

## RNA polymerase II pausing contributes to maintain chromatin organization in erythrocytes

Penagos-Puig Andrés<sup>1</sup>, Claudio-Galeana Sherlyn<sup>1</sup>, Stephenson-Gussinye Aura<sup>1</sup>, Jácome-López Karina<sup>1</sup>, Aguilar-Lomas Amaury<sup>1</sup>, Pérez-Molina Rosario<sup>1</sup> & Furlan-Magaril Mayra<sup>1\*</sup>

1. Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Mexico City, Z.C.04510

\* Correspondence to mfurlan@ifc.unam.mx

### Abstract

Chicken erythrocytes are nucleated cells 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. Here we characterized the changes in gene expression, chromatin accessibility, genome organization, and chromatin nuclear disposition during the terminal stages of erythropoiesis in chicken and found a complex chromatin reorganization at different genomic scales. We identified a robust decrease in transcription in erythrocytes. Nevertheless, a set of genes maintains their expression in erythrocytes, including genes involved in RNA pol II promoter-proximal pausing. Erythrocytes exhibit an inverted nuclear architecture and reposition euchromatin towards the nuclear periphery together with the paused RNA polymerase. In erythrocytes, chromatin domains are partially lost genome-wide except at mini domains retained around paused promoters. 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 polymerase in the maintenance of local chromatin organization.

### Introduction

Red blood cells (RBC) 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 HDACs<sup>1-3</sup>, and remodeling of the nuclear envelope<sup>4</sup> followed by the expulsion of the nucleus from the erythroblasts<sup>5</sup>.

Erythropoiesis in non-mammalian vertebrates like birds however, gives rise to circulating erythrocytes that retain their nucleus in a highly compacted state<sup>6</sup>.

Genome-wide chromatin condensation and gene expression silencing in nucleated RBC in chicken is promoted by the incorporation of the erythroid-specific linker histone H5 into chromatin<sup>7-9</sup>. In fact, chicken RBC undergo a dramatic drop in RNA synthesis and are regarded as transcriptionally inert cells<sup>10-12</sup>, 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 regulation<sup>13</sup>. However, the molecular mechanisms that mediate the extensive silencing of the erythroid genome are largely unknown and a study at high resolution of the remodeling of the chromatin and genome organization during this process is still lacking.

The three-dimensional organization of chromosomes in the interphase nucleus serves as a scaffold of chromatin interactions that ensures the execution of complex transcriptional programs<sup>14</sup>. In the last decades, Hi-C experiments unveiled multiple levels of genome organization occurring at different genomic scales<sup>15</sup>. 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 respectively<sup>16</sup>. At the sub-megabase scale, topologically associated domains (TADs) represent frequently interacting regions that are partly insulated from other neighboring sequences<sup>17,18</sup>. TADs enable proper promoter-enhancer communication inside the domain and prevent interference by regulatory elements outside the TAD and disruption of TAD boundaries leads to aberrant gene activation and disease<sup>19-21</sup>, 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<sup>22-24</sup> although the presence of CTCF-independent TADs has been reported<sup>22,25</sup>. 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<sup>26-28</sup>.

Here, we studied the changes in chromatin accessibility and genome organization that take place in the terminal erythroid differentiation of chicken RBC. We show that chicken erythrocytes undergo a complex multilayered remodeling of chromatin organization. First, we show that chicken RBC have a unique inverted nuclear architecture and relocate their open chromatin towards the nuclear periphery. Despite a genome-wide reduction in gene expression, chicken RBC 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 polymerase II in a paused state. Hi-C experiments show a strengthening of chromatin compartmentalization consistent with confocal microscopy imaging of euchromatin, transcription and heterochromatin foci. Moreover, the erythroid genomes are largely devoid of TADs but retain local organization at mini domains 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 persist in absence of transcription and propose that RNA polymerase II binding might be sufficient to maintain local organization.

## Results

### Chromatin accessibility is restricted to the nuclear periphery in aRBC

Chicken erythrocytes are regarded as mostly transcriptionally inactive cells that exhibit a dramatic drop in RNA synthesis<sup>29</sup>. To investigate the chromatin remodeling process during terminal erythroid differentiation in chicken we obtained primary embryonic (eRBC) and adult (aRBC) red blood cells from 10-day-old chicken embryos and 30 weeks old hens, respectively. We also analyzed HD3 cells<sup>30</sup> that correspond to an early erythroblast and primary fibroblasts were also used as a non-erythroid 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 accessible chromatin using ATAC-see<sup>31</sup> (Fig. 1a; Extended data 1a). ATAC-see signal intensity in eRBC and aRBC is lower than in erythroblasts suggesting an overall decrease in accessibility during erythroid terminal differentiation (Fig. 1a; Extended data 1b). However, chicken erythrocytes retain accessible chromatin even in absence of transcription.

ATAC-see of aRBC revealed a unique nuclear organization in which euchromatin is located predominantly at the periphery and less abundant at the center of the nucleus (Fig. 1a; arrowhead) as opposed to the conventional nuclear architecture observed in fibroblasts, erythroblasts, and eRBC. Next, we applied a Gaussian filter to the raw images using a threshold to obtain all pixels brighter than the mean in both ATAC-see and DAPI channels in each nucleus (see methods). The images generated confirmed that euchromatin is enriched at the nuclear periphery in aRBC while the DAPI signal is more intense at the nuclear center (Fig. 1b; Extended data 1c).

Heterochromatin is also relocated in chicken erythrocytes as shown by H3K9me3 immunostaining (Fig. 1c; arrow). In eRBC and aRBC H3K9me3 is preferentially segregated into foci located at the center of the nucleus instead of the conventional positioning of constitutive heterochromatin at the nuclear envelope (NE) seen in fibroblasts and erythroblasts.

