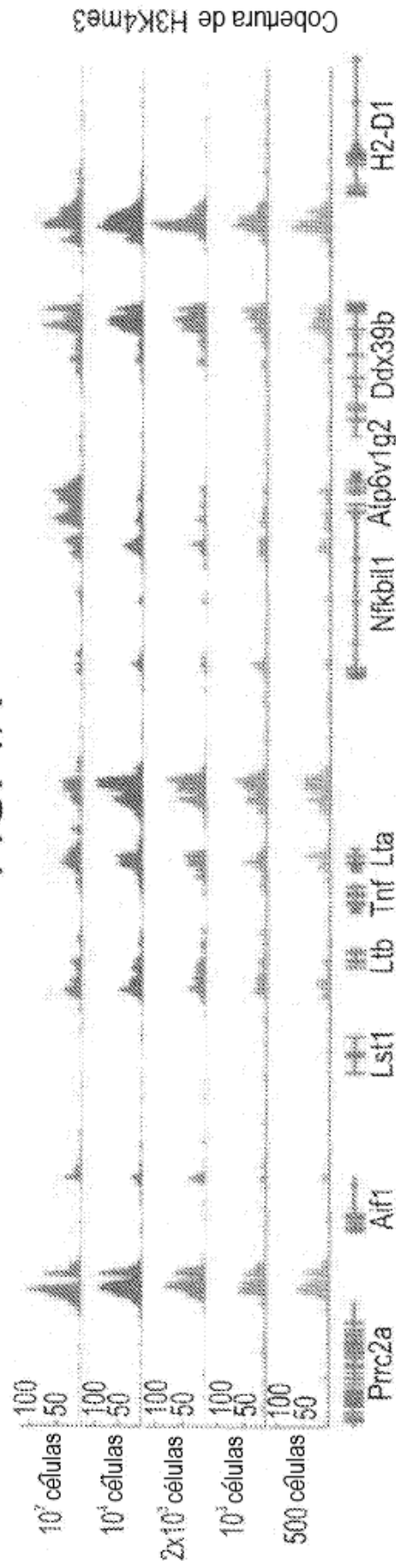


FIG. 1B

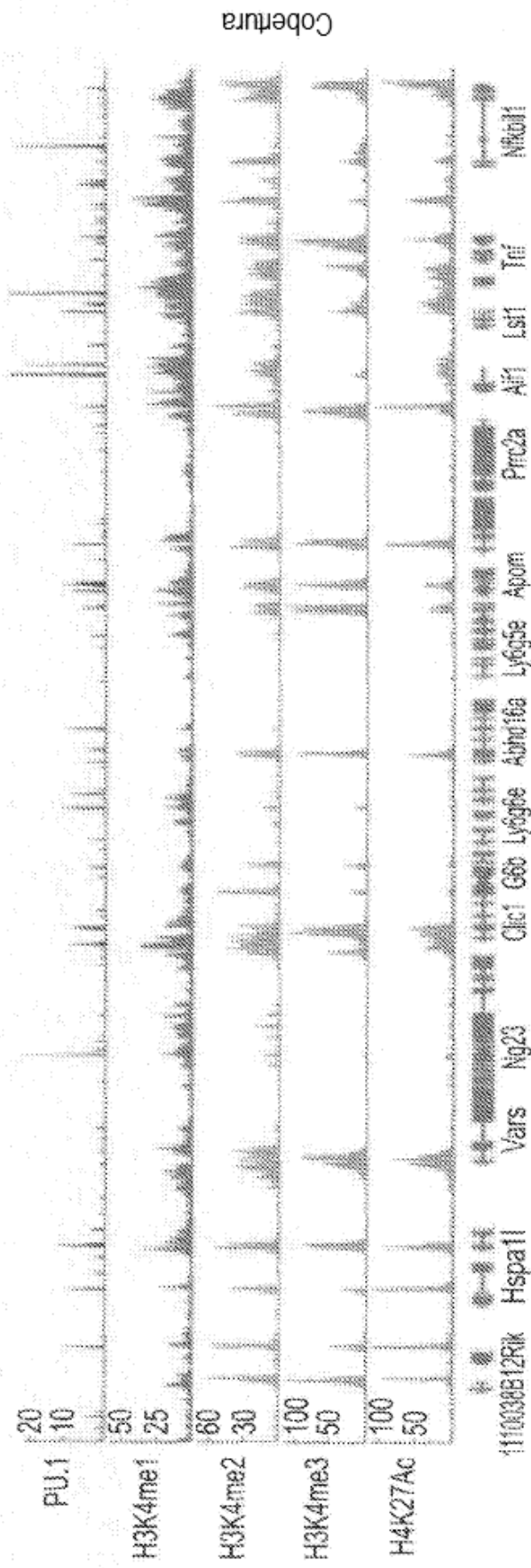


FIG. 1C

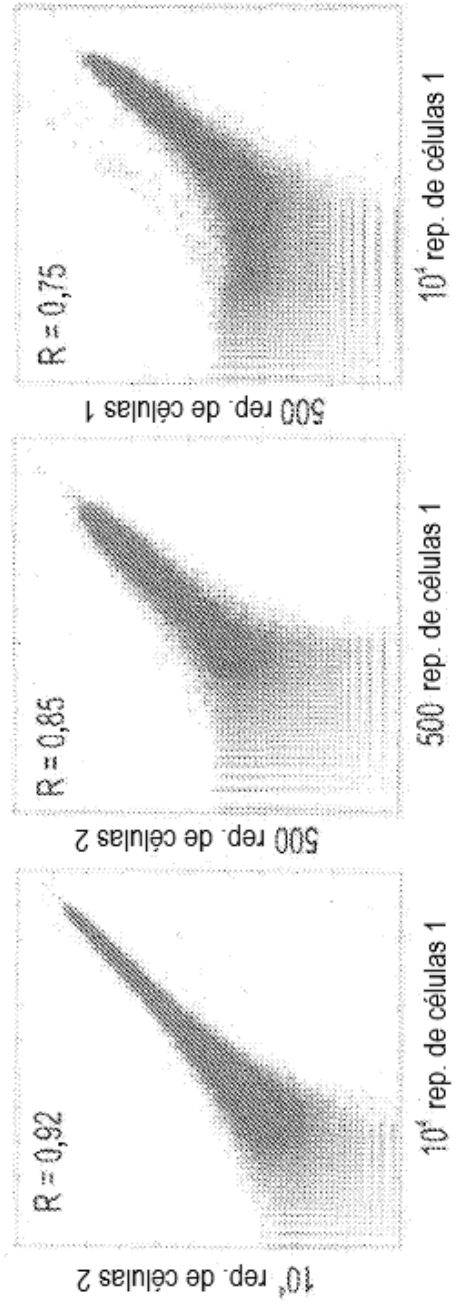


