



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Estudio de los cambios en la accesibilidad y organización tridimensional de la cromatina durante la diferenciación eritroide

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

Andrés Penagos Puig

TUTORA PRINCIPAL

Dra. Mayra Furlan Magaril

Instituto de Fisiología Celular, UNAM.

MIEMBROS DEL COMITÉ TUTOR

Dra. Viviana del Carmen Valadez Graham

Instituto de Biotecnología, UNAM.

Dr. Víctor Julián Valdés Rodríguez

Instituto de Fisiología Celular, UNAM.

Ciudad de México. Octubre, 2023



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



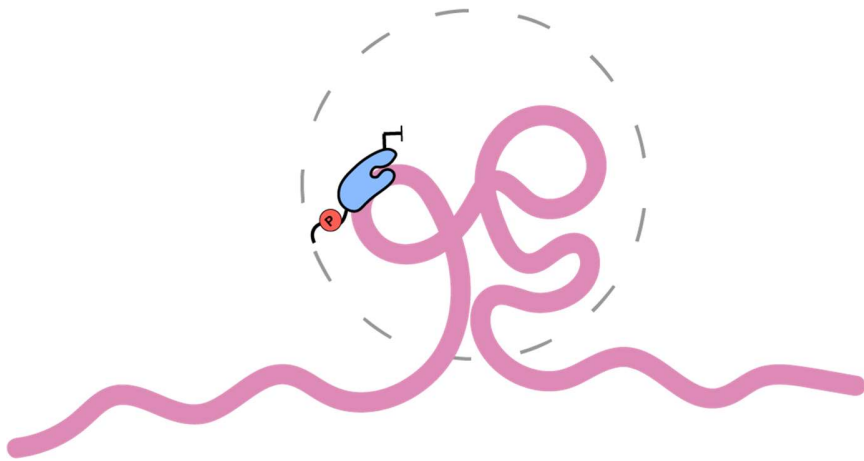
UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Estudio de los cambios en la accesibilidad y organización tridimensional de la cromatina durante la diferenciación eritroide



Andrés Penagos Puig

Agradecimientos académicos

Este trabajo se realizó bajo la dirección de la Dra. Mayra Furlan Magaril, en el laboratorio 202 Sur del Departamento de Genética Molecular de la División de Investigación Básica del Instituto de Fisiología Celular de la Universidad Nacional Autónoma de México (UNAM).

El comité tutor que asesoró el desarrollo de esta tesis estuvo conformado por:

Dra. Viviana del Carmen Valadéz Graham	Instituto de Biotecnología, UNAM
Dr. Víctor Julián Valdés Rodríguez	Instituto de Fisiología Celular, UNAM
Dra. Mayra Furlan Magaril	Instituto de Fisiología Celular, UNAM

Este trabajo se realizó con financiamiento del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica, de la Dirección General de Asuntos del Personal Académico de la UNAM (Donativos no. IN207319 y no. IN210323). Así como financiamiento otorgado por parte del Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT; no. 303068, no. 15758 y no. 137721).

Se agradece al CONAHCyT por la beca otorgada para realizar los estudios de doctorado (Número de CVU 822335).

Se reconoce a la Dra. Rosario Pérez Molina por su ayuda en el diseño y ejecución de los experimentos de inmunoprecipitación de la cromatina, su apoyo administrativo y logístico en la secuenciación de las bibliotecas genómicas y su labor como técnico académico del laboratorio.

Se reconoce a la M. en C. Giovana Sherlyn Claudio Galeana, quien realizó los experimentos de RNA-seq y ChIP-seq de este proyecto como parte de sus estudios de maestría. A las M. en C. Aura Stephenson Gussinye y Karina Jácome López, quienes realizaron los experimentos iniciales de Hi-C.

Se reconoce a la Dra. Ruth Rincón Heredia y el personal de la Unidad de Imagenología del Instituto de Fisiología Celular por su asistencia en los experimentos de microscopía y el entrenamiento en el uso del microscopio confocal.

Se reconoce al Dr. Augusto César Poot Hernández y al personal de la Unidad de Bioinformática del Instituto de Fisiología Celular por su asesoría en el almacenamiento, manejo y análisis de los datos genómicos generados.

Se reconoce al personal académico de la Unidad de Biología Molecular del Instituto de Fisiología Celular, la Dra. Laura Ongay Larios, Biól. Guadalupe Códiz Huerta y M. en C. Minerva Mora Cabrera por su apoyo en la síntesis de oligonucleótidos y asistencia para la electroforesis capilar de ADN de las bibliotecas genómicas.

Se reconoce al personal de la Unidad del Bioterio del Instituto de Fisiología Celular, particularmente a la M.V.Z. Claudia V. Rivera Cerecedo y el M.V.Z. Héctor A. Malagón Rivero, por su apoyo en el manejo apropiado de los animales empleados en este estudio.

Se reconoce a la Unidad de Supercómputo del Laboratorio Nacional de Ciencias de la Sostenibilidad (LANCIS) en el Instituto de Ecología (UNAM) y a Rodrigo García Herrera por su apoyo en el uso del cluster *Patung*.

Agradezco al Programa de Apoyo a los Estudios del Posgrado (PAEP) por su apoyo económico para asistir al XXXIII Congreso Nacional de Bioquímica (2022) en Mérida, Yucatán para la presentación del trabajo *“RNA polymerase II pausing contributes to maintain chromatin organization in the chicken erythrocytes”*.

Índice

Tabla de contenido

RESUMEN	6
ABSTRACT.....	7
I. PRINCIPIOS DE LA ORGANIZACIÓN TRIDIMENSIONAL DEL GENOMA.....	8
II. MECANISMOS DE ORGANIZACIÓN DEL GENOMA EN VERTEBRADOS ...	21
III. ANTECEDENTES	25
IV. PROPUESTA DE INVESTIGACIÓN	31
V. RESULTADOS	34
VI. DISCUSIÓN	58
VII. CONCLUSIONES	66
VIII. PERSPECTIVAS	67
IX. MÉTODOS.....	68
X. REFERENCIAS	79
XI. ANEXOS.....	85

Resumen

La eritropoyesis en vertebrados no mamíferos, como las aves, da lugar a eritrocitos circulantes que retienen el núcleo celular en un estado altamente compactado. Los eritrocitos circulantes en pollo, tanto embrionarios como adultos, experimentan una caída dramática en la tasa de síntesis de RNA y una extensa compactación de la cromatina que se puede apreciar en micrografías electrónicas. Sin embargo, aún es necesario caracterizar la remodelación de la cromatina y la organización del genoma en alta resolución durante este proceso, así mismo, es necesario ampliar nuestro entendimiento acerca de la severidad del silenciamiento transcripcional en estas células.

A través de la implementación de ensayos de microscopía y técnicas acopladas a secuenciación masiva caracterizamos los cambios en expresión génica, arquitectura nuclear, accesibilidad y topología de la cromatina a lo largo de la eritropoyesis terminal en pollo y reportamos una compleja remodelación de la organización del genoma. Mediante la identificación de la cromatina accesible en los eritrocitos adultos de pollo, se observó el reposicionamiento espacial de la cromatina abierta dentro del núcleo hacia la periferia nuclear. Dichos sitios accesibles en los genomas eritroides representan a los promotores de genes sin expresión que retienen a la RNA polimerasa II (RNA Pol II) en un estado pausado. A través de experimentos de RNA-seq se identificó a un grupo de genes que permanecen activos en los eritrocitos embrionarios y adultos, algunos de ellos relacionados con el control de la actividad y pausado de la RNA Pol II.

Finalmente, el estudio de la organización del genoma mediante la implementación de técnicas de captura (Hi-C) mostró que el genoma eritroide se encuentra altamente compartimentado y desprovisto de dominios topológicamente asociados excepto en regiones en donde la RNA Pol II pausada contribuye al plegamiento de la cromatina en mini dominios. Los resultados obtenidos sugieren que el pausado de la RNA Pol II participa en la regulación transcripcional del genoma eritroide y destacan la importancia de la RNA Pol II en el mantenimiento de la organización de la cromatina.

Abstract

Erythropoiesis in non-mammalian vertebrates such as birds gives rise to circulating erythrocytes that retain their nucleus in a highly compacted state. Chicken erythrocytes are often referred to as transcriptionally inactive, although the epigenetic changes and chromatin remodeling that mediate transcriptional repression and the extent of gene silencing during avian terminal erythroid differentiation are not fully understood. Chicken embryonic and adult red blood cells experience a dramatic drop in RNA synthesis rate and undergo extensive chromatin compaction as evidenced by electron micrography. However, a study of the chromatin remodeling and genome organization at high resolution during this process is still lacking.

Through the implementation of microscopy assays and experimental methodologies coupled to DNA sequencing we have studied the changes in gene expression, the nuclear architecture, chromatin accessibility, and genome topology that take place during avian terminal erythroid differentiation. Our results revealed a complex remodeling of genome organization. Chromatin accessibility profiling showed a unique positioning of the accessible chromatin inside the nucleus of adult red blood cells. Open chromatin in the erythrocytes comprises promoters of silent genes that retain the RNA polymerase II (RNA Pol II) in a paused state. We performed RNA-seq experiments and identified a set of genes that remain active in the terminally differentiated erythroid cells, some of them linked to the control of the RNA Pol II pausing.

Finally, we studied the organization of the erythroid genome with chromosome conformation capture technologies (Hi-C) and found it to be highly compartmentalized and practically devoid of topologically associated domains except for regions where the paused RNA Pol II contributes to chromatin folding into mini domains. Our results suggest that promoter-proximal pausing of the RNA Pol II participates in the transcriptional regulation of the erythroid genome and highlight the role of RNA Pol II in the maintenance of local chromatin organization.

I. Principios de la organización tridimensional del genoma

El biólogo del desarrollo Conrad Waddington acuñó, en 1953, el término *epigenética* para referirse a los mecanismos, desconocidos hasta entonces, que permiten a células con un mismo genotipo adquirir diferentes fenotipos durante el desarrollo embrionario¹. A 70 años de su concepción, el campo de la epigenética ha trascendido más allá de la embriología y ha evolucionado en respuesta a grandes avances en nuestro entendimiento sobre los mecanismos moleculares que gobiernan la expresión genética en eucariontes.

En la actualidad, el término epigenética comprende aquellos procesos moleculares que conllevan a cambios en los perfiles transcripcionales de la célula sin alterar la secuencia de ADN², que actúan regulando la manera en la que el ADN se empaqueta en la fibra de cromatina, la estructura de la misma y la organización espacial de la fibra dentro del núcleo celular. Estos procesos operan no sólo durante la diferenciación celular puesto que también proveen a la célula de la capacidad para responder ante diferentes estímulos ambientales y participan en la ejecución de complejos programas transcripcionales.

Estructura de la cromatina

El material genético de las células eucariontes se encuentra empaquetado en una estructura nucleoproteica denominada cromatina, formada por la unión del ADN con un grupo de proteínas ricas en aminoácidos con cargas positivas llamadas histonas. La unidad fundamental de la cromatina es el nucleosoma, compuesto por un octámero con dos de cada una de las histonas canónicas H2A, H2B, H3 y H4 alrededor del cual se enrollan cerca de 147 pares de bases de ADN³. Un segmento de ADN inter-nucleosomal de longitud variable conecta a un nucleosoma con el siguiente, lo que constituye la estructura primaria de la cromatina conocida como fibra de 10 nm. La incorporación de la histona H1 al segmento inter-nucleosomal

estabiliza el plegamiento de la fibra de 10 nm en una fibra de 30 nm, misma que puede continuar plegándose en estructuras secundarias subsecuentes⁴. A medida que la fibra de 30 nm se estructura en fibras de mayor tamaño la cromatina se condensa y reduce su volumen hasta alcanzar el grado máximo de compactación encontrado en los cromosomas metafásicos que presentan fibras de hasta 1400 nm de grosor^{5,6}.

La cromatina representa el sustrato de la maquinaria responsable de la transcripción génica, el metabolismo y la organización del ADN, por lo que su grado de compactación tiene implicaciones importantes en varios procesos celulares. Inicialmente la cromatina se clasificó en eucromatina y heterocromatina en función de la intensidad con la que los diferentes tipos de cromatina se tiñen en preparaciones con tinciones de ADN⁷, lo que refleja el estado de compactación de la cromatina. La heterocromatina se tiñe más fuertemente como cuando su estado es altamente compactado, lo que corresponde en gran medida a un ambiente cromosomal represivo que limita el acceso al ADN por parte de la maquinaria transcripcional y replicativa⁸. La eucromatina, por su parte, representa regiones de ADN en donde la cromatina adopta una estructura laxa y accesible que corresponde a las regiones transcripcionalmente activas y elementos que regulan la expresión génica como promotores y enhancers activos⁵.

El establecimiento de diferencias en la accesibilidad cromatínica entre regiones del ADN constituye un proceso orquestado por mecanismos epigenéticos. La región amino-terminal de las histonas sobresale de la estructura del nucleosoma y es sujeta a numerosas modificaciones post-traduccionales (MPTs) que cambian la estabilidad de los nucleosomas y la estructura de la cromatina al alterar las interacciones entre las histonas y el ADN. Modificaciones como la acetilación de aminoácidos específicos ubicados en dicha región neutraliza la carga neta de las histonas y disminuye su afinidad por el ADN, lo que resulta en una estructura más accesible. De manera inversa, la desacetilación de las histonas promueve la compactación de la cromatina⁹. Además, existen factores proteicos que reconocen

y median el efecto de perfiles específicos de MPTs de manera que asisten en el reposicionamiento nucleosomal y el reclutamiento de complejos remodeladores de la cromatina¹⁰.

El desarrollo de técnicas de inmunoprecipitación de la cromatina (ChIP) ha hecho posible correlacionar diferentes marcas epigenéticas con el establecimiento y mantenimiento de los distintos tipos de cromatina así como identificar los elementos genómicos asociados con las mismas. De manera general, se ha encontrado que la eucromatina es rica en marcas de acetilación de histonas y la coocurrencia de la acetilación del residuo de lisina 27 de la histona H3 (abreviado como H3K27ac) junto con la metilación de la lisina 4 (H3K4me) se utiliza comúnmente para identificar elementos enhancers activos, mientras que las marcas de H3K27ac y H3K4me₃ se encuentran en promotores de genes transcripcionalmente activos¹¹. En la heterocromatina se encuentra un estado general de hipoacetilación de histonas. La heterocromatina constitutiva, que comprende secuencias repetidas en las regiones teloméricas y centroméricas, es rica en la marca de H3K9me₃ y la presencia de las proteínas HP1 α/β que reconocen dicha MPT^{12,13}. En tanto, la heterocromatina facultativa se asocia con el silenciamiento transcripcional de genes y presenta la di- y tri-metilación de la lisina 27 de la histona H3 (H3K27me_{2/3}) mediada por las proteínas del complejo Polycomb¹⁴.

Los cambios en la accesibilidad de la cromatina mediados por la deposición de perfiles particulares de MPT de histonas, así como el reclutamiento de factores remodeladores de la cromatina que reconocen dichos perfiles asisten en la delimitación de elementos reguladores, como enhancers y promotores. Sin embargo, a menudo los elementos enhancer se localizan a distancias genómicas considerables de su gen blanco y su actividad no puede explicarse meramente por cambios locales en la accesibilidad de la cromatina, sino que requiere de la proximidad física entre ambas secuencias. En este contexto nació el campo de estudio de la topología del genoma el cual aborda la manera en la que la organización espacial de la fibra de cromatina al interior del núcleo y el

establecimiento de interacciones físicas entre secuencias genómicas participan en la regulación de diferentes procesos metabólicos del ADN y la precisa regulación espaciotemporal de la actividad génica.

Técnicas para el estudio de la topología genómica

El estudio de la organización de la cromatina dentro del núcleo en interfase ha sido posible gracias a dos ejes principales: el desarrollo de técnicas de microscopía diseñadas para evaluar la proximidad espacial entre secuencias de ADN, particularmente la hibridación fluorescente *in situ* (FISH), y el advenimiento de técnicas moleculares como la captura conformacional de cromosomas (3C) y tecnologías derivadas de esta.

En el experimento de FISH se utilizan sondas de ADN marcadas fluorescentemente que hibridan con las secuencias de interés, de manera que se puede obtener información acerca de la posición espacial de secuencias específicas e incluso de cromosomas completos¹⁵. Este tipo de ensayos ha aportado importantes observaciones sobre la organización de los cromosomas en territorios discretos que promueven la interacción entre secuencias ubicadas en el mismo cromosoma (es decir, en *cis*-) a la vez que la frecuencia de interacciones entre cromosomas diferentes (en *trans*-) es baja¹⁶. Asimismo, el marcaje fluorescente de ADN ha evidenciado la formación de asas de cromatina entre promotores y enhancers que coinciden con el momento de activación transcripcional de los genes involucrados en estos contactos^{17,18}, lo que provee de un modelo que explica el mecanismo de acción de los enhancers localizados a grandes distancias genómicas y apoya la noción de la organización del genoma como un nivel nuevo de regulación transcripcional.

Datos obtenidos por microscopía óptica y electrónica indican la segregación de la eucromatina y heterocromatina en diferentes compartimentos nucleares restringidos

a zonas particulares del núcleo. Por ejemplo, la heterocromatina suele encontrarse en la periferia nuclear o asociada a la zona perinucleolar y el confinamiento de un gen a estas regiones coincide con su silenciamiento transcripcional¹⁹⁻²¹, mientras que la eucromatina se localiza al centro del núcleo²¹. El uso de métodos de imagenología ha sido esencial para ampliar nuestro entendimiento de la estructura y plegamiento de la cromatina; sin embargo, el número de regiones que pueden ser evaluadas en un mismo experimento es limitado y las secuencias a evaluar deben ser definidas *a priori*, lo que limita la capacidad de estas técnicas de realizar observaciones genómicas.

Las técnicas de captura conformacional, por su parte, constituyen un grupo de metodologías bioquímicas utilizadas para revelar en alta resolución las interacciones entre segmentos de ADN y calcular la frecuencia en la que ocurren estos contactos en una población celular²². Las técnicas de captura explotan el hecho de que las interacciones y asas de cromatina están mediadas por proteínas, de modo que utilizan un primer paso de entrecruzamiento con formaldehído para fijar todas las interacciones proteína-proteína y proteína-ADN y así preservar la organización del genoma. Posteriormente, se digiere el genoma completo con una enzima de restricción que genera fragmentos de ADN con extremos cohesivos que sirven de sustrato para una reacción de ligación en condiciones subóptimas. Este último paso favorece la ligación de fragmentos que se encontraban cercanos espacialmente al momento del entrecruzamiento, mientras que las ligaciones entre fragmentos a mayores distancias son poco comunes. De esta manera, la ligación entre dos secuencias distantes en el genoma lineal da lugar a fragmentos de ADN híbridos cuyo origen sólo es posible si dichos fragmentos se encontraban formando un asa de cromatina dentro del núcleo²³.

Las tecnologías de captura surgieron con el desarrollo de la técnica de 3C, en la que la interacción entre dos secuencias se evidencia mediante la amplificación específica del fragmento híbrido resultante luego de la ligación, con el uso de oligonucleótidos específicos para las secuencias de interés²³. Es importante

recalcar que aunque la técnica de 3C únicamente evalúa la interacción entre un par de secuencias, el experimento genera millones de fragmentos que corresponden a las interacciones de todo el genoma. Por lo que posteriores modificaciones a la técnica de 3C dieron lugar a nuevos métodos experimentales capaces de revelar un número mayor de interacciones mediante el uso de secuenciación masiva de ADN²². Dentro de estas técnicas destaca la de Hi-C por su capacidad de recuperar de manera extensiva las interacciones de todo el genoma sin necesidad de definir previamente las secuencias de interés.

A diferencia del 3C, en un protocolo de Hi-C los extremos cohesivos dejados por la digestión con la enzima de restricción son posteriormente reparados y marcados con un nucleótido biotinilado y los extremos romos resultantes son ligados. De este modo, todas las moléculas que han sido exitosamente digeridas, reparadas y ligadas retienen al nucleótido biotinilado y son recuperadas mediante un *pull-down* con partículas paramagnéticas acopladas a estreptavidina (Fig. 1A). El experimento concluye con la preparación de una biblioteca que contiene a todos los fragmentos recuperados para su secuenciación pareada, de manera que el Hi-C permite evaluar en el mismo experimento la presencia de millones de fragmentos informativos que exploran la conformación de todo el genoma²⁴.

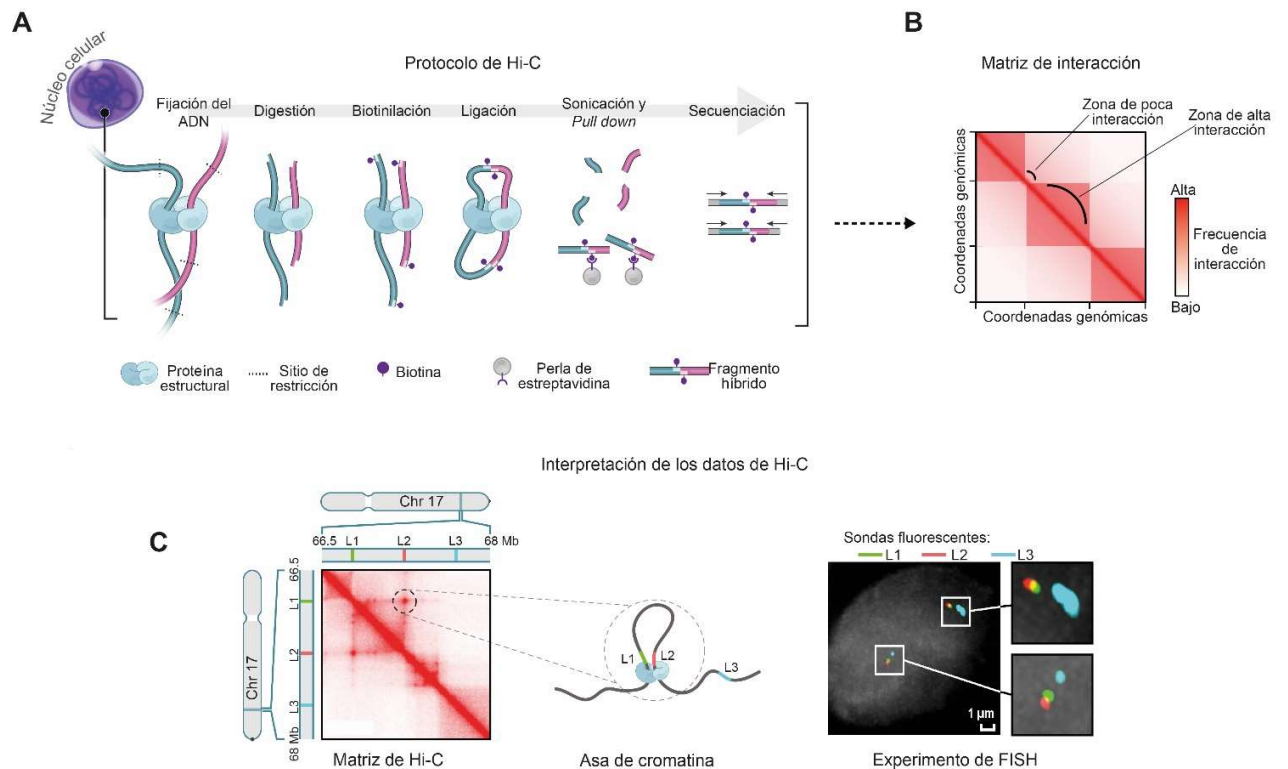


Figura 1. Estudio de la organización del genoma con la técnica de Hi-C. **A**, Protocolo del experimento de Hi-C. Las interacciones del genoma son fijadas y digeridas con una enzima de restricción. Posteriormente, se reparan los extremos dejados por la enzima utilizando un nucleótido biotinilado seguido de un paso de ligación en el que se generan fragmentos de ADN híbridos, que reflejan la cercanía espacial entre dos secuencias. **B**, Luego de la secuenciación de los fragmentos híbridos, los resultados de Hi-C se presentan como matrices de interacción que muestran, a través de un mapa de calor, la frecuencia en la que dos secuencias interactúan entre sí. **C**, Ejemplo de datos de Hi-C en células de humano que muestran la interacción específica entre dos regiones del cromosoma 17 (L1 y L2), mientras que la región denominada L3 del mismo cromosoma no participa en dicha interacción. La presencia de esta asa de cromatina se confirmó con experimentos de FISH²⁵.

El análisis de los resultados de Hi-C es complejo y requiere del uso de herramientas bioinformáticas capaces de alinear cada par de interacciones al genoma de referencia y calcular la frecuencia con la que cada interacción es detectada. La forma más común para visualizar esta información es construir una matriz de interacción como la que se muestra en la Fig. 1B.

Este tipo de matrices están divididas en celdas de tamaño constante, definidos por la resolución de la matriz. En el ejemplo presentado en la Fig. 1C se muestra una región que comprende de la megabase 66.5 a la 68 del cromosoma 17 de humano y la matriz está construida a una resolución de 10 Kb, de modo que cada celda en la matriz representa 10 Kb del cromosoma²⁵. Las matrices de Hi-C son simétricas y la diagonal que las atraviesa recapitula la secuencia lineal del cromosoma, mientras que a medida que nos alejamos de la diagonal podemos encontrar todas aquellas interacciones entre secuencias distantes en el genoma²⁴. Finalmente, la frecuencia de la interacción entre dos regiones es calculada con base en la abundancia de fragmentos híbridos secuenciados que soportan dicha interacción y es representada mediante un mapa de calor²⁴. Dada la naturaleza simétrica de la matriz, es común presentar únicamente la mitad de la misma.

El ejemplo de la Fig. 1C muestra una interacción detectada entre dos regiones ubicadas a 400 Kb de distancia entre ellas, denominadas L1 (Chr17: 66.8 Mb) y L2 (Chr17: 67.2 Mb), la interacción se aprecia en la matriz como una zona de color rojo en la intersección entre ambas coordenadas genómicas que denota una alta frecuencia de interacción y es interpretada como evidencia de un asa de cromatina entre L1 y L2²⁵. Por otra parte, la secuencia indicada como L3 se encuentra a la misma distancia genómica de L2 pero la intersección entre ambas coordenadas en la matriz refleja valores de interacción bajos²⁵.

En la última década el análisis de datos de Hi-C ha mostrado que el genoma se organiza en distintas estructuras, identificadas por patrones específicos observables en las matrices, a las que se les atribuye una alta capacidad regulatoria en el control transcripcional de las células cuya perturbación ha sido observada en defectos durante el desarrollo y procesos patológicos como el cáncer^{26,27}.

Niveles de organización tridimensional del genoma

Asas de cromatina

Se reconocen distintos niveles jerárquicos en la organización de la cromatina, el primero de ellos corresponde a la formación de asas o *loops* de cromatina que permiten el contacto de regiones de ADN distantes en la secuencia lineal del cromosoma²⁴. Las asas de cromatina mejor caracterizadas son aquellas entre enhancers y promotores, en cuyo caso la formación de la misma correlaciona con la activación transcripcional del gen blanco^{28,29}. De igual manera, existen algunos ejemplos de asas que conducen al silenciamiento transcripcional y cuya alteración resulta en la activación ectópica de los genes involucrados en dicho loop^{30,31}.

Además de las asas enhancer-promotor, existen ejemplos de asas formadas entre los promotores de genes que comparten el mismo estado transcripcional y coinciden en la regulación espaciotemporal de su expresión^{32,33}. Asimismo, se han descrito redes de interacciones entre promotores de genes silenciados por el complejo Polycomb³⁴, lo que sugiere que este tipo de estructuras podrían proveer de un andamiaje que asiste en la co-regulación transcripcional de múltiples genes.

Las asas de cromatina se identifican en las matrices de Hi-C como puntos que reflejan la alta frecuencia de interacción entre las secuencias de ADN ubicadas en las bases del loop (Fig. 2A). Numerosos reportes han mostrado que las bases de los loops en vertebrados suelen coincidir con sitios ocupados por proteínas estructurales, particularmente el factor de unión a CCCTC (CTCF) y el complejo de cohesina^{25,35,36}, por lo que se ha propuesto que las asas se forman gracias a la actividad conjunta de estas proteínas a través de un mecanismo conocido como *loop extrusion*, que será abordado en la siguiente sección.

Dominios topológicamente asociados

Una de las características más prominentes de las matrices de Hi-C es la presencia de estructuras conocidas como dominios topológicamente asociados (TADs), que se definen como regiones de ADN que albergan un gran número de interacciones entre secuencias dentro del mismo dominio, al mismo tiempo que se mantienen relativamente aisladas de las secuencias estructuradas en dominios adyacentes³⁷. En las matrices de Hi-C los TADs se aprecian como regiones continuas ricas en interacciones a lo largo de la diagonal, separadas por regiones de baja interacción que aíslan a un dominio del otro (Fig. 2B).

Los TADs están demarcados por regiones genómicas con capacidades aislantes, denominadas fronteras, que impiden el contacto entre secuencias en diferentes dominios. Al igual que las bases de las asas de cromatina, las fronteras de los TADs están ocupadas por CTCF y cohesina^{25,37}. Además, un gran número de fronteras están ocupadas por la RNA Pol II, coinciden con promotores de genes constitutivos y contienen marcas de histonas de eucromatina³⁷. Los TADs representan unidades estructurales de gran importancia en el genoma de los vertebrados. Por ejemplo, más del 90% del genoma de ratón se encuentra estructurado en dominios, que van de tamaños desde 40 Kb hasta 3 Mb, y suelen estar conservados entre tipos celulares e incluso entre especies^{25,38,39}.

La importancia de los TADs radica en su capacidad de proveer de una estructura genómica que facilita la interacción entre elementos reguladores a la vez que limitan el contacto entre otros. De modo que es más probable que un elemento regulatorio contacte a su promotor blanco si ambos elementos se encuentran dentro del mismo dominio, mientras que la confinación de estos elementos a diferentes TADs reduce la probabilidad de contacto entre ellos^{40,41}. De manera consistente con esto, mutaciones o variaciones estructurales que perturban a las fronteras entre TADs conducen a la formación de interacciones ectópicas y la activación aberrante de

genes, mismas que son frecuentemente observadas en procesos patológicos como el cáncer y defectos durante el desarrollo^{41,42}.

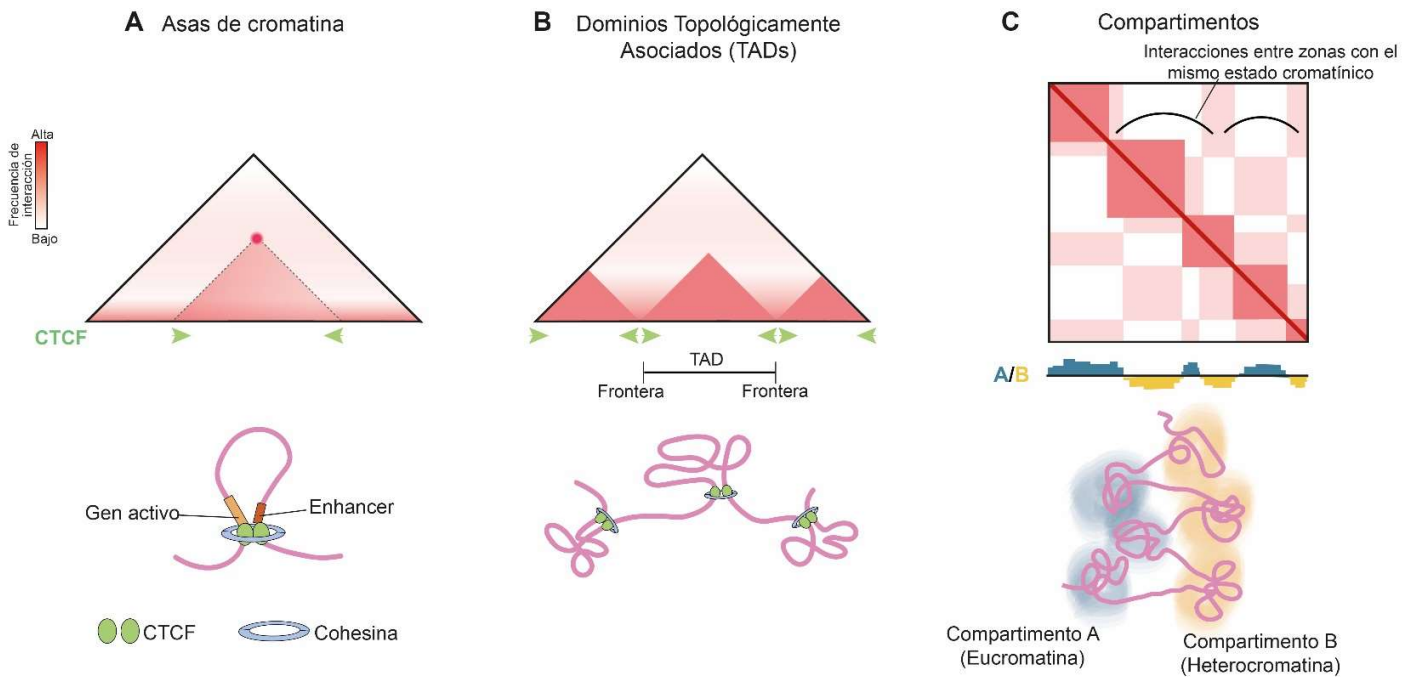


Figura 2. Niveles de organización tridimensional del genoma. Tipos de estructuras identificadas en las matrices de Hi-C (arriba) y el modelo del nivel de organización que representan (abajo). **A**, Representación de un asa de cromatina entre dos elementos ubicados a unas cuantas Kb de distancia. Las asas de cromatina se identifican por el patrón observado en la matriz y suelen tener a CTCF y el complejo de cohesina en la base. **B**, Los TADs se aprecian en los datos de Hi-C como triángulos continuos, cerca de la diagonal de matriz. Cada triángulo representa un TAD separado de los otros dominios por las regiones denominadas fronteras, ricas en CTCF y cohesina. **C**, A gran escala, las regiones del genoma que comparten el mismo estado de la cromatina se contactan espacialmente en un compartimento eucromatínico (Compartimento A, azul) y otro rico en heterocromatina (Compartimento B, amarillo). Este tipo de organización resulta en un patrón de tablero de ajedrez en las matrices de Hi-C.

Compartimentos nucleares

Como se mencionó en la sección anterior, dentro del núcleo ocurre la segregación espacial de la eucromatina y heterocromatina en diferentes zonas, a las que conocemos como compartimentos nucleares. La heterocromatina tiende a situarse en las regiones cercanas al nucléolo o hacia la periferia nuclear, de tal modo que los genes ubicados en la periferia nuclear se encuentran reprimidos y requieren ser relocalizados hacia zonas de eucromatina para poder ser expresados⁴³⁻⁴⁵. Las regiones ubicadas en la periferia nuclear son denominadas Dominios Asociados a la Lámina nuclear (LADs) y suelen ser zonas de baja densidad génica, con niveles de expresión bajos, ricas en la marca de histonas represiva H3K9me3 y la proteína de heterocromatina HP1^{46,47}. En células de mamífero se ha estimado que los LADs se encuentran en todos los cromosomas y pueden constituir hasta el 30% de todo el genoma⁴⁸, por lo que constituyen un compartimento heterocromático considerable.

La formación de los LADs depende de la actividad de proteínas puente, como HP1 α en mamíferos, capaces de interactuar con la cromatina marcada con H3K9me2/3 y con algunos componentes estructurales de la envoltura nuclear, particularmente las láminas A/C y B, así como de la participación del receptor de la lámina B (LBR)⁴⁹. La eliminación de los genes que codifican para la lámina A/C o el LBR resulta en la pérdida de los LADs, la relocalización aberrante de la heterocromatina hacia el centro del núcleo y en algunos casos en la activación transcripcional ectópica de genes previamente silenciados^{50,51}. Por otra parte, mutaciones en los genes que codifican para las láminas nucleares o enzimas involucradas en su procesamiento y maduración post-traduccionales dan lugar a un grupo de patologías denominadas laminopatías caracterizadas por defectos en el establecimiento de los LADs y en la expresión génica⁵².

Una excepción notable a la arquitectura nuclear clásica se encuentra en los bastones de la retina de mamíferos nocturnos, en donde la pérdida de LBR y las

láminas A/C ocurre de manera regulada durante el desarrollo y resulta en la inversión del patrón heterocromatina-eucromatina antes descrito^{50,53}. El reposicionamiento atípico de la heterocromatina hacia el centro del núcleo en este tipo de fotorreceptores responde a la necesidad fisiológica de optimizar la transmisión y captura de la luz y demuestra el estrecho vínculo entre la arquitectura nuclear y la función celular.