We validated the inverted organization of aRBC nuclei by plotting the normalized mean intensity of H3K9me3 and ATAC-see as a function of distance from the nuclear periphery (Fig. 1d). In fibroblasts and erythroblasts, H3K9me3 signal gradually decreases from the NE to the interior, where ATAC-see signal is higher. Consistently with a conventional nuclear architecture. In contrast, in eRBC and aRBC H3K9me3 increases from the nuclear periphery to the interior reflecting the position of the heterochromatic foci we observed in these cells. In agreement with the inversion of euchromatin observed in aRBC, ATAC-see signal intensity in these cells is higher at the nuclear periphery and decreases to the nuclear interior. Pearson's correlation analysis between signal intensity and distance from the nuclear periphery also recapitulates the inversion of euchromatin in aRBC (Extended data 1d). Thus, terminal erythroid differentiation results in repositioning of heterochromatin away from the NE in eRBC and aRBC, and the relocation of euchromatin towards the periphery in aRBC.

The unique inverted nuclear architecture observed in aRBC has been previously shown to occur in the rod photoreceptor cells of nocturnal mammals in response to downregulation of the Lamin B receptor (LBR) and absence of Lamin A/C expression <sup>32</sup>. We asked whether the chromatin inversion observed in aRBC could be explained by changes in the NE composition. Like 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. 1e). Lamin B1 was detected in all cells studied (Extended data 1e).

Taken together, these results show a complex reorganization of the nuclear architecture in chicken erythrocytes. Cells with a conventional nuclear architecture locate heterochromatin at the NE while euchromatin is preferentially positioned at the nuclear interior. During terminal differentiation, heterochromatin moves away from the NE at foci and loss of Lamin A/C in aRBC allows repositioning of euchromatin towards the nuclear periphery in an inverted architecture (Fig. 1f).

**Chicken erythrocytes show a global downregulation of gene expression except for small RNAs and erythroid-related 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 eRBC and aRBC (Fig. 2a) in agreement with a previous report of low levels of RNA synthesis in eRBC measured by radioactivity incorporation<sup>10</sup>. Furthermore, the electrophoretic profile of the RNA obtained shows changes in the diversity of RNA species in eRBC and aRBC with a gradual loss of the rRNA bands while small RNA molecules are still detectable (Fig. 2b).

To further characterize the transcriptional state of chicken erythrocytes we performed total RNA-seq experiments. A global downregulation of gene expression in the erythrocytes was evidenced by a marked decrease in FPKM values compared to other cell types (Fig. 2c). We identified all genes that were expressed (FPKM > 0.3) and found a small set of genes that remain active in the erythrocytes (1,354 in eRBC and 804 in aRBC) in agreement with RT-qPCR experiments (Fig. 2d-f; Extended data 2a-d). Of note, most FPKMs come from small RNA molecules such as snoRNAs and miscellaneous RNAs in eRBC and aRBC, consistent with the electrophoretic pattern of erythroid RNA extractions (Fig. 2g).

Differential expression analysis was conducted between the different erythroid stages. We found 810 and 428 differentially expressed genes between eRBC or aRBC and erythroblasts respectively (Extended data 2e). Gene ontology analysis of upregulated DEGs showed that chicken erythrocytes exhibit upregulation of genes related to oxygen-carrier activity, erythroid differentiation, and immune activity (Extended data 2f). 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). 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 labelling 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 open chromatin at the NE cannot be explained by active transcription since transcriptional foci were not preferentially located at the nuclear periphery in aRBC.

## The paused RNA polymerase retains accessibility at silent promoters in chicken erythrocytes

We performed 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 3a). Overall, we identified fewer ATAC-seq peaks in erythroid cells compared to 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, eRBC, and aRBC regardless of the genome-wide decrease in gene expression in the erythrocytes. Most of the peaks in the erythrocytes represent constitutively open regions or 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 eRBC and aRBC correspond to silenced genes that remain accessible at their promoters (Fig. 3c-e; Extended data 3b). Promoter-proximal pausing of the RNA polymerase II has been shown to maintain promoter chromatin accessibility and the gene's potential to be activated<sup>33</sup>. We hypothesized that RNA pol II pausing could be responsible for preserving the accessible promoters in the genome of eRBC and aRBC.

RNA pol II pausing is promoted by the inhibition of the positive elongation factor p-TEFB by the 7SK small nucleoprotein complex formed by the snRNA 7SK, HEXIM1/2, LARP1, and MePCE<sup>34,35</sup>. Interestingly, 7SK and HEXIM1 are highly 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; Extended data 3d), whereas the signal from the elongating CTD isoform (Ser2-P) is diminished in eRBC and aRBC (Extended data 3c, e). In addition, we found that RNA pol II Ser5-P is spatially located at the nuclear periphery in aRBC (Fig 3g; arrows) following the same distribution as the euchromatin detected by ATAC-see (Extended data 3f-g). Finally, we confirmed the presence of RNA pol II Ser-5P in the open promoter regions of genes without expression in the erythrocytes by ChIP-qPCR (Fig. 3h; Extended data 3h). 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.

## The genome of eRBC and aRBC is hyper-compartmentalized and structured in mini domains at open chromatin regions

Recent Hi-C experiments conducted in aRBC have evidenced a global rearrangement in their 3D genome organization characterized by an increase in long-range interactions and absence of Topologically Associated Domains (TADs)<sup>36</sup>. 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, eRBC, and aRBC (Extended data 4a).

High-order chromatin organization is defined by the segregation of genomic regions into A and B compartments corresponding to active and inactive chromatin, respectively<sup>16</sup>. 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 shows a more pronounced plaid pattern in eRBC and aRBC 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 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 by the presence of well-defined heterochromatin and transcriptional foci in these cells.