FIG. 1D

FIG. 2A

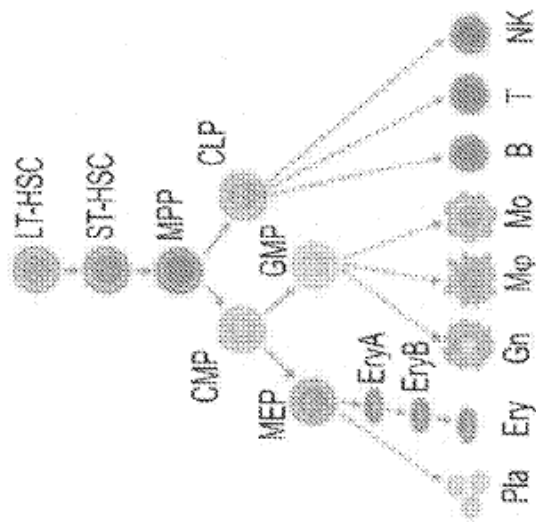


FIG. 2B

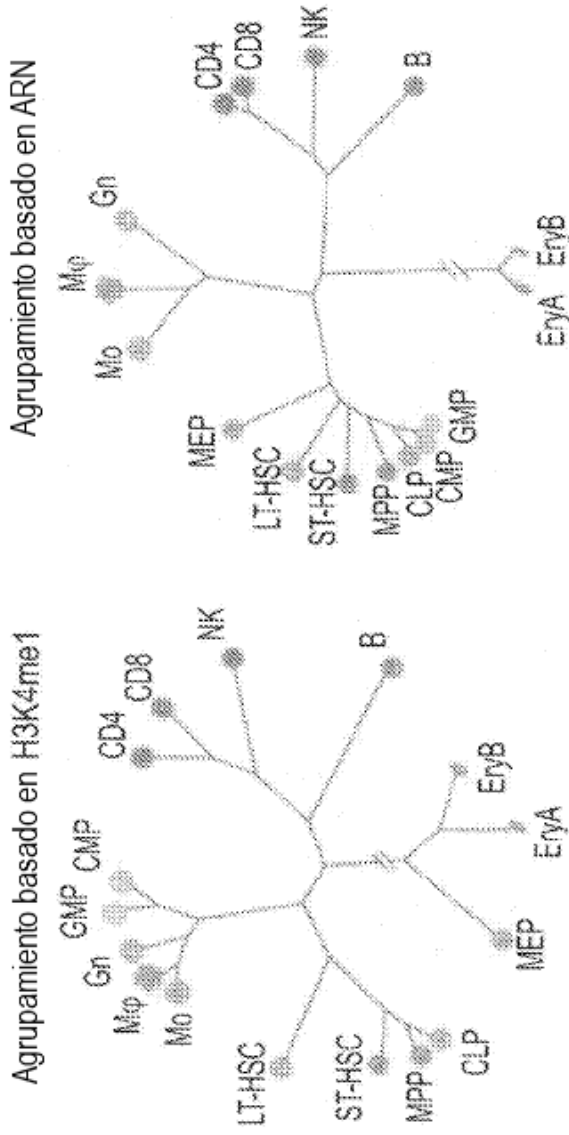


FIG. 2C