La formación de compartimentos promueve la interacción entre secuencias que comparten el mismo estado de compactación a la vez que evita las interacciones entre regiones de heterocromatina y eucromatina⁵⁴. Aunque los mecanismos moleculares detrás de la formación de compartimentos nucleares no han sido elucidados por completo, evidencia reciente sugiere que algunos de los componentes de la cromatina son capaces de inducir separación de fases líquido-líquido (LLPS)⁵⁵⁻⁵⁷. La LLPS resulta en la formación de compartimentos supramoleculares, similares a gotas, estabilizados por fuerzas hidrofóbicas débiles que limitan la difusión de proteínas hacia el interior del dominio⁵⁷. De esta manera los compartimentos representan dominios ricos en eucromatina o heterocromatina inmersos en una fase líquida más diluida (el nucleoplasma), de manera similar a gotas de aceite en agua⁵⁵.

La organización de la cromatina en compartimentos se observa en las matrices de Hi-C como un patrón característico en forma de tablero de ajedrez que refleja cómo regiones del mismo estado de compactación que se encuentran lejanas en el genoma lineal se contactan entre sí²⁴ (Fig. 2C). Con los datos de Hi-C pueden distinguirse dos grandes compartimentos a nivel de todo el genoma, el compartimento A que correlaciona con las regiones eucromáticas, ricas en genes activos y marcas de cromatina abierta; y el compartimento B que correlaciona con las marcas de heterocromatina y poca actividad transcripcional^{24,25}.

II. Mecanismos de organización del genoma en vertebrados

Desde el descubrimiento de las asas de cromatina, diferentes esfuerzos han tratado de elucidar los mecanismos implicados en la estructuración y regulación de la topología genómica. La caracterización de los loops y TADs, así como de las proteínas frecuentemente asociadas a las bases de dichas estructuras, han llevado a la postulación de un mecanismo por el cual ciertas proteínas estructurales asisten al plegamiento de la cromatina.

Proteínas estructurales y el modelo de extrusión de cromatina

Se conocen pocas proteínas que cumplen la función de organizar el genoma en vertebrados, la mejor estudiada de ellas es CTCF, una proteína expresada de manera constitutiva originalmente identificada como un represor transcripcional y una proteína insulator⁵⁸. La proteína CTCF posee una región central con 11 dedos de zinc, altamente conservada dentro del clado de los animales bilaterios⁵⁹, que le permiten reconocer una secuencia blanco en el genoma y unirse al ADN⁶⁰. Existen más de 60,000 sitios de unión a CTCF en el genoma humano predominantemente ubicados en regiones intergénicas y elementos reguladores, más del 60% de estos sitios se comparten entre distintos tipos celulares mientras que otros se encuentran únicamente ocupados en contextos celulares específicos⁶¹.

La integración de datos de captura y ChIP-seq ha demostrado que las bases de muchas asas de cromatina cuentan con motivos de unión a CTCF en orientación convergente ocupados por la proteína^{25,36} (Fig. 3), mientras que la eliminación o la inversión de la orientación de un sitio de CTCF ubicado en la base de un loop resulta en la disolución de dicha interacción³⁶. De igual manera, el abatimiento temporal de los niveles proteicos CTCF en células troncales embrionarias de ratón conduce a la pérdida de más del 80% de los TADs⁶². En conjunto estas evidencias demuestran

que CTCF juega un papel preponderante en la formación de asas de cromatina y TADs, aunque es probable que existan otras proteínas capaces de cumplir con un rol similar. Por ejemplo, recientemente se ha demostrado que la proteína Ying Yang 1 (YY1) también puede formar asas de cromatina entre *enhancers* y promotores de manera similar a CTCF⁶³.

CTCF generalmente se encuentra acompañado por el complejo de cohesina^{64,65}, conformado por las proteínas del Mantenimiento Estructural de Cromosomas (SMC, por sus siglas en inglés) SMC1 y SMC3, y Rad21⁶⁶. El complejo de cohesina posee la forma de un anillo a través del cual el ADN es extrudido, gracias a la actividad ATPasa de las subunidades SMC1 y SMC3⁶⁶. Los niveles de cohesina unidos al ADN están regulados por la actividad de la proteína NIPBL, encargada de montar el complejo al ADN, y de la proteína WAPL que remueve al complejo^{67,68}. El abatimiento de la subunidad Rad21 o de NIPBL impiden la unión de la cohesina con el ADN y generan la pérdida de las asas de cromatina y TADs en todo el genoma^{69,70}, lo que resalta su importancia en la organización del genoma.

Modelo de extrusión de cromatina
(Loop extrusion model)

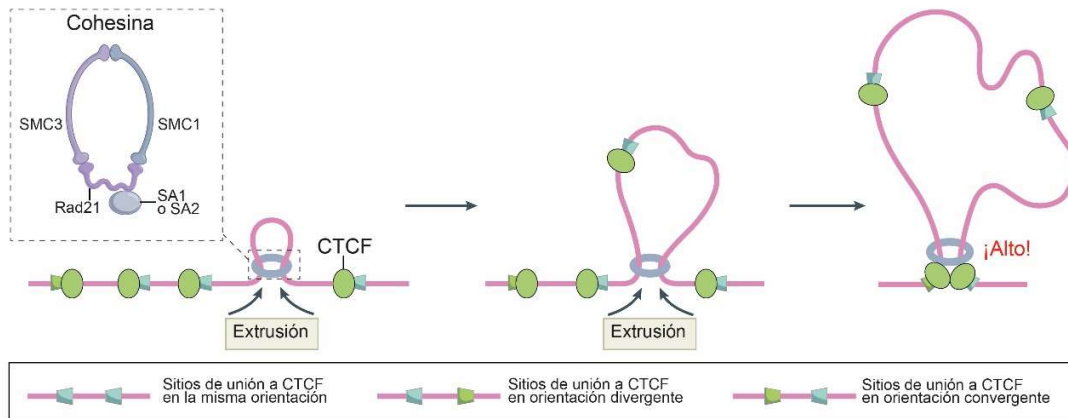


Figura 3. Modelo para la formación de asas de cromatina. De acuerdo con el modelo de extrusión, el complejo de cohesina (formado por las proteínas SMC1, SMC3 y Rad21) es cargado sobre la cromatina de manera que la fibra de cromatina pasa a través del anillo del complejo, dando lugar a un asa de cromatina (*izquierda*). Gracias a su actividad ATPasa, la cohesina es capaz de recorrer por la fibra de cromatina y hacer crecer el asa (*centro*). Finalmente, el asa de cromatina se estabiliza cuando la cohesina interactúa con alguna proteína capaz de frenar su movimiento. En el caso de CTCF, la estabilización del asa ocurre cuando la cohesina encuentra dos motivos de unión a CTCF orientados de manera convergente⁷¹ (*derecha*).

De acuerdo con el modelo de extrusión de cromatina, también conocido como *loop extrusion model*, las asas de cromatina se forman cuando el ADN es extrudido a través del anillo de cohesina. El movimiento de la cohesina es detenido cuando ésta se encuentra con alguna proteína capaz de frenarla, tal como CTCF (Fig. 3), lo que explica la presencia de CTCF en la base de los loops. En concordancia con este modelo, se ha descrito que la porción N-terminal de CTCF interactúa con la cohesina y es capaz de estabilizar su unión a la cromatina⁷².

El papel de la transcripción en la topología del genoma

La organización del genoma y el control de la transcripción se encuentran íntimamente relacionadas. Sin embargo, la forma en la que la maquinaria transcripcional influye o promueve la formación de asas de cromatina o TADs continúa siendo debatida.

Muchas de las fronteras de los TADs coinciden con zonas de alta actividad transcripcional y la RNA Pol II se encuentra enriquecida en la base de loops y TADs^{37,73}. Además, estudios dirigidos a evaluar la dinámica de la organización del genoma durante el desarrollo temprano en *Drosophila* y ratón han mostrado que el genoma embrionario carece de TADs durante las primeras etapas del desarrollo en las que encuentra transcripcionalmente inactivo. La aparición de los dominios topológicos coincide con la activación transcripcional del genoma, el reclutamiento masivo de factores transcripcionales y la RNA Pol II a la cromatina^{74,75}, lo que sugiere que la maquinaria transcripcional o el proceso de transcripción en sí mismos pueden propiciar la organización.

Estudios recientes han mostrado que el abatimiento de la RNA Pol II afecta a las asas de cromatina, particularmente aquellas entre elementos regulatorios⁷⁶, aunque la inhibición de la transcripción no parece tener algún efecto topológico en células K562, que corresponden a células eritroleucémicas de humano, ni en mESC⁷⁷. De modo que la unión de la RNA Pol II a la cromatina podría ser suficiente para promover la formación de asas y TADs, ya sea de manera directa o reclutando a la cohesina^{76,77}. En este sentido, resulta interesante explorar la organización del genoma en contextos celulares en donde la actividad transcripcional se encuentra alterada o disminuida.

III. Antecedentes

Los eritrocitos son células del torrente sanguíneo altamente especializadas en el transporte de oxígeno desde los pulmones a los tejidos, que son generadas a partir de progenitores hematopoyéticos a través del proceso de eritropoyesis. Las primeras etapas de la eritropoyesis llevan a la formación de células progenitoras eritroides conocidas como unidades formadoras de colonias eritroides (CFU-E)⁷⁸. La eritropoyesis terminal inicia cuando las CFU-E se diferencian en proeritroblastos, que a su vez atraviesan una serie de cambios morfológicos notables y un dramático silenciamiento transcripcional hasta dar lugar a los eritrocitos maduros⁷⁸.

Cambios en la estructura de la cromatina durante la eritropoyesis terminal en mamíferos

En mamíferos, la eritropoyesis terminal involucra la condensación generalizada de la cromatina, de manera concomitante con una reducción marcada del volumen nuclear y el silenciamiento transcripcional de todo el genoma⁷⁸. El proceso culmina cuando el eritroblasto tardío enuclea y expulsa el material genético condensado, un paso único de los eritrocitos de mamífero, de tal manera que los eritrocitos maduros son células carentes de núcleo^{79,80} (Fig. 4A).

La remodelación de la cromatina en la eritropoyesis en mamíferos está mediada, al menos en parte, por cambios en las MPTs de las histonas, particularmente en el estado de acetilación de las mismas. La expresión de las acetiltransferasas de histonas Gcn5, Elp3, Hat1 y Tip6 se reduce progresivamente durante la eritropoyesis y la expresión ectópica de Gcn5 evita la condensación de la cromatina en eritroblastos de ratón⁸¹. A su vez, las deacetilasas de histona (HDACs) 1, 2, 3 y 5 se expresan altamente en los blastos y la inhibición farmacológica de la actividad de la HDAC2 bloquea por completo la enucleación^{82,83}. Como consecuencia de esta pérdida de acetilación, el núcleo de los proeritroblastos de humano puede reducir su volumen hasta 10 veces su tamaño original antes de enuclear⁸⁴.

Además de la pérdida de acetilación, hay pérdida de marcas de histona asociadas con la elongación de la transcripción⁸⁵. El eritroblasto más tardío muestra el silenciamiento generalizado de todo el genoma excepto en 474 genes que continúan expresándose, particularmente los genes globina, proteínas de la membrana eritroide y genes involucrados en el transporte de oxígeno⁸⁵, de manera que el silenciamiento extensivo del genoma está mediado tanto por la ganancia de heterocromatina como por la regulación de la actividad de la RNA Pol II.

Algunos reportes han descrito que previo a la enucleación ocurre la migración parcial de las histonas desde el compartimento nuclear hacia el citosol, donde son degradadas⁸⁶. Dicho proceso se ha atribuido a la apertura transitoria de la envoltura nuclear y la remodelación de los componentes de la misma, mediadas por el corte proteolítico de la lámina B por parte de la caspasa 3. La inhibición de la caspasa 3 o la expresión de una versión mutante de la lámina B que no es susceptible al corte en eritroblastos previenen la condensación de la cromatina, la enucleación y detienen el proceso de diferenciación⁸⁷. Finalmente, durante la enucleación el núcleo de los eritroblastos tardíos se dirige hacia uno de los extremos de la célula y ocurre la formación de un anillo contráctil que termina por separar al eritrocito maduro del núcleo⁸⁴.

La reducción del volumen nuclear y compactación de la cromatina son características comunes en los procesos de diferenciación eritroide de todos los vertebrados; sin embargo, los eritrocitos definitivos de los vertebrados no mamíferos retienen el núcleo celular en un estado condensado y aparentemente transcripcionalmente inactivo⁸⁸ (Fig. 4B). Los cambios aquí descritos no han sido explorados en otros vertebrados, lo que impide evaluar la conservación de los mecanismos involucrados en la condensación de la cromatina y sus diferencias con los de mamíferos.

Cambios en la estructura de la cromatina durante la diferenciación eritroide en pollo

El desarrollo embrionario del pollo (*Gallus gallus*) tiene una duración aproximada de 21 días al final de los cuales el embrión eclosiona. Las primeras ondas de diferenciación del linaje eritroide ocurren dentro del saco vitelino dentro de las primeras 36 horas de desarrollo y dan origen al linaje de eritrocitos denominados primitivos, dado que expresan los genes de globinas embrionarias α -Pi, β -Rho y β -Épsilon. A partir del día 5 de desarrollo ocurre un recambio de globinas en el que los eritrocitos primitivos comienzan a expresar las hemoglobinas adultas α -A, α -D y β -A, a este linaje eritroide se le conoce como eritrocitos definitivos⁸⁹. Posteriormente, alrededor de los días 12-15 embrionarios, la médula ósea comienza a fungir como el nicho hematopoyético más importante en el organismo⁹⁰.

Los linajes primitivo y definitivo progresan a través de varios estadios de diferenciación similares a la eritropoyesis en humano y ratón, pasando por el proeritroblasto, eritroblasto y eritroblasto basófilo hasta concluir su diferenciación hacia eritrocitos nucleados⁹⁰. Estas distinciones se han realizado con base en la morfología celular y nuclear, la proporción entre el tamaño nuclear respecto al tamaño celular y la abundancia general de organelos intracelulares.

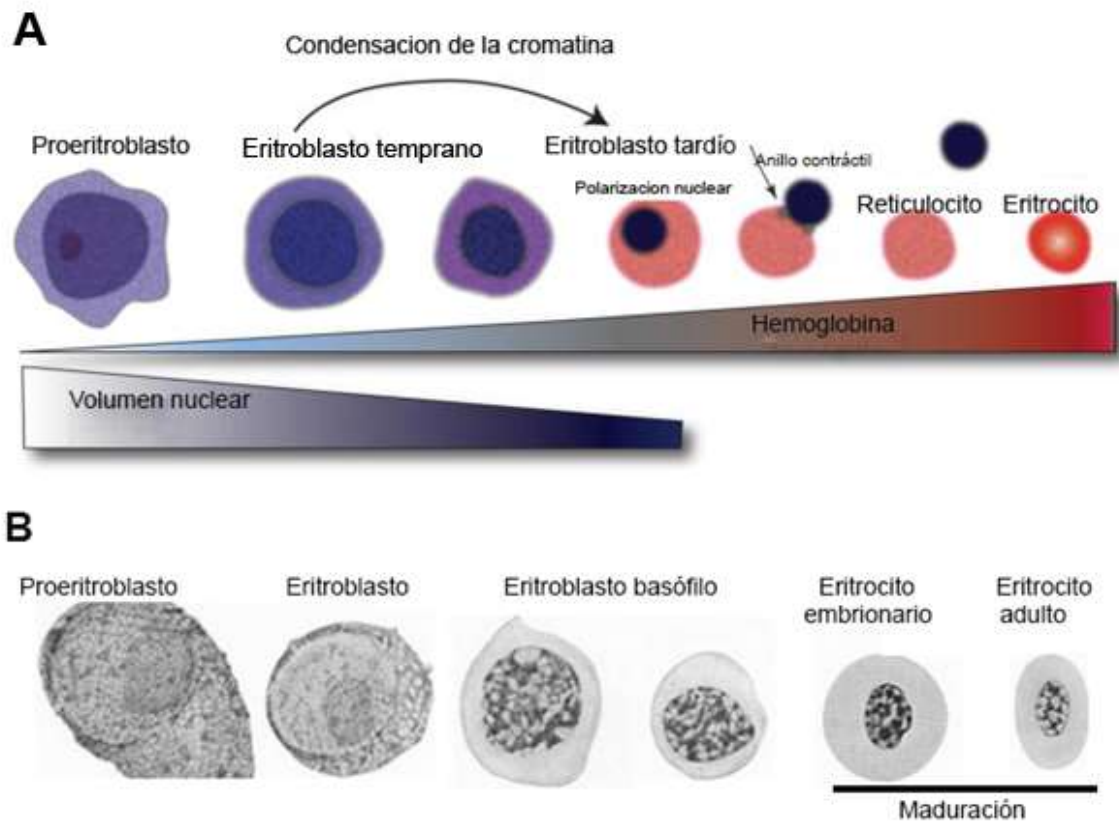


Figura 4. Durante la eritropoyesis ocurren grandes cambios en la estructura de la cromatina. A. En los mamíferos, el volumen nuclear disminuye conforme avanza la eritropoyesis hasta que en el eritroblasto tardío el núcleo se polariza y la célula enuclea⁸⁴. **B.** Micrografías del proceso de eritropoyesis en pollo. En las aves no ocurre la enucleación y los eritrocitos definitivos mantienen el núcleo en un estado altamente compactado, que se asevera en los eritrocitos de organismos adultos⁹¹.

La información disponible sobre los mecanismos que conducen a la remodelación de la cromatina en la eritropoyesis aviar es escasa. En particular, los eritrocitos inmaduros y maduros de ave expresan a la variante de histona H5, una forma alterna de la histona H1 que representa hasta el 60% del contenido total de histonas inter-nucleosomales en los eritrocitos maduros^{92,93}. La histona H5 tiene una mayor proporción de residuos de arginina y lisina que la histona canónica, lo que le confiere mayor afinidad por el ADN, además de que puede formar dímeros con otras H5^{94,95}. Estas características sugieren que juega un papel fundamental en la compactación de la cromatina en la eritropoyesis. De hecho, la sobreexpresión de la H5 es

suficiente para disminuir significativamente la transcripción y la tasa replicativa incluso en células de mamífero⁹⁶.

La extrapolación de datos obtenidos a través de la cristalización de una partícula nucleosomal con la histona H5 presente y del modelado computacional de esta interacción sugieren que la incorporación de la variante H5 aumenta la longitud del ADN protegido por el nucleosoma y permite el plegamiento de la cromatina en una estructura más compacta^{97,98}. Sin embargo, la evidencia experimental de esta observación ha sido contradictoria hasta el momento^{99,100}.

Cambios en la estructura de la cromatina durante la maduración eritroide en pollo

Aunque los eritrocitos definitivos embrionarios (eRBC) son células terminalmente diferenciadas, existen diferencias morfológicas y estructurales entre estos y los eritrocitos circulantes de las aves adultas (aRBC). Por ejemplo, los eRBC tiene un volumen celular mayor que el de los aRBC⁹¹. Las diferencias entre ambos tipos de eritrocitos son atribuidas a un proceso de maduración eritroide, en el que las capacidades biosintéticas de los eritrocitos se ven progresivamente restringidas⁹¹ (Fig. 4B).

Se ha descrito que los eRBC mantienen una actividad transcripcional baja que disminuye a medida que la maduración progresa, a la fecha la evidencia experimental reportada es poca. Se ha medido el contenido total de RNA extraído de eRBC desde el día 3 de desarrollo hasta el día 17, que muestra una caída a partir del día 8 de más del 90% en el total de RNA extraído¹⁰¹. Un ensayo independiente midió la incorporación de uridina marcada radioactivamente al RNA naciente en aRBC y mostró niveles muy bajos de síntesis de RNA y un contenido total de aproximadamente 0.02-0.04 picogramos de RNA por célula¹⁰². Los eritrocitos de

pollo tienen una vida media de 30 días en circulación, a lo largo de los cuáles retienen el núcleo en un estado inactivo⁹⁰.

Por otra parte, existe un reporte de patrones de localización celular diferencial de CTCF entre los eritrocitos terminalmente diferenciados y células de linaje no eritroide, así mismo entre los eritrocitos y eritroblastos (Fig. 5). De acuerdo con este estudio, CTCF detectable por inmunotinción se encuentra ampliamente distribuido al interior del núcleo celular en fibroblastos y eritroblastos de pollo, mientras que en el núcleo de eRBC la distribución de CTCF se reduce a pequeños compartimentos que coinciden con zonas eucromáticas, de acuerdo con la tinción de ácidos nucleicos con DAPI¹⁰³. En los aRBC, los cambios en la localización nuclear de CTCF se acentúan, a la vez que también se aprecia la formación de compartimentos ricos en la RNA Pol II¹⁰³. Dada la importancia de CTCF y la RNA Pol II en la organización del genoma, estos datos sugieren que el proceso de maduración eritroide podría estar acompañado por cambios en la organización del núcleo además de los cambios transcripcionales descritos.

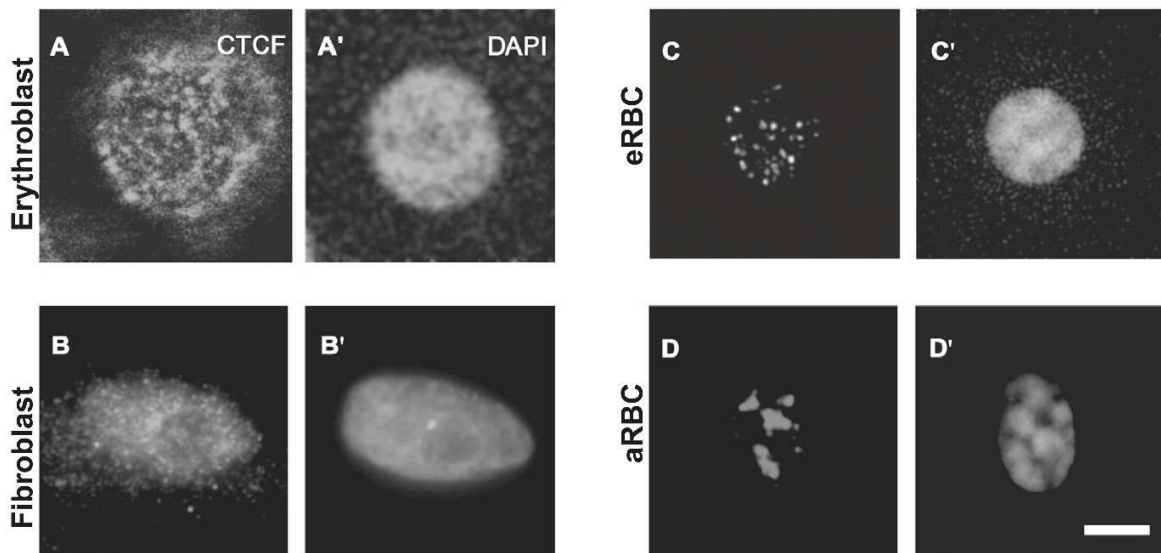


Figura 5. Cambios en la localización de CTCF en el núcleo celular durante la maduración eritroide en pollo. A-A'. En los eritroblastos de pollo, CTCF se localiza uniformemente dentro del núcleo celular. **B-B'.** De igual manera, CTCF se distribuye

homogéneamente dentro del núcleo de fibroblastos. **C-C'**. En los eritrocitos embrionarios (eRBC), CTCF se acumula en compartimentos principalmente localizados cerca de la membrana nuclear. **D-D'**. Por último, en los eritrocitos adultos (aRBC) CTCF se distribuye en compartimentos de gran tamaño ubicados en regiones eucromáticas¹⁰³. En todos los casos se muestra la detección de CTCF por inmunofluorescencia y la tinción nuclear con DAPI.

Aunque se describe a los eritrocitos nucleados como células transcripcionalmente inactivas, los mecanismos epigenéticos detrás del silenciamiento y la extensión del mismo no se han estudiado a profundidad. Además, dada la estrecha relación entre la topología del genoma y la regulación transcripcional, los eritrocitos de pollo constituyen un gran modelo para explorar el efecto del silenciamiento global del genoma sobre la estructuración de asas de cromatina y TADs.

IV. Propuesta de investigación

La cromatina se organiza dentro del núcleo en TADs y compartimentos que contribuyen a la regulación de la expresión génica y a la ejecución de programas transcripcionales específicos. La organización del genoma esta orquestada por mecanismos tales como los perfiles de marcas de histonas, la presencia de proteínas estructurales y la RNA Pol II en las bases de las asas y dominios. Sin embargo, a pesar de la estrecha relación entre la topología del genoma y el control de la transcripción, la fecha se desconoce el estado de la cromatina y cómo esta se reorganiza topológicamente durante un proceso de silenciamiento transcripcional de todo el genoma.

El proceso de diferenciación y maduración eritroide de *Gallus gallus* involucra el silenciamiento global del genoma, la reducción del volumen nuclear y la redistribución de proteínas que estructuran al ADN dentro del núcleo por lo que provee de un modelo único de cambios grandes a nivel transcripcional, estructura primaria de la cromatina y su organización dentro del núcleo durante la diferenciación celular. El estudio de estos niveles de regulación durante la eritropoyesis podría resultar en la identificación de elementos o agentes

moleculares que median dicho proceso a la vez que permiten evaluar el impacto del silenciamiento sobre la cromatina y su organización tridimensional.

Hipótesis

Los eritrocitos en pollo son considerados células transcripcionalmente inactivas que retienen el núcleo celular en un estado altamente compactado. Dada la estrecha relación entre la organización del genoma y la regulación de la expresión génica, la caída en la actividad transcripcional en los eritrocitos estará acompañada de la desestructuración generalizada del genoma y cambios en la arquitectura nuclear. Por otra parte, algunos genes y elementos genómicos que continúen siendo importantes para la fisiología eritroide permanecerán organizados.

Objetivo general

I. Caracterizar el estado transcripcional y los cambios en la accesibilidad y organización tridimensional de la cromatina durante la diferenciación terminal eritroide en *Gallus gallus*, comparando estas características entre eritroblastos tempranos y eritrocitos terminalmente diferenciados.

II. Estudiar el estado transcripcional de los eritrocitos terminalmente diferenciados en pollo y su relación con cambios en la accesibilidad y organización tridimensional de la cromatina durante la maduración eritroide, mediante el análisis de eritrocitos de embriones de pollo de 10 días de desarrollo (eRBC) y de organismos adultos (aRBC).

Objetivos particulares

Los siguientes objetivos plantean caracterizar los eventos ocurridos durante dos eventos fundamentales de la eritropoyesis en aves. La diferenciación terminal, mediante el estudio de eritroblastos y eritrocitos terminalmente diferenciados, y la maduración eritroide que ocurre entre eRBC y aRBC.

I. Evaluar por microscopía los cambios en la localización y abundancia de la eucromatina, marcas de histona asociadas a heterocromatina y la maquinaria transcripcional durante la diferenciación y maduración eritroide en pollo.

II. Explorar los cambios en la composición de la envoltura nuclear durante la diferenciación y la maduración eritroide.

III. Identificar el estado transcripcional de los eritrocitos mediante experimentos de RNA-seq total e identificar a los genes que continúan expresándose en eRBC y aRBC.

IV. Evaluar la capacidad transcripcional de los eritrocitos, el estado de fosforilación de la RNA Pol II y su ocupación en la cromatina.

V. Determinar en alta resolución los cambios en la accesibilidad de la cromatina durante la diferenciación y maduración eritroide. Anotar aquellos sitios genómicos permanezcan abiertos en relación con su estado transcripcional.

VI. Investigar la presencia de los distintos niveles de organización tridimensional del genoma en los eritrocitos y relacionar su presencia y/o ausencia con los datos de accesibilidad y expresión.

V. Resultados

Para estudiar los cambios epigenéticos que tienen lugar durante la eritropoyesis terminal y maduración eritroide en el pollo, se empleó la línea celular HD3, que corresponde a eritroblastos tempranos de pollo, se obtuvieron eRBC procedentes de embriones de 10 días de desarrollo y aRBC de gallinas adultas. De esta manera, la comparación entre eritroblastos, eRBC y aRBC nos proporcionará información sobre cambios en la estructura de la cromatina durante la diferenciación terminal eritroide. Por otra parte, aunque eRBC y aRBC representan eritrocitos terminalmente diferenciados cuya transcripción se encuentra severamente disminuida, los eRBC continúan expresando los genes globina y genes eritroides mientras que en los aRBC el silenciamiento es más severo. De esta manera, la comparación entre eRBC y aRBC nos permitirá estudiar los cambios ocurridos durante la maduración eritroide. Por último, se obtuvieron fibroblastos primarios de pollo como un control que representa a un linaje no eritroide.

La cromatina accesible es reposicionada hacia la periferia nuclear en aRBC

La compactación de la cromatina y reducción del volumen nuclear son características esenciales del proceso de diferenciación eritroide. En mamíferos, esta compactación implica la remodelación de la cromatina y la pérdida de marcas de acetilación asociadas a la eucromatina⁸². En cambio, se desconoce cómo se modifica la accesibilidad de la cromatina durante la diferenciación eritroide en aves y cuáles son los mecanismos epigenéticos que median esta compactación.

Para conocer el panorama general de accesibilidad de la cromatina en los eritrocitos se implementó un ensayo de cromatina accesible a la transposasa acoplado a microscopía (ATAC-see por sus siglas en inglés), el cuál utiliza a la transposasa Tn5

para cortar y etiquetar fluorescentemente los sitios de cromatina accesible en todo el genoma¹⁰⁴. De esta manera se puede evaluar por microscopía el contenido global y localización de la cromatina accesible dentro del espacio nuclear.

En las imágenes de ATAC-seq obtenidas es observable la disminución en la cantidad de cromatina accesible detectada en eRBC y aRBC comparada con los núcleos de eritroblastos (Fig. 6a y Fig. 7a-b), lo que concuerda con el proceso de condensación de la cromatina reportado en estas células. Sin embargo, en ambos tipos de eritrocitos fue detectable la presencia de cromatina accesible, lo que muestra que algunas zonas del genoma eritroide permanecen abiertas aún en este ambiente tan represivo.

De manera notable, en aRBC las regiones que permanecen accesibles parecen localizarse predominantemente en la periferia nuclear (Fig. 6a, flecha), de manera contraria a la arquitectura nuclear convencional observada en eRBC, eritroblastos y fibroblastos. Se identificaron las regiones del núcleo en donde la señal de ATAC-seq o de la tinción de DAPI es más intensa (ver Métodos) y las imágenes resultantes muestran que en fibroblastos, eritroblastos y eRBC las zonas más condensadas (ricas en DAPI) se localizan hacia la periferia del núcleo mientras que la cromatina accesible (rica en ATAC-seq) se posiciona de forma más central (Fig. 6b). En contraste, en aRBC las zonas más ricas en DAPI se encuentran al centro del núcleo, rodeados por la cromatina más accesible (Fig. 7c).

Los datos anteriores sugerirían el reposicionamiento de la cromatina accesible hacia la periferia nuclear en aRBC. De ser este el caso, la heterocromatina constitutiva que generalmente se encuentra asociada a la envoltura nuclear debería cambiar su localización en aRBC. Este fenómeno se corroboró mediante la realización de inmunofluorescencias para detectar la marca de heterocromatina H3K9me3, misma que se encuentra enriquecida en la periferia nuclear en fibroblastos y eritroblastos pero es segregada a *foci* predominantemente posicionados en el centro del núcleo en eRBC y aRBC (Fig., 6c).

En conjunto, estas evidencias muestran que la eritropoyesis terminal resulta en una reorganización de la arquitectura nuclear en los eritrocitos, en donde la heterocromatina es relocalizada hacia el centro nuclear, a la vez que en aRBC la cromatina accesible se posiciona en la periferia.

En aquellas células que muestran la arquitectura nuclear convencional, el anclaje de la heterocromatina en la periferia nuclear es mediado por su interacción con proteínas de la envoltura nuclear, particularmente las láminas nucleares⁵². La pérdida de estas proteínas, en especial de la lámina A/C y el receptor LBR, resulta en la inversión completa del patrón heterocromatina/eucromatina⁵⁰. En concordancia con el reposicionamiento observado, la lámina A/C es detectable por inmunofluorescencia en todos los tipos celulares excepto en aRBC (Fig. 6f), mientras que no se detectaron cambios en la presencia de otra de las láminas nucleares (Fig. 7e).

En conjunto, estos resultados muestran una compleja reorganización de la arquitectura nuclear en los aRBC en la que la pérdida de la lámina A/C coincide con el reposicionamiento de la heterocromatina constitutiva hacia el centro nuclear y la reubicación de la cromatina accesible hacia la periferia (Fig. 6g).

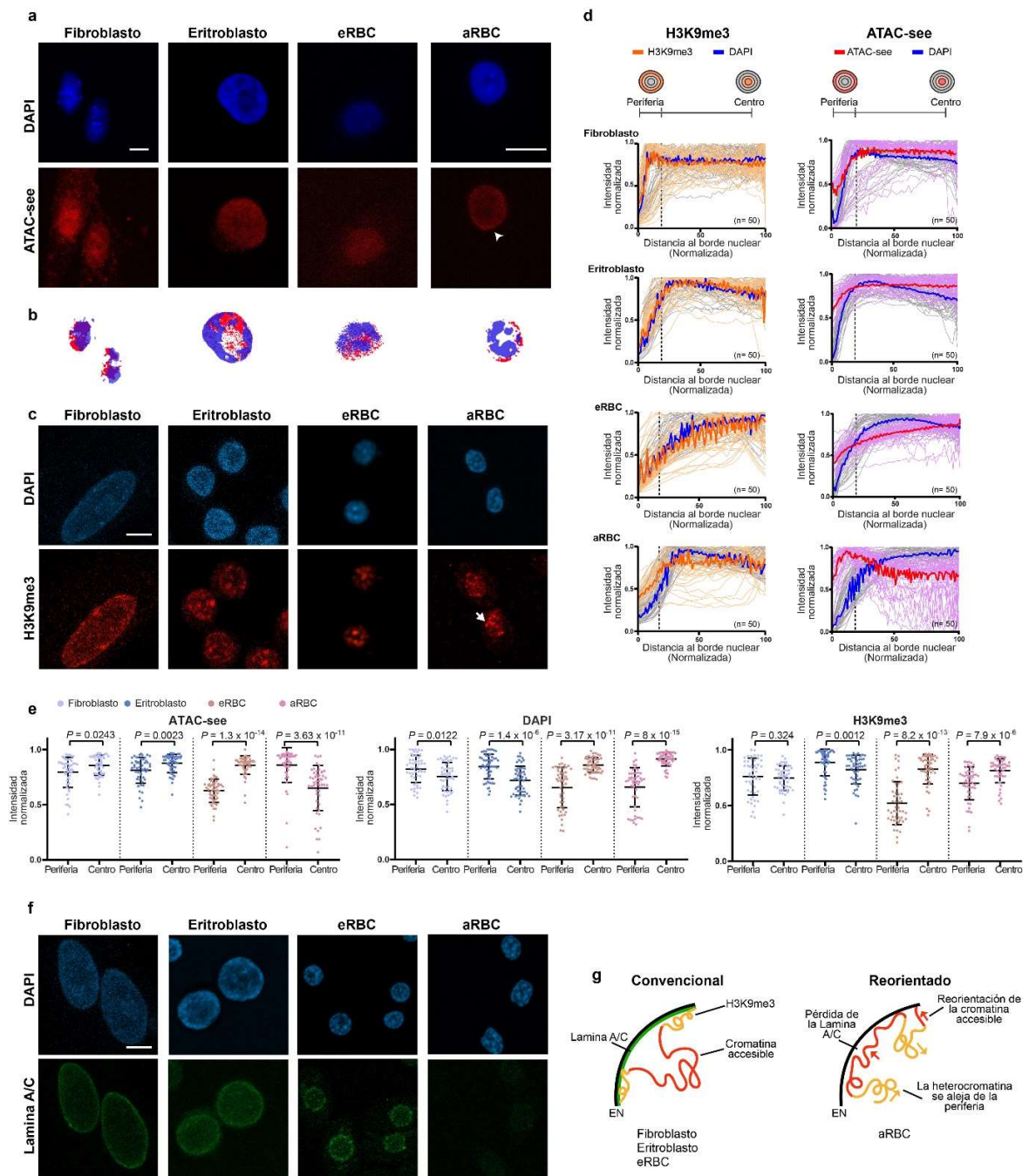


Figura 6. Durante la eritropoyesis terminal la arquitectura nuclear es alterada. a, Detección de la eucromatina accesible con ATAC-see y la señal de DAPI en fibroblastos y células eritroides. En aRBC, la cromatina accesible es reposicionada hacia la periferia nuclear (flecha). En todas las microfotografías de este panel se muestra una barra de calibración = 5 μ m y se presentan resultados representativos de dos réplicas independientes. **b,** Representación de las regiones más intensas de ATAC-see (rojo) y DAPI (azul). **c,** Inmunofluorescencia de la marca de cromatina H3K9me3 y DAPI. **d,** Para

cada tipo celular se presenta la intensidad de la señal de H3K9me3, ATAC-see y DAPI en función de su distancia a la periferia nuclear. Cada gráfica muestra en colores claros trazos individuales de 50 núcleos y el promedio se muestra con un trazo sólido. **e**, Intensidad promedio de ATAC-see, DAPI y H3K9me3 en la periferia o el centro nuclear (n=50 núcleos). Las gráficas muestran el promedio +/- S.D. Se estimaron diferencias significativas con la prueba de U de Mann-Whitney. **f**, Inmunofluorescencia para la lámina A/C. **g**, Modelo de la reorientación de la cromatina accesible hacia la periferia nuclear en aRBC.