Identification of A and B compartments showed a hyper-compartmentalization pattern in the genome of the erythrocytes compared to 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 to erythroblasts and fibroblasts (Fig. 4e-f). This hyper-compartmentalized 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 eRBC and aRBC (Fig. 5a) and it has been suggested that chicken erythrocytes lack TADs<sup>36</sup>. TAD calling using insulation index<sup>37</sup> 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; Extended data 4b). These domains are fewer and smaller (with an average size of 241 Kb) than TADs observed in erythroblasts and fibroblasts (Fig. 5c-d), hence we called these structures min-domains. Of note, only a small fraction of the genome remains structured in mini domains (24.18% in eRBC and 20.4% in aRBC) (Fig. 5e).

Topological domains insulation depends on the presence of the architectural proteins CTCF and cohesin at the domain boundaries<sup>24,38,39</sup> and auxin-inducible depletion of CTCF or the cohesin-loading factor Nipbl causes genome-wide loss of topological domains<sup>22,23</sup>. Interestingly, CTCF and Rad21 are downregulated in eRBC and aRBC (Fig. 5f). Thus, downregulation of architectural proteins might explain the loss of topological domains in chicken erythrocytes.

Notably, mini domains demarcate regions in 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 genome of eRBC and aRBC, euchromatic regions remain structured in mini domains.

### Chicken RBC genomes are organized around paused promoters

We hypothesized that the presence of local structure at mini domains reflected transcriptional activity inside erythroid mini domains. Indeed, mini domains are enriched in RNA-seq reads (Extended data 4c). However, visualization of Hi-C maps together with ATAC-seq and RNA-seq tracks showed that mini domains are often found around open regions of the genome that lack transcription (Fig. 6a; Extended data 4b).

We classified topological domains by the transcriptional and chromatinic state of their TSSs (Fig. 6b). Domains within the 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 are depleted of open TSSs. This classification evidenced that the presence of paused promoters is a common feature in mini domains as more than 80% of mini domains fall into either class I or III.

Interestingly, paused genes that reside within mini domains in eRBC (3,458 genes) and aRBC (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*<sup>40,41</sup>, and the nuclear export protein *Pcid2*<sup>42</sup> (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 mini domains 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 challenges<sup>43</sup>. In summary, we show that paused promoters of genes involved in gene expression and RNA metabolism remain organized into mini domains despite a global loss of TADs in chicken eRBC and aRBC.

## Discussion

Here we analyzed thoroughly the changes in chromatin accessibility and organization that accompany terminal erythroid differentiation and downregulation of gene expression in the nucleated chicken erythrocytes. We found that loss of Lamin A/C results in a complete inversion of the nuclear architecture in aRBC. Furthermore, the erythrocyte nucleus is a highly compartmentalized organelle that confines euchromatin, 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 mini domains 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.

Chromatin condensation and gene silencing are common traits of mammalian and non-mammalian erythropoiesis<sup>5,9,44</sup>. Nucleated erythrocytes are commonly described as transcriptionally inert cells<sup>12</sup> and defects in heterochromatin stability or gene silencing are considered hallmarks of some myelodysplastic syndromes and anemia in mammals and birds<sup>45-47</sup>. Here, we provide evidence that both eRBC and aRBC actively transcribe a reduced number of genes (1,354 in eRBC, 804 in aRBC) including hemoglobin 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 repression<sup>48</sup>. 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 eRBC and aRBC. In aRBC, loss of Lamin A/C leads to a complete inversion of nuclear architecture in which heterochromatin detaches from the nuclear envelope and moves towards the center of the nucleus. The atypical spatial organization of euchromatin 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 LBR and absence of Lamin A/C<sup>32</sup>.

In recent years, heterochromatin has emerged as a key factor mediating genome organization and heterochromatin repositioning has important effects on cellular function<sup>49</sup>. 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 neutrophiles prepares the cells for initiation of NETosis<sup>31</sup>. Moreover, heterochromatin detachment and relocation cause gene misexpression in laminopathies and senescence<sup>49</sup>. An increase in heterochromatin-mediated long-range interactions is observed in the last nucleated stage of human erythropoiesis<sup>50</sup> and Lamin B cleavage by caspase-3 and transient nuclear opening has been shown to promote chromatin condensation in the erythroblasts of mouse and human<sup>4,45</sup>. Together with our results, these evidences 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<sup>23,51</sup>. Thus, compartmentalization may be regulated by the binding of cohesin to chromatin and loop-extrusion. Here we found Rad21 to be downregulated in eRBC and aRBC which provides a possible explanation for the hyper compartmentalized state of their genomes.

Recent Hi-C experiments in chicken aRBC and the immature polychromatic erythrocyte using a six-cutter restriction enzyme showed a genome-wide depletion of TADs in these cells<sup>36</sup>. Although 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 structured in the terminal differentiated eRBC and aRBC. 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 boundaries<sup>52</sup>. Ablation of CTCF results in a global loss of TADs except at regions where low levels of CTCF are able to maintain TAD boundaries<sup>22,53</sup>. Similarly, degradation of Rad21 and Nipbl eliminates chromatin domains<sup>23,51</sup>. Here we show that transcription of CTCF and Rad21 decreases 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 compartmentalization<sup>50</sup>. 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 genome<sup>28</sup>. Inhibition of transcriptional elongation weakens TAD insulation but does not preclude TAD emergence<sup>19,28,54</sup>. Hence, transcription is not necessary for the establishment of TADs but might play a role in maintaining chromatin organization by other associated processes like chromatin opening, the assembly of the pre-initiation complex, and recruitment of transcription factors and the transcriptional machinery<sup>28</sup>. Moreover, RNA pol II dictates cohesin loading into chromatin and rapid degradation of RNAPII prevents the reestablishment of chromatin folding in cells following exit from mitosis<sup>55</sup>. Our results show that mini domains in the chicken erythrocytes frequently contain paused promoters occupied by the RNA pol II-Ser5, 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 cues<sup>56</sup>. Here we present evidence that suggest that RNA pol II pausing represents a key mechanism by which chicken nucleated erythrocytes maintain low levels of transcription. We show that eRBC and aRBC sustain elevated levels of transcription of components of the snRNP 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 erythrocytes<sup>57</sup>. A recent study suggests that regulation of RNA pol II activity is important during terminal erythroid maturation in humans<sup>58</sup>.