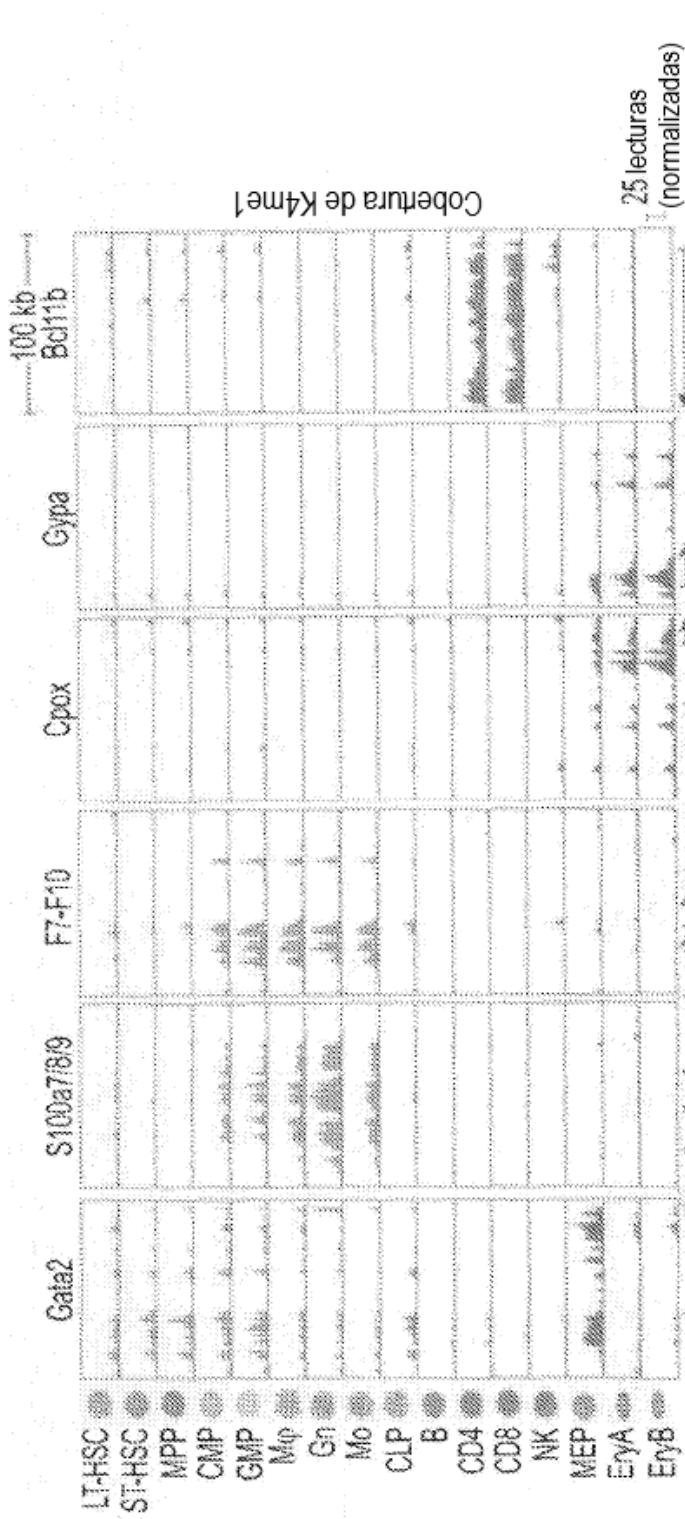


FIG. 2D

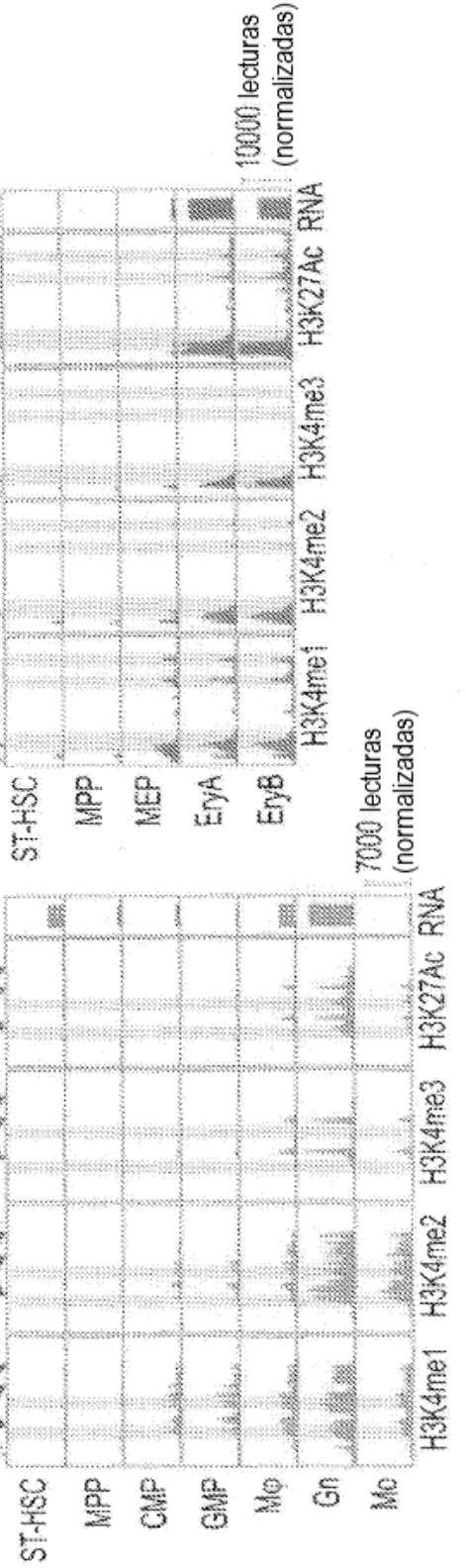


FIG. 3A

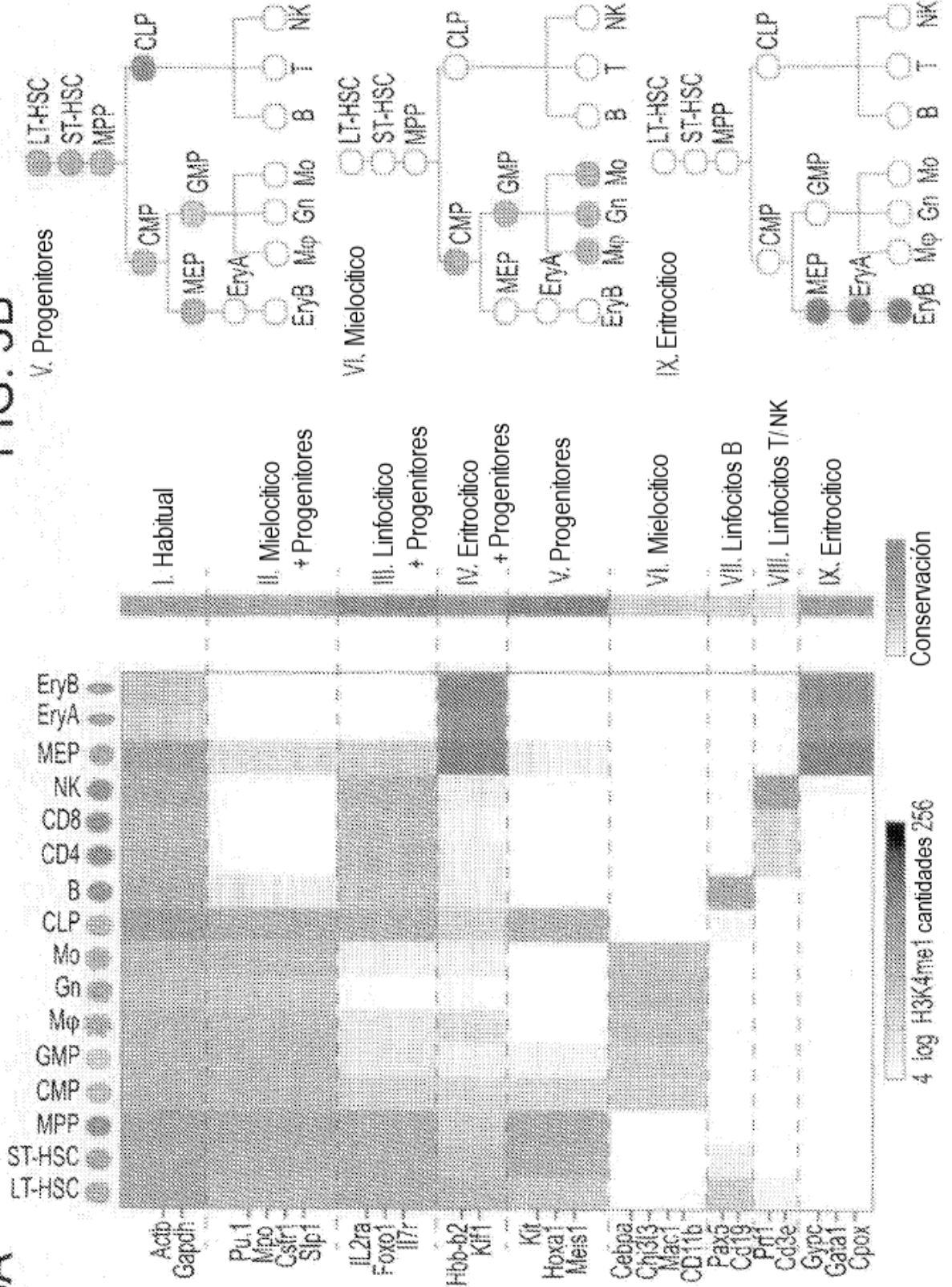