De igual manera, se confirmó este patrón graficando la intensidad promedio de ATAC-see, DAPI y H3K9me3 en función de su distancia a la envoltura nuclear (Fig., 6d), la cual muestra que en aRBC la señal de ATAC-see es más intensa en las zonas adyacentes a la periferia nuclear, mientras que la señal de H3K9me3 y DAPI es más intensa en la zona del centro. Análisis de correlación de Pearson sobre estos datos (Fig. 7d) y de la intensidad promedio de cada marca en la periferia vs el centro nuclear (Fig. 6e) apoyan dicha observación.

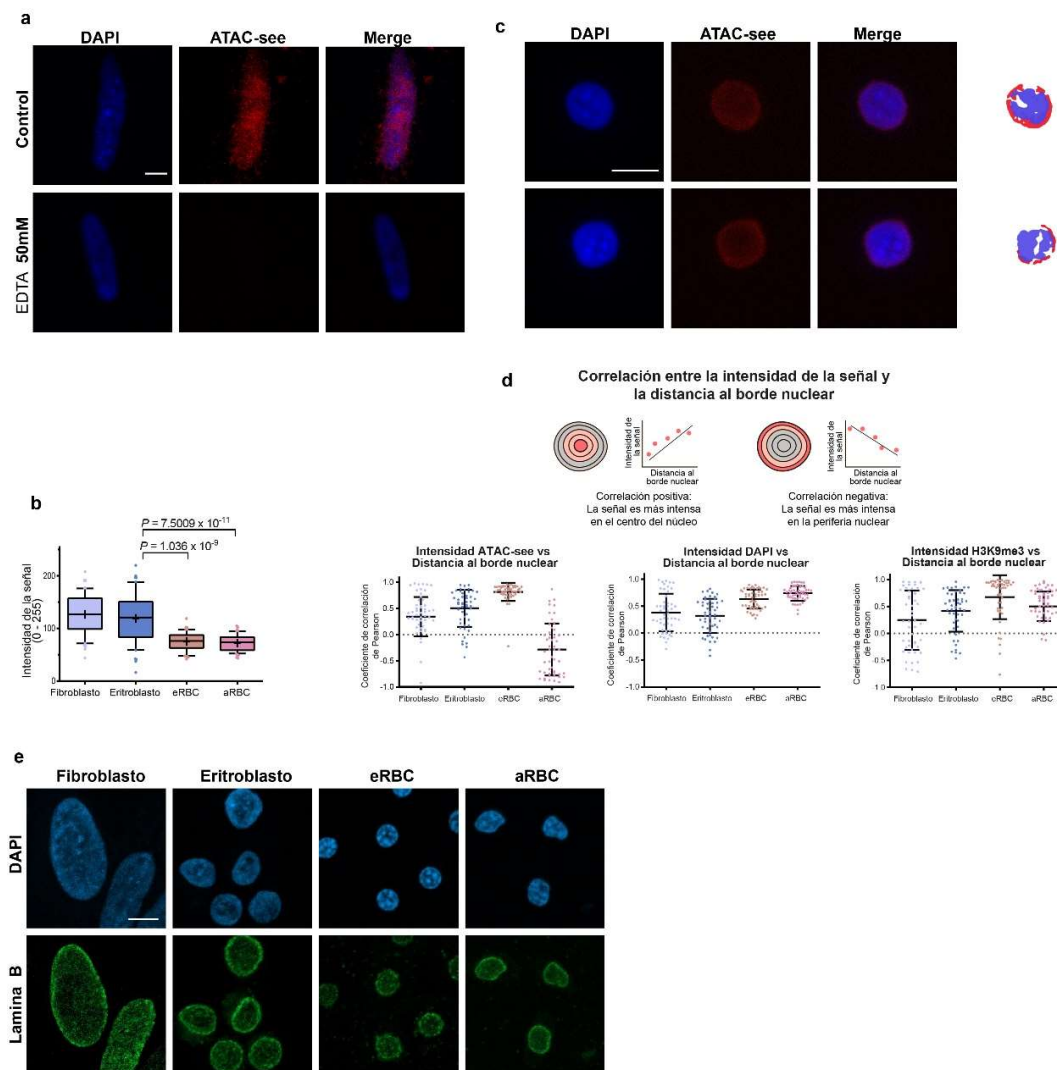


Figura 7. Validación de los resultados de ATAC-see y la reorganización de la cromatina en aRBC. **a**, Resultados de la reacción de ATAC-see en fibroblastos de pollo. Se realizó un control negativo de la reacción en presencia de EDTA 50 mM, lo que inhibe a la transposasa y causa la ausencia de la marca fluorescente. **b**, Promedio de la intensidad de la señal de ATAC-see para cada tipo celular (n=50 núcleos). Las gráficas muestran la mediana (línea horizontal), promedio (+), las cajas van del primer al tercer cuartil y los bigotes del percentil 10 y 90. Se muestran los valores de P obtenidos tras una prueba de ANOVA de una vía seguida de la prueba post-hoc de Tukey. **c**, Imágenes representativas de dos experimentos de ATAC-see en aRBC. Las zonas más intensas de cada fluoróforo se muestran a la derecha. **d**, Coeficiente de correlación de Pearson entre la intensidad de la señal de ATAC-see, DAPI o H3Kme3 y la distancia a la periferia nuclear. Las gráficas muestran el valor promedio +/- SD de 50 núcleos. **e**, Inmunofluorescencia para la lámina B. En todas las micrografías de este panel se muestran resultados representativos de dos réplicas independientes y una barra de calibración = 5 μ m.

Los eritrocitos de pollo expresan RNAs pequeños y genes implicados en la función eritroide

La presencia de cromatina accesible en los genomas eritroides abre la posibilidad de que estas zonas reflejen la presencia de sitios activos transcripcionalmente. De manera consistente con reportes previos¹⁰², la cantidad de RNA en nanogramos obtenido por millón de células decrece significativamente en eRBC y es aún menor en aRBC (Fig. 8a). Además, el perfil electroforético de los RNAs presentes en los eritrocitos muestran la gradual pérdida de los RNAs ribosomales y la permanencia de moléculas de RNA pequeñas (Fig. 8b).

Para caracterizar el estado transcripcional de los eritrocitos, se realizaron ensayos por duplicado de RNA-seq en estas células (en colaboración con la M. en C. Sherlyn Claudio Galeana). Las réplicas biológicas de los experimentos mostraron buena concordancia entre sí (Fig. 9a), por lo que se continuó con el análisis. A partir de estos datos se calculó la abundancia relativa de todos los transcritos utilizando la normalización por Fragmentos Por Kilobases del transcrito, por Millón de lecturas (FPKM). Los valores globales de FPKM en los eritrocitos muestran que la abundancia global de los transcritos está severamente disminuida (Fig. 8c). Sin embargo, utilizando un valor de corte de 0.3 FPKMs, se detectó que un subgrupo de genes que continúa expresándose en eRBC (1,304 genes) y aRBC (804 genes) (Fig. 8d; Fig. 9b).

Los datos de RNA-seq muestran buena concordancia con mediciones realizadas con experimentos de RT-qPCR. Por ejemplo, los datos de RNA-seq muestran la disminución de la expresión del gen *GAPDH* en eritrocitos y el aumento en la expresión de la hemoglobina *HBBA* en eRBC (Fig. 8e-f), siguiendo un patrón similar al observado por RT-qPCR (Fig. 9c). De la misma manera, en la Fig. 9d se muestra la comparación entre los datos de RNA-seq y RT-qPCR de los genes codificantes para las proteínas estructurales de la membrana eritroide *EPB41* y *EPB42*.

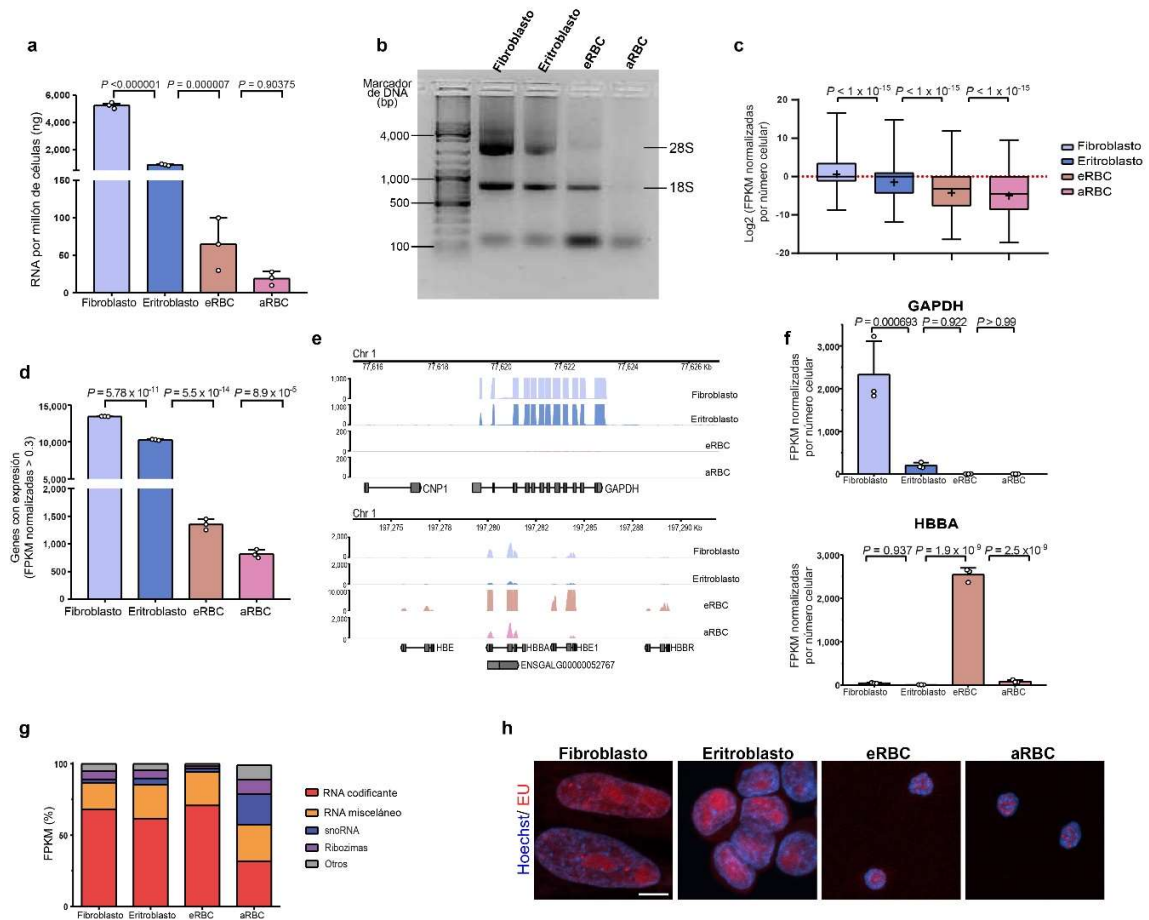


Figura 8. Los eritrocitos de pollo sufren una dramática disminución transcripcional pero continúan expresando RNAs pequeños y genes eritroides. **a**, Cuantificación en ng de RNA total obtenido en fibroblastos, eritroblastos, eRBC y aRBC. Se muestran el promedio +/- SD de 3 réplicas independientes. Los valores de P mostrados provienen de una ANOVA de una vía seguida de la prueba *post-hoc* de Fisher LSD. **b**, Perfil electroforético del RNA total de las células estudiadas. **c**, Log2 de los FPKM promedio de tres experimentos de RNA-seq. Las gráficas muestran la mediana (línea horizontal), promedio (+), primer y tercer cuartil (caja) y los bigotes denotan los valores mínimos y máximos de la distribución. Los valores de P mostrados provienen de una ANOVA de una vía seguida de la prueba *post-hoc* de Dunn. **d**, Número de genes expresados (FPKM > 0.3). Se presenta el valor promedio +/- SD de tres réplicas de RNA-seq. Las diferencias estadísticas se estimaron con una ANOVA de una vía seguida de la prueba *post hoc* de Tukey. **e**, Señal de RNA-seq para *GAPDH* (arriba) y el gen beta-globina *HBBA* (abajo). **f**, Valores de FPKM de *GAPDH* y *HBBA*. Se muestra el promedio +/- SD de tres réplicas de RNA-seq y las diferencias significativas calculadas con una ANOVA de una vía y la prueba *post hoc* de Tukey. **g**, Estimación de la diversidad de RNAs para cada tipo celular expresado como el porcentaje del total de FPKM que corresponde a genes codificantes para proteínas, RNAs misceláneos, RNAs pequeños nucleolares (snoRNA), ribozimas u otros. **h**, Marcaje

fluorescente del RNA naciente mediante la incorporación de 5-etinil uridina (EU). Se muestran resultados representativos de triplicados independientes. Barra de calibración = 5 μ m

La anotación de los genes expresados en eritrocitos de acuerdo con el tipo de gen al que corresponden apoya el cambio en la diversidad de RNAs observados en el perfil electroforético. Particularmente en aRBC la proporción de genes codificantes para proteínas que continúan expresados disminuye a la vez que aumenta la proporción de RNAs pequeños nucleolares (snoRNAs) y RNAs misceláneos aumenta (Fig. 8g). Además, de acuerdo con el análisis de expresión diferencial llevado a cabo entre los eritroblastos y los eritrocitos, tanto eRBC como aRBC muestran la sobreexpresión de un número limitado de genes (165 en eRBC y 137 en aRBC) (Fig. 9f), los cuales están enriquecidos en funciones tales como el transporte de oxígeno, el metabolismo de hierro y la diferenciación eritroide (Fig. 9g).

Por último, se confirmó que tanto eRBC como aRBC mantienen una tasa transcripcional detectable mediante la incubación de las células por 6 horas con el análogo de uridina, 5'-etinil-uridina (EU). La EU es incorporada al RNA recién sintetizado y puede ser posteriormente detectado por fluorescencia, de modo que su detección refleja directamente la presencia de transcripción en las células. Aunque la fluorescencia promedio de EU después de 6 horas de incubación es reducida en eRBC y aRBC, comparado con los eritroblastos (Fig. 9e), la incorporación de EU muestra que los eritrocitos continúan siendo células transcripcionalmente activas (Fig. 8h). De manera que estos datos demuestran que, a pesar de mantener una baja tasa transcripcional, los eritrocitos continúan expresando genes involucrados en funciones eritroides y, de manera interesante, RNAs pequeños no codificantes.

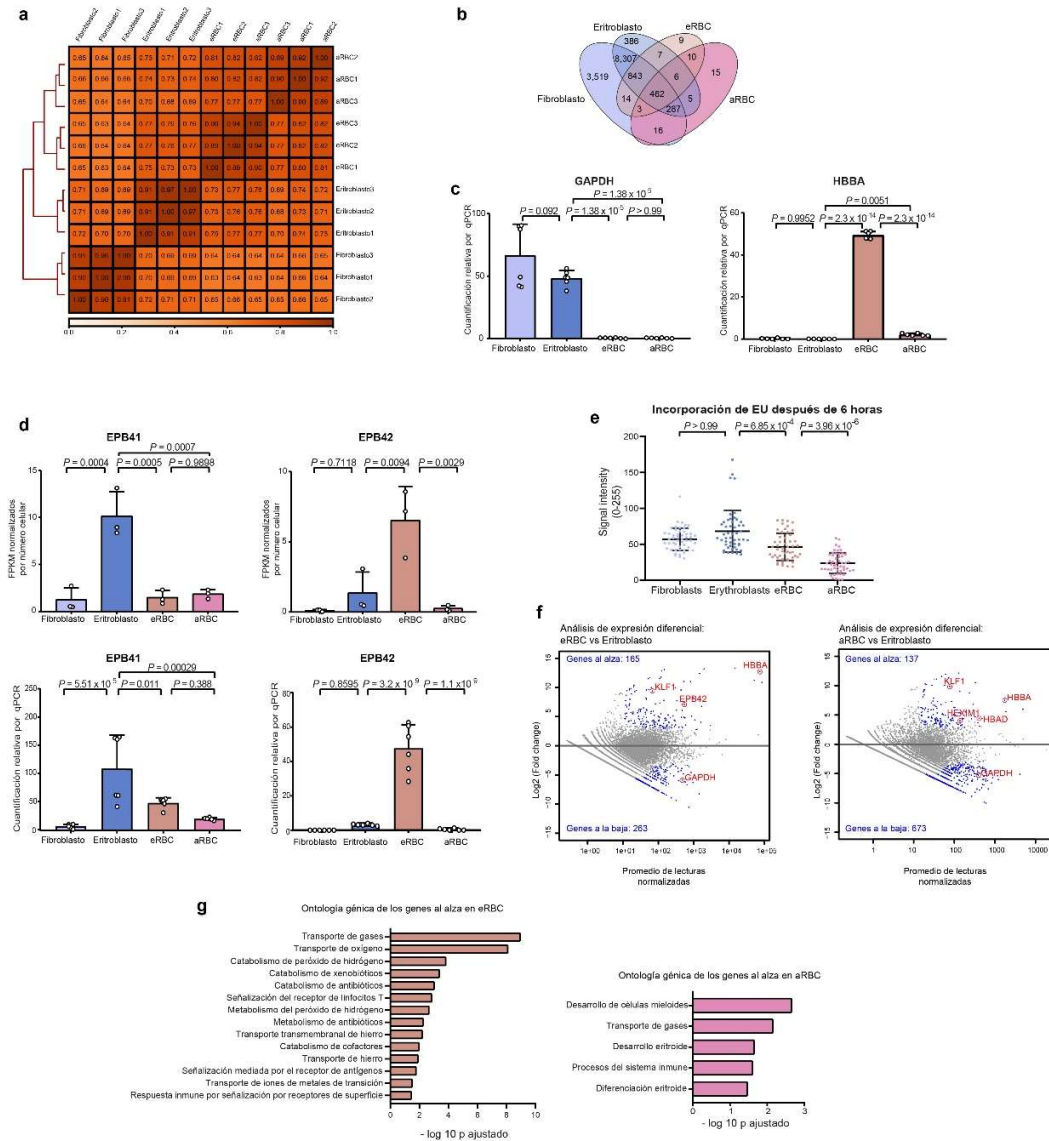


Figura 9. Caracterización del transcriptoma eritroide. **a**, Mapa de correlación entre las réplicas biológicas de RNA-seq. **b**, Diagrama de Venn que muestra los genes expresados en fibroblastos y las células eritroides. **c**, Expresión relativa de *GAPDH* y *HBBA* estimada por RT-qPCR. Se muestran triplicados de dos réplicas biológicas. Los datos se presentan con el promedio +/- SD y las diferencias estadísticas estimadas con una ANOVA de una vía seguida de la prueba *post hoc* de Tukey. **d**, Niveles de expresión de los genes eritroides *EPB41* y *EPB42* estimados por RNA-seq (arriba, n= 3 experimentos independientes) y por RT-qPCR (abajo, n= 6). Se muestra el promedio +/- SD y las diferencias estadísticas estimadas con una ANOVA de una vía seguida de la prueba *post hoc* de Tukey. **e**, Incorporación de EU medida por la intensidad de fluorescencia detectada después de 6 horas de incubación (n = 50 núcleos de tres réplicas independientes). Los valores de P mostrados provienen de una ANOVA de una vía seguida de la prueba *post hoc* de Kruskal-Wallis. **f**, Análisis de expresión diferencial entre eRBC y eritroblastos (izquierda) o aRBC y

eritroblastos (derecha). **g**, Términos ontológicos enriquecidos en los genes sobre expresados en eRBC y aRBC.

La RNA Pol II pausada ocupa los promotores de genes silenciados en los eritrocitos

La detección de EU en los eritrocitos sugiere que, si bien aún son detectables sitios transcripcionalmente activos, estos no se localizan en la periferia nuclear de aRBC (Fig. 8h). De manera que la cromatina accesible en la periferia no se relaciona con transcripción en estas regiones. Para identificar en alta resolución las secuencias genómicas accesibles en los eritrocitos, se realizaron ensayos de accesibilidad de la cromatina (ATAC-seq) en dos réplicas biológicas de cada tipo celular (Fig. 11a).

El análisis de los datos de ATAC-seq muestra que existen picos de cromatina accesible en todos los tipos celulares estudiados. En general, los fibroblastos muestran un mayor número de sitios accesibles que todas las células del linaje eritroide (Fig. 10a), lo que sugiere que los eritroblastos podrían ya haber sufrido la pérdida de cromatina accesible. De hecho, el número de picos de ATAC-seq es muy similar entre eritroblastos, eRBC y aRBC a pesar de las grandes diferencias en el nivel transcripcional entre estas células, y muchos de estos sitios se mantienen accesibles en eritroblastos y eritrocitos (Fig. 10a-b). Es decir, se trata de zonas que retienen su accesibilidad a lo largo de la diferenciación terminal.

La anotación genómica de los picos de ATAC-seq muestra que mientras que en fibroblastos y eritroblastos la mayoría de los sitios accesibles corresponden a promotores y exones de genes transcripcionalmente activos, en eritrocitos la gran mayoría de sitios accesibles corresponden a promotores de genes inactivos (Fig. 10c). En las Figs. 10d y 11b se muestran ejemplos de genes desprovistos de transcripción en eritrocitos que retienen accesibilidad a nivel de su promotor. Así mismo, la integración de los datos de ATAC-seq y RNA-seq muestran cómo, de

manera global, el promotor de muchos genes permanece abierto en eRBC y aRBC a pesar de no presentar niveles detectables de transcripción (Fig. 10e).

La relativa abundancia de este tipo de promotores podría ser explicada por efecto del control de la procesividad de la RNA Pol II. En la fase inicial de la transcripción la RNA Pol II se une al promotor del gen a transcribir y es fosforilada en el residuo de serina 5 (Ser5-P) del dominio carboxilo-terminal. Sin embargo, para que la polimerasa proceda a elongar requiere de la actividad de factores de elongación y ser fosforilada en el residuo de serina 2 (Ser2-P). Cuando esto no ocurre, la polimerasa permanece pausada posicionada en un sitio proximal al promotor¹⁰⁵. Estudios previos han mostrado que el pausado de la RNA Pol II es capaz de mantener accesible el promotor sobre el que se encuentra posicionada y retiene el potencial del gene a ser activado¹⁰⁶. De manera interesante, dos de los factores que promueven el pausado de la polimerasa, el gen *HEXIM1* y el RNA misceláneo *7SK*, mantienen su expresión elevada en eRBC y aRBC (Fig.10f).

La Ser5-P de la RNA Pol II es detectable por inmunofluorescencia en todos los tipos celulares estudiados, sin mostrar cambios en la intensidad de la señal, mientras que la señal correspondiente a la elongación activa de la RNA Pol II (Ser2-P) está disminuida en eRBC y aRBC (Fig. 11d-e). Además, la señal de Ser5-P de la RNA Pol II se encuentra enriquecida en la periferia nuclear de aRBC (Fig. 10g; flechas), siguiendo un patrón similar al de la cromatina accesible (Fig.11g-h). Por su parte, la señal de Ser2-P de la RNA Pol II se encuentra presente en el centro nuclear en todos los tipos celulares (Fig.11f-i), lo que coincide con la señal de EU observada previamente. De modo que la señal de la RNA Pol II en un estado pausado es abundante en los eritrocitos y coincide espacialmente con la cromatina accesible en aRBC.

una vía seguida de la prueba *post hoc* de Tukey. **g**, Detección de la RNA Pol II Ser5-P por inmunofluorescencia. En aRBC, la señal se ubica en la periferia nuclear (flecha). Se muestran imágenes representativas de tres experimentos independientes. Barra de calibración = 5 μ m. **h**, CHIP-seq para la RNA Pol II Ser5-P en eritroblastos, eRBC y aRBC. Para cada tipo celular primero se muestra la señal en CPM centrada en los TSSs +/- 1.5 Kb de todos los genes inactivos y luego la señal +/- 1.5 Kb en los TSSs pausados. **i**, Ejemplo de un promotor pausado ocupado por la RNA Pol II Ser5-P en aRBC. El gen *TAF8* codifica para una subunidad integral del factor general de transcripción TFIID.

Para corroborar que la RNA Pol II Ser5-P se encuentra posicionada en los promotores accesibles de los eritrocitos, se realizaron ensayos de CHIP-seq para esta proteína (en colaboración con la M. en C. Sherlyn Claudio). Se detectaron 1,136 y 1,358 picos de RNA Pol II Ser5-P en promotores en eRBC y aRBC, respectivamente. De estos, más del 95% representan promotores de genes inactivos (Fig. 10h). En la Fig. 10i se muestra el promotor del factor transcripcional *TAF8* ocupado por la RNA Pol II Ser5-P en aRBC.

Finalmente, se corroboró la presencia de la RNA Pol II Ser5-P en el promotor de otros genes sin expresión mediante CHIP-qPCR (Fig. 12a-b). El promotor inaccesible del gen *Sox2* sirvió como control negativo del experimento y no fue detectable la presencia de la RNA Pol II Ser5-P. Estos resultados muestran que, a pesar de sufrir un silenciamiento transcripcional considerable, los eritrocitos terminalmente diferenciados retienen la accesibilidad de la cromatina alrededor de los promotores génicos, probablemente gracias al pausado de la RNA Pol II en estos sitios.

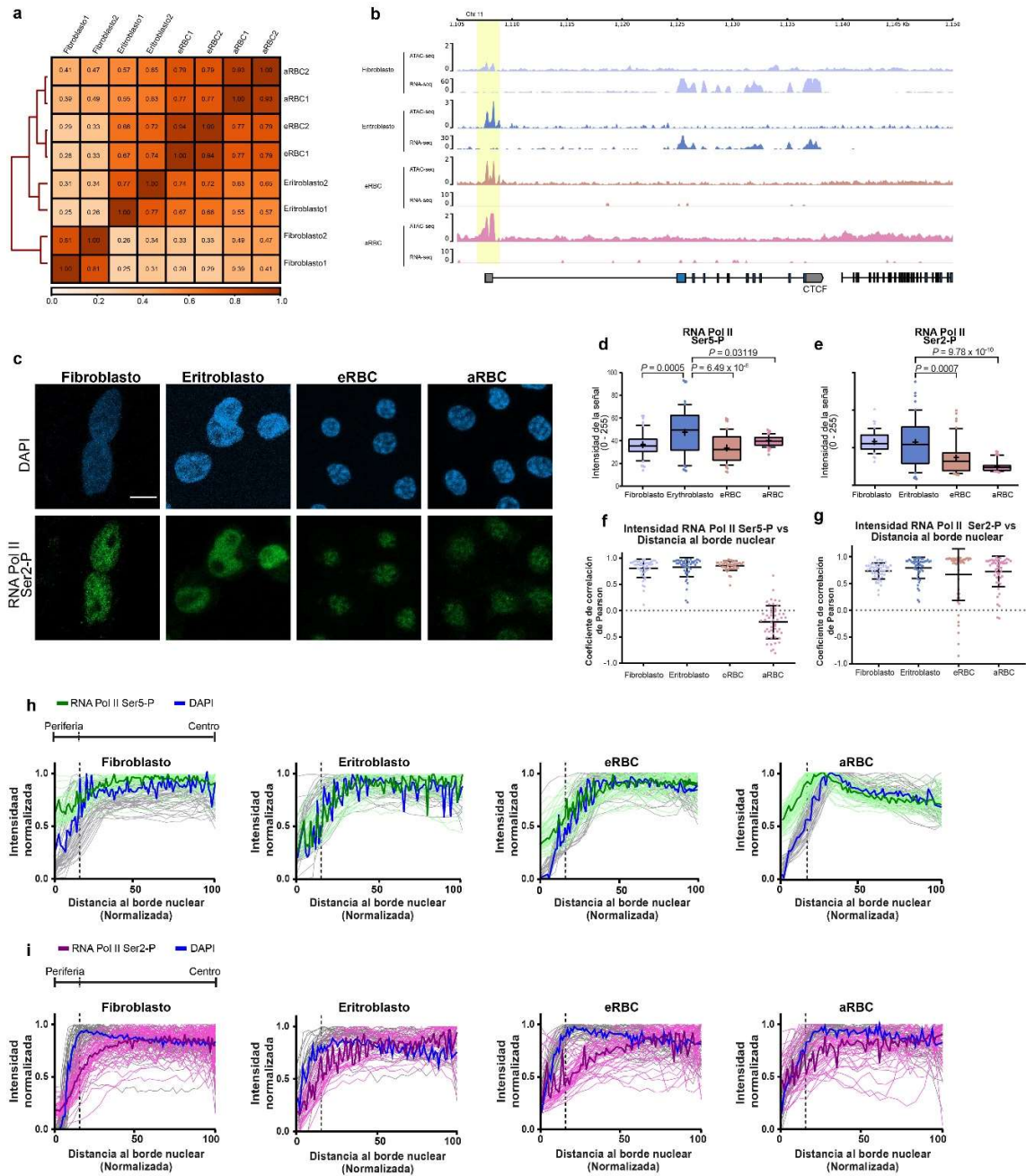


Figura 11. Los eritrocitos de pollo retienen la accesibilidad de la cromatina alrededor de promotores pausados. **a**, Mapa de correlación entre las réplicas biológicas de ATAC-seq. **b**, El promotor de CTCF continúa accesible a pesar de que no se detecta su expresión de acuerdo con los datos de RNA-seq. **c**, Detección de la RNA Pol II Ser2-P por inmunofluorescencia. Se muestran imágenes representativas de dos réplicas independientes. Se muestra también la intensidad promedio de la RNA Pol II Ser5-P (**d**) y Ser2-P (**e**), en 50 núcleos. Las diferencias estadísticas se estimaron con una ANOVA de una vía seguida de la prueba *post hoc* de Tukey. **f-g**, Se muestra el coeficiente de

correlación de Pearson entre la intensidad de la RNA Pol II Ser5-P o Ser2-P y la distancia a la periferia nuclear (n=50 núcleos). **h-i**, Intensidad de la señal de RNA Pol II Ser5-P y Ser2-P en función de la distancia a la periferia nuclear en 50 núcleos. Los trazos individuales se muestran en colores claros mientras que el trazo promedio se muestra en un color sólido.

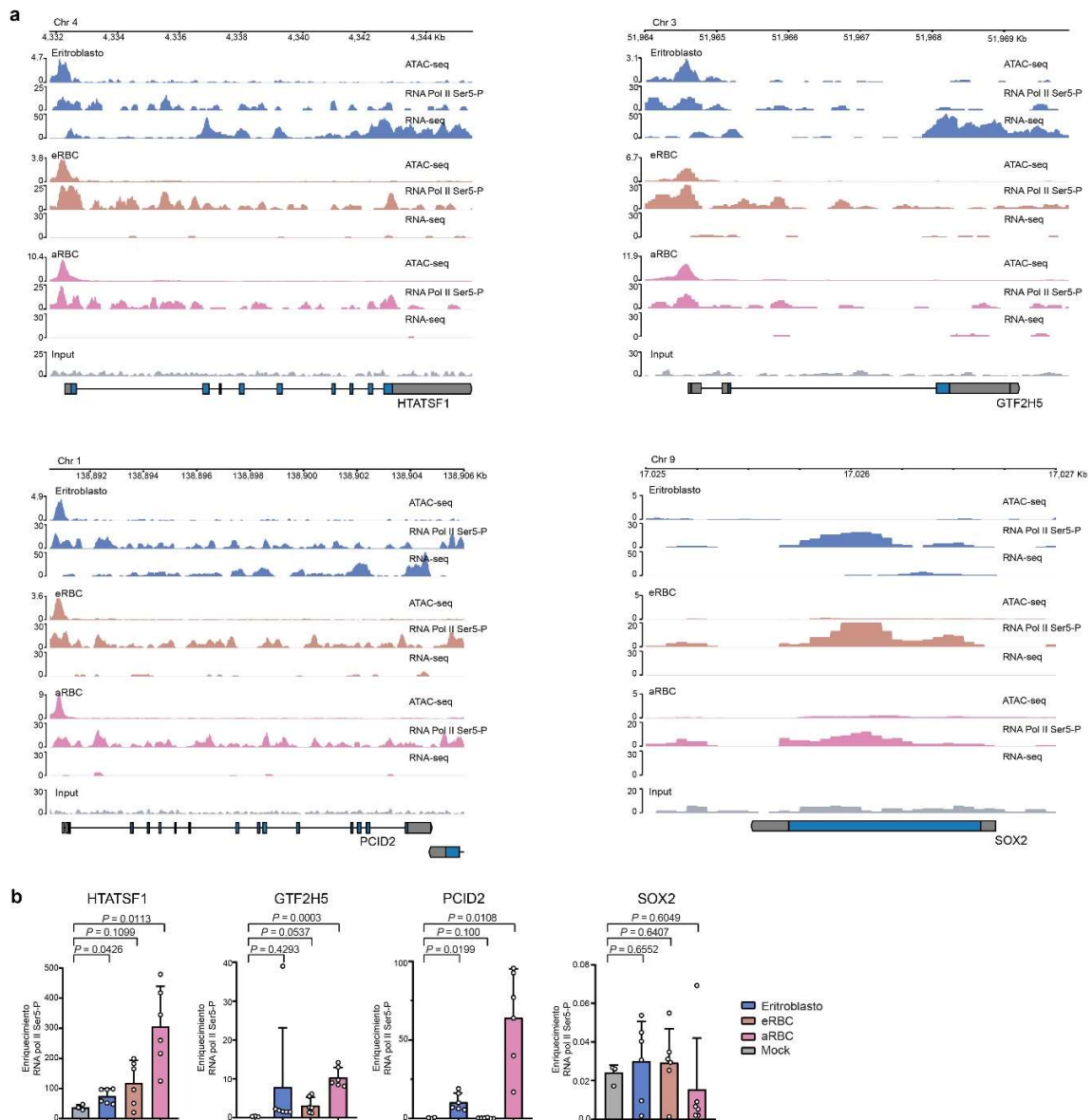


Figura 12. Presencia de la RNA Pol II Ser5-P en promotores pausados evaluada por ChIP-qPCR. a, Señal de ATAC-seq, ChIP-seq para la RNA Pol II Ser5-P y RNA-seq en los genes *HTATSF1*, *GTF2H5*, *PCID2* y *SOX2*. **b**, Enriquecimiento de la RNA Pol II Ser5-P en los promotores de *HTATSF1*, *GTF2H5*, *PCID2* y *SOX2* evaluada por ChIP-qPCR. Se

muestran los valores de 4 réplicas independientes. Las diferencias estadísticas contra el mock se estimaron con pruebas de t de dos colas.

El genoma de los eritrocitos de pollo está hiper- compartimentado

Los resultados obtenidos sugieren que el genoma de los eritrocitos de pollo está altamente reorganizado, tal como demuestran los cambios en la arquitectura nuclear observados mediante la técnica de ATAC-see, la formación de *foci* de heterocromatina y la segregación espacial de la RNA Pol II. Para estudiar a mayor profundidad la organización del genoma eritroide se realizaron experimentos de Hi-C por duplicado en fibroblastos, eritroblastos, eRBC y aRBC y se construyeron las matrices de interacción.

La inspección visual de las matrices de interacción muestra claras diferencias entre los eritrocitos terminalmente diferenciados y los eritroblastos a nivel de la formación de compartimentos nucleares (Fig.13a). Como se mencionó antes, la segregación de la eucromatina al compartimento A y la heterocromatina al compartimento B se aprecia en las matrices de Hi-C como un patrón de tablero de ajedrez. Los datos de Hi-C de los eritrocitos presentan un patrón de compartimentos muy marcado en comparación a las matrices de eritroblastos y fibroblastos (Fig.13a) y la comparación entre tipos celulares muestra que los eritrocitos ganan interacciones entre secuencias a más de 10 Mb de distancia en el genoma lineal (Fig. 13b-c).

El patrón de compartimentos de las matrices de Hi-C apunta a que el genoma de eRBC y aRBC está hiper-compartimentado, es decir, las regiones de eucromatina se encuentran espacialmente alejadas de las zonas heterocromáticas dentro del núcleo y existe poco contacto entre ambos tipos de cromatina. Así mismo, la distribución de la heterocromatina en *foci* lejos de la cromatina accesible observada por microscopía apoyan esta noción.

Se llevó a cabo la identificación de las zonas del genoma en el compartimento A y B mediante un análisis de componentes principales sobre los datos de Hi-C. De esta

manera, es posible evaluar de manera indirecta el estado de compactación de la cromatina. Esta inspección reveló cambios en la compactación del genoma eritroide.