Finally, paused promoters in eRBC and aRBC 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 stimulus like immune challenges<sup>43,59</sup>.

Our results reveal a dramatic reorganization of the genome at different scales during the formation of nucleated erythrocytes in chicken including the inversion of euchromatin regions to the nuclear periphery in aRBC 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 mount a transcriptional response.

## Methods

### Blood collection and cell culture

All applied procedures were approved considering the animal welfare practices according to the IFC animal facility board. Blood of 10-day old chicken embryos or 28-38 weeks-old White-leghorn hens was collected in cold phosphate saline buffer with 50 mM EDTA. Cells were centrifuged at 1,000 rpm, 4°C for 10 minutes and washed three times and counted on a Neubauer chamber. Chicken fibroblasts were prepared from 10-day old chicken embryos

Chicken fibroblasts and the erythroblasts HD3 cell line were cultured in Dulbecco's modified Eagle's medium (DMEM; GIBCO) supplemented with 10% fetal bovine serum (Biowest), 2% chicken serum and penicillin/streptomycin (Invitrogen).

### ATAC-seq library preparation

ATAC-seq experiments were performed as previously described<sup>60</sup> with minor modification for chicken cells. 150,000 cells lysed on lysis buffer (10 mM Tris-Cl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.01% Igepal) for 15 minutes on ice. Fibroblasts were sieved through a 70 µm cell-strainer and dounce homogenized before lysis. Nuclei were centrifuged at 600 x g, 4°C for 10 minutes and washed with 50 µL of lysis buffer. Nuclei were incubated with the Tn5 transposase (Illumina) and tagmentation buffer (Illumina) for exactly 30 minutes at 37°C. After tagmentation,

fragments were purified using MinElute columns (Qiagen). ATAC-seq libraries were amplified by 10 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<sup>61</sup> 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 markup) and reads longer than 150 bp (samtools view) were removed. Signal tracks were obtained using deeptools (bamCoverage –normalizeUsing CPM) and visualized with pyGenomeTracks<sup>62</sup>.

Accessible peaks were identified on each replicate using MACS2<sup>63</sup> callpeak function with options -f=BAMPE and -q 0.01. Consensus peaks conserved in between the two biological replicates were conserved for subsequent analysis.

### **RNA purification and 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 1µg 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.

### **RNA-seq data processing**

RNA-seq data was analyzed using the SnakePipes pipeline<sup>64</sup>. Briefly, RNA-seq reads were mapped to the GRCg6a genome using STAR<sup>65</sup>, raw counts were obtained using featureCounts<sup>66</sup>, and FPKM were calculated. Because RNA content in chicken erythrocytes is significantly lower than in the other cell types, raw counts and FPKM were normalized for the number of cells required to obtain 1 µg of RNA. Genes with  $\geq 0.3$  FPKM were considered as expressed genes. Differential expression analysis was performed with DESeq2<sup>67</sup> using an FDR value of 0.5 and FC of 2.

### **Gene Ontology analysis**

Gene ontology and pathway enrichment analysis were done using g:Profiler<sup>68</sup> using 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 described<sup>69</sup>. Briefly, 10 x 10<sup>6</sup> cells were fixed with formaldehyde 2% for 10 minutes at room temperature. The genomes were digested with MboI (NEB) overnight and 5' overhangs were filled with Biotin-14-dATP (Invitrogen) and religated in nuclei. DNA fragments were purified and shared 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-Pro<sup>70</sup>. 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 ICE normalized matrices at 20 Kb, 50 Kb and 100 Kb bin resolution. Correlation between Hi-C replicates and counts vs distance plots were generated with HiCExplorer<sup>62</sup> 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 HiCPlotter<sup>71</sup>.

### **TAD calling**

Dense Hi-C matrices generated with HiC-Pro (sparseToDense.py) and TAD calling was conducted with TADtool<sup>37</sup> 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 HiCExplorer (hicPCA) at 50 Kb resolution. RNA-seq tracks were used to assign the sign of PCA results.

## Antibodies

The following antibodies were used for immunofluorescence experiments: anti-H3K9me3 (Millipore, #07-442), anti-RNA pol II Ser5-P (Santa Cruz, #SC-47701), anti-RNA pol II Ser2-P (Abcam, #ab5095), Lamin B1 (Abcam, #ab16048), anti-Lamin A/C (Santa Cruz, #SC-376248), anti-rabbit Alexa Fluor 594-conjugated (Life Technologies, #R37117), and anti-mouse FITC-conjugated (Millipore, AP192F).

## 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% bovine serum albumin (BSA) for 1 hour and 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 hour and stained with DAPI for 20 minutes, then mounted using Fluoromount (Sigma). Image acquisition was performed in a Zeiss LSM800 confocal microscope.