FIG. 3C

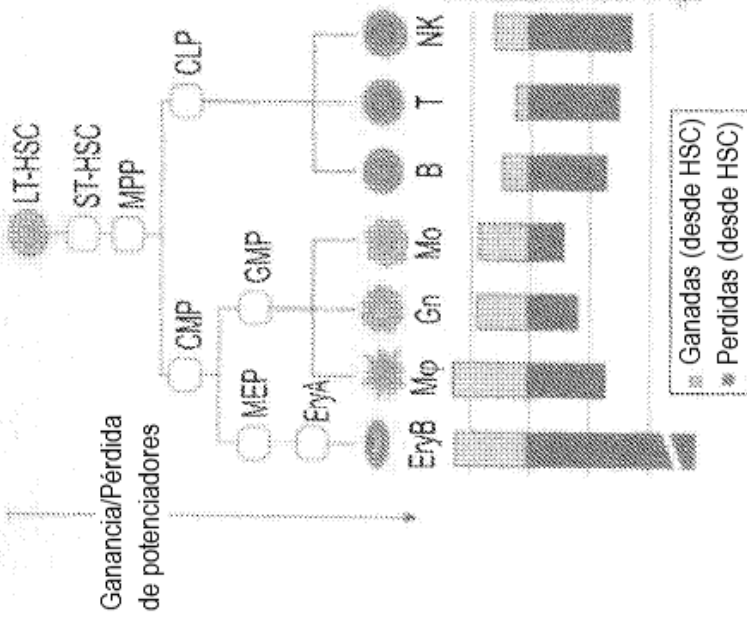


FIG. 3D

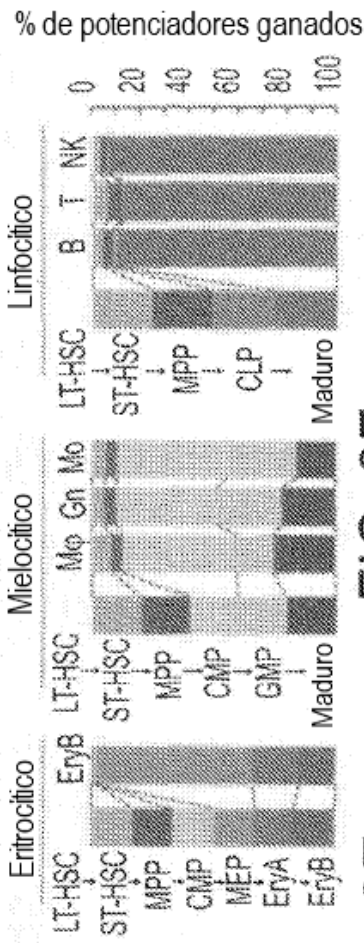


FIG. 3E