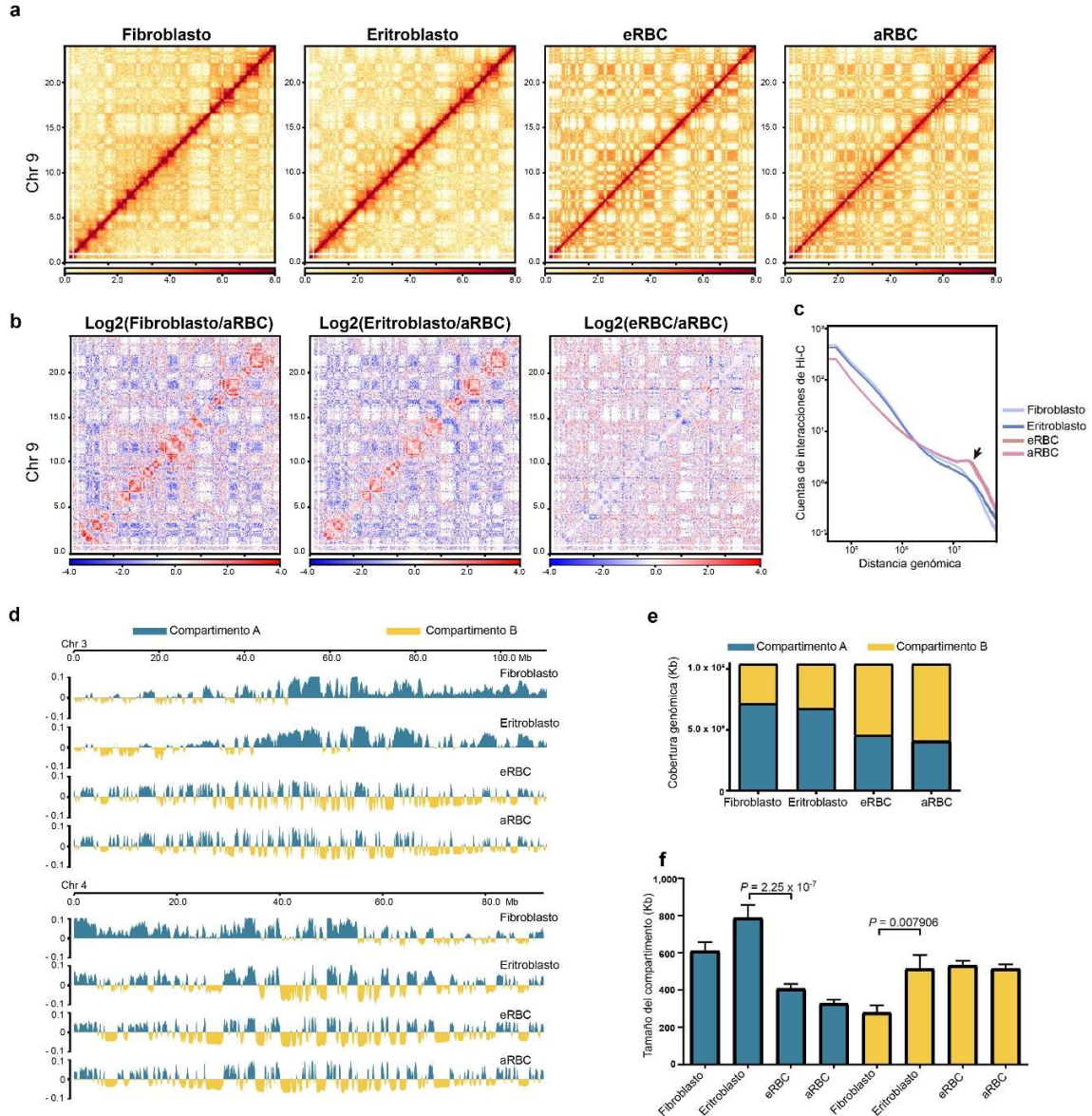


Figura 13. Los eritrocitos de pollo tienen genomas hiper-compartmentados. **a**, Matrices de interacción de Hi-C del cromosoma 9 (Mb 0 - 24). **b**, Comparación de las matrices de interacción en **a**, respecto a la matriz de aRBC. El mapa de calor indica las regiones que ganan interacciones en aRBC (azul) y las que pierden contacto (rojo). **c**, Conteo de los contactos de Hi-C en función de la distancia genómica. El incremento en el número de interacciones a larga distancia en eRBC y aRBC se indica con una flecha. **d**, Identificación del compartimento A (azul) y B (amarillo) en los fibroblastos y células

eritroides. Se presenta el cromosoma 3 (arriba) y 4 (abajo). **e**, Porcentaje del genoma dentro del compartimento A o B. **f**, Tamaño promedio de las regiones dentro del compartimento A o B. La gráfica de barras muestra el promedio \pm SEM y las diferencias estadísticas se calcularon con una ANOVA de una vía y la prueba *post hoc* de Bonferroni.

En eritroblastos y fibroblastos, el genoma cuenta con regiones extensas de eucromatina, de 800 Kb en promedio (Fig. 13d-f) que reflejan la actividad transcripcional de estas células. Sin embargo, en eRBC y aRBC el compartimento heterocromático comienza a invadir al compartimento A (Fig. 13d), de modo que las zonas de eucromatina en los eritrocitos son menos abundantes y de menor extensión (alrededor de 300 Kb en promedio) (Fig. 13f). Importantly, la mayor parte del genoma eritroide se encuentra dentro del compartimento B (Fig. 13e), es decir, altamente condensado.

El genoma de los eritrocitos de pollo está organizado alrededor de promotores pausados

Como se mencionó antes, los TADs representan estructuras genómicas, de menor tamaño que los compartimentos, que favorecen las interacciones entre elementos reguladores y genes ubicados en su interior y limitan el contacto entre elementos ubicados en dominios diferentes, por lo que tienen gran relevancia en la regulación transcripcional⁴¹.

La presencia de TADs en el genoma se identifica en las matrices de Hi-C como triángulos ubicados sobre la diagonal de la matriz. Sin embargo, el marcado patrón de compartimentos en las matrices de Hi-C de eRBC y aRBC dificulta la identificación visual de los TADs en estos genomas (Fig. 14a). Se realizó la identificación de los TADs en las matrices de Hi-C haciendo uso del programa TADtool y el índice de aislamiento (ver Métodos). Mientras que en el genoma de fibroblastos y eritroblastos la presencia de TADs es fácilmente apreciable en las matrices (Fig. 14b), el genoma de eRBC y aRBC se encuentra ampliamente

desprovisto de este nivel de organización y únicamente se detectaron regiones pequeñas con un enriquecimiento local de interacciones, similares a los TADs (Fig. 14b; Fig. 15b).

Los dominios identificados en eritrocitos son menores en número que los TADs de los blastos (Fig. 14c) y también son más pequeños (Fig. 14d). De hecho, menos del 25% del genoma de eRBC y aRBC permanece estructurado en estos mini dominios, mientras que los eritroblastos y fibroblastos cuentan con TADs en todo el genoma (Fig. 14e).

La formación de TADs en vertebrados está mediada por la actividad conjunta de las proteínas estructurales CTCF y el complejo de cohesina, formado por SMC1, SMC3 y Rad21⁶⁴. Previamente se han reportado cambios en la localización nuclear de la proteína estructural CTCF en el núcleo de eRBC y aRBC¹⁰³. En los datos de RNA-seq aquí presentados se puede observar una marcada disminución en la expresión de *CTCF* y *Rad21* (Fig. 14f), de igual manera se observó una reducción en los niveles proteicos de CTCF y SMC1 por ensayos de Western Blot (Fig. 15c). En suma, estos datos muestran que el genoma eritroide pierde estructuración a nivel de TADs, probablemente por la ausencia total o parcial de las proteínas estructurales que median su formación, mientras que una porción pequeña del genoma de eRBC y aRBC continúa estructurado en mini dominios.

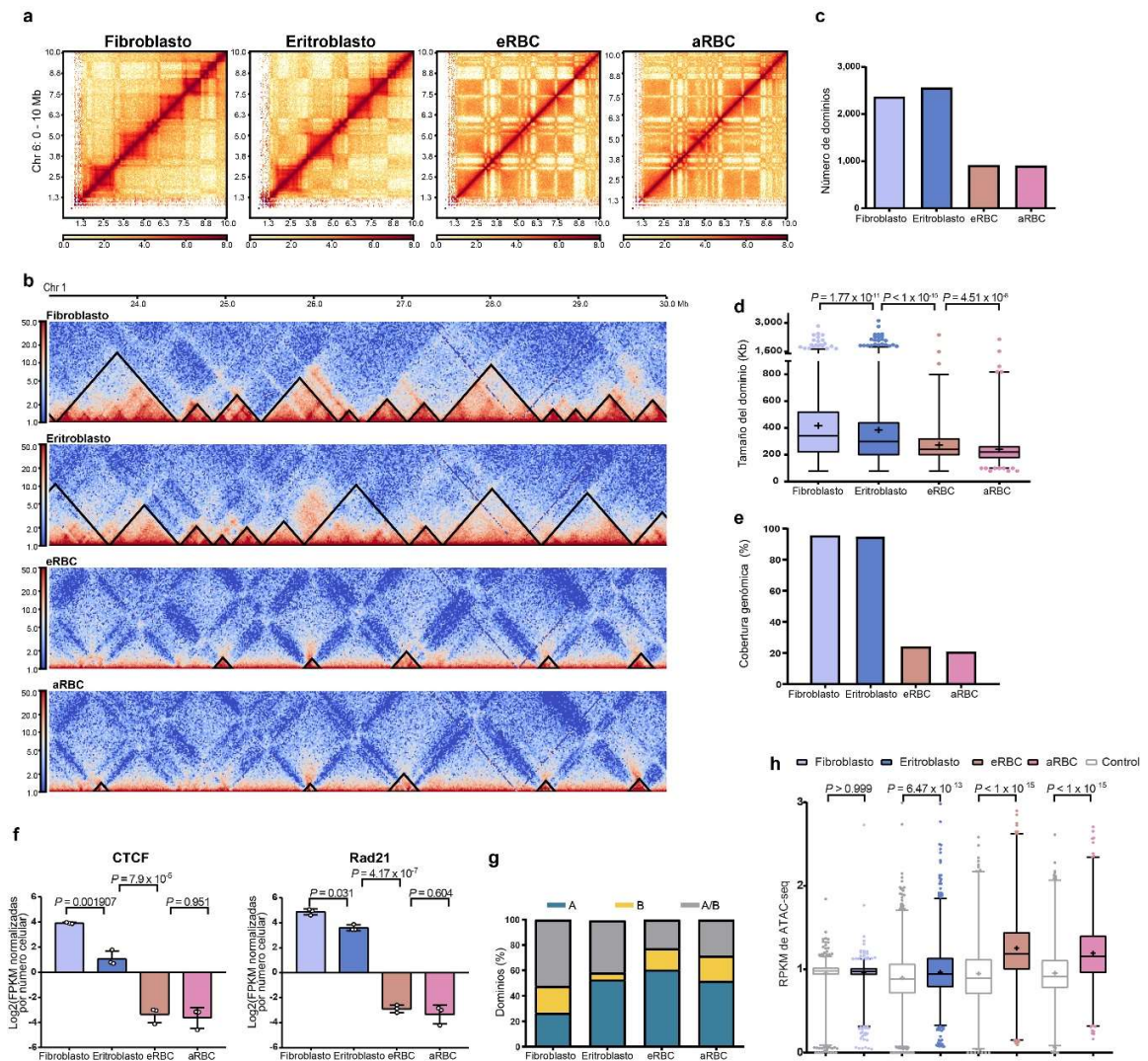


Figura 14. El genoma de los eritrocitos pierde los TADs a nivel global pero retiene dominios en regiones de cromatina abierta. a, Matrices de contacto de Hi-C del cromosoma 6 (0-10 Mb). Los TADs son claramente distinguibles en fibroblastos y eritroblastos, en los eritrocitos la señal más prominente corresponde a los compartimentos. **b,** Dominios identificados en cada tipo celular demarcados por triángulos negros. Número **(c)** y tamaño promedio **(d)** de los dominios en fibroblastos ($n=2,373$ dominios), eritroblastos ($n=2,562$), eRBC ($n=916$) y aRBC ($n=903$). Las diferencias significativas se estimaron con una prueba de ANOVA una vía seguida de la prueba *post hoc* de Tukey. **e,** Porcentaje del genoma estructurado en dominios. **f,** Niveles de expresión de las proteínas estructurales ($n=3$). Las diferencias estadísticas se calcularon con una prueba de ANOVA de una vía seguida de la prueba *post hoc* de Tukey. **g,** Porcentaje de los dominios que contienen exclusivamente regiones del compartimento A (azul), del compartimento B (amarillo) o que contienen zonas en ambos compartimentos (gris). **h,** Enriquecimiento de ATAC-seq dentro de los dominios expresado en Lecturas por Kilobase por Millón de lecturas (RPKM). Como

control se empleó un grupo de regiones aleatorias del mismo tamaño (Fibroblastos n= 2,373, eritroblastos n= 2,562, eRBC n= 916, aRBC n= 903). Los valores de P mostrados provienen de una ANOVA de una vía seguido de la prueba *post hoc* de Kruskal-Wallis.

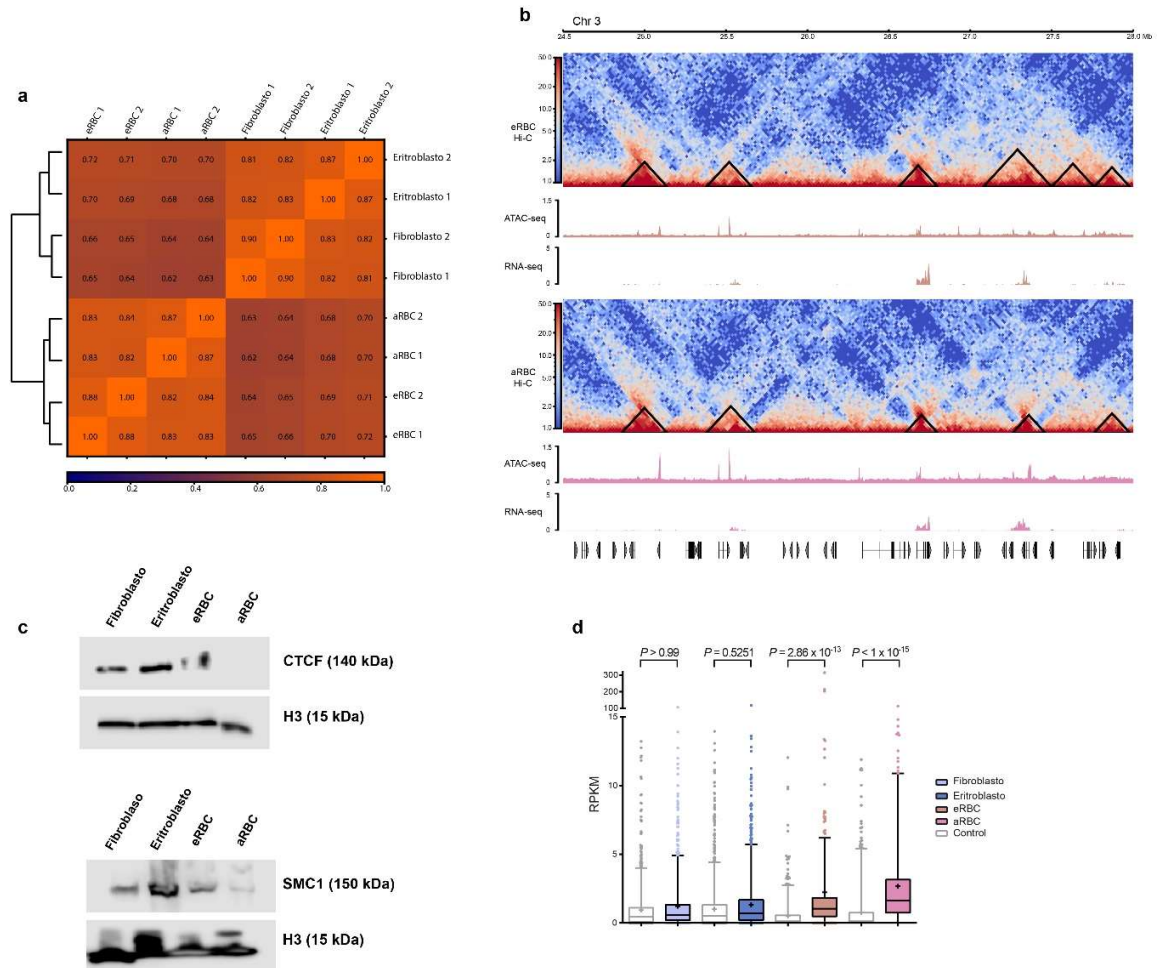


Figura 15. Características cromatínicas de los mini dominios eritroides. **a**, Coeficiente de correlación de Pearson entre las réplicas de Hi-C a 50 Kb de resolución. **b**, Matrices de contacto de Hi-C en eRBC y aRBC (Chr 3: 24.5-28 Mb) y ejemplos de los mini dominios detectados. Debajo se muestra la señal de ATAC-seq y RNA-seq. **c**, Niveles proteicos de CTCF y SMC1 detectados por Western Blot. La Histona 3 (H3) sirvió como control de carga. Se muestran resultados representativos de dos experimentos independientes. **d**, Enriquecimiento de RNA-seq dentro de los dominios expresado en RPKMs. Como control se empleó un grupo de regiones aleatorias del mismo tamaño (Fibroblastos n= 2,373, eritroblastos n= 2,562, eRBC n= 916, aRBC n= 903). Los valores de P mostrados provienen de una ANOVA de una vía seguido de la prueba *post hoc* de Kruskal-Wallis.

En cuanto a los mini dominios presentes en los eritrocitos, este tipo de estructuras coincide con regiones de cromatina abierta, se encuentran dentro del compartimento A y están enriquecidos en lecturas de ATAC-seq y RNA-seq (Fig. 14g-h; Fig. 15d).

A raíz de estas observaciones, se planteó explorar si los mini dominios estructuran regiones transcripcionalmente activas y por tanto pudieran explicarse por la presencia de la maquinaria transcripcional (Fig. 16a).

Se clasificó a los dominios de acuerdo con el estado transcripcional de los genes en su interior en cuatro clases (Fig. 16b):

- Clase I: Dominios que contienen a genes pausados (Fig. 16b, rojo).
- Clase II: Dominios que contienen a genes expresados (Fig. 16b, azul).
- Clase III: Dominios que contienen tanto genes pausados como genes activos (Fig. 16b, amarillo).
- Clase IV: Dominios que sólo contienen genes inactivos, sin accesibilidad a nivel del promotor (Fig. 16b, gris).

Este análisis mostró que la presencia de promotores pausados es una característica común en más del 80% de los mini dominios en eRBC y aRBC. Además, los genes que permanecen pausados y estructurados en mini dominios están enriquecidos en categorías ontológicas relacionadas con el control de la transcripción, el procesamiento y metabolismo de RNA y el transporte de transcritos maduros hacia el citoplasma (Fig. 16c; Fig. 12a). De tal manera que el genoma eritroide mantiene la formación de mini dominios alrededor de los promotores pausados de genes implicados en la regulación transcripcional.

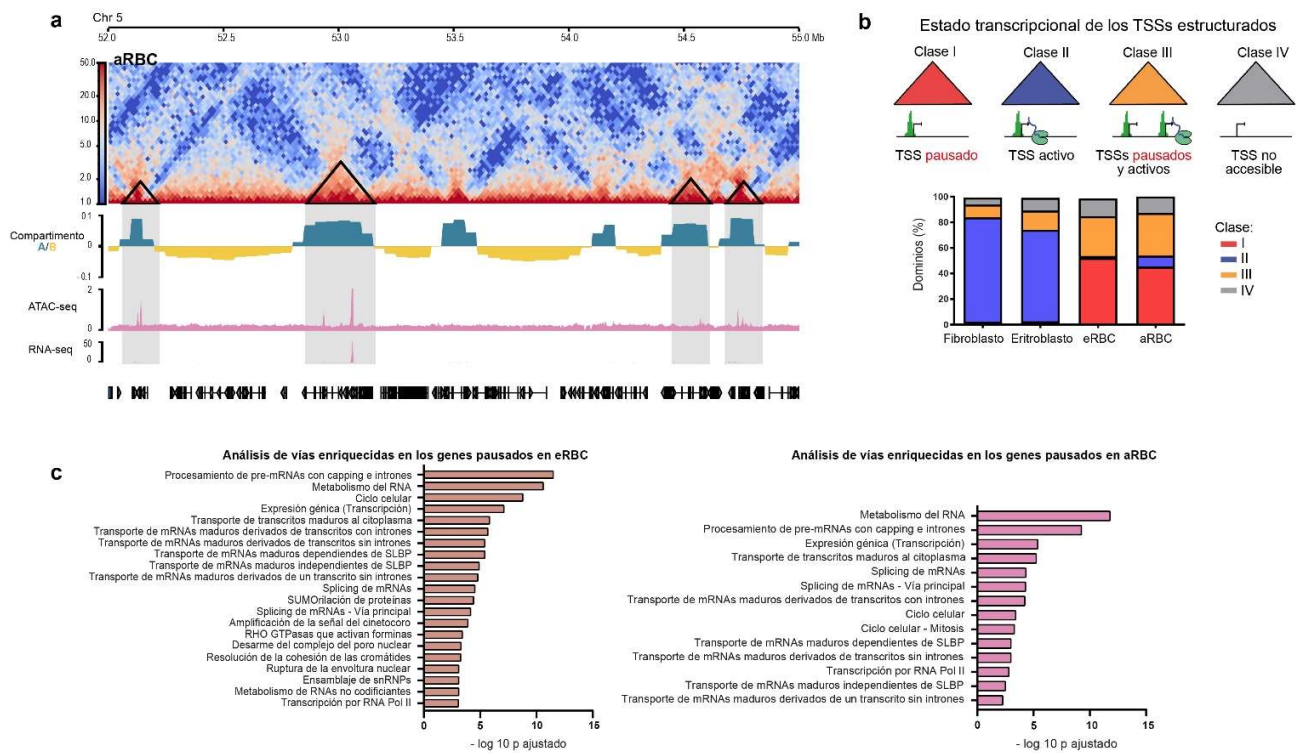


Figura 16. Los mini dominios contienen y organizan a los promotores pausados en el genoma eritroide. a, Matriz de contacto de Hi-C en aRBC (Chr 5: 52 – 55 Mb), los dominios identificados se presentan como triángulos negros. Debajo se presenta la señal de los compartimentos, ATAC-seq y RNA-seq. En gris se muestran los dominios que contienen picos de ATAC-seq. **b,** Distribución de los mini dominios de acuerdo con el estado transcripcional de los TSS que contienen. Se muestran cuatro categorías: Clase I (TSSs pausados), Clase II (TSSs activos), Clase III (TSSs pausados y activos) y Clase IV (TSSs cerrados). **c,** Términos ontológicos enriquecidos en los genes pausados dentro de los mini dominios de eRBC y aRBC.

VI. Discusión

El presente trabajo analiza a profundidad los cambios en la accesibilidad de la cromatina y la organización del genoma que acompañan al silenciamiento transcripcional en los eritrocitos nucleados de pollo. La integración de los datos obtenidos muestra que los eRBC y aRBC atraviesan por una compleja reorganización a diferentes escalas genómicas (Fig. 17). El núcleo de los eritrocitos es un organelo altamente compartimentado que confina la cromatina accesible, la transcripción y la heterocromatina en distintas zonas. Además, en aRBC ocurre la pérdida de la lámina A/C y la reorientación de la cromatina accesible hacia la periferia nuclear.

La reducción de los niveles de las proteínas estructurales CTCF y cohesina afecta severamente la estructuración de TADs. Sin embargo, el pausado de la RNA Pol II retiene la accesibilidad de un grupo de promotores y mantiene la estructuración local alrededor de los genes pausados, formando mini dominios que contienen genes involucrados para la síntesis y metabolismo del RNA.

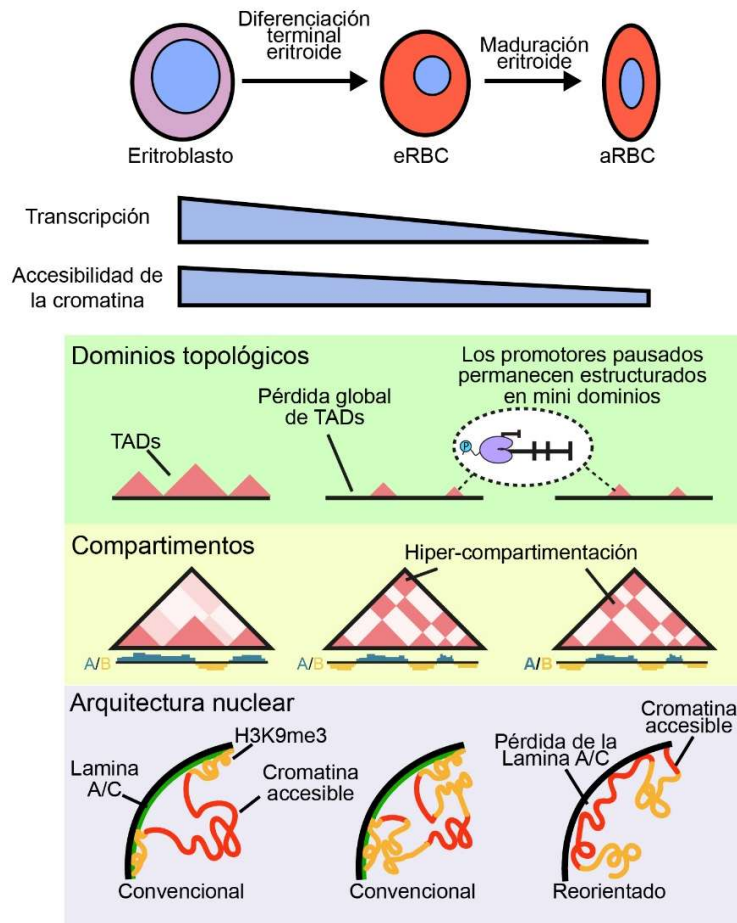


Figura 17. Remodelación de la cromatina durante la eritropoyesis en pollo.

La diferenciación terminal eritroide resulta en un decremento dramático en la accesibilidad y tasa transcripcional que es exacerbada en los aRBC. Los eritrocitos nucleados pierden a los TADs en todo el genoma pero retienen mini dominios estructurados alrededor de promotores ocupados por la RNA Pol II pausada. A nivel de los compartimentos, el genoma se encuentra hiper-compartimentado. Por último, en aRBC la ausencia de la lámina A/C causa el reposicionamiento de la cromatina abierta.

Los eritrocitos terminalmente diferenciados de gallinas adultas sufren una inversión en la localización nuclear de la cromatina

En su conjunto, los resultados obtenidos hasta el momento muestran que el severo silenciamiento transcripcional que acompaña a la eritropoyesis involucra por una parte la disminución de la accesibilidad de la cromatina y la ganancia de heterocromatina.

Los experimentos de ATAC-seq mostraron que, aunque los eRBC son células aun transcripcionalmente activas, su contenido de cromatina abierta disminuye al compararse con los eritroblastos. El silenciamiento transcripcional se agrava en los aRBC¹⁰², lo que concuerda con los resultados de ATAC-seq en los que la cromatina accesible parece ser más escasa en estas células comparadas con los eRBC.

Otro punto importante es que en eRBC y aRBC las regiones más ricas de H3K9me3 se encuentran más dispersas dentro del núcleo y no necesariamente aledañas a la periferia nuclear. A su vez, en aRBC la cromatina accesible se localiza en la periferia nuclear, lo que contrasta con la arquitectura nuclear convencional. La inversión del patrón eucromatina/heterocromatina es un evento que sólo se ha reportado en los bastones de la retina de mamíferos nocturnos y se ha atribuido a cambios en el contenido de las proteínas de la envoltura nuclear⁵³. Al igual que en los bastones, la lámina A/C no fue detectada en la envoltura nuclear de aRBC, por lo que probablemente un mecanismo muy similar sea el que causa la inversión de la cromatina en este linaje celular.

En la eritropoyesis en mamífero se han reportado el corte proteolítico de la lámina B y la remodelación transitoria de las proteínas en la envoltura nuclear⁸⁷. Inhibir esta remodelación evita que el núcleo se condense y retrasa la diferenciación eritroide^{87,107}, sin embargo, no parece ocurrir la relocalización de la cromatina accesible. Por lo que la pérdida de las láminas nucleares podría representar un paso común durante la eritropoyesis conservado en otras especies y/o un fenómeno necesario para reestructurar y compartimentar de manera global el genoma en linajes muy especializados y terminalmente diferenciados.

Tanto la señal de la RNA Pol II asociada con la elongación, junto con la localización del RNA nascente marcado con EU sugieren la periferia nuclear no es una región activa transcripcionalmente. Además, la señal de ATAC-seq en los TSS muestra que una proporción importante de los TSS se mantiene abierta, aunque la transcripción haya

cesado, de modo que la porción de cromatina abierta aledaña a la periferia pudiera corresponder a TSS abiertos en donde la RNA Pol II se mantiene pausada. Así mismo, la RNA Pol II Ser5-P, que corresponde a la polimerasa de inicio o pausada, se localiza en la periferia nuclear en aRBC.

Los eritrocitos terminalmente diferenciados son transcripcionalmente activos

Si bien la evidencia experimental del silenciamiento transcripcional del genoma en los eritrocitos es escasa, se había asumido que estas células carecen de actividad transcripcional debido al bajo contenido de RNA que presentan. En este trabajo fue posible detectar la presencia de la RNA Pol II fosforilada, tanto en la Ser5 (inicio/pausa de la transcripción) como Ser2 (elongación de la transcripción), en eRBC y aRBC, además de detectar la presencia de RNAs recién sintetizados gracias al marcaje fluorescente de RNA con EU, permitiéndonos afirmar que estas células son transcripcionalmente activas aunque en efecto, la actividad transcripcional en eRBC (1,352 genes) es baja comparada con el eritroblasto y es aún más reducida en los aRBC (818 genes); sin embargo, en estos experimentos se puso en evidencia que ambos tipos de eritrocitos circulantes cuentan con los componentes necesarios para la transcripción y mantienen la expresión activa de un número reducido de genes.

Dentro de los genes que continúan expresándose en los eRBC y aRBC destacan aquellos directamente involucrados en el transporte de oxígeno y el metabolismo de hierro, ambos procesos propios de la fisiología eritroide. Esta observación coincide con lo descrito para el último estadio nucleado de la eritropoyesis fetal de ratón, que ocurre en el hígado, en donde se mantiene la expresión de 474 genes asociados con la síntesis de hemoglobina y el transporte de oxígeno hasta antes de enuclear⁸⁵. Sin embargo, destaca que en los eritrocitos nucleados la transcripción de estos genes continúa aun cuando las células se encuentran en circulación, lo que llama a reconsiderar la idea de que los eritrocitos nucleados son células inactivas que sólo transportan oxígeno de manera pasiva.

En mamíferos se conoce la participación de micro RNAs (miRNAs; <25 pb) durante la eritropoyesis y enucleación¹⁰⁸. Los eritrocitos circulantes son ricos en este tipo de transcritos y se ha propuesto que actúan como un reservorio de miRNAs capaz de señalizar a otros

tipos celulares presentes en la sangre¹⁰⁹. La diversidad y tipo de transcritos también cambia a lo largo del proceso de eritropoyesis en pollo.

A diferencia del enriquecimiento en miRNAs en mamífero, los eritrocitos terminalmente diferenciados están particularmente enriquecidos en RNAs pequeños no codificantes (<200 pb), especialmente RNAs clasificados como misceláneos y nucleolares. Esta característica única del transcriptoma del eritrocito de pollo podría reflejar la necesidad de los eritrocitos nucleados de regular procesos epigenéticos asociados con la transcripción, así como la estabilidad y procesamiento de RNAs recién sintetizados, en contraste con los eritrocitos de mamífero desprovistos de material genético.

El pausado de la RNA Pol II es un proceso extendido en los genomas eritroides

Como se mencionó antes, la presencia de cromatina accesible en la periferia nuclear no puede ser adjudicada a la actividad transcripcional remanente en los eritrocitos. No obstante, fue posible detectar la presencia de la RNA Pol II Ser5-P en la periferia nuclear de aRBC, lo que indica que estos sitios están ocupados por la polimerasa, aunque esta no se encuentre activamente elongando. Por otra parte, aunque la señal de la RNA Pol II Ser2-P está disminuida en los eritrocitos, no se detectaron diferencias significativas en la intensidad de la señal de la RNA Pol II Ser5-P entre eritrocitos y blastos. Estos resultados muestran que el control de la actividad de la RNA Pol II, particularmente el pausado de la misma, podría estar implicado en el mantenimiento de la accesibilidad de la cromatina.

Durante la transcripción, después del ensamblaje del complejo de pre-iniciación en el TSS, la polimerasa inicia la transcripción de aproximadamente 20-60 nucleótidos y se detiene antes de proceder a la elongación del transcrito, por efecto de los factores negativos de la elongación NELF y DSIF¹¹⁰. Para que la RNA Pol II proceda a elongar se requiere que el factor positivo P-TEFb fosforile a la polimerasa en la serina 2 y la libere del complejo pausado¹¹⁰, de manera que la actividad de la polimerasa puede ser limitada promoviendo su pausado.

En apoyo a esta idea, aquí se reporta que las regiones accesibles del genoma eritroide corresponden, en su mayoría, a promotores de genes carentes de transcripción y se encuentran ocupados por la RNA Pol II Ser5-P de acuerdo con experimentos de ChIP.

Además, los eritrocitos de pollo mantienen niveles elevados de expresión del RNA no codificante 7SK y la proteína HEXIM1, ambos conocidos factores que secuestran al factor de elongación PTEF-b y promueven el pasaje de la polimerasa¹⁰⁵. Al respecto, un estudio hecho anteriormente muestra que los aRBC retienen a la RNA Pol II en un estado que puede ser estimulado durante ensayos de transcripción *in vitro*, lo que confirma que el estado pausado de la polimerasa es abundante en el genoma eritroide¹¹¹.

Asimismo, recientemente se reportó que la actividad de HEXIM1 es esencial para la diferenciación de progenitores eritroides humanos¹¹² y que los eritroblastos tardíos de ratón pierden las marcas de histona asociadas a la elongación transcripcional⁸⁵. Por lo que el pasaje de la polimerasa podría ser un mecanismo común durante la diferenciación eritroide para reducir de manera importante la tasa transcripcional.

Una consecuencia importante del pasaje de la polimerasa es que mantiene la accesibilidad de los promotores en los que se encuentra y retiene el potencial del gen a ser activado, por lo que la presencia de la polimerasa pausada es abundante en genes de respuesta a estrés¹⁰⁶. Dado que los eritrocitos nucleados permanecen en circulación cerca de 30 días, resulta interesante explorar si el mecanismo de pasaje les brinda la posibilidad de reactivar la transcripción ante estímulos ambientales.

El genoma de los eritrocitos está hiper-compartimentado

El análisis de los resultados de Hi-C indican que ocurren grandes cambios en la organización tridimensional del genoma en los eritrocitos. Resulta interesante que, a pesar de tener la cromatina invertida, los aRBC son muy similares a los eRBC en cuanto a las estructuras observadas en el Hi-C. Esto apunta a que la inversión de la cromatina no afecta a otros niveles de organización como TADs y compartimentos, lo que tiene sentido considerando que el anclaje de la cromatina a la periferia nuclear difiere de los TADs en los mecanismos implicados en su formación. De manera similar, experimentos de Hi-C en timocitos WT y knock-out para el LBR no muestran diferencias en la organización del genoma después de un proceso de inversión¹¹³, de modo que la pérdida general de interacción entre la cromatina y la envoltura nuclear no parece modificar la formación de TADs o compartimentos.

En los genomas de los eritrocitos es notable la ganancia de interacciones de larga distancia, así como de la acentuación de la señal de los compartimentos en las matrices de Hi-C. Estas observaciones posiblemente estén relacionadas, ya que la segregación de la cromatina en compartimentos contribuye a formar interacciones entre regiones a más de 20 Kb de distancia¹¹⁴. Consecuentemente, la formación de compartimentos dicta gran parte de la organización del genoma en estas células como se evidencia en la formación de *foci* claramente apreciables de heterocromatina y la segregación espacial de la cromatina accesible a la periferia.

De forma consistente con el proceso de silenciamiento que atraviesan los eritrocitos, el análisis de compartimentos muestra el que el compartimento B crece conforme la diferenciación y maduración avanza. Los compartimentos son estructuras dinámicas y el cambio de compartimento correlaciona con cambios en el estado transcripcional o de compactación de la cromatina³⁹. En los aRBC más de la mitad del genoma está dentro del compartimento B, lo que refleja una amplia reorganización del genoma, así como la inactivación de gran parte del mismo.