## ChIP-qPCR

For ChIP-qPCR experiments, four biological replicates per cell type were fixed in 1% formaldehyde for 10 min. Samples were sonicated for 10 cycles of 30sec on/60seg off with 30% amplitude using a Cole-Parmer 500-watt ultrasonic Processor. Chromatin immunoprecipitation was performed using 5µg of anti-RNA pol II Ser5-P antibody (Santa Cruz, #SC-47701) 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 to a negative control without protein occupation (SOX2 gene). Statistically significant differences were analyzed using one-way ANOVA followed by the Fisher's LSD post hoc test.

## ATAC-see

ATAC-see experiments were performed following the published protocol<sup>31</sup>. Briefly, following fixation and lysis, cells were incubated with the Tn5-ATTO-590N transposase (donated by Dr. Xingqi Chen) for 30 minutes 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 manufacturer's instructions. Cells were fed with 5-ethynyl uridine 1 mM for 6 hours, fixated with formaldehyde 3.7%, and permeabilized with Triton X-100 0.5%. 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 Fiji<sup>72</sup>. Nuclei were identified and segmented 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-see signal was filtered out using the nuclear outlines generated. A total of 50 nuclei were analyzed. Mean intensity per nucleus was measured and 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. 50 individual nuclei were analyzed per experiment and average signal was calculated. Correlation analysis between signal intensity and distance from the nuclear rim was performed in R.

### **Data accessibility**

All data sets have been deposited in the GEO repository with accession number GSE206194. Access to the data sets will be provided upon paper acceptance in a peer-reviewed journal. For more information please contact the corresponding author.

### **Acknowledgements**

We thank Dr. Xingqi Chen for his generous donation of the Tn5-ATTO-590N enzyme. We thank Dr. Félix 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 and in particular to Dr. Ruth Rincón-Heredia and Dr. Augusto Cesar Poot-Hernández for advice on

image processing and genomic data analysis respectively. We thank the Animal Facility at the IFC in particular Dr. Claudia Rivera-Cerecedo and Dr. Hector Malagón-Rivera. We thank the Supercomputing Unit at LANCIS and in particular to M.Sc. Rodrigo García-Herrera for access to the computer cluster. We thank Dr. Ricardo Saldaña-Meyer and Dr. Martín Escamilla Del Arenal for helpful discussion on the results obtained in this work.

## Funding

Andrés Penagos-Puig is supported by CONACyT scholarship 822335. This work was supported by grants PAPIIT IN207319 and IA201817 and CONACyT grant 303068.

## Authors Approval

All authors have seen and approved the manuscript. This manuscript hasn't been accepted or published elsewhere.