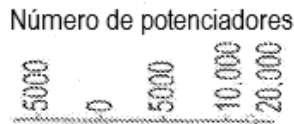


FIG. 3F

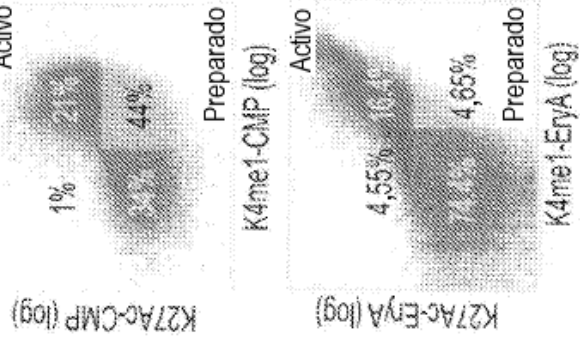
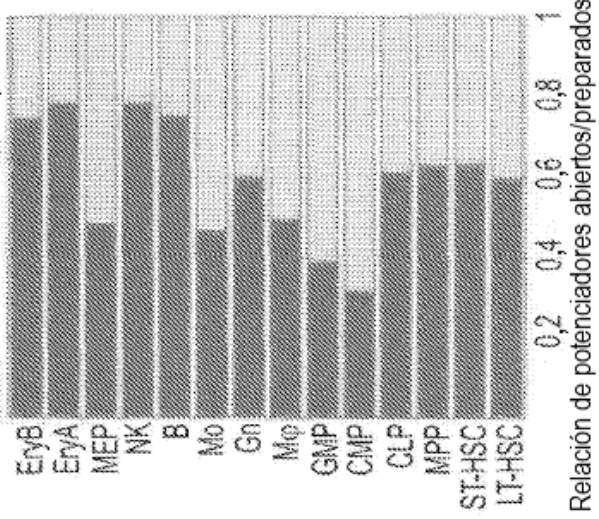