Un estudio reciente describe el aumento en la compartimentación del genoma y la pérdida de TADs después del abatimiento del factor NIPBL, encargado de cargar el complejo de cohesina a la cromatina⁷⁰. En el presente trabajo se reporta un fenómeno similar en el que los compartimentos se ven reforzados a la vez que se pierden otras estructuras topológicas, probablemente atribuible a la reducción en los niveles de expresión y proteicos de los componentes del anillo de cohesina.

El genoma eritroide está organizado alrededor de los promotores pausados

La inspección visual de las matrices de Hi-C supone que el genoma de eRBC y aRBC carece de TADs. La identificación de dominios con TADtool mostró que el genoma de los eritrocitos retiene algunas estructuras similares a dominios, aunque de menor tamaño, que consistentemente se encuentran dentro del compartimento A, son ricas en cromatina abierta y contienen a los promotores pausados. Lo que sugiere que aquellas regiones del genoma eritroide que retienen la presencia de la RNA Pol II permanecen estructuradas.

La pérdida de TADs en todo el genoma se asemeja al fenómeno observado en modelos experimentales en donde se pierde la unión de cohesina a la cromatina o se ocasiona la pérdida de CTCF en un modelo de degradación inducible por auxina⁶². A diferencia de NIPBL, la ausencia de CTCF no parece alterar los compartimentos, por lo que no explica por completo lo observado en los eritrocitos. Los experimentos de RNA-seq muestran que CTCF y cohesina están silenciados en los eritrocitos y los niveles totales de las proteínas están severamente disminuidos. Sin embargo, cuando la ausencia de CTCF no es completa los niveles residuales de la proteína son capaces de retener algunos dominios⁶².

Aunque el rol de la transcripción como una fuerza que coordina la organización espacial del genoma ha sido muy debatido, reportes recientes muestran que la RNA Pol II es capaz de formar asas de cromatina e interacciones de corto alcance⁷⁶. En contra de esta idea, se ha mostrado que la inhibición de la transcripción no parece tener efecto en la topología del genoma⁷⁷. Coincidentemente, la evidencia aquí presentada muestra que más del 80% de los mini dominios de los eritrocitos cuentan con promotores pausados, lo que indica que la presencia de la RNA Pol II pausada es capaz de promover la organización local del genoma. Otros mecanismos que podrían asistir en este proceso es el reclutamiento de factores de transcripción, complejos remodeladores de la cromatina y las proteínas estructurales por parte de la polimerasa.

Por último, los genes que permanecen pausados y estructurados en mini dominios están enriquecidos en términos ontológicos relacionados con el control de la transcripción, el procesamiento del RNA y el transporte de los transcritos maduros hacia el citosol. A la luz de esto, es factible proponer que la RNA Pol II pausada mantiene la accesibilidad de la cromatina y la organización genómica alrededor de este grupo de genes para poder reactivar el proceso de transcripción y metabolismo del RNA en respuesta a un estímulo ambiental. Por ejemplo, se ha observado que los eritrocitos de pollo pueden ser estimulados ante un reto inmunológico y montar una respuesta transcripcional que involucra la expresión de receptores inmunitarios y la producción de citocinas^{115,116}.

Este trabajo ahonda, por primera vez, en la reorganización del genoma de los eritrocitos nucleados de pollo y su relación con el silenciamiento transcripcional. La exploración del transcriptoma eritroide llama a reconsiderar la noción de los eritrocitos como células completamente inactivas y apunta al pausado de la RNA Pol II como un mecanismo importante para la regulación transcripcional y estructuración del genoma eritroide.

VII. Conclusiones

- I. Los aRBC sufren una inversión de la arquitectura nuclear. La cromatina accesible se repositona hacia la periferia nuclear mientras que la heterocromatina rica en H3K9me3 se dispone al centro del núcleo.
- II. La envoltura nuclear de los núcleos de aRBC carece de la Lamina A/C.
- III. Aunque eRBC y aRBC experimentan una dramática caída en su contenido de RNA, ambos tipos celulares son transcripcionalmente activos.
- IV. En aRBC el silenciamiento transcripcional es más severo que los eRBC. Sin embargo, ambos tipos celulares expresan genes relacionados con el transporte de oxígeno, el metabolismo de hierro y la diferenciación eritroide.
- V. eRBC y aRBC expresan en niveles elevados al gen *HEXIM1* y el RNA pequeño 7SK, los cuales están involucrados en el pausado proximal de la RNA Pol II.
- VI. Un número importante de promotores de genes sin expresión continúan accesibles y ocupados por la RNA Pol II Ser5-P en el genoma de eRBC y aRBC.
- VII. La expresión y niveles proteicos de la proteína estructural CTCF están abatidos en eRBC y aRBC, mientras que Rad21 está severamente disminuido.
- VIII. Únicamente el 24.18% del genoma de eRBC y 20.9% del genoma de aRBC continúa organizado en mini dominios.
- IX. Los mini dominios estructuran promotores pausados ricos en accesibilidad y la RNA Pol II Ser5-P involucrados en procesos como la regulación transcripcional y el metabolismo, procesamiento y transporte del RNA.

VIII. Perspectivas

La caracterización de la organización del genoma y su relación con la transcripción en los eritrocitos nucleados de pollo aquí realizada pone sobre la mesa nuevas preguntas relativas a la función eritroide y su regulación transcripcional. Se ha sugerido que los cambios topológicos observados en los eritrocitos de pollo están conservados en otras especies de vertebrados¹¹⁷, por lo que será importante extender estas observaciones a otros grupos de vertebrados, ya que es probable que el pausado de la polimerasa sea un mecanismo conservado en la diferenciación eritroide de otras especies que presentan eritrocitos nucleados.

El estudio realizado comprende únicamente un estadio tardío de la diferenciación eritroide y a los eritrocitos terminalmente diferenciados. Un estudio más amplio que involucre a estadios más tempranos de la diferenciación extendería el entendimiento de la remodelación de la cromatina durante este proceso y brindaría una perspectiva más amplia acerca de la regulación temporal de la remodelación.

Por otra parte, será importante poner a prueba el papel de la RNA Pol II pausada en la capacidad de los eritrocitos nucleados de montar una respuesta transcripcional ante estímulos ambientales, particularmente aquellos que ya se sabe que son capaces de inducir la expresión de algunos genes, así como evaluar si inhibir el pausado de la polimerasa suprime la capacidad de respuesta de los eritrocitos.

Finalmente, las consecuencias topológicas de la modulación de la actividad de la RNA Pol II requieren mayor exploración. La manera en la que la RNA Pol II mantiene la estructura de los mini dominios eritroides no es clara, por lo que habrá que caracterizar la ocupación de las proteínas estructurales CTCF y cohesina en estas células y explorar su relación con la polimerasa. Así como explorar las consecuencias topológicas de perturbar la actividad de la RNA Pol II al estimular o inhibir su pausado.

IX. Métodos

Colecta de sangre y cultivo celular

Se obtuvo la sangre de embriones de pollo de 10 días de desarrollo por medio de la punción de las venas extraembrionarias. Los eritrocitos de organismos adultos se obtuvieron de gallinas de 28-38 semanas de edad. La sangre se colectó en buffer de fosfatos salino (PBS, NaCl 137 mM, KCl 2.7 mM, Na₂HPO₄ 10 mM, KH₂PO₄ 1.8 mM) con EDTA 50 mM y se mantuvo en hielo. La sangre colectada se centrifugó a 1,000 rpm por 10 minutos a 4°C, el pellet celular se lavó tres veces y se contaron las células en una cámara de Neubauer.

Se empleó la línea celular HD3, que deriva de eritroblastos de pollo adulto transformados con el virus de la eritroblastosis aviar. Las células HD3 se cultivaron en suspensión con medio DMEM suplementado con suero fetal bovino (10%) y suero de pollo (2%) en una atmósfera humidificada de CO₂ al 5% a 37°C. También se obtuvieron fibroblastos de pollo a partir de embriones de 10 días y se cultivaron bajo las condiciones ya descritas.

Preparación de librerías de ATAC-seq

Las reacciones de ATAC-seq se realizaron a partir de 150,000 células que fueron transferidas a un tubo Lo-Bind y se centrifugaron a 1,600 rpm por 5 minutos a 4°C. Se retiró el sobrenadante y se añadieron 50 µL de buffer de lisis (10 mM Tris-Cl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.01% Igepal), se incubó por 15 minutos en hielo. En el caso de los experimentos con fibroblastos, antes de realizar la lisis el pellet fue resuspendido en 600 µL de buffer de lisis y filtrado con una membrana de 70 µm. El material filtrado fue tratado con un homogeneizador de vidrio. Posteriormente se realizó la lisis de la manera descrita anteriormente. Se realizaron dos réplicas biológicas de cada tipo celular.

Posteriormente, los núcleos obtenidos se centrifugaron a 600 x g por 10 minutos a 4°C y se retiró el sobrenadante. Se realizó la reacción de etiquetado e incubación de los núcleos en un tubo Lo-Bind con 25 µL del buffer TD 2X (Illumina), 22.5 µL de

agua inyectable y 2.5 μ L de la transposasa TDE1 (Illumina) por exactamente 30 minutos a 37°C. Inmediatamente después de la reacción se purificaron las librerías con columnas MinElute (Qiagen) y las librerías se amplificaron por 10 ciclos de PCR utilizando el sistema de índices de Nextera. Se estimó el tamaño de los fragmentos obtenidos en un equipo TapeStation 4200 (Agilent). Las librerías se secuenciaron de manera pareada en un equipo HiSeq2500 (Illumina) o NextSeq (Illumina),

Análisis de datos de ATAC-seq

Las lecturas de ATAC-seq fueron alineadas al genoma de pollo (GRCg6a) con Bowtie2, con los parámetros por defecto para lecturas pareadas. Posteriormente se generaron los archivos BAM (samtools view), se filtraron (samtools sort) e indexaron (samtools index). Se filtraron las dúplicas de PCR y las lecturas mayores de 150 pb. Se generaron tracks genómicos de las lecturas filtradas normalizadas por CPM con deeptools (bamCoverage -normalizeUsing CPM) y la visualización se realizó con pyGenomeTracks.

Para la identificación de los sitios de cromatina abierta se utilizó MACS2 para llamar picos de accesibilidad (-q val = 0.01). Únicamente los picos consenso, conservados entre las dos réplicas de cada tipo celular, se conservaron para los análisis siguientes.

Preparación de librerías de RNA-seq

Se realizaron extracciones de RNA total con TRIzol siguiendo las instrucciones del fabricante y la integridad del RNA se revisó por electroforesis. Las librerías de RNA-seq se prepararon a partir de 1 μ g de RNA utilizando el kit TruSeq Stranded Total RNA Ribo-Zero H/M/R Gold (Illumina). Posteriormente, se estimó el tamaño de los fragmentos obtenidos en un equipo TapeStation 4200 (Agilent). Se prepararon tres réplicas biológicas por tipo celular y se secuenciaron en un equipo HiSeq2500 por secuenciación pareada.

Análisis de datos de RNA-seq

El análisis de datos de RNA-seq se realizó siguiendo el procesamiento de SnakePipes. Brevemente, las lecturas de RNA-seq se mapearon al genoma del pollo GRCg6a con STAR, se obtuvieron las cuentas crudas con featureCounts y se calcularon los valores de FPKM. Después de la normalización por FPKM, se realizó una normalización adicional basada en el número de células necesarias para obtener 1 µg de RNA, dado que el contenido de RNA total en los eritrocitos es muy bajo.

Debido a que el contenido total de RNA por célula en los eritrocitos es muy bajo, el número de células necesario para obtener 1 µg de RNA es muy variado dependiendo del tipo celular. Para corregir este sesgo experimental se realizó una normalización adicional en la que los FPKM calculados se normalizaron al número celular más bajo utilizado (190,000 en fibroblastos). La fórmula utilizada es la siguiente: FPKMs normalizadas por número celular = $(190,000 \times \text{FPKM}) / (\text{Número celular requerido})$. Se estableció un valor de corte de 0.3 FPKM para considerar un gen como expresado.

El análisis de expresión diferencial se realizó con DESeq2 v.1.28.1 utilizando los valores de cuentas crudas normalizados de la misma manera que los FPKM. Finalmente, se estableció un valor de corte de False Discovery Rate = 0.5 y Fold Change = 2.

RT-qPCR

Se realizaron extracciones de RNA total con TRIzol siguiendo las instrucciones del fabricante y la integridad del RNA se revisó por electroforesis. Posteriormente se llevó a cabo la reacción de transcripción reversa con random primers y las qPCRs se realizaron con el kit de KAPA SYBR Fast qPCR kit (Merck, KR0389-V9.13). La secuencia de los oligonucleótidos utilizados se presenta en la Tabla 1. La cuantificación relativa se estimó por el método de curva estándar, con puntos de 50, 100 y 150 ng de DNA. Se realizaron dos réplicas biológicas con triplicados técnicos.

Tabla 1. Secuencia de los oligonucleótidos empleados para los experimentos de expresión.

Blanco	Orientación	Secuencia 5'-3'
GAPDH	Forward	CCGTCCTCTCTGGCAAAGTCCA
	Reverse	GTGAAGTCCTTGCTGAAGTGGG
HBBA	Forward	TGCATTGTGACAAGCTGCATGT
	Reverse	CAGCATGGGGAGATCTATGGGG
EPB41	Forward	TTTTATCCTCCGGATCCCGCTC
	Reverse	AGCTCAGACTGCACTGTGTAGG
EPB42	Forward	GGGCAGCCTTTCGTTATCACAG
	Reverse	CCTGGGTTCCGTCTGCTTTAGA

Análisis de ontología de genes

El análisis de ontología y enriquecimiento de vías se realizó con la herramienta g:Profiler utilizando los parámetros por defecto: la prueba de hipótesis múltiples (g:SCS) y la prueba exacta de Fisher. Todos los procesos biológicos designados como enriquecidos tienen un P value < 0.01.

Preparación de librerías de Hi-C

Para la preparación de las librerías de Hi-C se siguió el protocolo de Hi-C in nuclei. Para cada experimento se fijaron 10 millones de células en formaldehído 2% por 10 minutos a temperatura ambiente. Las células se centrifugaron a 800 rpm por 9 minutos a 4°C. Los pellets se congelaron en nitrógeno líquido y se preservaron a -80°C hasta su uso. Se realizó la lisis celular en 5 mL de buffer de lisis (10mM Tris-HCl pH 8, 0.2% NP 40/Igepal, 10mM NaCl, Inhibidor de proteasas Complete 1X) por 30 minutos. En el caso de los fibroblastos, las células se resuspendieron en 1 mL de buffer de lisis y se utilizó un cell strainer seguido de un homogeneizador antes de iniciar la lisis. La cromatina se digirió con 250 U de la enzima Mbol (NEB) overnight y posteriormente se añadieron 50 U más por 2 horas. Posteriormente se realizó el rellenado empleando un nucleótido biotinilado (Biotin-14-dATP; Invitrogen) y se religaron los fragmentos.

Las librerías se sonicaron a un tamaño promedio de los fragmentos de 500 pb en un equipo COVARIS. Se llevó a cabo la reparación de los extremos y selección de fragmentos de un tamaño 250-550 pb antes de proceder con el pull-down utilizando partículas paramagnéticas acopladas a estreptavidina (MyOne Streptavidin C1 DynaBeads; Invitrogen). Se ligaron los adaptadores para secuenciación TruSeq (Illumina) y se amplificaron las librerías con 5-6 ciclos de PCR. Se estimó el tamaño de los fragmentos obtenidos en un equipo TapeStation 4200 (Agilent). Se prepararon dos réplicas biológicas por tipo celular y se secuenciaron en un equipo HiSeq2500 o NextSeq por secuenciación pareada.

Análisis de datos de Hi-C

El mapeo y filtrado de las secuencias obtenidas, así como la normalización y construcción de matrices de interacción se realizó con el programa HiC-Pro utilizando los parámetros por defecto. Se mapearon las lecturas a la versión GRCg6a del genoma de pollo utilizando Bowtie2 y se filtraron los pares de interacción válidos para construir matrices de interacción normalizadas por el método de Corrección Iterativa (ICE) a resoluciones de 10, 20, 50, 100 y 500 Kb.

El coeficiente de correlación de Pearson entre réplicas biológicas se computó con HiCExplorer a una resolución de 50 Kb y las réplicas fueron combinadas en un sólo archivo. Posteriormente se homogeneizó el número de lecturas válidas entre cada tipo celular para descartar un sesgo experimental originado por diferencias en la profundidad de secuenciación. De igual manera, se generaron matrices de interacción que fueron comparadas entre sí con HiCPlotter.

Llamado de TADs

Se generaron matrices densas con HiCPro (sparseToDense.py). La identificación de TADs se realizó en TADtool con el algoritmo de índice de aislamiento a una resolución de 20 Kb y una ventana deslizante de 100 Kb. Fue necesario extraer los cromosomas 16, 30 y 32 que, debido a su tamaño, no cuentan con el número

suficiente de lecturas para llamar TADs. Se corroboró que los TADs identificados concordaran con la inspección visual de las matrices.

Análisis de compartimentos

Se realizó un Análisis de Componentes Principales de las matrices de HiC utilizando la función -hicPCA de HiCExplorer y se obtuvieron los valores del primer y segundo componente (PC1 y PC2). Para asignar el signo correspondiente al análisis, es decir, identificar el compartimento A y B, se proporcionaron las lecturas alineadas en formato bigwig de los experimentos de RNA-seq realizados en cada tipo celular.

ChIP-qPCR y librerías de ChIP-seq

Para los experimentos de ChIP se fijaron cuatro réplicas biológicas con formaldehído al 1% por 10 minutos. Las muestras se sonicaron por 10 ciclos (30 s on/60 s off) con 30% de amplitud utilizando un ultrasonicador Cole-Parmer 500-W. Se realizó un spike-in con 1% de cromatina de células de S2R+ de *Drosophila* y se hizo la inmunoprecipitación con 5 µg de anticuerpo anti-RNA Pol II Ser5-P (Santa Cruz, no. SC-47701 para el ChIP-qPCR o Abcam no. Ab5131 para ChIP-seq) y Protein G DynaBeads (Invitrogen). El DNA obtenido se purificó por columnas (Zymo Research).

Se realizaron las reacciones de qPCR de acuerdo con las instrucciones del proveedor (Jena Bioscience). Se diseñaron oligonucleótidos dirigidos al pico de ATAC-seq detectados en los promotores de genes de interés en aRBC y se calculó el fold change respecto a un control negativo que no presenta pico de ATAC-seq (el promotor de SOX2). La lista de oligonucleótidos utilizados se muestra en la Tabla 3. Las librerías de ChIP-seq se prepararon utilizando el kit NEBNext Multiplex Oligos primer set (Illumina). Se secuenciaron dos réplicas biológicas de cada tipo celular en un equipo NovaSeq por secuenciación pareada.

Tabla 3. Secuencia de los oligonucleótidos empelados para el ChIP-qPCR contra la RNA Pol II Ser5-P.

Blanco	Orientación	Secuencia 5'-3'
Promotor <i>HTATSF1</i>	Forward	GGGTAAGATCACGCTGCCTT
	Reverse	CATCTCAGCCAATGAGACGC
Promotor <i>PCID2</i>	Forward	GGGTAAGATCACGCTGCCTT
	Reverse	CATCTCAGCCAATGAGACGC
Promotor <i>GTF2H5</i>	Forward	TACCAGTACCTTCCGCTTCCTA
	Reverse	CTCCTCCTATTGCCCTCCTGA
Promotor <i>SOX2</i>	Forward	GAAGTCCCAAGCCGGAC
	Reverse	TGATCTCGATCCAAGTGTGG

Análisis de datos de ChIP-seq

Las lecturas pareadas de ChIP-seq fueron alienadas al genoma de pollo (GRCg6a) empleando Bowtie2 con los parámetros por defecto. Las lecturas alineadas se normalizaron utilizando el spike-in proveniente de células S2R+, se filtraron las lecturas duplicadas (samtools markdup) y se generaron los archivos BAM con samtools. El llamado de picos se realizó en cada réplica con MACS2 con un valor de P de 0.001. Los picos conservados entre ambas réplicas fueron utilizados para los subsecuentes análisis. Las lecturas se normalizaron para su visualización con deeptools (bamCoverage -normalizeUsing RPGC) y se graficaron con pyGenomeTracks. Los heatmaps presentados se generaron deeptools (computeMatrix y plotHeatmap). En la Tabla 4 se muestran todos los programas empleados para el análisis de datos de secuenciación.

Tabla 4. Herramientas bioinformáticas y versión de los programas utilizados para el análisis bioinformáticas.

Programa	Versión
Bowtie2	v2.3.5
Deeptools	3.5.0
DeSeq2	1.28.1
featureCounts	1.6.3
Fiji	2.1.0
gProfiler	e107_eg54_p17
HiCExplorer	3.2
HiCPlotter	0.6.6
HiC-Pro	v2.11.1
MACS2	v2.1.2
pyGenomeTracks	3.1.2
Samtools	v1.9
SeqMonk	1.48.1
SnakePipes	1.2.2
STAR	v2.7.2b

Inmunofluorescencias

Para los ensayos de inmunofluorescencia se sembraron 50,000-100,000 células en cubreobjetos previamente tratados con Poli-L-Lisina. Las células se fijaron con formaldehído al 1%, lavadas tres veces con PBS y se permeabilizaron con Igepal 0.1% y se incubaron por 1 h en una solución de bloqueo con albúmina de suero bovino (BSA) al 5%. Posteriormente se incubaron toda la noche con el anticuerpo primario de interés diluido en BSA 0.2%. Al día siguiente se realizaron tres lavados con PBS y se incubó por 1 h con el anticuerpo secundario. Se incubó con DAPI (300 nM) por 20 minutos y se montaron las laminillas con Fluoromount (Sigma). Los anticuerpos y dilución empleados se proporcionan en la Tabla 2. Las preparaciones se observaron en el microscopio confocal invertido Zeiss LSM 800 con un objetivo de inmersión en aceite 63X. Se emplearon los láseres UV-405 nm, Argón- 488 nm, He/Ne-543 nm y se realizaron reconstrucciones en Z de 30 cortes ópticos de 0.2 μm .

Tabla 2. Anticuerpos y diluciones utilizadas para las inmunofluorescencias.

Blanco	No. Catálogo	Dilución empleada
H3K9me3	Millipore No.04-442	1:700
RNA Pol II Ser5-P	Santa Cruz No.SC-47701	1:500
RNA Pol II Ser2-P	Abcam No. ab5095	1:800
Lámina B1	Abcam No. ab16048	1:500
Lámina A/C	Santa Cruz No.SC-376248	1:500
Anti-Rabbit Alexa Fluor 594	Life Technologies No. R37117	1:500
Anti-Mouse FITC	Millipore No. AP192F	1:500

Western Blot

Se obtuvieron extractos nucleares realizando un paso de lisis celular en un buffer de alta sacarosa (10 mM HEPES, 10 mM KCl, 1.5 mM MgCl₂, 0.34 M sacarosa, 10% glicerol, 0.1% Triton X-100). Los núcleos aislados se lisaron utilizando el buffer RIPA (10 mM Tris-HCl, 1 mM EDTA, 0.5 mM EGTA, 0.5% Triton X-100, 0.1% SDS, 140 mM NaCl) y las proteínas nucleares se cuantificaron con el método de Bradford.

Después, las proteínas se separaron en un gel de poliacrilamida y se transfirieron a una membrana de PDVF. Después de la transferencia, la membrana se bloqueó por 2 h con 5% de leche e incubada con el anticuerpo anti-CTCF (Donado por el Dr. Félix Recillas) 1:1,000 ó anti-SMC1 (ThermoFisher, A300-055A) 1:1,000. Finalmente, las membranas se incubaron por 2 horas con el anticuerpo secundario acoplado a HRP apropiado y la presencia de las proteínas se detectó con el sustrato quimioluminiscente (Millipore) en un equipo C-DiGitBlot (Li-COR). Los ensayos de Western Blot se realizaron por duplicado.

ATAC-seq

Se utilizó la transposasa hiperactiva Tn5 cargada con adaptadores fluorescentes con la molécula Atto-590N en el extremo 5' de cada adaptador. Se partió de 150,000 células previamente adheridas a laminillas con poli-L-lisina y fijadas. Se permeabilizaron con buffer de lisis (Tris-Cl pH 7.4 10mM, NaCl 10 mM, MgCl₂ 3mM, Igepal 0.01%) por 10 minutos a temperatura ambiente y se lavaron dos veces con PBS. Se añadió el buffer de transposición

(25 μ L de BufferTD 2X, 100 nM Tn5-ATTO-590N, 25 μ L dH₂O) y se incubó durante 30 minutos a 37°C en una cámara húmeda. Se realizó a la par un control negativo con EDTA 50 mM presente en el buffer de transposición, lo cual inhibe la actividad de la transposasa Tn5. Al terminar la reacción se realizaron tres lavados a 55°C con PBS + SDS 0.01% + EDTA 50 mM para inactivar a la enzima. Posteriormente se incubaron las laminillas con DAPI 300 nM por 5 minutos y se lavó una vez más con PBS. Las preparaciones se montaron con medio de montaje.

Las preparaciones se observaron en el microscopio confocal vertical LEICA TCS SP5 con un objetivo de inmersión en aceite 100X y un zoom digital de 10X. Se emplearon los láseres UV-405 nm y He/Ne-543 nm y se realizaron reconstrucciones en Z de 30 cortes ópticos de 0.2 μ m a través del software LEICA Application Suite X. Se realizaron dos réplicas del experimento y se analizaron 50 células en total para cada tipo celular.

Marcaje fluorescente de RNA naciente

Se detectó la síntesis de RNA mediante el marcaje del RNA naciente con un nucleósido modificado utilizando el kit Click-iT RNA imaging (Thermo Fisher). Se llevaron a cabo la reacción incubando con EU 1 mM por 6 horas. En el caso de las células HD3 y fibroblastos, previo a la incubación se promovió la adhesión de las células a cubreobjetos de vidrio pretratados con poli-L-lisina. Los eritrocitos embrionarios y adultos se mantuvieron en suspensión durante toda la incubación, posteriormente se centrifugaron y adhirieron a un cubreobjetos de vidrio.

Inmediatamente después de la incubación con EU, se retiró el medio y se fijaron las células con formaldehído al 3.7% por 15 minutos y posteriormente permeabilizadas con Triton X-100 0.5% por 15 minutos para proceder a la detección del RNA marcado, en la que se utiliza CuSO₄ como catalizador de una reacción de cicloadición entre la EU y un azida fluorescente (Alexa Fluor 594) por 30 minutos. Se tiñó el núcleo con Hoechst 33342 por 15 minutos y se montaron los cubreobjetos para su inspección bajo el microscopio confocal.

Procesamiento digital de imágenes

Todo el procesamiento digital se realizó en el software libre Fiji. Se delimitaron los márgenes de cada núcleo mediante la aplicación de un filtro Gaussiano sobre la imagen obtenida para la señal de DAPI. Con los márgenes generados se filtró la señal de ATAC-seq proveniente del ADN mitocondrial. Se estimó la fluorescencia promedio para núcleo y se empleó un filtro para detectar aquellos píxeles más brillantes que el promedio.

Con el fin de evaluar la distribución preferencial de las marcas estudiadas respecto a su posición hacia la periferia nuclear se generaron mapas de distancia que segmentan a los núcleos en círculos radiales de 1 píxel de ancho desde la periferia hasta el centro del núcleo y se estimó la intensidad promedio en cada sección. Se graficó la intensidad promedio de cada sección en función de su distancia a la periferia nuclear en RStudio. Para ello, se normalizó la intensidad de cada fluoróforo a su intensidad máxima y se normalizó el tamaño nuclear. Para cada ensayo se analizaron 50 núcleos y se computó un trazo promedio. Los análisis de correlación de Pearson entre la intensidad de señal y la distancia a la periferia se realizaron en R.

X. Referencias

1. Waddington, C. H. *Sym. Soc. Exp. Biol.* 7, 186; Principles of Embryology (G. Allen & Unwin Ltd., London, 1956).
2. Holliday, R. Epigenetics: a historical overview. *Epigenetics* 1, 76-80 (2006).
3. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251-60 (1997).
4. Kepper, N., Foethke, D., Stehr, R., Wedemann, G. & Rippe, K. Nucleosome geometry and internucleosomal interactions control the chromatin fiber conformation. *Biophys J* 95, 3692-705 (2008).
5. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* 421, 448-53 (2003).
6. Mora-Bermudez, F. & Ellenberg, J. Measuring structural dynamics of chromosomes in living cells by fluorescence microscopy. *Methods* 41, 158-67 (2007).
7. Penagos-Puig, A. & Furlan-Magaril, M. Heterochromatin as an Important Driver of Genome Organization. *Front Cell Dev Biol* 8, 579137 (2020).
8. Martin, R.M. & Cardoso, M.C. Chromatin condensation modulates access and binding of nuclear proteins. *FASEB J* 24, 1066-72 (2010).
9. Morales, V. & Richard-Foy, H. Role of histone N-terminal tails and their acetylation in nucleosome dynamics. *Mol Cell Biol* 20, 7230-7 (2000).
10. Bannister, A.J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* 21, 381-95 (2011).
11. Trojer, P. & Reinberg, D. Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* 28, 1-13 (2007).
12. Bulut-Karslioglu, A. et al. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol Cell* 55, 277-90 (2014).
13. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410, 116-20 (2001).
14. Francis, N.J., Kingston, R.E. & Woodcock, C.L. Chromatin compaction by a polycomb group protein complex. *Science* 306, 1574-7 (2004).
15. Lichter, P., Cremer, T., Borden, J., Manuelidis, L. & Ward, D.C. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum Genet* 80, 224-34 (1988).
16. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2, 292-301 (2001).
17. Chen, H. et al. Dynamic interplay between enhancer-promoter topology and gene activity. *Nat Genet* 50, 1296-1303 (2018).
18. Williamson, I. et al. Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev* 28, 2778-91 (2014).
19. Davies, H.G. Fine structure of heterochromatin in certain cell nuclei. *Nature* 214, 208-10 (1967).
20. Reddy, K.L., Zullo, J.M., Bertolino, E. & Singh, H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* 452, 243-7 (2008).

21. Girelli, G. et al. GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat Biotechnol* **38**, 1184-1193 (2020).
22. Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. *Nat Rev Genet* **21**, 207-226 (2020).
23. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-11 (2002).
24. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
25. Rao, S.S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).
26. Anania, C. & Lupianez, D.G. Order and disorder: abnormal 3D chromatin organization in human disease. *Brief Funct Genomics* **19**, 128-138 (2020).
27. Mota-Gomez, I. & Lupianez, D.G. A (3D-Nuclear) Space Odyssey: Making Sense of Hi-C Maps. *Genes (Basel)* **10**(2019).
28. Montavon, T. et al. A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132-45 (2011).
29. Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453-65 (2002).
30. Luo, Z., Rhie, S.K., Lay, F.D. & Farnham, P.J. A Prostate Cancer Risk Element Functions as a Repressive Loop that Regulates HOXA13. *Cell Rep* **21**, 1411-1417 (2017).
31. Eagen, K.P., Aiden, E.L. & Kornberg, R.D. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc Natl Acad Sci U S A* **114**, 8764-8769 (2017).
32. Li, G. et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98 (2012).
33. Furlan-Magaril, M. et al. The global and promoter-centric 3D genome organization temporally resolved during a circadian cycle. *Genome Biol* **22**, 162 (2021).
34. Schoenfelder, S. et al. Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat Genet* **47**, 1179-1186 (2015).
35. Guo, Y. et al. CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice. *Proc Natl Acad Sci U S A* **109**, 21081-6 (2012).
36. Guo, Y. et al. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900-10 (2015).
37. Dixon, J.R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-80 (2012).
38. Vietri Rudan, M. et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep* **10**, 1297-309 (2015).
39. Dixon, J.R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-6 (2015).
40. Zuin, J. et al. Nonlinear control of transcription through enhancer-promoter interactions. *Nature* **604**, 571-577 (2022).
41. Lupianez, D.G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).
42. Xu, Z. et al. Structural variants drive context-dependent oncogene activation in cancer. *Nature* **612**, 564-572 (2022).
43. Meister, P., Towbin, B.D., Pike, B.L., Ponti, A. & Gasser, S.M. The spatial dynamics of tissue-specific promoters during *C. elegans* development. *Genes Dev* **24**, 766-82 (2010).

44. Osborne, C.S. et al. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* **36**, 1065-71 (2004).
45. Towbin, B.D., Gonzalez-Sandoval, A. & Gasser, S.M. Mechanisms of heterochromatin subnuclear localization. *Trends Biochem Sci* **38**, 356-63 (2013).
46. Guelen, L. et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948-51 (2008).
47. Pickersgill, H. et al. Characterization of the *Drosophila melanogaster* genome at the nuclear lamina. *Nat Genet* **38**, 1005-14 (2006).
48. Meuleman, W. et al. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res* **23**, 270-80 (2013).
49. Gesson, K. et al. A-type lamins bind both hetero- and euchromatin, the latter being regulated by lamina-associated polypeptide 2 alpha. *Genome Res* **26**, 462-73 (2016).
50. Solovei, I. et al. LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell* **152**, 584-98 (2013).
51. Gonzalez-Sandoval, A. et al. Perinuclear Anchoring of H3K9-Methylated Chromatin Stabilizes Induced Cell Fate in *C. elegans* Embryos. *Cell* **163**, 1333-47 (2015).
52. Palka, M. et al. Laminopathies: what can humans learn from fruit flies. *Cell Mol Biol Lett* **23**, 32 (2018).
53. Solovei, I. et al. Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* **137**, 356-68 (2009).
54. van Berkum, N.L. et al. Hi-C: a method to study the three-dimensional architecture of genomes. *J Vis Exp* (2010).
55. Rippe, K. Liquid-Liquid Phase Separation in Chromatin. *Cold Spring Harb Perspect Biol* **14**(2022).
56. Strom, A.R. et al. Phase separation drives heterochromatin domain formation. *Nature* **547**, 241-245 (2017).
57. Larson, A.G. et al. Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236-240 (2017).
58. Arzate-Mejia, R.G., Recillas-Targa, F. & Corces, V.G. Developing in 3D: the role of CTCF in cell differentiation. *Development* **145**(2018).
59. Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. & Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc Natl Acad Sci U S A* **109**, 17507-12 (2012).
60. Filippova, G.N. et al. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol* **16**, 2802-13 (1996).
61. Chen, H., Tian, Y., Shu, W., Bo, X. & Wang, S. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS One* **7**, e41374 (2012).
62. Nora, E.P. et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e22 (2017).
63. Weintraub, A.S. et al. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**, 1573-1588 e28 (2017).
64. Chien, R. et al. Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression. *J Biol Chem* **286**, 17870-8 (2011).
65. Hou, C., Dale, R. & Dean, A. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc Natl Acad Sci U S A* **107**, 3651-6 (2010).

66. Makrantonis, V. & Marston, A.L. Cohesin and chromosome segregation. *Curr Biol* **28**, R688-R693 (2018).
67. Gillespie, P.J. & Hirano, T. Scc2 couples replication licensing to sister chromatid cohesion in *Xenopus* egg extracts. *Curr Biol* **14**, 1598-603 (2004).
68. Kueng, S. et al. Wapl controls the dynamic association of cohesin with chromatin. *Cell* **127**, 955-67 (2006).
69. Rao, S.S.P. et al. Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e24 (2017).
70. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56 (2017).
71. Zhang, Y., Zhang, X., Dai, H.Q., Hu, H. & Alt, F.W. The role of chromatin loop extrusion in antibody diversification. *Nat Rev Immunol* **22**, 550-566 (2022).
72. Nora, E.P. et al. Molecular basis of CTCF binding polarity in genome folding. *Nat Commun* **11**, 5612 (2020).
73. Zhang, Y. et al. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* **504**, 306-310 (2013).
74. Ke, Y. et al. 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell* **170**, 367-381 e20 (2017).
75. Hug, C.B., Grimaldi, A.G., Kruse, K. & Vaquerizas, J.M. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169**, 216-228 e19 (2017).
76. Zhang, S., Ubelmesser, N., Barbieri, M. & Papanonis, A. Enhancer-promoter contact formation requires RNAPII and antagonizes loop extrusion. *Nat Genet* **55**, 832-840 (2023).
77. Barshad, G. et al. RNA polymerase II dynamics shape enhancer-promoter interactions. *Nat Genet* (2023).
78. Mei, Y., Liu, Y. & Ji, P. Understanding terminal erythropoiesis: An update on chromatin condensation, enucleation, and reticulocyte maturation. *Blood Rev* **46**, 100740 (2021).
79. McGrath, K.E. et al. Enucleation of primitive erythroid cells generates a transient population of "pyrenocytes" in the mammalian fetus. *Blood* **111**, 2409-17 (2008).
80. Kingsley, P.D., Malik, J., Fantauzzo, K.A. & Palis, J. Yolk sac-derived primitive erythroblasts enucleate during mammalian embryogenesis. *Blood* **104**, 19-25 (2004).
81. Jayapal, S.R. et al. Down-regulation of Myc is essential for terminal erythroid maturation. *J Biol Chem* **285**, 40252-65 (2010).
82. Ji, P., Yeh, V., Ramirez, T., Murata-Hori, M. & Lodish, H.F. Histone deacetylase 2 is required for chromatin condensation and subsequent enucleation of cultured mouse fetal erythroblasts. *Haematologica* **95**, 2013-21 (2010).
83. Popova, E.Y. et al. Chromatin condensation in terminally differentiating mouse erythroblasts does not involve special architectural proteins but depends on histone deacetylation. *Chromosome Res* **17**, 47-64 (2009).
84. Ji, P., Murata-Hori, M. & Lodish, H.F. Formation of mammalian erythrocytes: chromatin condensation and enucleation. *Trends Cell Biol* **21**, 409-15 (2011).
85. Wong, P. et al. Gene induction and repression during terminal erythropoiesis are mediated by distinct epigenetic changes. *Blood* **118**, e128-38 (2011).
86. Hattangadi, S.M. et al. Histones to the cytosol: exportin 7 is essential for normal terminal erythroid nuclear maturation. *Blood* **124**, 1931-40 (2014).
87. Zhao, B. et al. Nuclear Condensation during Mouse Erythropoiesis Requires Caspase-3-Mediated Nuclear Opening. *Dev Cell* **36**, 498-510 (2016).