## Competing Interests

The authors declare no competing interests

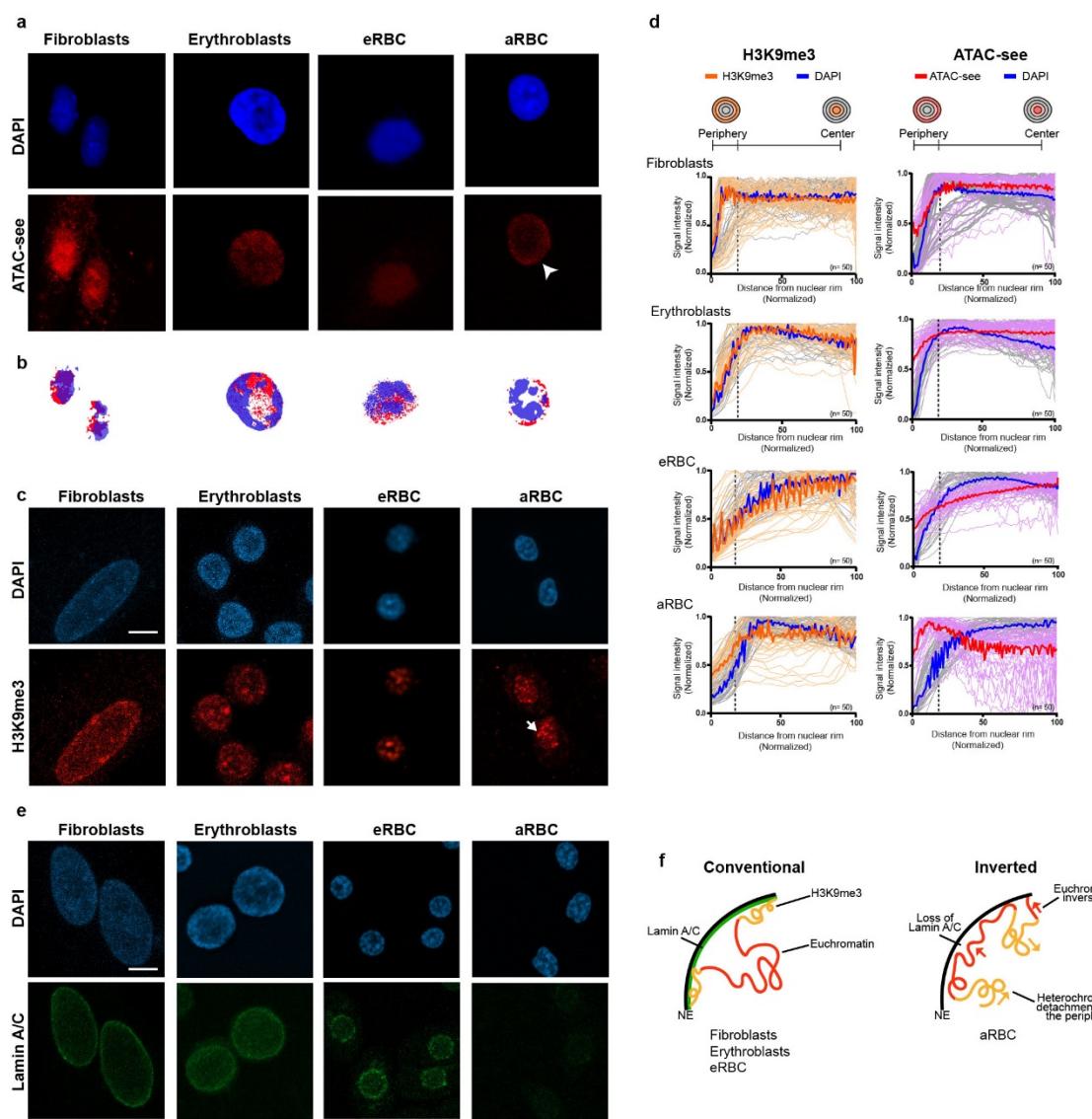
## References

1. Jaypal, S.R. et al. Down-regulation of Myc is essential for terminal erythroid maturation. *J Biol Chem* **285**, 40252-65 (2010).
2. 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).
3. 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).
4. Zhao, B. et al. Nuclear Condensation during Mouse Erythropoiesis Requires Caspase-3-Mediated Nuclear Opening. *Dev Cell* **36**, 498-510 (2016).
5. Ji, P., Murata-Hori, M. & Lodish, H.F. Formation of mammalian erythrocytes: chromatin condensation and enucleation. *Trends Cell Biol* **21**, 409-15 (2011).
6. Hartenstein, V. Blood cells and blood cell development in the animal kingdom. *Annu Rev Cell Dev Biol* **22**, 677-712 (2006).
7. 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).
8. 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).
9. 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).

10. 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).
11. 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).
12. Cameron, I.L. & Prescott, D.M. RNA and protein metabolism in the maturation of the nucleated chicken erythrocyte. *Exp Cell Res* **30**, 609-12 (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**(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**, 289-93 (2009).
17. Dixon, J.R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-80 (2012).
18. Nora, E.P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-5 (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**, 900-10 (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** ( Pt 1), 257-65 (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**, 422-31 (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**, 584-98 (2013).

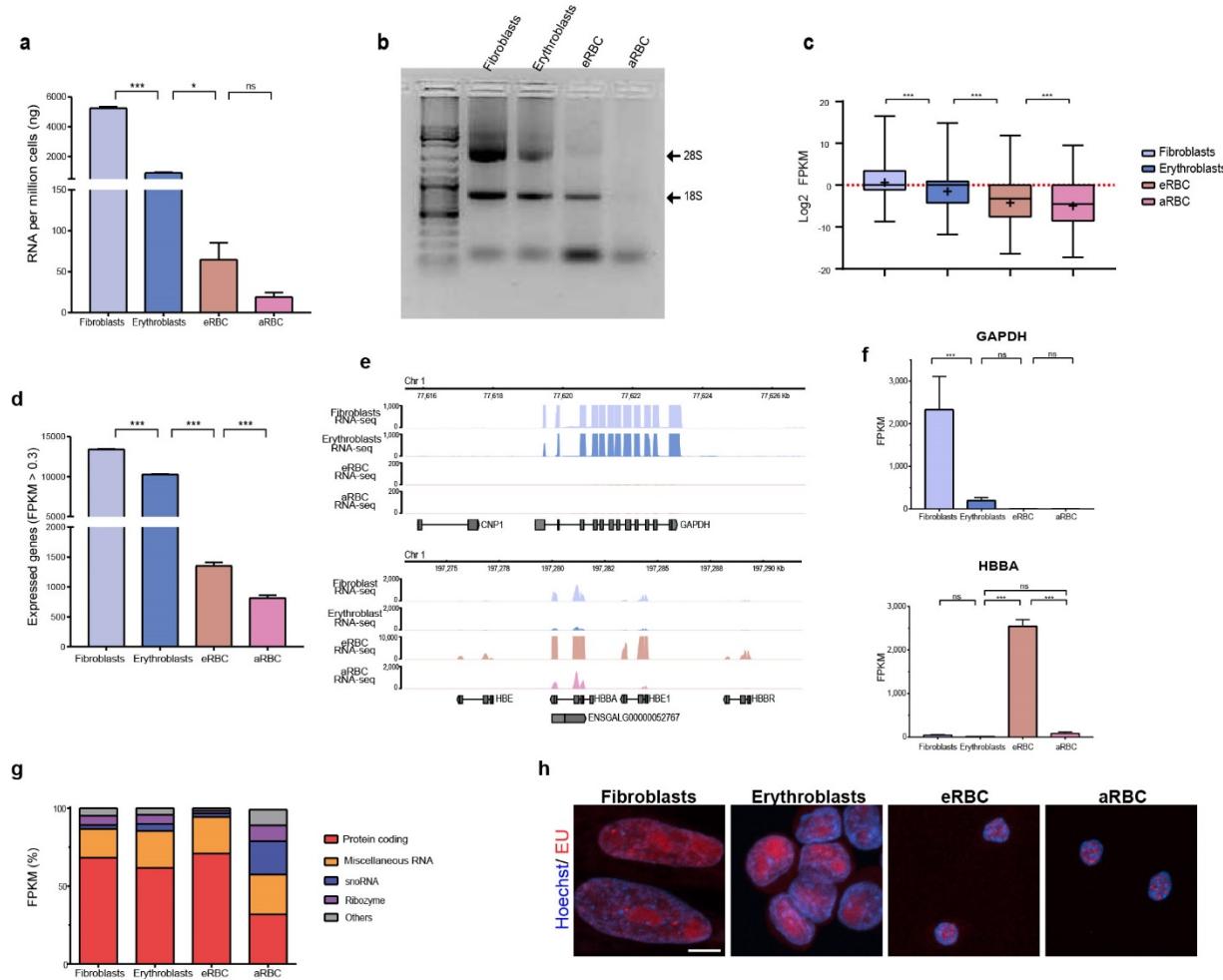
33. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev* **33**, 960-982 (2019).
34. Peterlin, B.M., Brogie, J.E. & Price, D.H. 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. *Wiley Interdiscip Rev RNA* **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**, 1665-80 (2014).
39. 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).
40. Singh, A., Compe, E., Le May, N. & Egly, J.M. TFIIH 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**, 929-33 (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**, 4562-73 (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**, 5540-50 (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**, 382-99 (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., Wu, F., Zhou, S., Huang, X.-J. & Lee, H.-Y. Heterochromatin rewiring and domain disruption-mediated chromatin compaction during erythropoiesis. 2021.08.12.456090 (2021).
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**, 2038-49 (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**, 502-11 (2011).