FIG. 4B

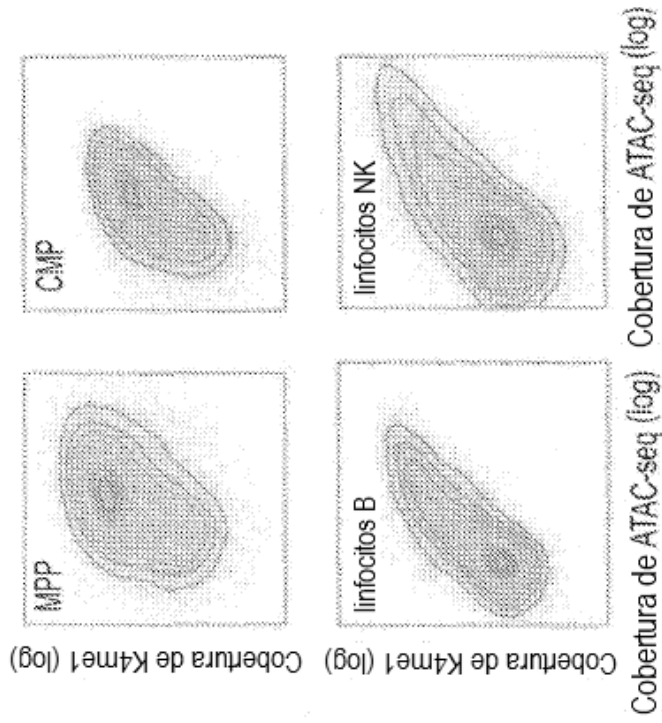


FIG. 4A

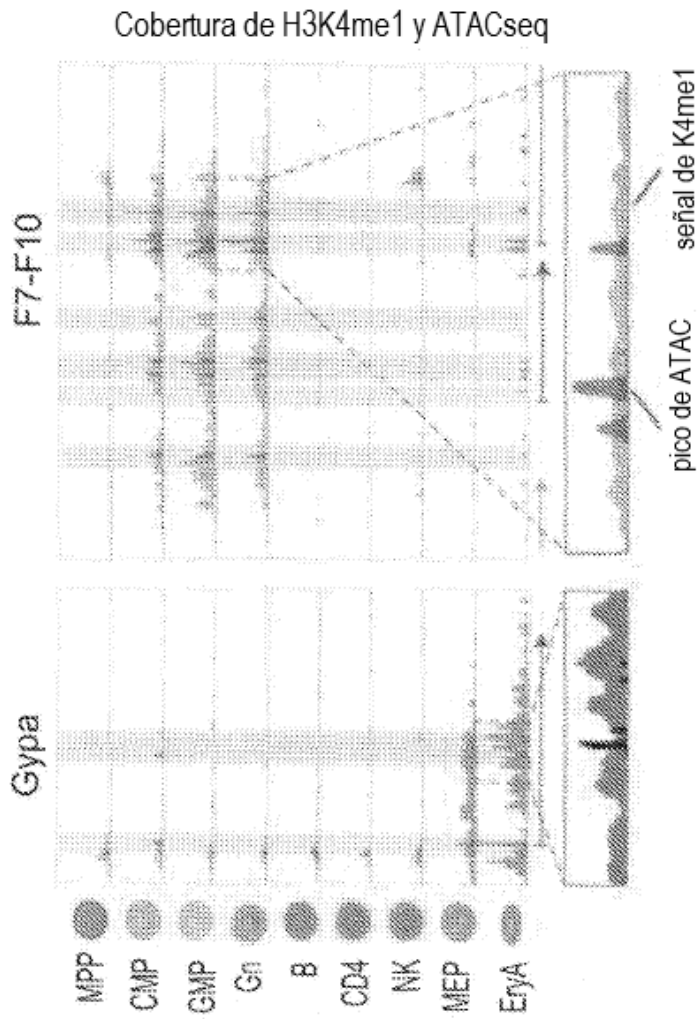