88. Hartenstein, V. Blood cells and blood cell development in the animal kingdom. *Annu Rev Cell Dev Biol* **22**, 677-712 (2006).
89. Nagai, H. & Sheng, G. Definitive erythropoiesis in chicken yolk sac. *Dev Dyn* **237**, 3332-41 (2008).
90. Sheng, G. Primitive and definitive erythropoiesis in the yolk sac: a bird's eye view. *Int J Dev Biol* **54**, 1033-43 (2010).
91. Glomsky, C. A., y Pica, A. (2016). *The Avian Erythrocyte: Its Phylogenetic Odyssey*. CRC Press: 644 pp.
92. Koutzamani, E., Loborg, H., Sarg, B., Lindner, H.H. & Rundquist, I. Linker histone subtype composition and affinity for chromatin in situ in nucleated mature erythrocytes. *J Biol Chem* **277**, 44688-94 (2002).
93. Rousseau, S., Asselin, M., Renaud, J. & Ruiz-Carrillo, A. Transcription of the histone H5 gene is regulated by three differentiation-specific enhancers. *Mol Cell Biol* **13**, 4904-17 (1993).
94. Carter, G.J. & van Holde, K. Self-association of linker histone H5 and of its globular domain: evidence for specific self-contacts. *Biochemistry* **37**, 12477-88 (1998).
95. Orrego, M. et al. Differential affinity of mammalian histone H1 somatic subtypes for DNA and chromatin. *BMC Biol* **5**, 22 (2007).
96. Bergman, M.G., Wawra, E. & Winge, M. Chicken histone H5 inhibits transcription and replication when introduced into proliferating cells by microinjection. *J Cell Sci* **91 (Pt 2)**, 201-9 (1988).
97. Fan, L. & Roberts, V.A. Complex of linker histone H5 with the nucleosome and its implications for chromatin packing. *Proc Natl Acad Sci U S A* **103**, 8384-9 (2006).
98. Zhou, B.R. et al. Structural Mechanisms of Nucleosome Recognition by Linker Histones. *Mol Cell* **59**, 628-38 (2015).
99. Haye, K.R. & Schlegel, R.A. Influence of histone H5 on mononucleosome structure during differentiation in the avian erythroid series. *Exp Cell Res* **157**, 504-10 (1985).
109. Sun, J.M., Ali, Z., Lurz, R. & Ruiz-Carrillo, A. Replacement of histone H1 by H5 in vivo does not change the nucleosome repeat length of chromatin but increases its stability. *EMBO J* **9**, 1651-8 (1990).
101. Baumann, R., Gotz, R. & Dragon, S. NTP pattern of avian embryonic red cells: role of RNA degradation and AMP deaminase/5'-nucleotidase activity. *Am J Physiol Regul Integr Comp Physiol* **284**, R771-9 (2003).
102. Zentgraf, H., Scheer, U. & Franke, W.W. Characterization and localization of the RNA synthesized in mature avian erythrocytes. *Exp Cell Res* **96**, 81-95 (1975).
103. Kantidze, O.L., Iarovaia, O.V., Philonenko, E.S., Yakutenko, II & Razin, S.V. Unusual compartmentalization of CTCF and other transcription factors in the course of terminal erythroid differentiation. *Biochim Biophys Acta* **1773**, 924-33 (2007).
104. Chen, X. et al. ATAC-se reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat Methods* **13**, 1013-1020 (2016).
105. Peterlin, B.M., Brogie, J.E. & Price, D.H. 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. *Wiley Interdiscip Rev RNA* **3**, 92-103 (2012).
106. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev* **33**, 960-982 (2019).
107. Zhao, B. et al. Disruption of erythroid nuclear opening and histone release in myelodysplastic syndromes. *Cancer Med* **8**, 1169-1174 (2019).
108. Doss, J.F. et al. A comprehensive joint analysis of the long and short RNA transcriptomes of human erythrocytes. *BMC Genomics* **16**, 952 (2015).

109. Sun, L., Yu, Y., Niu, B. & Wang, D. Red Blood Cells as Potential Repositories of MicroRNAs in the Circulatory System. *Front Genet* **11**, 442 (2020).
110. Abuhashem, A., Garg, V. & Hadjantonakis, A.K. RNA polymerase II pausing in development: orchestrating transcription. *Open Biol* **12**, 210220 (2022).
111. Gariglio, P., Bellard, M. & Chambon, P. Clustering of RNA polymerase B molecules in the 5' moiety of the adult beta-globin gene of hen erythrocytes. *Nucleic Acids Res* **9**, 2589-98 (1981).
112. Murphy, Z.C. et al. Regulation of RNA polymerase II activity is essential for terminal erythroid maturation. *Blood* **138**, 1740-1756 (2021).
113. Falk, M. et al. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395-399 (2019).
114. Belaghzal, H. et al. Liquid chromatin Hi-C characterizes compartment-dependent chromatin interaction dynamics. *Nat Genet* **53**, 367-378 (2021).
115. Morera, D. et al. RNA-Seq reveals an integrated immune response in nucleated erythrocytes. *PLoS One* **6**, e26998 (2011).
116. St Paul, M., Paolucci, S., Barjesteh, N., Wood, R.D. & Sharif, S. Chicken erythrocytes respond to Toll-like receptor ligands by up-regulating cytokine transcripts. *Res Vet Sci* **95**, 87-91 (2013).
117. Ryzhkova, A., Taskina, A., Khabarova, A., Fishman, V. & Battulin, N. Erythrocytes 3D genome organization in vertebrates. *Sci Rep* **11**, 4414 (2021).

XI. Anexos

nature structural & molecular biology

Article

<https://doi.org/10.1038/s41594-023-01037-0>

RNA polymerase II pausing regulates chromatin organization in erythrocytes

Received: 14 June 2022

Accepted: 16 June 2023

Published online: 27 July 2023

 Check for updates

Andrés Penagos-Puig, Sherlyn Claudio-Galeana, Aura Stephenson-Gussinye, Karina Jácome-López, Amaury Aguilar-Lomas, Rosario Pérez-Molina & Mayra Furlan-Magaril  

Chicken erythrocytes are nucleated cells often considered to be transcriptionally inactive, although the epigenetic changes and chromatin remodeling that would mediate transcriptional repression and the extent of gene silencing during avian terminal erythroid differentiation are not fully understood. Here, we characterize the changes in gene expression, chromatin accessibility, genome organization and chromatin nuclear disposition during the terminal stages of erythropoiesis in chicken and uncover complex chromatin reorganization at different genomic scales. We observe a robust decrease in transcription in erythrocytes, but a set of genes maintains their expression, including genes involved in RNA polymerase II (Pol II) promoter-proximal pausing. Erythrocytes exhibit a reoriented nuclear architecture, with accessible chromatin positioned towards the nuclear periphery together with the paused RNA Pol II. In erythrocytes, chromatin domains are partially lost genome-wide, except at minidomains retained around paused promoters. Our results suggest that promoter-proximal pausing of RNA Pol II contributes to the transcriptional regulation of the erythroid genome and highlight the role of RNA polymerase in the maintenance of local chromatin organization.

Red blood cells (RBCs) are highly specialized cells that transport oxygen from the lungs to all tissues in the body. In mammals, terminal stages of erythroid differentiation involve a reduction in nuclear size, extensive chromatin condensation mediated by histone deacetylases¹⁻³ and remodeling of the nuclear envelope⁴, followed by the expulsion of the nucleus from the erythroblasts⁵. Erythropoiesis in nonmammalian vertebrates such as birds, however, gives rise to circulating erythrocytes that retain their nucleus in a highly compacted state⁶.

Genome-wide chromatin condensation and gene expression silencing in nucleated RBCs in chicken are promoted by the incorporation of the erythroid-specific linker histone H5 into chromatin⁷⁻⁹. In fact, chicken RBCs undergo a dramatic drop in RNA synthesis and are regarded as transcriptionally inert cells¹⁰⁻¹², although the exact extent of this repression has not been studied. The chicken nucleated erythrocyte has been useful in the study of chromatin structure and

transcriptional regulation¹³. However, the molecular mechanisms that mediate the extensive silencing of the erythroid genome are largely unknown and a study at high resolution of the chromatin remodeling and genome organization during this process is still lacking.

The three-dimensional (3D) organization of chromosomes in the interphase nucleus serves as a scaffold of chromatin interactions that ensures the execution of complex transcriptional programs¹⁴. In the last decades, Hi-C experiments unveiled multiple levels of genome organization occurring at different genomic scales¹⁵. At the megabase scale, the genome is partitioned into two compartments, A and B, which correspond to predominantly active and inactive regions of the genome, respectively¹⁶. At the submegabase scale, topologically associated domains (TADs) represent frequently interacting regions that are partly insulated from other neighboring sequences^{17,18}. TADs enable proper promoter–enhancer communication inside the domain

Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Mexico City, Mexico.

 e-mail: mfurlan@ifc.unam.mx

Nature Structural & Molecular Biology

and prevent interference by regulatory elements outside the TAD, and disruption of TAD boundaries leads to aberrant gene activation and disease^{19–21}, highlighting the role of TADs in the spatio-temporal regulation of gene expression. In vertebrates, TADs are formed by the loop-extrusion activity of the cohesin complex and CTCF^{22–24}, although the presence of CTCF-independent TADs has been reported^{22,25}. The existence of other factors that contribute to TAD formation is still under investigation, and it has been suggested that transcription or at least certain components of the transcriptional machinery might mediate TAD establishment^{26–28}.

Here, we studied the changes in chromatin accessibility and genome organization that take place in the terminal erythroid differentiation of chicken RBCs. We show that chicken erythrocytes undergo a complex multilayered remodeling of chromatin organization. First, we show that chicken RBCs have a unique reoriented nuclear architecture and relocate most of their accessible chromatin towards the nuclear periphery. Despite a genome-wide reduction in gene expression, chicken RBCs retain transcription of erythroid-function-related genes and genes involved in RNA Pol II promoter-proximal pausing. Furthermore, chromatin accessibility is maintained at promoters of inactive genes and colocalizes with the presence of RNA Pol II in a paused state. Hi-C experiments show a strengthening of chromatin compartmentalization consistent with confocal microscopy imaging of accessible chromatin, transcription and constitutive heterochromatin foci. Moreover, the erythroid genomes are largely devoid of TADs but retain local organization at minidomains around paused genes. Together, our data provide an in-depth characterization of chromatin remodeling during erythropoiesis in chicken. We identified a type of domain organization that persists in absence of transcription and propose that RNA Pol II binding might be sufficient to maintain local organization.

Results

Accessible chromatin locates at the nuclear periphery in adult RBCs

Chicken erythrocytes are regarded as mostly transcriptionally inactive cells that exhibit a dramatic drop in RNA synthesis²⁹. To investigate the chromatin remodeling process during terminal erythroid differentiation in chicken, we obtained primary embryonic RBCs (eRBCs) and adult RBCs (aRBCs) from 10-day-old chicken embryos and 30-week-old hens, respectively. In terms of transcription, these two populations represent definitive erythrocytes in which global transcription is already diminished and chromatin is compacted. However, in eRBCs active gene expression of erythroid genes and adult globin genes is more abundant. We also analyzed HD3 cells³⁰, which correspond to an early erythroblast, and primary fibroblasts were also used as a nonerythroid cell control.

First, we sought to explore the effect of this genome-wide silencing process on chromatin accessibility and nuclear architecture by labeling and imaging occupied accessible chromatin using the assay for transposase-accessible chromatin with visualization (ATAC-seq)³¹ (Fig. 1a and Extended Data Fig. 1a). ATAC-seq signal intensity in eRBCs

and aRBCs is lower than in erythroblasts, suggesting an overall decrease in accessibility during erythroid terminal differentiation (Fig. 1a and Extended Data Fig. 1b). However, chicken erythrocytes retain accessible chromatin even in the absence of transcription.

ATAC-seq of aRBCs revealed a nuclear organization in which accessible chromatin is located predominantly at the periphery and is less abundant at the center of the nucleus (Fig. 1a, arrowhead), as opposed to the conventional nuclear architecture observed in fibroblasts, erythroblasts and eRBCs. We applied a Gaussian filter to the raw images using a threshold to obtain all pixels brighter than the mean in both ATAC-seq and 4,6-diamidino-2-phenylindole (DAPI) channels in each nucleus (Methods). The images generated confirmed that accessible chromatin is enriched at the nuclear periphery in aRBCs while the DAPI signal is more intense towards the nuclear center (Fig. 1b and Extended Data Fig. 1c). In agreement with the reorientation of occupied accessible chromatin observed in aRBCs, ATAC-seq signal intensity in these cells is higher at the nuclear periphery and decreases to the nuclear interior (Fig. 1d). Next, we compared the signal intensity only at the nuclear periphery versus the nuclear center (Methods). We corroborated that the signal is enriched at the nuclear periphery, with intensity values significantly higher than at the nuclear center (Fig. 1e). Finally, Pearson's correlation analysis between signal intensity and distance from the nuclear rim also recapitulated the peripheral location preference of accessible chromatin in aRBCs (Extended Data Fig. 1d). Together, these results show that most accessible chromatin locates towards the nuclear periphery in aRBCs.

H3K9me3 histone modification seems to relocate in chicken erythrocytes as shown by immunostaining (Fig. 1c, arrow). In eRBCs and aRBCs, H3K9me3 is preferentially segregated into large foci located towards the center of the nucleus, instead of the conventional positioning of H3K9me3-rich constitutive heterochromatin being more enriched at the nuclear envelope, as seen in fibroblasts and erythroblasts.

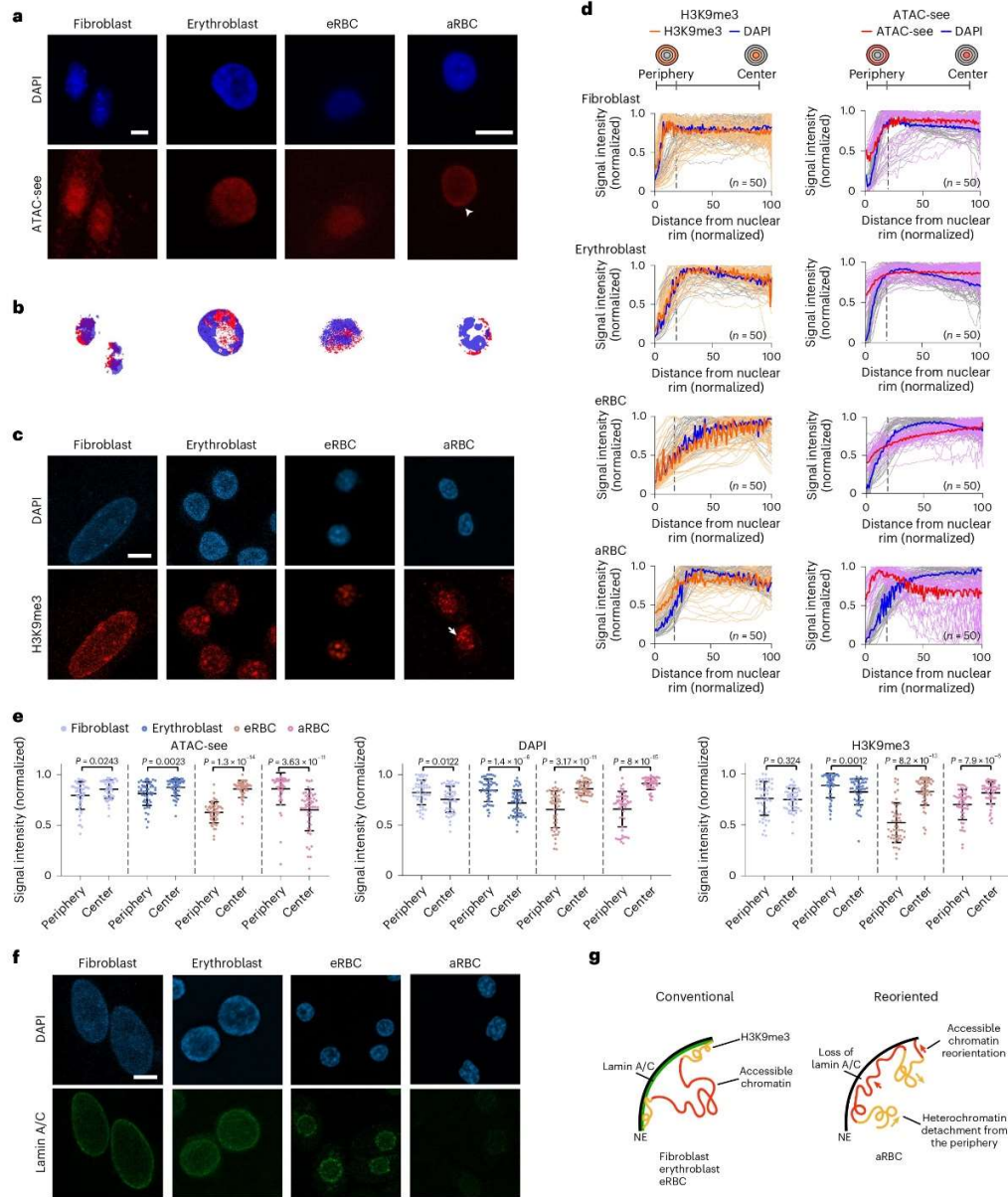
We validated the reoriented chromatin organization of aRBC nuclei by plotting the normalized mean intensity of H3K9me3 and ATAC-seq as a function of distance from the nuclear periphery (Fig. 1d). In fibroblasts and erythroblasts, H3K9me3 signal gradually decreases from the nuclear envelope to the interior where ATAC-seq signal is higher, consistent with a conventional nuclear architecture. In contrast, in eRBCs and aRBCs, H3K9me3 increases from the nuclear periphery to the interior, reflecting the position of the heterochromatic foci we observed in these cells. We next measured the H3K9me3 signal intensity only at the periphery and the nuclear center and confirmed that the signal is higher at the center in RBCs compared with fibroblasts and erythroblasts, in which the signal is more intense at the nuclear periphery (Fig. 1e). Pearson's correlation analysis between the H3K9me3 signal intensity and distance from the nuclear rim also recapitulated that the majority of H3K9me3 signal occupies a more central position in RBCs (Extended Data Fig. 1d). Finally, we measured the signal distribution of the DAPI staining. DAPI measures double-stranded DNA, but its accumulation is a good proxy of compact chromatin. DAPI signal distribution is also found to be enriched at the center of the nuclei compared

Fig. 1 | Chromatin remodeling during chicken erythropoiesis results in a reoriented nuclear architecture with accessible chromatin at the nuclear rim. a, Imaging of open chromatin by ATAC-seq (red) and DAPI (blue, nuclei) in fibroblasts and erythroid cells. Repositioning of the accessible chromatin towards the nuclear periphery in aRBCs is indicated with an arrowhead. Scale bar, 5 μ m. Representative results from two independent experiments are shown. **b**, Brightest pixels enriched in ATAC-seq (red) or DAPI (blue). **c**, Immunofluorescence imaging of H3K9me3 (red) and DAPI. In aRBCs the heterochromatin is relocated towards the nuclear center (arrows). Scale bar, 5 μ m. Representative results from two independent experiments are shown. **d**, For each cell type, H3K9me3, ATAC-seq and DAPI intensity are presented as a function of distance to the nuclear periphery. Individual tracks for 50 nuclei are

presented in lighter colors and the mean intensity is presented in bold. **e**, Signal intensity for the ATAC-seq, DAPI and H3K9me3 measured only at the nuclear periphery and center for all cell types ($n = 50$ nuclei from two independent experiments). Scatter dot plot shows mean \pm s.d. Significant differences between peripheral and central intensities were estimated with a Mann-Whitney *U* test. **f**, Fluorescence imaging of Lamin A/C proteins (green) and DAPI. In every case, maximal projections of 3D stacks are presented. Scale bar, 5 μ m. Representative results from two independent experiments are shown. **g**, In contrast to the conventional nuclear architecture in fibroblasts, erythroblasts and eRBCs, in aRBCs the loss of the nuclear Lamin A/C leads to heterochromatin detachment from the nuclear periphery and relocation towards the nuclear center. NE, nuclear envelope.

with the periphery in RBCs, in contrast with the signal enriched at the nuclear periphery in fibroblasts and erythroblasts by the same analysis performed for the ATAC-seq and H3K9me3 signals (Fig. 1d,e and Extended Data Fig. 1d).

Thus, terminal erythroid differentiation results in repositioning most of the H3K9me3 away from the nuclear envelope in eRBCs and aRBCs, and the location of most accessible chromatin towards the periphery in aRBCs.



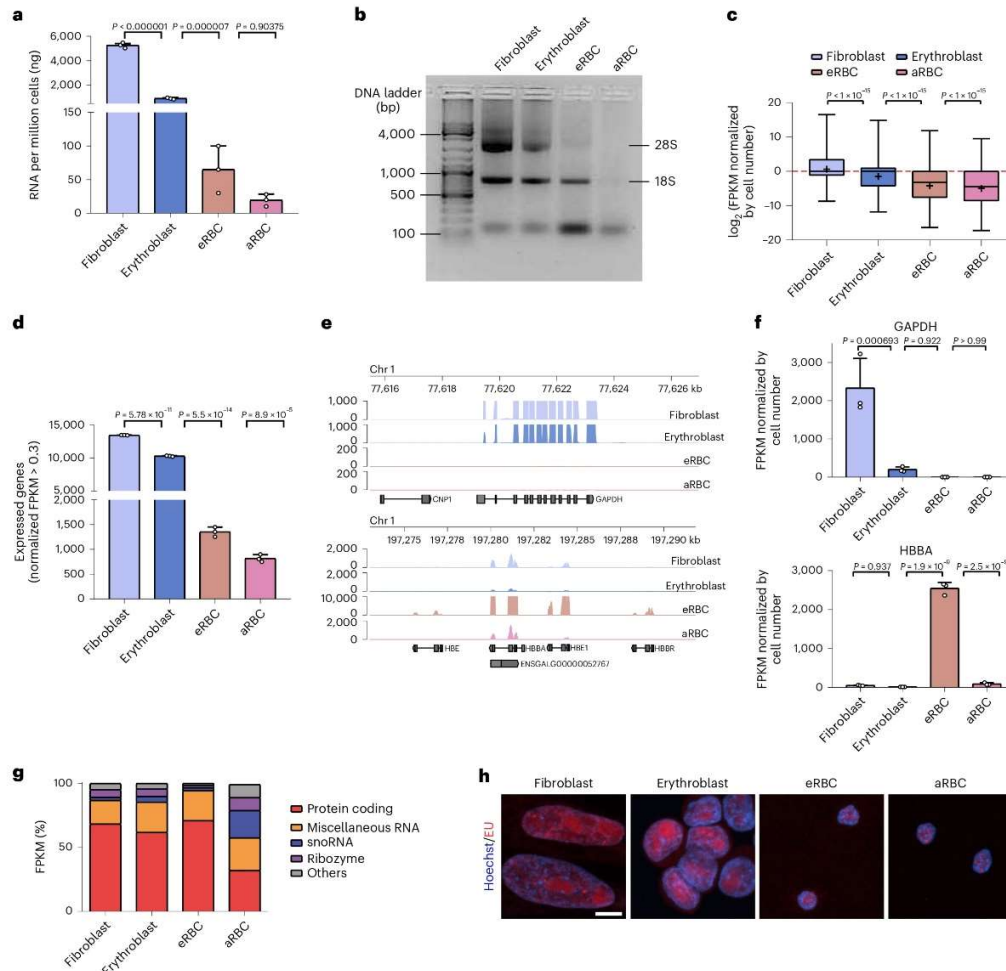


Fig. 2 | Chicken erythrocytes show a dramatic decrease in transcription but continue expressing small RNAs and erythroid genes. **a**, Quantification of total RNA obtained from chicken fibroblasts, erythroblasts, eRBCs and aRBCs. Shown are the mean \pm s.d. from three independent replicates. *P* values based on one-way ANOVA followed by Fisher's least significant difference post hoc test are stated. **b**, Representative picture of an agarose gel electrophoresis of total RNA. **c**, Mean log₂ FPKM values of three RNA-seq experiments. Boxplots show median (horizontal line), mean (+), and 25th and 75th percentiles (boxes), while whiskers denote the minima and maxima of the distribution. *P* values after a one-way ANOVA followed by Dunn's post hoc test are stated. **d**, Number of expressed genes (FPKM > 0.3)

according to three independent RNA-seq experiments. Data are presented as mean \pm s.d. *P* values and statistical differences based on one-way ANOVA followed by Tukey's post hoc test are shown. **e**, RNA-seq tracks for GAPDH (top) and the beta-globin locus (bottom). **f**, FPKM values for GAPDH (top) and HBBA (bottom) genes; shown are the values of three replicates with mean \pm s.d. *P* values after one-way ANOVA followed by Tukey's post hoc tests are presented. **g**, RNA diversity expressed as the percentage of total FPKM that corresponds to protein coding genes, miscellaneous RNA, snoRNA, ribozyme or others. **h**, Fluorescent labeling of nascent RNA with EU (red). Representative maximal projections from triplicates are presented. Scale bar, 5 μ m. ANOVA, analysis of variance.

The reoriented nuclear architecture observed in aRBCs has similarities with what has been previously shown to occur in the rod photoreceptor cells of nocturnal mammals in response to downregulation of the Lamin B receptor and absence of Lamin A/C expression¹². We asked whether the chromatin relocation observed in aRBCs could

be explained by changes in the nuclear envelope composition. Similar to the nuclear inversion in rod cells, Lamin A/C is present in the cells with conventional nuclear architecture but lost in the aRBC nuclei (Fig. 1f). Lamin B1 was detected in all cells studied (Extended Data Fig. 1e).

Taken together, these results show a complex reorganization of the nuclear architecture in chicken erythrocytes. Cells with a conventional nuclear architecture locate constitutive heterochromatin at the nuclear envelope, while accessible chromatin is preferentially positioned at the nuclear interior. During terminal differentiation, H3K9me3-marked heterochromatin moves away from the nuclear envelope concomitant with the loss of Lamin A/C in aRBCs, and the majority of accessible chromatin locates towards the nuclear periphery (Fig. 1g).

RBCs express small RNAs and erythroid genes

We wondered if the accessible chromatin detected by ATAC-seq in the chicken erythrocytes reflected sites of ongoing transcription in these cells. Quantitation of total RNA extracted from erythrocytes showed a dramatic decrease in RNA content in eRBCs and aRBCs (Fig. 2a) in agreement with a previous report of low levels of RNA synthesis in eRBCs measured by radioactivity incorporation¹⁰. Furthermore, the electrophoretic profile of the RNA obtained shows changes in the diversity of RNA species in eRBCs and aRBCs, with a gradual loss of the ribosomal RNA bands while small RNA molecules are still detectable (Fig. 2b).

To further characterize the transcriptional state of chicken erythrocytes, we performed total RNA sequencing (RNA-seq) experiments. A global downregulation of gene expression in the erythrocytes was evidenced by a marked decrease in fragments per kilobase of transcript per million reads (FPKM) values compared with other cell types (Fig. 2c). We identified all genes that were expressed (FPKM > 0.3; Methods) and found a small set of genes that remained active in the erythrocytes (1,354 in eRBCs and 804 in aRBCs) in agreement with quantitative PCR (qPCR) with reverse transcription (RT-qPCR) experiments (Fig. 2d–f and Extended Data Fig. 2a–d). Of note, most FPKMs come from small RNA molecules such as small nucleolar RNAs (snoRNAs) and miscellaneous RNAs in eRBCs and aRBCs, consistent with the electrophoretic pattern of erythroid RNA extractions (Fig. 2g).

Differential expression analysis was conducted between the different erythroid stages (Methods). We found 810 and 428 differentially expressed genes between eRBCs or aRBCs and erythroblasts, respectively (Extended Data Fig. 2f). Gene ontology analysis of upregulated differentially expressed genes showed that chicken erythrocytes exhibit upregulation of genes related to oxygen-carrier activity, erythroid differentiation and immune activity (Extended Data Fig. 2g). The presence of active transcription in chicken erythrocytes was corroborated *in vivo* by incorporation of 5-ethynyl uridine (EU) to nascent RNA and subsequent detection and imaging of EU (Fig. 2h, Extended Data Fig. 2e and Methods). Hence, chicken erythrocytes show a global downregulation of gene expression but retain transcription of a small number of genes associated with erythroid function.

Nascent RNA labeling allowed us to study the spatial location of transcription. As expected, RNA synthesis in fibroblasts and erythroblasts was distributed throughout the nucleus but more prominent at the center (Fig. 2h). In contrast, transcription in the erythrocytes was spatially restricted into foci, indicating a high degree of compartmentalization inside the nucleus. However, the presence of accessible chromatin at the nuclear envelope cannot be explained by active transcription since transcriptional foci were not preferentially located at the nuclear periphery in aRBCs.

Paused RNA polymerase occupies silent gene promoters in aRBCs

We performed assay for transposase-accessible chromatin using sequencing (ATAC-seq) experiments to identify the genomic regions that remain accessible in the chicken erythrocytes. We verified that ATAC-seq replicates of each cell type were highly correlated ($R > 0.92$; Extended Data Fig. 3a). Overall, we identified fewer ATAC-seq peaks in erythroid cells compared with fibroblasts, suggesting that chromatin accessibility is already diminished at the erythroblast stage (Fig. 3a). In fact, we identified similar numbers of ATAC-seq peaks in erythroblasts,

eRBCs and aRBCs regardless of the genome-wide decrease in gene expression in the erythrocytes. Most of the peaks in the erythrocytes represent constitutively open regions or regions shared with the erythroblasts, indicating that these regions retain their accessibility throughout terminal erythroid differentiation (Fig. 3b).

We classified the ATAC-seq peaks according to their genomic location and transcriptional state and found that many peaks in eRBCs and aRBCs correspond to silenced genes that remain accessible at their promoters (Fig. 3c–e and Extended Data Fig. 3b). Promoter-proximal pausing of RNA Pol II has been shown to maintain promoter chromatin accessibility and the gene's potential to be activated³⁵. We hypothesized that RNA Pol II pausing could be responsible for preserving the accessible promoters in the genome of eRBCs and aRBCs.

RNA Pol II pausing is promoted by inhibition of the positive elongation factor p-TEFb by the 7SK small nucleoprotein complex formed by the small nuclear RNA 7SK, HEXIM1/2, LARP1 and MePCE^{34,35}. Interestingly, 7SK and HEXIM1 are expressed in the erythrocytes (Fig. 3f). Furthermore, immunofluorescence of the RNA Pol II engaged in transcriptional initiation (Ser5-P) was readily detected in all cell types with no significant changes in signal intensity (Fig. 3g and Extended Data Fig. 3d), whereas the signal from the elongating CTD isoform (Ser2-P) is diminished in eRBCs and aRBCs (Extended Data Fig. 3c,e). In addition, we found that RNA Pol II Ser5-P is spatially located at the nuclear periphery in aRBCs (Fig. 3g, arrows), following the same distribution as the accessible chromatin detected by ATAC-seq (Extended Data Fig. 3f,h). However, the localization of RNA Pol II Ser2 is similar in all cell types, with the signal more enriched towards the nuclear interior (Extended Data Fig. 3g,i).

We next performed chromatin immunoprecipitation (ChIP) followed by sequencing (ChIP-seq) experiments against RNA Pol II Ser5 in erythroblasts, eRBCs and aRBCs. We detected 7,740 and 9,215 peaks in eRBCs and aRBCs, respectively, and from these 1,136 and 1,358 coincided with gene promoters. Most of these RNA Pol II Ser5 peaks (96% in eRBCs and 97% in aRBCs) correspond to promoters of genes without transcription (Fig. 3h). An example of a paused gene promoter in aRBCs is shown in Fig. 3i for the *TAF8* gene, a subunit of the general transcription factor complex TFIID. More examples are also shown in Extended Data Fig. 4a. The results show that paused RNA Pol II is indeed occupying a proportion of promoters of genes that are not actively transcribing in both eRBCs and aRBCs.

Finally, we confirmed the presence of RNA Pol II Ser5-P in the open promoter regions of genes without expression in the erythrocytes by ChIP-qPCR (Extended Data Fig. 4a,b). In contrast, no enrichment of RNA Pol II Ser5-P was detected on the inaccessible *Sox2* gene promoter. These results indicate that despite global reduction of gene expression, chicken erythrocytes retain expression of genes involved in RNA Pol II pausing and accessibility at gene promoters due to the presence of RNA Pol II in a promoter-proximal paused state.

RBC chromatin organizes in hypercompartments and minidomains

Recent Hi-C experiments conducted in aRBCs have evidenced a global rearrangement in their 3D genome organization characterized by an increase in long-range interactions and absence of TADs³⁶. However, characterization of the chromosome organization at distinct stages of erythroid differentiation is still lacking. We sought to explore how the observed changes in nuclear architecture, chromatin accessibility and gene expression relate to 3D genome organization, and generated Hi-C contact maps for chicken fibroblasts, erythroblasts, eRBCs and aRBCs (Extended Data Fig. 5a).