57. 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).
58. Murphy, Z.C. et al. Regulation of RNA polymerase II activity is essential for terminal erythroid maturation. *Blood* **138**, 1740-1756 (2021).
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).
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**, 357-9 (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**, 923-30 (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. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-82 (2012).



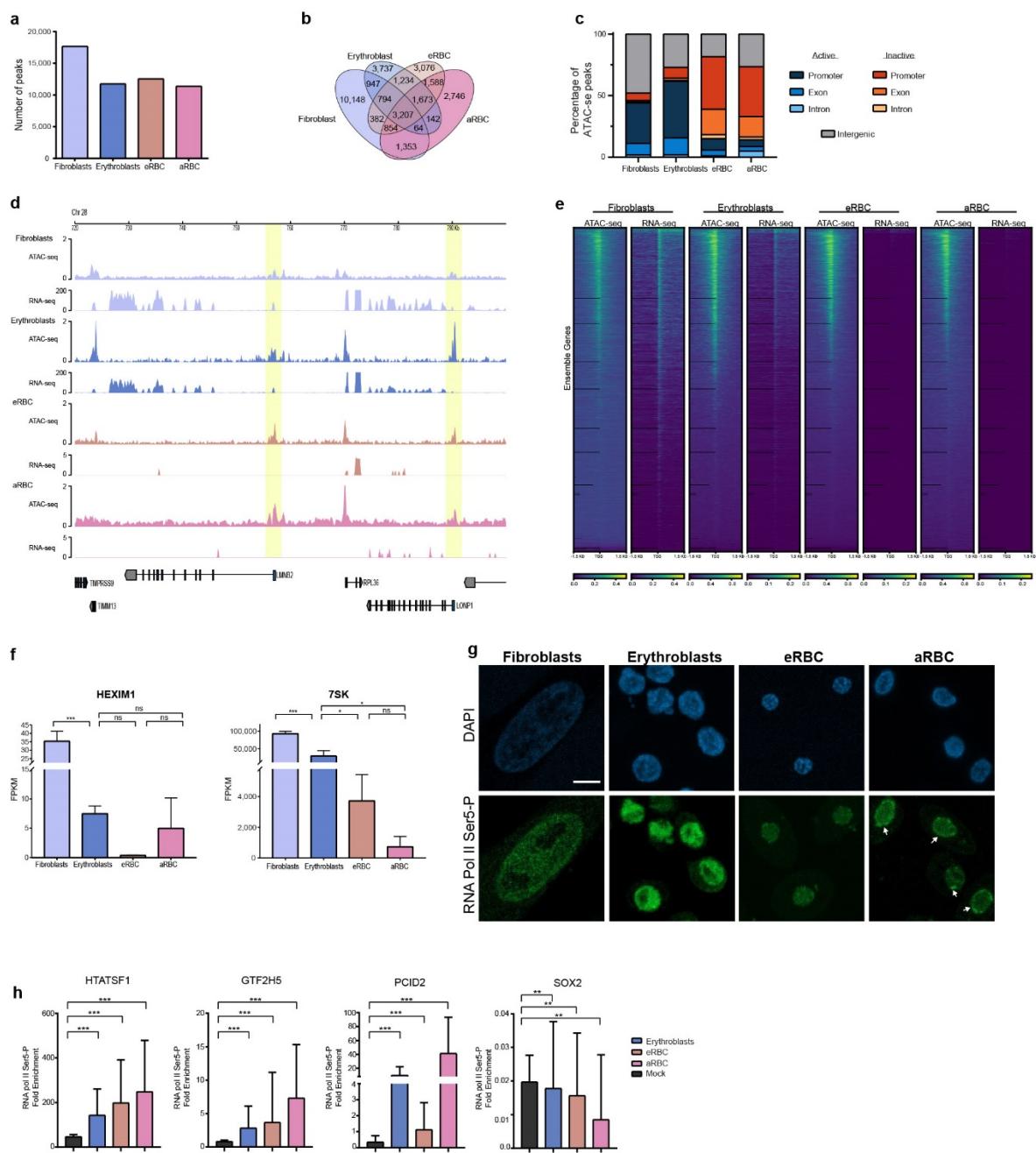
**Fig. 1. Chromatin remodeling during chicken erythropoiesis results in an inverted nuclear architecture.**

**a**, Imaging of open chromatin by ATAC-see (red) and DAPI (blue, nuclei) in fibroblasts and erythroid cells. Repositioning of the accessible chromatin towards the nuclear periphery in aRBC is indicated with an arrowhead. **b**, Brightest pixels enriched in ATAC-see (red) or DAPI (blue). **c**, Immunofluorescence imaging of H3K9me3 (red) and DAPI. In aRBC the heterochromatin is relocated towards the nuclear center (arrows). **d**, For each cell type, H3K9me3, ATAC-see 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. **e**, Fluorescence imaging of Lamin A/C proteins (green) and DAPI. In every case, maximal projections of three-dimensional (3D) stacks are presented. Scale bar = 5  $\mu$ m. **f**, In contrast to the conventional nuclear architecture in fibroblast, erythroblasts, and eRBC in aRBC the loss of the nuclear lamin A/C leads to heterochromatin detachment from the nuclear periphery and euchromatin inversion.