FIG. 4D

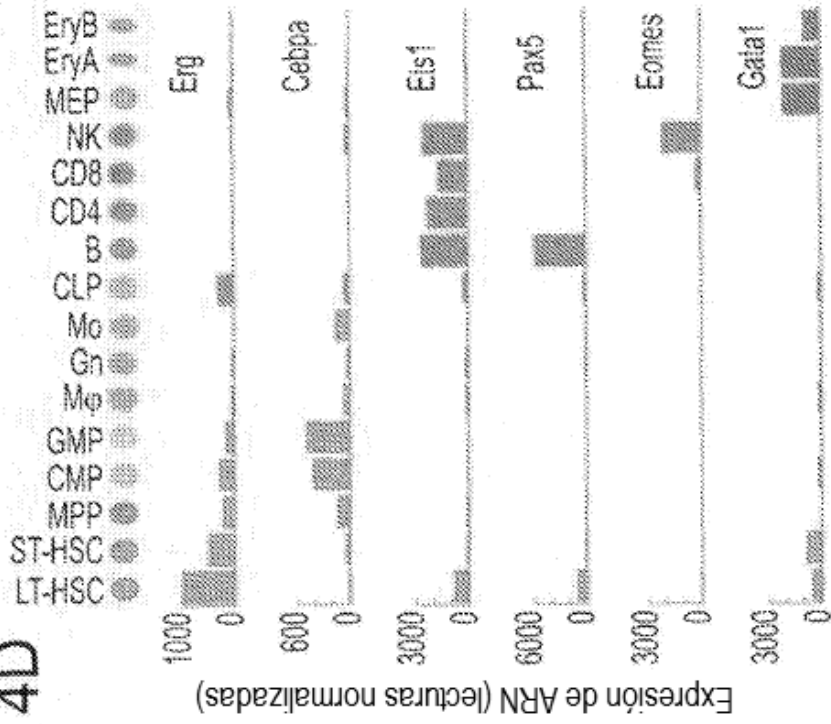
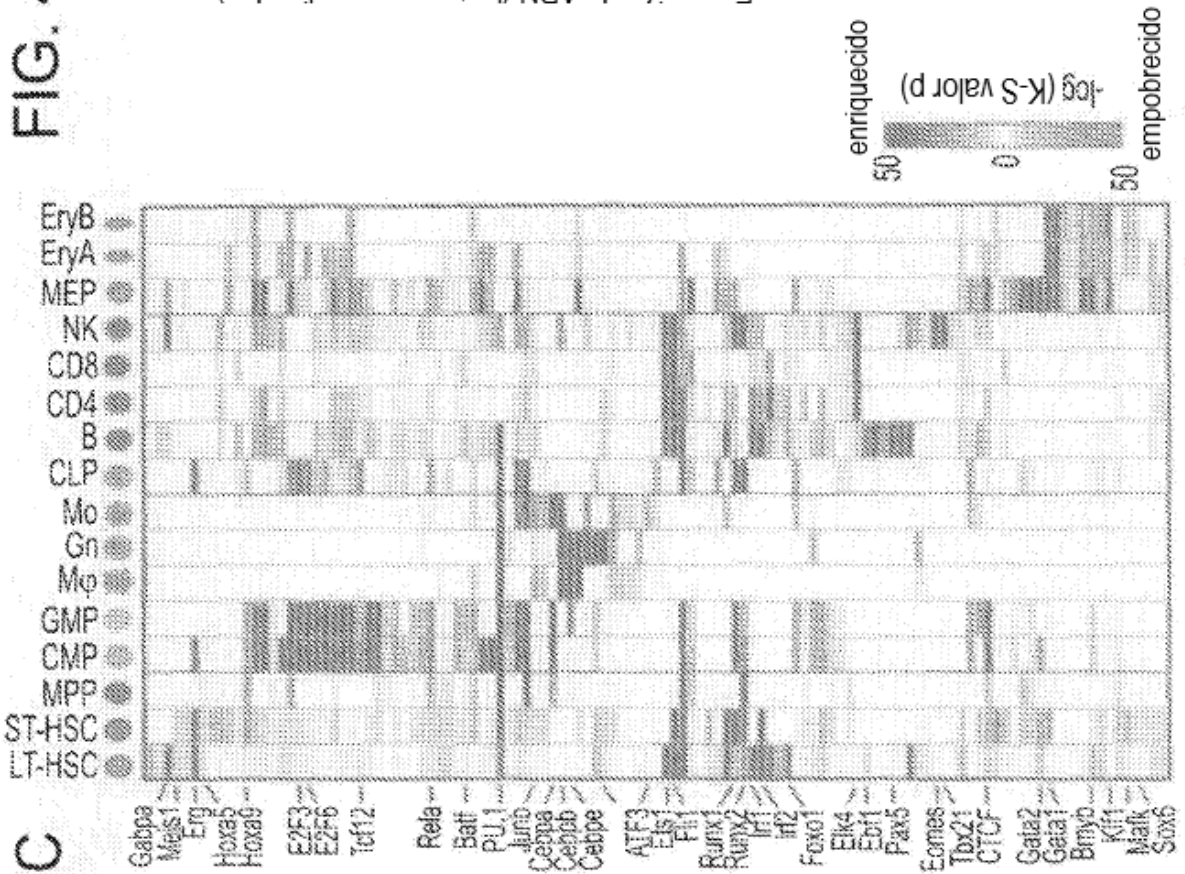