High-order chromatin organization is defined by the segregation of genomic regions into A and B compartments corresponding to active and inactive chromatin, respectively¹⁶. These compartments are observed in whole-chromosome Hi-C contact maps as a plaid pattern of long-range interactions. Visual inspection of Hi-C maps

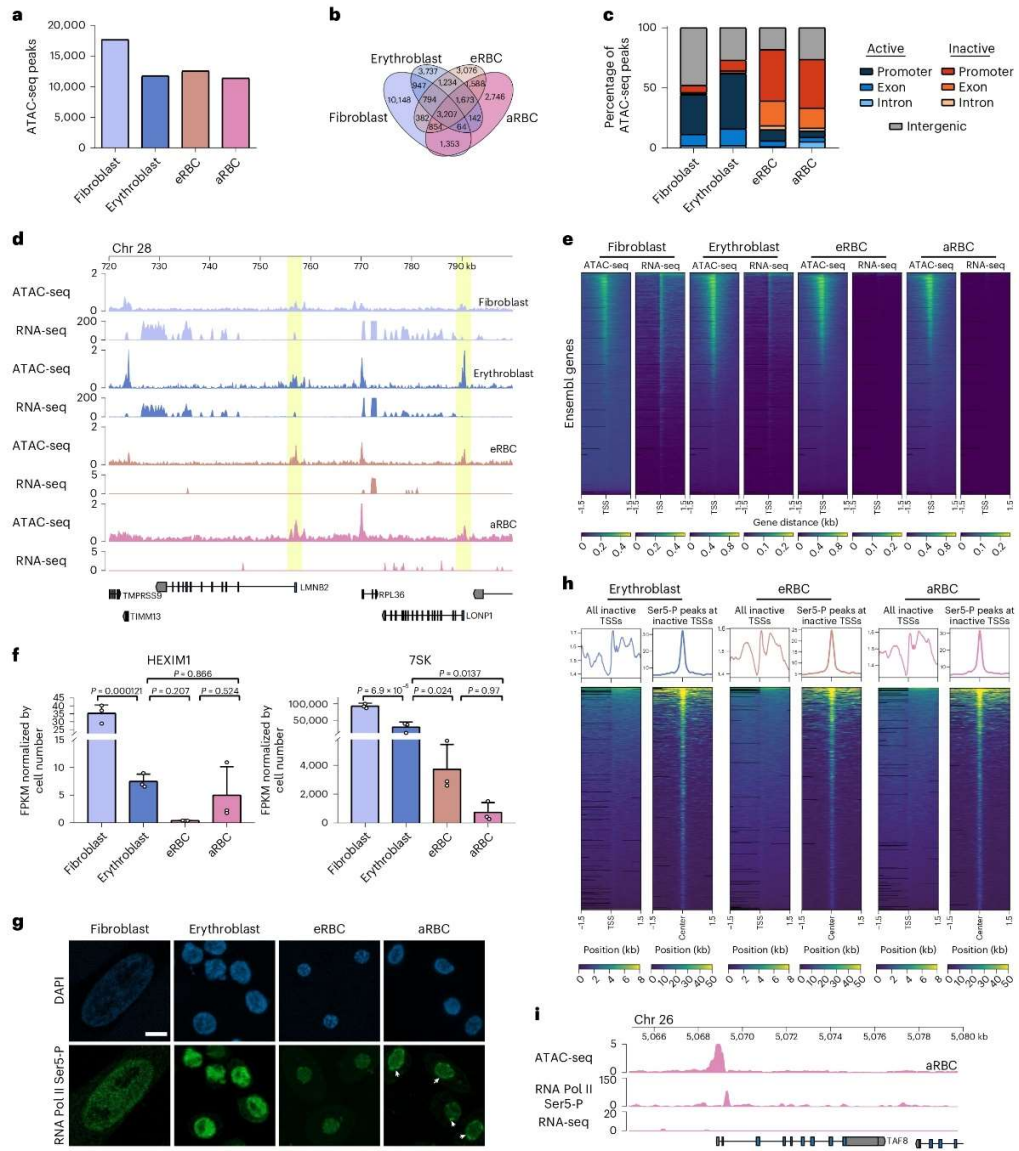


Fig. 3 | Open chromatin in the chicken erythrocytes comprises paused promoters of silenced genes. **a**, Number of ATAC-seq peaks detected in each cell type. **b**, Venn diagram of ATAC-seq peaks in each cell type. **c**, Annotation of ATAC-seq peaks at intergenic regions, promoters, exons and introns of transcriptionally active or inactive genes. **d**, ATAC-seq and RNA-seq tracks showing an example of inactive genes in eRBCs and aRBCs that retain accessibility at their promoter (yellow) in absence of transcription. **e**, ATAC-seq and RNA-seq CPM \pm 1.5 kb of annotated TSSs. **f**, Expression levels of HEXIM1 and 7SK, part of the 7SK small nuclear ribonucleoprotein involved in RNA polymerase

pausing ($n = 3$ independent RNA-seq experiments). Significant differences based on a one-way ANOVA followed by Tukey's post hoc test are stated. **g**, Imaging of the RNA Pol II Ser5-P. In aRBCs the RNA Pol II Ser5-P accumulates at the nuclear periphery (arrow). Maximal projections are presented. Representative results from three experiments are shown. Scale bar, 5 μ m. **h**, ChIP-seq against RNA Pol II Ser5-P in erythroblasts, eRBCs and aRBCs. CPM \pm 1.5 kb of TSSs of all inactive genes or \pm 1.5 kb of Ser5-P peaks in paused TSSs. **i**, Example of an RNA Pol II Ser5-P-occupied paused gene promoter in aRBCs. The TAF8 gene is an integral subunit of the general transcription factor complex TFIID.

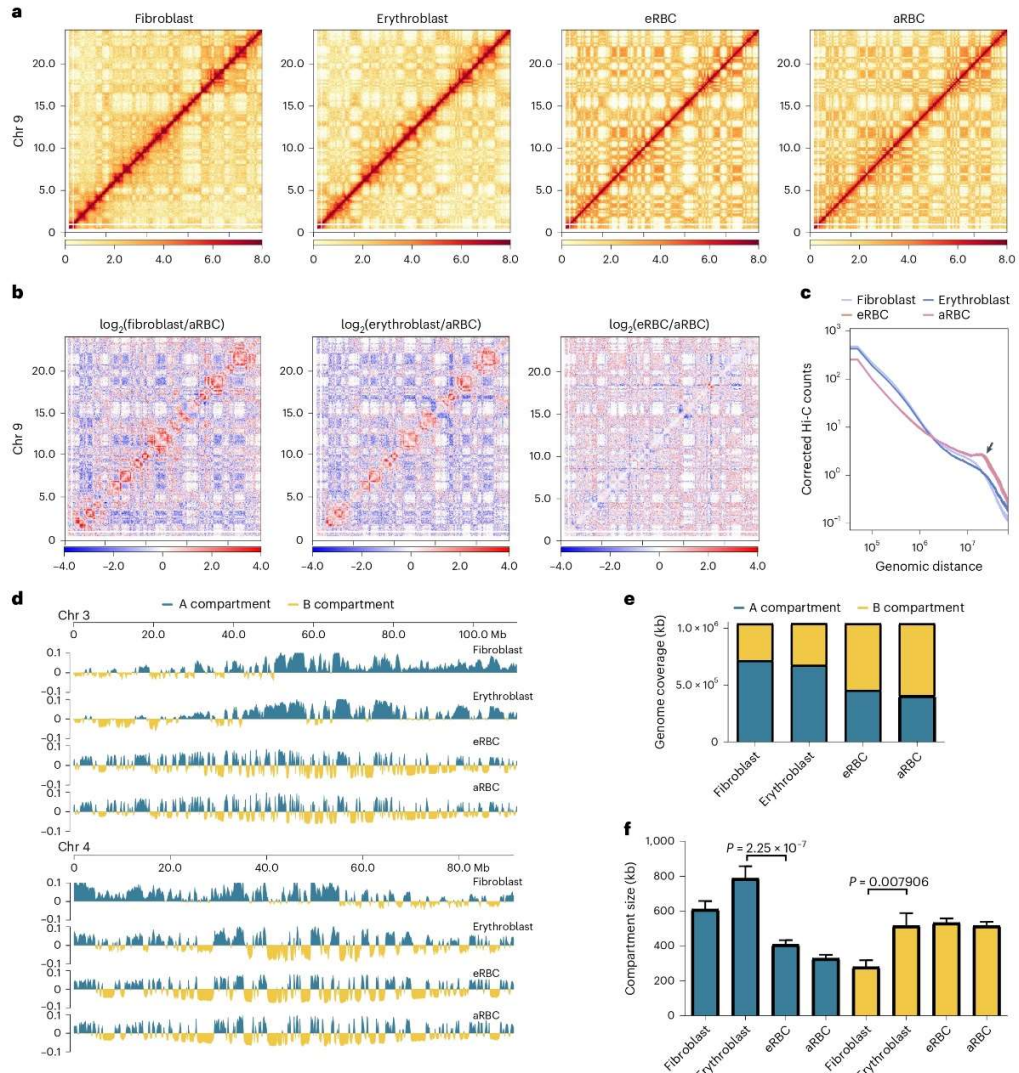


Fig. 4 | Chicken erythrocytes have hypercompartmentalized genomes. a, Hi-C contact matrices of chromosome 9 (Mb 0–24). **b**, Hi-C contact matrices showing the log₂ fold change between each cell type and aRBCs. **c**, Hi-C contacts count as a function of genomic distance. An increase in long-range interaction in eRBCs and aRBCs is indicated with an arrow. **d**, A (blue) and B (yellow) compartment identification by PCA analysis in fibroblasts, erythroblasts, eRBCs and aRBCs

in chromosomes 3 (top) and 4 (bottom). **e**, Genome coverage of the A and B compartments. **f**, Average size of A and B compartment regions; the genome was partitioned in 21,244 bins of 50 kb and continuous bins within the same compartment were grouped to estimate compartment size. Data are presented as mean ± s.e.m. *P* values based on one-way ANOVA followed by Bonferroni's post hoc test are indicated.

shows a more pronounced plaid pattern in eRBCs and aRBCs than in erythroblasts (Fig. 4a,b), indicative of the strengthening of chromatin segregation into A and B compartments in these cells. Comparison of Hi-C maps also shows a decrease in interactions along the diagonal.

Accordingly, we detected an increase in long-range (>10-megabase (Mb)) interactions in the erythrocytes while short-range interactions are less frequent (Fig. 4c). These results indicate that the genome of chicken erythrocytes is strongly compartmentalized, as evidenced

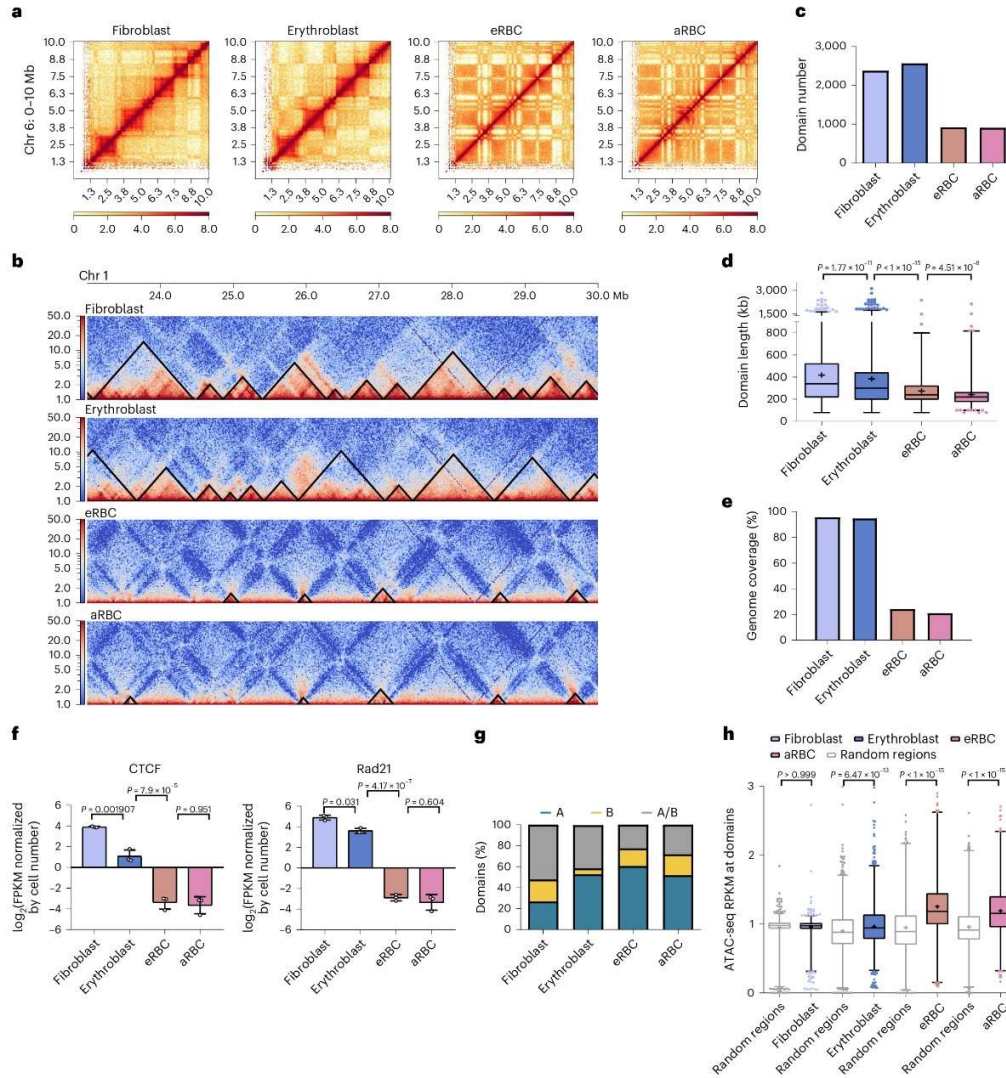


Fig. 5 | Chicken erythrocytes present global TAD loss but retain structure at minidomains enriched in open chromatin. a, Hi-C contact matrices of the first 10 Mb of chromosome 6. TADs are clearly visible in fibroblasts and erythroblasts while the most prominent signal in erythrocytes corresponds to the plaid pattern of compartments. **b**, Domains identified in each cell type indicated by black triangles. **c, d**, Number (**c**) and average size (**d**) of the domains in fibroblasts ($n = 2,373$ domains), erythroblasts ($n = 2,562$), eRBCs ($n = 916$) and aRBCs ($n = 903$) are presented. Boxplots show median (horizontal line), mean (+), and 25th and 75th percentiles. **e**, Percentage of the genome structured in domains. **f**, Expression levels of architectural proteins based on RNA-seq data from three independent experiments. Data are presented as mean \pm s.d. **g**, Percentage of domains that contain only A compartment regions (blue), only B compartment regions (yellow) or both (gray). **h**, ATAC-seq RPKM counts inside chromatin domains or a set of random regions of the same size in fibroblasts ($n = 2,373$ domains), erythroblasts ($n = 2,562$), eRBCs ($n = 916$) and aRBCs ($n = 903$). Boxplots show median (horizontal line), mean (+), and 25th and 75th percentiles (boxes), and whiskers denote the 1st and 99th percentiles. **P** values based on one-way ANOVA followed by Kruskal–Wallis post hoc test are stated.

f, Expression levels of architectural proteins based on RNA-seq data from three independent experiments. Data are presented as mean \pm s.d. **g**, Percentage of domains that contain only A compartment regions (blue), only B compartment regions (yellow) or both (gray). **h**, ATAC-seq RPKM counts inside chromatin domains or a set of random regions of the same size in fibroblasts ($n = 2,373$ domains), erythroblasts ($n = 2,562$), eRBCs ($n = 916$) and aRBCs ($n = 903$). Boxplots show median (horizontal line), mean (+), and 25th and 75th percentiles (boxes), and whiskers denote the 1st and 99th percentiles. **P** values based on one-way ANOVA followed by Kruskal–Wallis post hoc test are shown.

by the presence of well-defined heterochromatin and transcriptional foci in these cells.

Identification of A and B compartments showed a hypercompartmentalization pattern in the genome of the erythrocytes compared with erythroblasts, characterized by the disruption of long A-type regions by the emergence of regions in the B compartment (Fig. 4d). In fact, most of the genome falls into the B compartment in the erythrocytes, as regions in the A compartment are reduced in size compared with erythroblasts and fibroblasts (Fig. 4e,f). This hypercompartmentalized pattern with the B compartment ‘invading’ the A compartment in the erythrocytes reflects the genome-wide reduction in transcription in these cells.

Strong compartmentalization and decrease in short-range interactions make visual identification of topological domains difficult in eRBCs and aRBCs (Fig. 5a), and it has been suggested that chicken erythrocytes lack TADs³⁶. TAD calling using insulation index³⁷ uncovered a profound loss of TAD structures in chicken erythrocytes except at regions where a closer inspection reveals a local enrichment of short-range interactions (Fig. 5b and Extended Data Fig. 5b). These domains are fewer and smaller (with an average size of 241 kilobases (kb)) than TADs observed in erythroblasts and fibroblasts (Fig. 5c,d), and hence we called these structures minidomains. Of note, only a small fraction of the genome remains structured in minidomains (24.18% in eRBCs and 20.4% in aRBCs) (Fig. 5e).

Topological domain insulation depends on the presence of the architectural proteins CTCF and cohesin at the domain boundaries^{24,38,39}, and auxin-inducible depletion of CTCF or the cohesin-loading factor Nipbl causes genome-wide loss of topological domains^{22,23}. Interestingly, CTCF and Rad21 are downregulated in eRBCs and aRBCs (Fig. 5f) and CTCF and SMC1 protein levels are also reduced (Extended Data Fig. 5c). Thus, downregulation of architectural proteins might explain the loss of topological domains in chicken erythrocytes.

Notably, minidomains demarcate regions mostly overlapping the A compartment and are enriched in ATAC-seq reads (Fig. 5g,h). Taken together, these results indicate that despite a global loss of TADs in the genomes of eRBCs and aRBCs, euchromatic regions remain structured in minidomains.

Chicken RBC genomes are organized around paused promoters

We hypothesized that the presence of local structure at minidomains reflected transcriptional activity inside erythroid minidomains. Indeed, minidomains are enriched in RNA-seq reads (Extended Data Fig. 5d). However, visualization of Hi-C maps together with ATAC-seq and RNA-seq tracks showed that minidomains are often found around open regions of the genome that lack transcription (Fig. 6a and Extended Data Fig. 5b).

We classified topological domains by the transcriptional and chromatinic states of their transcription start sites (TSSs) (Fig. 6b). Domains within class I contain open promoters of genes without expression, herein referred to as ‘paused’ promoters, while TSSs within class II domains are open and transcriptionally active. Class III domains contain both paused and active promoters and class IV domains are depleted of open TSSs. This classification evidenced that the presence of paused promoters is a common feature in minidomains, as more than 80% of minidomains fall into either class I or class III.

Interestingly, paused genes that reside within minidomains in eRBCs (3,458 genes) and aRBCs (2,804 genes) are involved in regulation of gene expression, RNA processing and transport of mature RNAs to the cytoplasm (Fig. 6c). Examples of such paused genes are the transcriptional elongation factors *Htatsf1* and *Gtf2h5* (refs. 40,41), and the nuclear export protein *Pcid2* (ref. 42) (Fig. 3h). Paused genes in erythroblasts (726 genes) and fibroblasts (791 genes) were not enriched in any pathway. One possible explanation for this observation is that the preserved minidomains together with the paused RNA Pol II bookmark

the necessary components of the transcriptional machinery and RNA processing complexes to rapidly resume gene expression in response to environmental cues. For example, transcriptional activation of the erythroid genome has been observed in the nucleated erythrocytes of fish and birds in response to immune challenges⁴³. In summary, we show that paused promoters of genes involved in gene expression and RNA metabolism remain organized into minidomains despite a global loss of TADs in chicken eRBCs and aRBCs.

Discussion

Here, we analyzed thoroughly the changes in chromatin accessibility and organization that accompany terminal erythroid differentiation and downregulation of gene expression in nucleated chicken erythrocytes. We found that loss of Lamin A/C results in a reorientation of the nuclear architecture in aRBCs. Furthermore, the erythrocyte nucleus is a highly compartmentalized organelle that confines accessible chromatin, transcription and heterochromatin into readily observable distinct foci. Downregulation of architectural proteins greatly affects TADs genome-wide. However, RNA Pol II pausing retains accessibility at open promoters and supports short-range chromatin interactions, thus forming minidomains around genes important in RNA synthesis and RNA metabolism. Our results uncover a complex reorganization process at different genomic scales with important implications in the physiology of nucleated erythrocytes and the mechanisms that govern genome organization (Fig. 6d).

Chromatin condensation and gene silencing are common traits of mammalian and nonmammalian erythropoiesis^{39,44}. Nucleated erythrocytes are commonly described as transcriptionally inert cells⁴⁵ and defects in heterochromatin stability or gene silencing are considered hallmarks of some myelodysplastic syndromes and anemia in mammals and birds^{45–47}. Here, we provide evidence that both eRBCs and aRBCs actively transcribe a reduced number of genes (1,354 in eRBCs, 804 in aRBCs), including hemoglobin genes and genes involved in gas exchange and transport. These results challenge the notion of nucleated erythrocytes as transcriptionally inactive cells and suggest a functional role of the nucleus in the erythroid physiology.

By direct imaging of RNA synthesis in chicken erythrocytes, we found that transcription is confined into discrete active foci. In fact, segregation of euchromatin and heterochromatin into different compartments is exacerbated, as evidenced by imaging and Hi-C data. Chromatin compartmentalization prevents abnormal contacts between euchromatic and heterochromatic regions that may result in heterochromatin spreading and ectopic gene repression⁴⁸. The confinement of transcription into foci in erythrocytes probably reflects the requirement of bringing together the remaining active regions of the genome to favor transcription in an environment where transcriptional factors are scarce.

Heterochromatin is also relocated in eRBCs and aRBCs. In aRBCs, loss of Lamin A/C leads to a reorientation of nuclear architecture in which heterochromatin detaches from the nuclear envelope and moves towards the center of the nucleus. A similar spatial organization of open chromatin at the nuclear rim was described in the rod photoreceptors of nocturnal mammals, where loss of heterochromatin tethering to the nuclear envelope is triggered by downregulation of Lamin B receptor and absence of Lamin A/C³².

In recent years, heterochromatin has emerged as a key factor mediating genome organization, and heterochromatin repositioning has important effects on cellular function⁴⁹. For example, the inverted nuclear architecture in rod cells is thought to maximize light transmission in the retina, and disposition of accessible chromatin at the nuclear rim in neutrophils prepares the cells for initiation of NETosis³¹. Moreover, heterochromatin detachment and relocation cause gene misexpression in laminopathies and senescence⁴⁹. An increase in heterochromatin-mediated long-range interactions is observed in the last nucleated stage of human erythropoiesis⁵⁰, and Lamin B

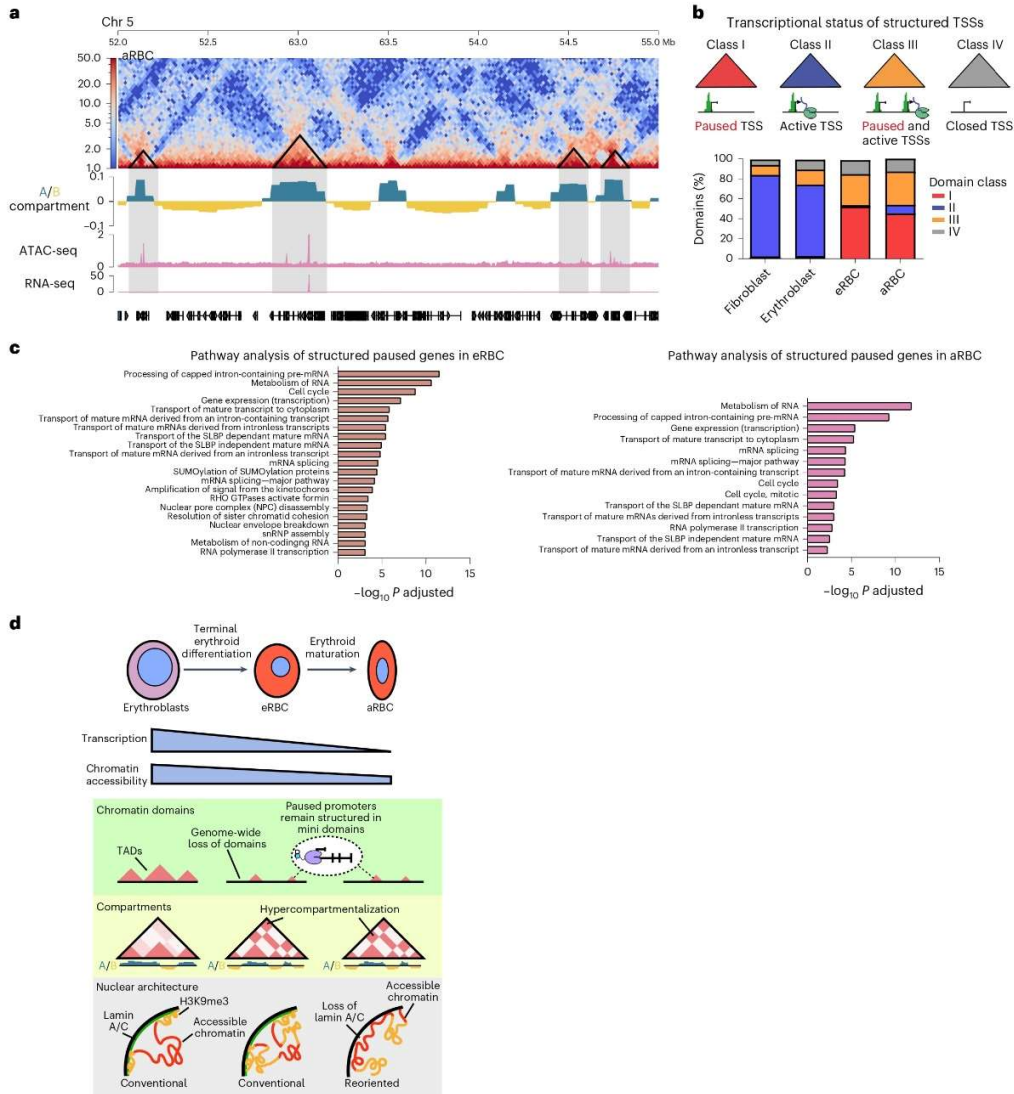


Fig. 6 | Minidomains contain and organize paused promoters in chicken erythrocytes. a, Hi-C contact matrix from aRBCs (5: 52–55 Mb), the minidomains identified (black triangles), the corresponding compartment signal, and ATAC-seq and RNA-seq tracks. aRBC minidomains contain ATAC-seq peaks independent from transcription (gray). **b**, Distribution of minidomains according to the transcriptional status of the contained TSSs into class I (paused TSSs), class II (active TSSs), class III (both paused and active TSSs) or class IV (closed TSSs). **c**, Gene ontology terms of paused TSSs structured in minidomains. Analysis was performed using Fisher’s exact test and the default multiple-hypotheses

testing method (gSCS) on gProfiler. **d**, Terminal erythroid differentiation in chicken results in a dramatic drop in transcription and chromatin accessibility that is exacerbated during erythroid maturation. Erythrocyte genomes lose TADs genome-wide but retain minidomains structured around open paused promoters with presence of the stalled RNA polymerase. At the compartment level the erythrocyte genome is hypercompartmentalized and rich in long-range interactions. Lastly, in aRBCs the absence of Lamin A/C leads to heterochromatin detachment from the nuclear periphery and relocation towards the nuclear center.

cleavage by caspase-3 and transient nuclear opening have been shown to promote chromatin condensation in the erythroblasts of mouse and human^{4,45}. Together with our results, this evidence suggests that nuclear envelope remodeling, compartment strengthening and heterochromatin relocation may be a conserved mechanism to ensure proper chromatin compaction.

The molecular forces driving genome compartmentalization are not entirely understood. Interestingly, knock-out of the cohesin-loading complex Nipbl or the cohesin subunit Rad21 causes a general enhancement of compartmentalization^{23,51}. Thus, compartmentalization may be regulated by the binding of cohesin to chromatin and loop-extrusion. Here, we found Rad21 to be downregulated in eRBCs and aRBCs which provides a possible explanation for the hypercompartmentalized state of their genomes.

Recent Hi-C experiments in chicken aRBCs and immature polychromatic erythrocytes using a six-cutter restriction enzyme showed a genome-wide depletion of TADs in these cells³⁶. Although a strong compartment signal obscures the presence of chromatin domains in the Hi-C matrices, here we provide compelling evidence of local interacting mini domains that persist as structured in the terminally differentiated eRBCs and aRBCs. TADs are formed when the loop-extrusion activity of the cohesin complex is blocked by the presence of CTCF bound at convergently oriented binding-sites located at TAD boundaries⁵². Ablation of CTCF results in a global loss of TADs except at regions where low levels of CTCF are able to maintain TAD boundaries^{22,53}. Similarly, degradation of Rad21 and Nipbl eliminates chromatin domains^{23,51}. Here, we show that transcription of CTCF and Rad21 as well as protein levels of CTCF and SMCI decrease in the nucleated erythrocytes of chicken, which may account for a loss of chromatin domains. TAD disturbance has been recently shown to occur in human erythroblasts without evident changes in chromatin compartmentalization⁵⁰. Thus, global genome reorganization represents a conserved feature of terminal erythropoiesis that reflects the particular transcriptional state of these cells.

The contribution of RNA polymerase to the establishment and maintenance of genome organization is still a matter of debate. The emergence of TADs during development coincides with the transcriptional activation of the zygotic genome²⁸. Inhibition of transcriptional elongation weakens TAD insulation but does not preclude TAD emergence^{19,28,54}. Hence, transcription is not necessary to establish TADs but might play a role in maintaining chromatin organization through other associated processes such as chromatin opening, the assembly of the preinitiation complex, and recruitment of transcription factors and the transcriptional machinery²⁸. Moreover, RNA Pol II dictates cohesin-loading into chromatin and rapid degradation of RNA Pol II prevents the reestablishment of chromatin folding in cells following exit from mitosis⁵⁵. Our results show that minidomains in the chicken erythrocytes frequently contain paused promoters occupied by the RNA Pol II Ser5-P, suggesting that RNA Pol II binding is enough to maintain local organization and insulation of small domains.

Promoter-proximal pausing of RNA Pol II has been shown to serve as a marker that primes silenced genes to be rapidly activated during development and in response to environmental cues⁵⁶. Here, we present evidence that suggests that RNA Pol II pausing represents a key mechanism by which chicken nucleated erythrocytes maintain low levels of transcription. We show that eRBCs and aRBCs sustain elevated levels of transcription of components of the small nuclear ribonucleoprotein 7SK complex involved in polymerase pausing, and observed the presence of RNA Pol II Ser5-P in the nucleus of these cells despite a global decrease in transcription, in agreement with early run-on experiments that identified the presence of a stalled form of the RNA Pol II in the beta-globin genes in hen erythrocytes⁵⁷. A recent study suggests that regulation of RNA Pol II activity is important during terminal erythroid maturation in humans⁵⁸.

Finally, paused promoters in eRBCs and aRBCs are enriched in categories concerning transcriptional elongation, RNA processing and nucleocytoplasmic transport of mature RNAs. We hypothesize that Pol II pausing prepares chicken nucleated erythrocytes to restore the expression of components of the transcriptional and RNA processing machineries to mount an efficient response to stimuli such as immune challenges^{43,59}.

Our results reveal a dramatic reorganization of the genome at different scales during the formation of nucleated erythrocytes in chicken, including the reorientation of most accessible chromatin to the nuclear periphery in aRBCs, and highlight the role of RNA Pol II in the maintenance of insulated domains in this highly repressive chromatin environment. We show that nucleated erythrocytes maintain a collection of accessible promoters by promoter-proximal pausing of the RNA Pol II at genes important to mounting a transcriptional response.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41594-023-01037-0>.

References

- Jayapal, S. R. et al. Down-regulation of Myc is essential for terminal erythroid maturation. *J. Biol. Chem.* **285**, 40252–40265 (2010).
- Popova, E. Y. et al. Chromatin condensation in terminally differentiating mouse erythroblasts does not involve special architectural proteins but depends on histone deacetylation. *Chromosome Res.* **17**, 47–64 (2009).
- Ji, P., Yeh, V., Ramirez, T., Murata-Hori, M. & Lodish, H. F. Histone deacetylase 2 is required for chromatin condensation and subsequent enucleation of cultured mouse fetal erythroblasts. *Haematologica* **95**, 2013–2021 (2010).
- Zhao, B. et al. Nuclear condensation during mouse erythropoiesis requires Caspase-3-mediated nuclear opening. *Dev. Cell* **36**, 498–510 (2016).
- Ji, P., Murata-Hori, M. & Lodish, H. F. Formation of mammalian erythrocytes: chromatin condensation and enucleation. *Trends Cell Biol.* **21**, 409–415 (2011).
- Hartenstein, V. Blood cells and blood cell development in the animal kingdom. *Annu. Rev. Cell Dev. Biol.* **22**, 677–712 (2006).
- Koutzamani, E., Loborg, H., Sarg, B., Lindner, H. H. & Rundquist, I. Linker histone subtype composition and affinity for chromatin in situ in nucleated mature erythrocytes. *J. Biol. Chem.* **277**, 44688–44694 (2002).
- Bergman, M. G., Wawra, E. & Winge, M. Chicken histone H5 inhibits transcription and replication when introduced into proliferating cells by microinjection. *J. Cell Sci.* **91**, 201–209 (1988).
- Fan, L. & Roberts, V. A. Complex of linker histone H5 with the nucleosome and its implications for chromatin packing. *Proc. Natl Acad. Sci. USA* **103**, 8384–8389 (2006).
- Zentgraf, H., Scheer, U. & Franke, W. W. Characterization and localization of the RNA synthesized in mature avian erythrocytes. *Exp. Cell Res.* **96**, 81–95 (1975).
- Baumann, R., Gotz, R. & Dragon, S. NTP pattern of avian embryonic red cells: role of RNA degradation and AMP deaminase/5'-nucleotidase activity. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **284**, R771–R779 (2003).
- Cameron, I. L. & Prescott, D. M. RNA and protein metabolism in the maturation of the nucleated chicken erythrocyte. *Exp. Cell Res.* **30**, 609–612 (1963).

13. Beacon, T. H. & Davie, J. R. The chicken model organism for epigenomic research. *Genome* **64**, 476–489 (2021).
14. Mota-Gomez, I. & Lupianez, D. G. A (3D-nuclear) space odyssey: making sense of Hi-C maps. *Genes (Basel)* **10**, 415 (2019).
15. Jerkovic, I. & Cavalli, G. Understanding 3D genome organization by multidisciplinary methods. *Nat. Rev. Mol. Cell Biol.* **22**, 511–528 (2021).
16. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
17. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
18. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
19. Arzate-Mejia, R. G., Josue Cerecedo-Castillo, A., Guerrero, G., Furlan-Magaril, M. & Recillas-Targa, F. In situ dissection of domain boundaries affect genome topology and gene transcription in *Drosophila*. *Nat. Commun.* **11**, 894 (2020).
20. Lupianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
21. Andrey, G. & Mundlos, S. The three-dimensional genome: regulating gene expression during pluripotency and development. *Development* **144**, 3646–3658 (2017).
22. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 e22 (2017).
23. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
24. Guo, Y. et al. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
25. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 e24 (2017).
26. van Steensel, B. & Furlong, E. E. M. The role of transcription in shaping the spatial organization of the genome. *Nat. Rev. Mol. Cell Biol.* **20**, 327–337 (2019).
27. Rowley, M. J. et al. Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* **67**, 837–852 e7 (2017).
28. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* **169**, 216–228 e19 (2017).
29. Chen, H. Y., Sun, J. M., Hendzel, M. J., Rattner, J. B. & Davie, J. R. Changes in the nuclear matrix of chicken erythrocytes that accompany maturation. *Biochem. J.* **320**, 257–265 (1996).
30. Grdisa, M. & White, M. K. Molecular and biochemical events during differentiation of the HD3 chicken erythroblastic cell line. *Int. J. Biochem. Cell Biol.* **35**, 422–431 (2003).
31. Chen, X. et al. ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
32. Solovei, I. et al. LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell* **152**, 584–598 (2013).
33. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* **33**, 960–982 (2019).
34. Peterlin, B. M., Brogie, J. E. & Price, D. H. 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. *Wiley Interdiscip. Rev. RNA* **3**, 92–103 (2012).
35. Brogie, J. E. & Price, D. H. Reconstitution of a functional 7SK snRNP. *Nucleic Acids Res.* **45**, 6864–6880 (2017).
36. Fishman, V. et al. 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. *Nucleic Acids Res.* **47**, 648–665 (2019).
37. Kruse, K., Hug, C. B., Hernandez-Rodriguez, B. & Vaquerizas, J. M. TADtool: visual parameter identification for TAD-calling algorithms. *Bioinformatics* **32**, 3190–3192 (2016).
38. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
39. Hou, C., Dale, R. & Dean, A. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. USA* **107**, 3651–3656 (2010).
40. Singh, A., Compe, E., Le May, N. & Egly, J. M. TFIH subunit alterations causing xeroderma pigmentosum and trichothiodystrophy specifically disturb several steps during transcription. *Am. J. Hum. Genet.* **96**, 194–207 (2015).
41. Fong, Y. W. & Zhou, Q. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414**, 929–933 (2001).
42. Jani, D. et al. Functional and structural characterization of the mammalian TREX-2 complex that links transcription with nuclear messenger RNA export. *Nucleic Acids Res.* **40**, 4562–4573 (2012).
43. Morera, D. et al. RNA-Seq reveals an integrated immune response in nucleated erythrocytes. *PLoS ONE* **6**, e26998 (2011).
44. Gilbert, N. et al. Formation of facultative heterochromatin in the absence of HP1. *EMBO J.* **22**, 5540–5550 (2003).
45. Zhao, B. et al. Disruption of erythroid nuclear opening and histone release in myelodysplastic syndromes. *Cancer Med.* **8**, 1169–1174 (2019).
46. Lefevre, C., Bondu, S., Le Goff, S., Kosmider, O. & Fontenay, M. Dyserythropoiesis of myelodysplastic syndromes. *Curr. Opin. Hematol.* **24**, 191–197 (2017).
47. Kabat, D. & Attardi, G. Synthesis of chicken hemoglobins during erythrocyte differentiation. *Biochim. Biophys. Acta* **138**, 382–399 (1967).
48. Hildebrand, E. M. & Dekker, J. Mechanisms and functions of chromosome compartmentalization. *Trends Biochem. Sci.* **45**, 385–396 (2020).
49. Penagos-Puig, A. & Furlan-Magaril, M. Heterochromatin as an important driver of genome organization. *Front. Cell Dev. Biol.* **8**, 579137 (2020).
50. Li, D. et al. Heterochromatin rewiring and domain disruption-mediated chromatin compaction during erythropoiesis. *Nat. Struct. Mol. Biol.* **30**, 463–474 (2023).
51. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 e24 (2017).
52. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
53. Khoury, A. et al. Constitutively bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. *Nat. Commun.* **11**, 54 (2020).
54. Barutcu, A. R., Blencowe, B. J. & Rinn, J. L. Differential contribution of steady-state RNA and active transcription in chromatin organization. *EMBO Rep.* **20**, e48068 (2019).
55. Zhang, S. et al. RNA polymerase II is required for spatial chromatin reorganization following exit from mitosis. *Sci. Adv.* **7**, eabg8205 (2021).
56. Levine, M. Paused RNA polymerase II as a developmental checkpoint. *Cell* **145**, 502–511 (2011).
57. Gariglio, P., Bellard, M. & Chambon, P. Clustering of RNA polymerase B molecules in the 5' moiety of the adult β -globin gene of hen erythrocytes. *Nucleic Acids Res.* **9**, 2589–2598 (1981).
58. Murphy, Z. C. et al. Regulation of RNA polymerase II activity is essential for terminal erythroid maturation. *Blood* **138**, 1740–1756 (2021).

Article

<https://doi.org/10.1038/s41594-023-01037-0>

59. St Paul, M., Paolucci, S., Barjesteh, N., Wood, R. D. & Sharif, S. Chicken erythrocytes respond to Toll-like receptor ligands by up-regulating cytokine transcripts. *Res. Vet. Sci.* **95**, 87–91 (2013).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023

Methods

Blood collection and cell culture

All applied procedures were approved by the animal welfare practices according to the Institute of Cellular Physiology-National Autonomous University of Mexico animal facility board. Blood of 10-day-old chicken embryos or 28–38-week-old White-leghorn hens (*Gallus gallus*) was collected in cold phosphate saline buffer with 50 mM EDTA. Cells were centrifuged at 1,000 r.p.m., 4 °C for 10 min, washed three times and counted on a Neubauer chamber. Chicken fibroblasts were prepared from 10-day-old chicken embryos.

Chicken fibroblasts and the erythroblast HD3 cell line were cultured in DMEM (GIBCO) supplemented with 10% FBS (Biowest), 2% chicken serum and penicillin/streptomycin (Invitrogen).

ATAC-seq library preparation

ATAC-seq experiments were performed as previously described⁶⁰ with minor modifications for chicken cells. We lysed 150,000 cells in lysis buffer (10 mM Tris-Cl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.01% Igepal) for 15 min on ice. Fibroblasts were sieved through a 70-µm cell-strainer and Dounce homogenized before lysis. Nuclei were centrifuged at 600g, 4 °C for 10 min and washed with 50 µl of lysis buffer. Nuclei were incubated with the Tn5 transposase (Illumina) and tagmentation buffer (Illumina) for exactly 30 min at 37 °C. After tagmentation, fragments were purified using MinElute columns (Qiagen). ATAC-seq libraries were amplified by ten PCR cycles using the Nextera index system (Illumina). Libraries were quantified in a Qubit system (ThermoFisher) and fragment size was determined on a TapeStation 4200 system (Agilent). Paired-end sequencing of biological replicates was conducted in a HiSeq2500, NextSeq or NovaSeq machine.

ATAC-seq data processing

ATAC-seq reads were aligned against the chicken genome (GRCg6a) using Bowtie2 (ref. 61) with the default parameters for paired-end reads. BAM files were generated using samtools (samtools view), sorted (samtools sort) and indexed (samtools index). Duplicates (samtools markdup) and reads longer than 150 bp (samtools view) were removed. Signal tracks were normalized to counts per million (CPM) using deeptools (bamCoverage –normalizeUsing CPM) and visualized with pyGenomeTracks⁶².

Accessible peaks were identified on each replicate using MACS2 (ref. 63) callpeak function with options -f=BAMPE and -q 0.01. Consensus peaks conserved between the two biological replicates were preserved for subsequent analysis.

RNA-seq library preparation

Total RNA was extracted using TRIzol following the manufacturer's instructions and the integrity of the material was evaluated by electrophoresis. RNA-seq libraries were prepared starting from 1 µg of RNA using the TruSeq Stranded Total RNA Ribo-Zero H/M/R Gold kit (Illumina). Libraries were quantified in a Qubit system (ThermoFisher) and fragment size was determined on a TapeStation 4200 system (Agilent). Three biological replicates per cell type were sequenced on a HiSeq2500 machine, paired-end 50 bp.

RT-qPCR

Total RNA was extracted using TRIzol following the manufacturer's instructions and the integrity of the material was evaluated by electrophoresis. Reverse transcription was carried out using random primers and qPCRs were performed using the KAPA SYBR Fast qPCR kit (Merck, KR0389-V9.13). The sequences of primers used for each reaction are provided in the Supplementary Information. Relative quantitation was estimated with the relative standard curve method, with points of 50, 100 and 150 ng of complementary DNA. Biological replicates were performed, each with technical triplicates.

RNA-seq data processing

RNA-seq data were analyzed using the SnakePipes pipeline⁶⁴. Briefly, RNA-seq reads were mapped to the GRCg6a genome using STAR⁶⁵, raw counts were obtained using featureCounts⁶⁶ and FPKM values were calculated. After normalization by FPKM, an additional normalization was performed based on the number of cells necessary to obtain 1 µg of RNA, the input concentration for the RNA-seq protocol, as total RNA content is reduced in RBCs. For this, the FPKM counts of all cell types were adjusted to 190,000 cells since this was the lowest number of cells used, corresponding to fibroblasts. This normalization was done using the following formula: FPKM normalized by cell number = (190,000 × FPKM)/(cell number).

Finally, the distributions of the cell number-normalized FPKM values were plotted, and a threshold of 0.3 was established to determine that a gene was expressed.

The differential expression analysis was performed with DESeq2 (ref. 67) v.1.28.1 using the raw counts. Before DESeq2 analysis, crude counts were normalized by cell number in the same way that cell number-normalized FPKM was obtained. Finally, the differential expression analysis was conducted by establishing a threshold of 0.5 false discovery rate and twofold change.

Gene ontology analysis

Gene ontology and pathway enrichment analyses were done using g:Profiler⁶⁸ with default parameters. All significant biological processes and pathways have a *P* value < 0.01.

Hi-C library preparation

Hi-C libraries were generated as previously described⁶⁹. Briefly, 10 × 10⁶ cells were fixed with formaldehyde 2% for 10 min at room temperature. The genomes were digested with MboI (NEB) overnight and 5' overhangs were filled with Biotin-14-dATP (Invitrogen) and re-ligated in nuclei. DNA fragments were purified and sheared to an average size of 500 bp using a Covaris machine. The sheared DNA was end-repaired, adenine-tailed and size-selected. Biotinylated fragments were pulled down using MyOne Streptavidin C1 DynaBeads (Invitrogen) and ligated to Truseq adapters (Illumina). Hi-C libraries were amplified using 5–6 PCR amplification cycles. Libraries were quantified in a Qubit system (ThermoFisher) and fragment size was determined on a TapeStation 4200 system (Agilent). Two biological replicates per cell type were paired-end sequenced using a HiSeq2500 or NextSeq machine.

Hi-C data processing

Mapping, filtering, correction and generation of Hi-C matrices was done using HiC-Pro⁷⁰. Briefly, read pairs were mapped against the chicken genome (GRCg6a) using Bowtie2 with HiC-Pro default parameters. Valid pairs were used to generate raw and normalized by iterative correction and eigenvector decomposition (ICE) matrices at 20-kb, 50-kb and 100-kb bin resolution. Correlations between Hi-C replicates and counts versus distance plots were generated with HiCExplorer⁶² with a bin size of 50 kb and replicates were merged (samtools merge). Matrices were normalized to the smallest number of valid pairs to correct for differences in sequencing depth. Heatmaps of normalized contact matrices and matrix comparisons were performed with HiCPlotter⁷¹.

TAD calling

Dense Hi-C matrices were generated with HiC-Pro (sparseToDense.py) and TAD calling was conducted with TADtool⁷² using the insulation score algorithm at a 20-kb resolution with a window size of 100 kb. TADs identified showed good agreement with visual inspection of Hi-C matrices.

Compartment analysis

Chromatin compartments were identified using the principal component analysis (PCA) function of HiCExplorer (hicPCA) at 50-kb

resolution. RNA-seq tracks were used to assign the sign of PCA results. Read counts at domains were performed on SeqMonk, www.bioinformatics.babraham.ac.uk/projects/seqmonk/.

Antibodies

The following antibodies were used for immunofluorescence experiments: anti-H3K9me3 (Millipore, no. 07-442) in a 1:700 dilution, anti-RNA Pol II Ser5-P (Santa Cruz, no. SC-47701) in a 1:500 dilution, anti-RNA Pol II Ser5-P antibody (Abcam, no. ab5095) in a 1:800 dilution, Lamin B1 (Abcam, no. ab16048) in a 1:500 dilution, anti-Lamin A/C (Santa Cruz, no. SC-376248) in a 1:500 dilution, anti-SMCI (A300-055A) in a 1:1,000 dilution, anti-CTCF primary antibody (kindly donated by F. Recillas-Targa) in a 1:3,000 dilution, anti-rabbit Alexa Fluor 594-conjugated (Life Technologies, no. R37117 <https://www.ncbi.nlm.nih.gov/nucore/R37117>) in a 1:500 dilution and anti-mouse FITC-conjugated (Millipore, AP192F) in a 1:500 dilution.

Immunofluorescence

Cells were plated and cultured on coverslips pretreated with poly-L-lysine (Electron Microscopy Sciences), then fixed with formaldehyde 1% and permeabilized with Igepal 0.1%. Fixed cells were washed with PBS and preincubated with 5% BSA for 1 h, then incubated with primary antibodies in PBS with 0.2% BSA overnight at 4 °C. After three washes with PBS, cells were incubated with the secondary antibody for 1 h and stained with DAPI for 20 min, then mounted using Fluoromount (Sigma). Image acquisition was performed in a Zeiss LSM800 confocal microscope.

Western blot

Nuclear extracts were obtained through cell lysis and centrifugation using a high-sucrose buffer (10 mM HEPES, 10 mM KCl, 1.5 mM MgCl₂, 0.34 M sucrose, 10% glycerol, 0.1% Triton X-100). Isolated nuclei were lysed using RIPA (10 mM Tris-HCl, 1 mM EDTA, 0.5 mM EGTA, 0.5% Triton X-100, 0.1% SDS, 140 mM NaCl), and nuclear proteins were quantified by Bradford assay (BioRad).

Subsequently, proteins were separated on a polyacrylamide gel and transferred to a PVDF membrane. After transfer, the membrane was blocked for 2 h with 5% nonfat milk and incubated with anti-CTCF primary antibody (kindly donated by F. Recillas-Targa), 1:1,000, and anti-SMCI (A300-055A), 1:1,000. Finally, membranes were incubated for 2 h with an appropriate HRP-conjugated secondary antibody and proteins were detected with a chemiluminescence substrate (P36599 <https://www.uniprot.org/uniprot/P36599>, Millipore) on a C-DiGit Blot Scanner (LI-COR). Western blots were performed for two independent biological replicates (Extended Data Fig. 5 and Source Data for Extended Data Fig. 5).

ChIP-seq and ChIP-qPCR

For ChIP experiments, four biological replicates per cell type were fixed in 1% formaldehyde for 10 min. Samples were sonicated for ten cycles of 30 s on/60 s off with 30% amplitude using a Cole-Parmer 500-W ultrasonic processor. We added 1% S2R+ spike-in to each sample and performed ChIP using 5 µg of anti-RNA Pol II Ser5-P antibody (Santa Cruz, no. SC-47701 for ChIP-qPCR, or Abcam, no. ab5095 for ChIP-seq) and Protein G DynaBeads (Invitrogen, 10003D). DNA was purified using DNA purification columns (Zymo Research, D5205).

Quantitative protein occupancy was performed by qPCR according to the manufacturer's instructions (Jena Bioscience, PCR-369). Mean values for the analyzed regions were expressed as fold enrichment compared with a negative control (SOX2 gene promoter). Statistically significant differences were analyzed using unpaired *t*-test. All oligonucleotides used are listed in the Supplementary Information.

ChIP-seq libraries were prepared using the NEBNext Multiplex Oligos primer set (Illumina, E7600S). Libraries were quantified in a

Qubit system (ThermoFisher) and fragment size was determined on a TapeStation 4200 system (Agilent). Two replicates per cell type were sequenced on a NovaSeq machine, paired-end 150 bp.

ChIP-seq data processing

The paired-end reads were quality-checked with FastQC and mapped to the chicken genome (GRCg6a) using Bowtie²³ with the default parameters for paired-end reads. Mapped reads were normalized using the S2R+ spike-in, duplicates were removed (samtools markdup) and BAM files were generated using samtools. Peak calling was performed using MACS2 (ref. 63) callpeak function with *P* value 0.001. Signal tracks were generated using deeptools (bamCoverage -normalizeUsing RPGC) and visualized with pyGenomeTracks⁶⁵. Conserved peaks between replicates were obtained and used for subsequent analysis. Heatmaps were generated using deeptools (computeMatrix and plotHeatmap).

ATAC-seq

ATAC-seq experiments were performed following the published protocol⁷³. Briefly, following fixation and lysis, cells were incubated with the Tn5-ATTO-590N transposase (donated by X. Chen) for 30 min at 37 °C. Cells were washed at 55 °C with PBS, SDS 0.01% and EDTA 50 mM to stop the tagmentation reaction, then counterstained with DAPI and mounted with Fluoromount. A negative control was conducted by adding EDTA 50 mM to the tagmentation mix to inhibit Tn5 activity. Image acquisition was performed in a LEICA TCS SP5 or Zeiss LSM800 microscope.

Nascent RNA labeling

Nascent RNA imaging was performed using the Click-iT RNA imaging kit (ThermoFisher) following the manufacturer's instructions. Cells were fed with EU 1 mM for 6 h, fixed with formaldehyde 3.7% and permeabilized with Triton X-100 0.5%. Treatment for 6 h was selected as in RBCs, and nascent transcription was not detected at earlier time points in contrast to erythroblasts (Extended Data Fig. 2e). Click-iT reaction with Alexa Fluor 594 was carried out and nuclei were stained with Hoechst 33342. Image acquisition was performed in a Zeiss LSM800 microscope.

Digital image processing

Image processing was performed using Fiji⁷⁴. Nuclei were identified and segmented by using a Gaussian filter on raw DAPI images and then applying an intensity threshold to generate nuclear outlines. All nuclear outlines were verified visually, and mitochondrial ATAC-seq signal was filtered out using the nuclear outlines generated. A total of 50 nuclei were analyzed. Mean intensity per nucleus was measured and the brightest regions were identified using an intensity threshold to retain only the pixels with intensity values higher than the mean.

Distance maps of individual nuclei were generated in Fiji to measure signal intensity as a function of radial distance to the nuclear periphery. Mean intensity was measured in a series of one-pixel-wide annuli from the outermost pixel to the nuclear center. Each channel was normalized to its mean intensity per nucleus and nuclear size was scaled in R. Fifty individual nuclei were analyzed per experiment and the average signal was calculated. Correlation analysis between signal intensity and distance from the nuclear rim was performed in R.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All reads generated were mapped against the chicken genome version GRCg6a. All datasets generated in this study have been deposited as a super series at GEO with accession number [GSE206194](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE206194). Source data are provided with this paper.

References

60. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).
61. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
62. Ramirez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).
63. Zhang, Y. et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
64. Bhardwaj, V. et al. snakePipes: facilitating flexible, scalable and integrative epigenomic analysis. *Bioinformatics* **35**, 4757–4759 (2019).
65. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
66. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
67. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
68. Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
69. Rubin, A. J. et al. Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* **49**, 1522–1528 (2017).
70. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
71. Akdemir, K. C. & Chin, L. HiCPlotter integrates genomic data with interaction matrices. *Genome Biol.* **16**, 198 (2015).
72. Kruse, K., Hug, C. B., Hernandez-Rodriguez, B. & Vaquerizas, J. M. TADtool: visual parameter identification for TAD-calling algorithms. *Bioinformatics* **32**, 3190–3192 (2016).
73. Chen, X. et al. ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
74. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

Acknowledgements

A.P.-P. is supported by CONAHCyT scholarship no. 822335. This work was supported by PAPIIT grants no. IN207319 and no. IN210323 and by CONAHCyT grants no. 303068, no. 15758 and no. 137721. The funders had no role in study design, data collection and analysis, decision

to publish or preparation of the manuscript. We thank X. Chen for his generous donation of the Tn5-ATTO-590N enzyme. We thank F. Recillas-Targa for kind donation of the HD3 erythroblast cell line. We thank the Molecular Biology Unit, the Microscopy Unit and the Bioinformatics Unit at the IFC, in particular R. Rincón-Heredia and A. Cesar Poot-Hernández for advice on image processing and genomic data analysis, respectively. We thank the Animal Facility at the IFC, in particular C. Rivera-Cerecedo and H. Malagón-Rivera. We thank the Supercomputing Unit at LANCIS, in particular R. García-Herrera for access to the computer cluster. We thank R. Saldaña-Meyer and M. Escamilla Del Arenal for helpful discussion on the results obtained in this work.

Author contributions

A.P.-P. and M.F.-M. designed the project, analyzed and discussed the data, and wrote the manuscript with comments from all authors. A.P.-P. performed and analyzed ATAC-seq, ATAC-seq and Hi-C experiments and immunostainings. S.C.-G. performed and analyzed the RNA-seq and ChIP-seq as well as the ChIP-qPCR experiments and western blots. A.S.-G. performed and analyzed the initial Hi-C and RT-qPCR experiments. A.A.-L. and K.J.-L. helped with data analysis and R.P.-M. helped with experiments and provided laboratory assistance to all authors. All authors contributed to the discussion of the results.

Competing interests

The authors declare no competing interests.

Additional information

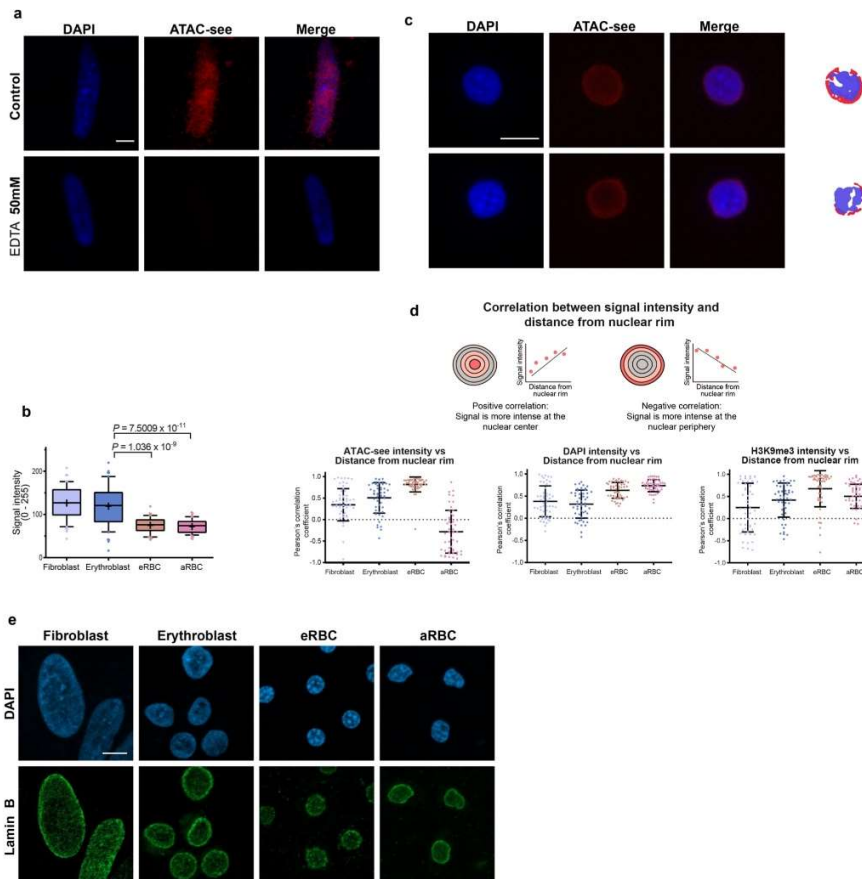
Extended data is available for this paper at <https://doi.org/10.1038/s41594-023-01037-0>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41594-023-01037-0>.

Correspondence and requests for materials should be addressed to Mayra Furlan-Magaril.

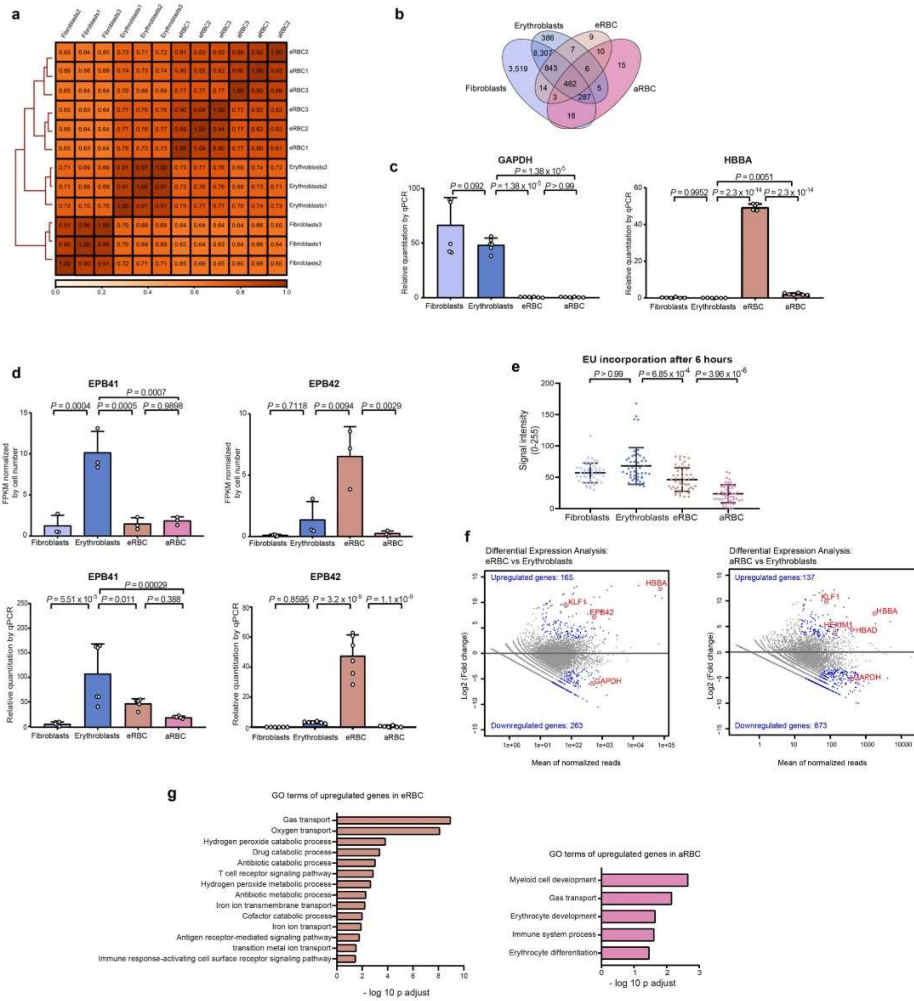
Peer review information *Nature Structural & Molecular Biology* thanks Rob Beagrie and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Sara Osman and Dimitris Typas were the primary editors on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.



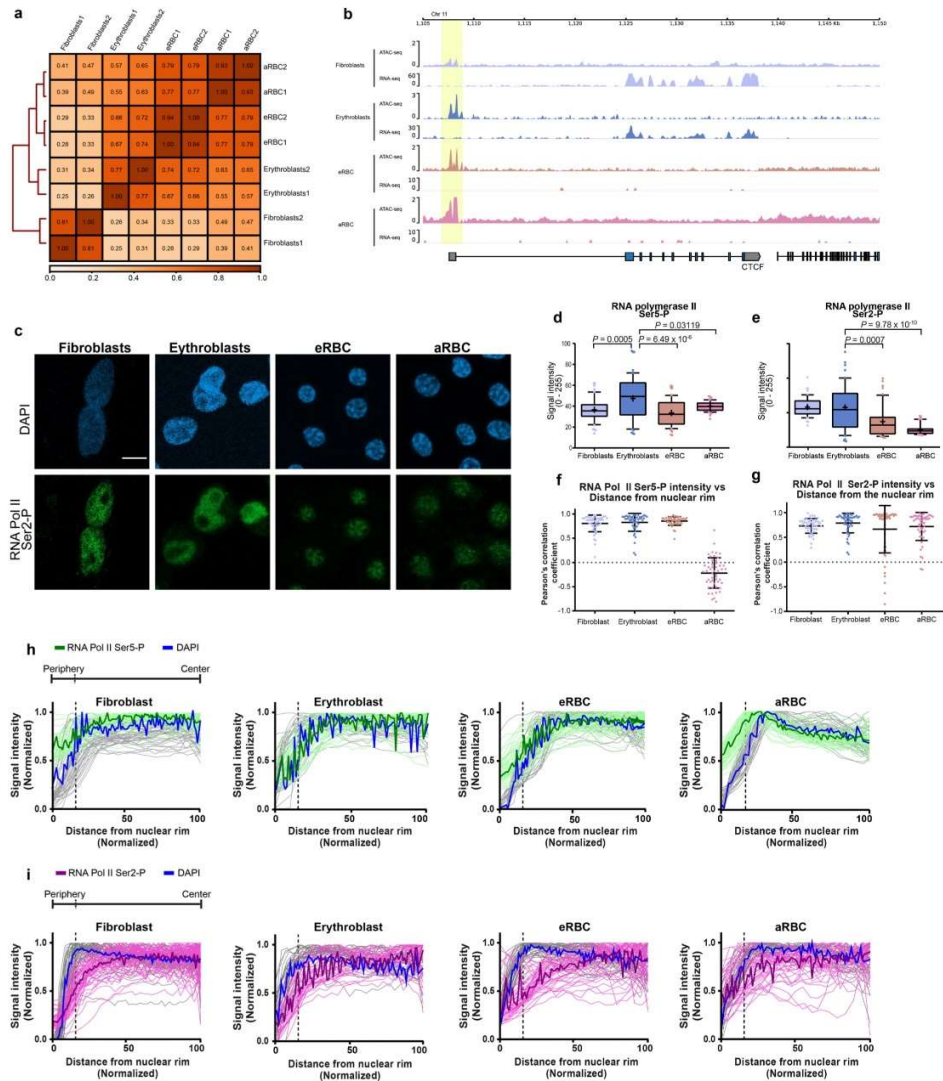
Extended Data Fig. 1 | Validation of ATAC-seq and chromatin reorientation in aRBC. **a**, Imaging of open chromatin by ATAC-seq (red) and DAPI (blue, nuclei) in chicken fibroblasts. A negative control of ATAC-seq in presence of EDTA 50 mM shows no incorporation of the fluorophore. Representative images from two independent experiments are shown. **b**, ATAC-seq signal average intensity per nucleus in chicken erythrocytes. Boxplots with median (horizontal line), mean (+), 25th and 75th percentiles (boxes), and whiskers denoting the 10th and 90th percentile from 50 nuclei from two independent experiments are shown. P values based on one-way ANOVA followed by Tukey's post-hoc test are indicated.

c, Representative images from two ATAC-seq experiments in aRBC. Brightest pixels enriched in ATAC-seq (red) or DAPI (blue) are shown on the right. **d**, Pearson's correlation score between ATAC-seq, DAPI or H3K9me3 signal intensity and the distance to the nuclear rim ($n = 50$ nuclei from two independent replicates). Scatter dot plots show mean values \pm SD. **e**, Fluorescence imaging of Lamin B1 (green) and DAPI. In every case, maximal projections of three-dimensional (3D) stacks are presented. Representative images from two replicates are shown. Scale bar = 5 μ m.



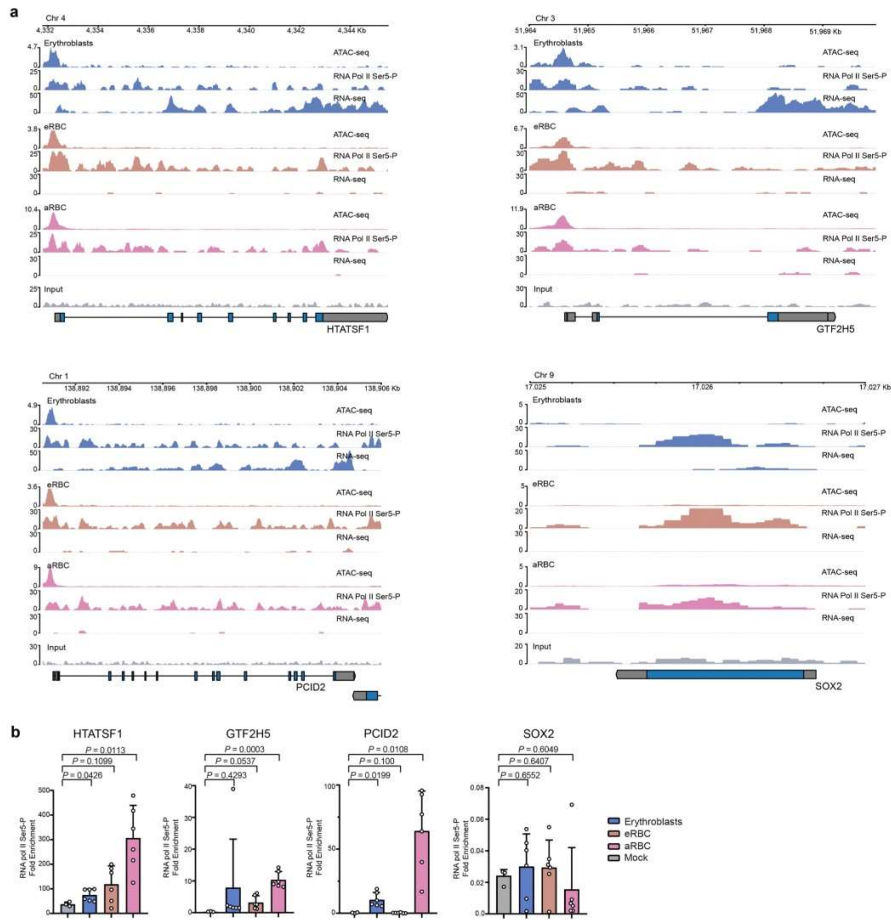
Extended Data Fig. 2 | Characterization of the erythroid transcriptome. **a**, Spearman correlation heatmap of RNA-seq replicates. **b**, Venn diagram of expressed genes in fibroblasts and erythroid cells. **c**, Relative expression of GAPDH and HBBA measured by RT-qPCR, triplicates from two biological replicates are presented. For aRBC and fibroblasts this means independent samples from two different animals, for eRBC this means two independent RBC collection from approximately 20 embryos, and for erythroblasts this means two independent cell cultures. Data is presented as means \pm SD. P values indicated are based on one-way ANOVA analysis followed by Tukey's post-hoc test **d**, Expression levels of the erythroid genes EPB41 and EPB42 measured by RNA-seq

(up; $n = 3$ independent RNA-seq libraries) and RT-qPCR (down; $n =$ same as **c**). Data is presented as means \pm SD with P values based on one-way ANOVA followed by Tukey's post-hoc test. **e**, EU incorporation measured by fluorescence intensity after 6 hours of incubation with EU ($n = 50$ nuclei from three independent Click-iT experiments). P values based on one-way ANOVA analysis followed by Kruskal-Wallis post hoc test are presented. **f**, MA plots of differentially expressed genes between eRBC and erythroblasts, and aRBC and erythroblasts. **g**, GO terms enriched in upregulated genes from **f**. Analysis was performed using Fisher's exact test and the default multiple-hypotheses testing method (g:SCS) on gProfiler.



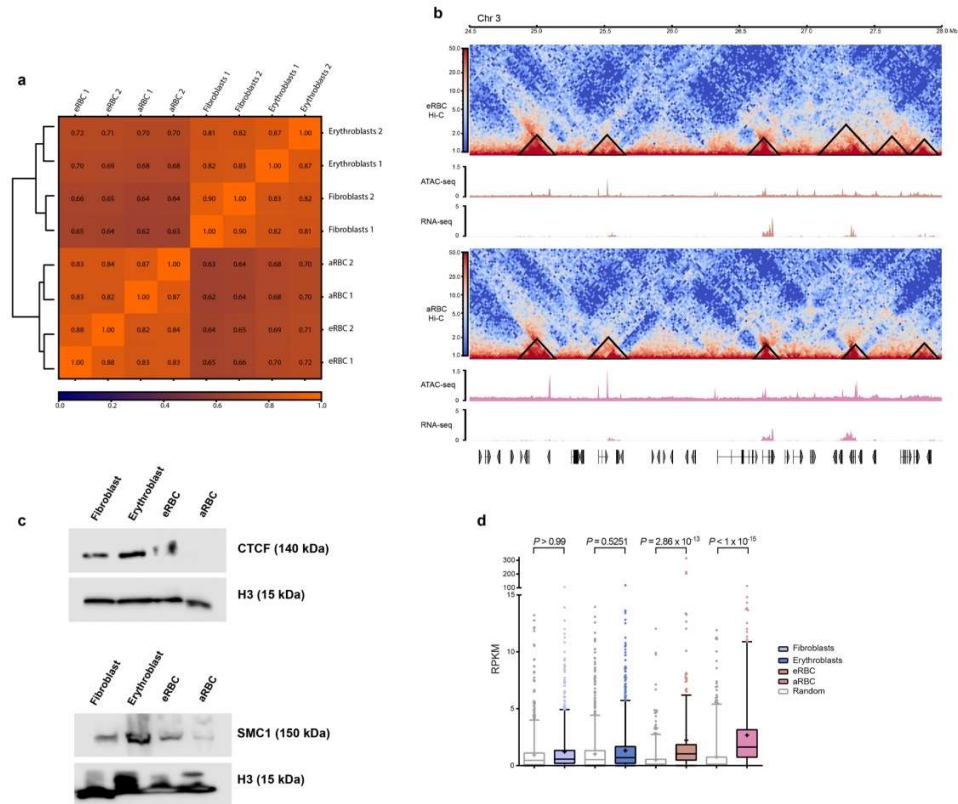
Extended Data Fig. 3 | Chicken erythrocytes retain accessible paused promoters. **a**, Spearman correlation heatmap of ATAC-seq replicates. **b**, Example of the preserved accessibility of the promoter of CTCF despite no transcriptional activity detected by RNA-seq. **c**, Representative images of DAPI and RNA pol II Ser2-P immunofluorescence from two replicates. Maximal projections of three-dimensional (3D) stacks are presented. Scale bar = 5 μ m. Average signal intensity of RNA pol II Ser5-P (**d**) and Ser2-P (**e**) signal average intensity per nucleus in chicken erythrocytes. Boxplots with median (horizontal line), mean (+), 25th and 75th percentiles (boxes), and whiskers denoting the 10th and 90th percentile from 50 nuclei from two independent experiments are shown. P values based

on one-way ANOVA followed by Tukey's post-hoc test are presented. Pearson's correlation score between RNA Pol II Ser-5 (**f**) and Ser-2 (**g**) signal intensity and the distance to the nuclear rim ($n = 50$ nuclei from two independent replicates). Scatter dot plots show mean values \pm SD. **h**, RNA pol II Ser5-P and DAPI intensity is presented as a function of distance to the nuclear periphery. Individual tracks for 50 nuclei are presented in lighter colors and the mean intensity is presented in bold. **i**, RNA pol II Ser2-P and DAPI intensity is presented as a function of distance to the nuclear periphery. Individual tracks for 50 nuclei are presented in lighter colors and the mean intensity is presented in bold.



Extended Data Fig. 4 | ChIP-qPCR confirms RNA Pol II Ser5-P occupancy of promoters of genes that are not expressed in aRBC. a, ATAC-seq, RNA Pol II Ser5-P ChIP-seq, and RNA-seq tracks of HTATSF1, GTF2H5, PCID2, and SOX2. **b**, ChIP-qPCR against RNA Pol II Ser5-P. ATAC-seq peaks at the promoter of HTATSF1,

GTF2H5 and PCID2 were analyzed. Fold enrichment was calculated against amplification in the promoter of SOX2. SOX2 enrichment was calculated against input. Mean \pm SD from 4 replicates is presented. Statistical differences against a mock control were calculated with a two-tailed t-test.



Extended Data Fig. 5 | Erythroid mini domains retain local structure around open chromatin and paused genes. **a**, Pearson correlation heatmap between Hi-C replicates at 50 Kb resolution. **b**, Hi-C contact matrices from eRBC and aRBC (3:24.5–28 Mb) and examples of mini domains identified. ATAC-seq and RNA-seq tracks are plotted below. **c**, Western blot for CTCF and SMC1 proteins for all cell types. Histone 3 (H3) was used as the loading control. Representative results

from two experiments are presented. **d**, RNA-seq RPKM count inside chromatin domains or a set of random regions of the same size in fibroblasts (n = 2,373 domains), erythroblasts (n = 2,562), eRBC (n = 916), and aRBC (903). Boxplots show median (horizontal line), mean (+), 25th and 75th percentiles (boxes), and whiskers denote the 1st and 99th percentiles. P values based on one-way ANOVA followed by Kruskal-Wallis post hoc test are stated.