FIG. 4C



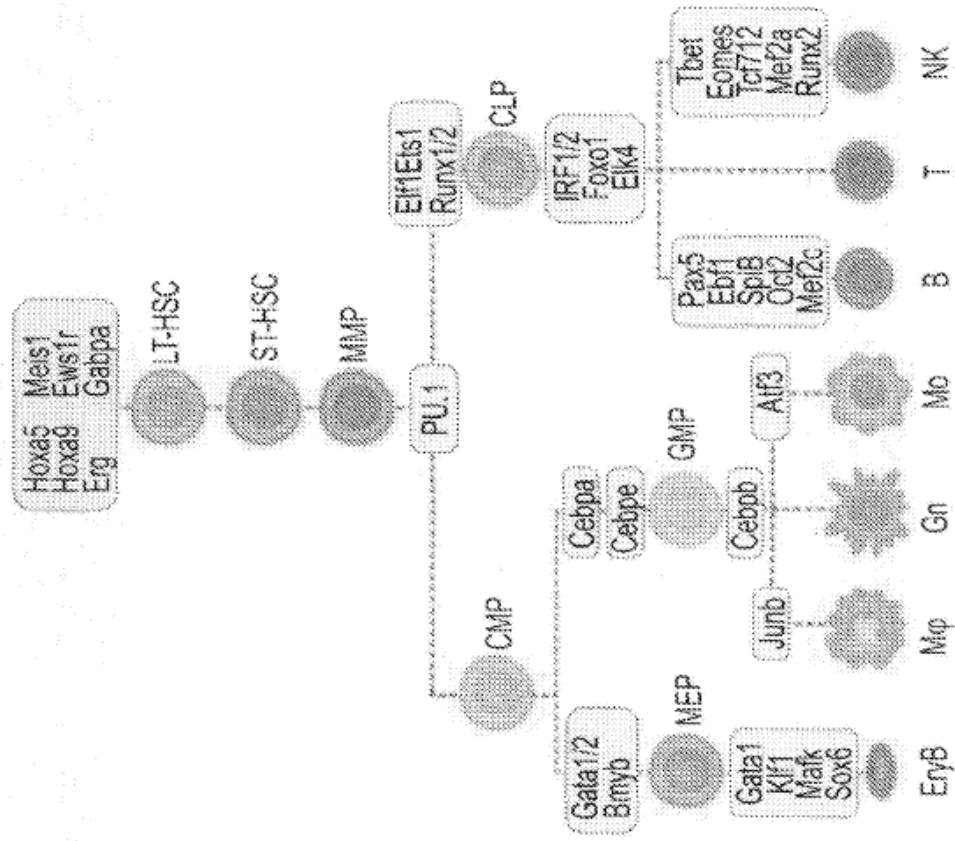


FIG. 4E

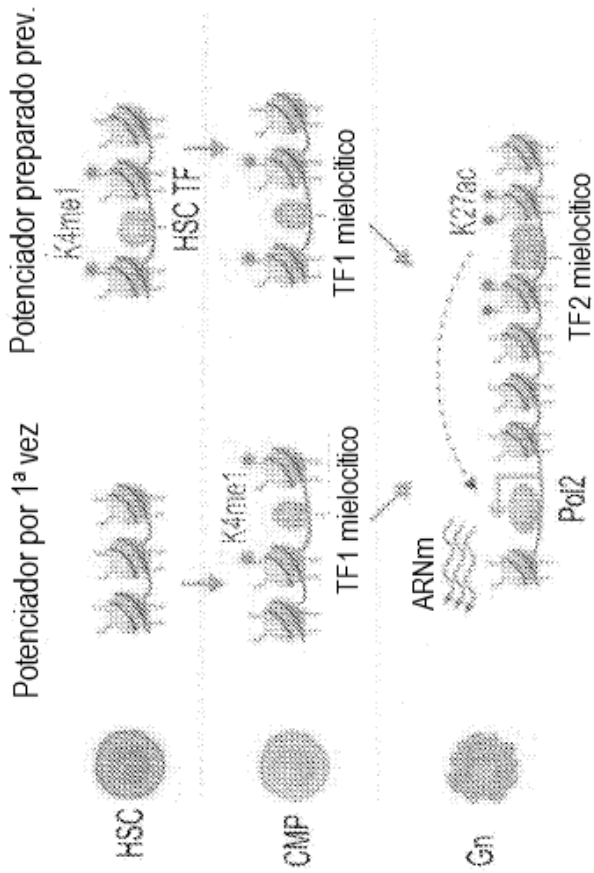


FIG. 4F