**Fig. 2. Chicken erythrocytes show a dramatic decrease in transcription but continue expressing small RNAs and erythroid genes.**

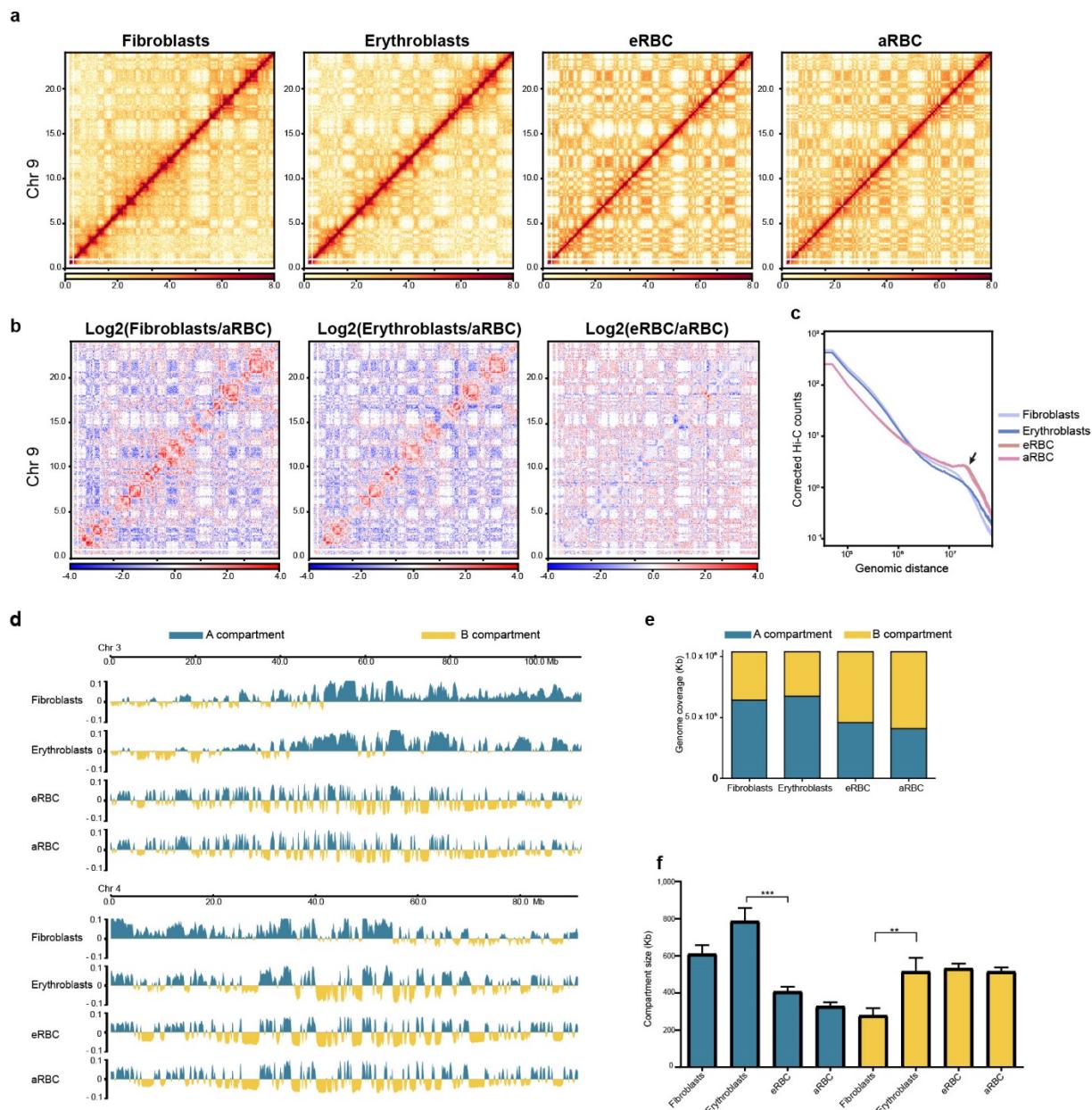
**a**, Quantification of total RNA obtained from chicken fibroblasts, erythroblasts, eRBC, and aRBC. Shown are values and means of 3 independent replicates. Asterisks indicate P values based on one-way ANOVA followed by Fisher's LSD post-hoc test. \*P < 0.05; \*\*\*P < 0.001; ns not significant differences. **b**, Representative picture of an agarose gel electrophoresis of total RNA. **c**, Mean Log2 FPKM values of three RNA-seq experiments. Mean is shown as +. Statistical differences were determined based on a one-way ANOVA followed by Dunn's post hoc test. \*\*\* P < 0.01. **d**, Number of expressed genes (FPKM > 0.3) according to RNA-seq experiments. Asterisk indicate statistical differences based on one-way ANOVA followed by Tukey's post hoc test. \*\*\* P < 0.01 **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 standard deviation. Asterisk indicate statistical differences based on one-way ANOVA followed by Tukey's post hoc test. \*\* P < 0.05; \*\*\* P < 0.01; ns not significant differences. **g**, RNA diversity expressed as the percentage of total FPKM that correspond to protein coding genes, Miscellaneous RNA, snoRNA, ribozyme or others. **h**, Fluorescent labeling of nascent RNA with 5-ethynil uridine (EU, red). Maximal projections are presented. Scale bar = 5  $\mu$ m



**Fig. 3. Open chromatin in the chicken erythrocytes comprises paused promoters of silenced genes.**

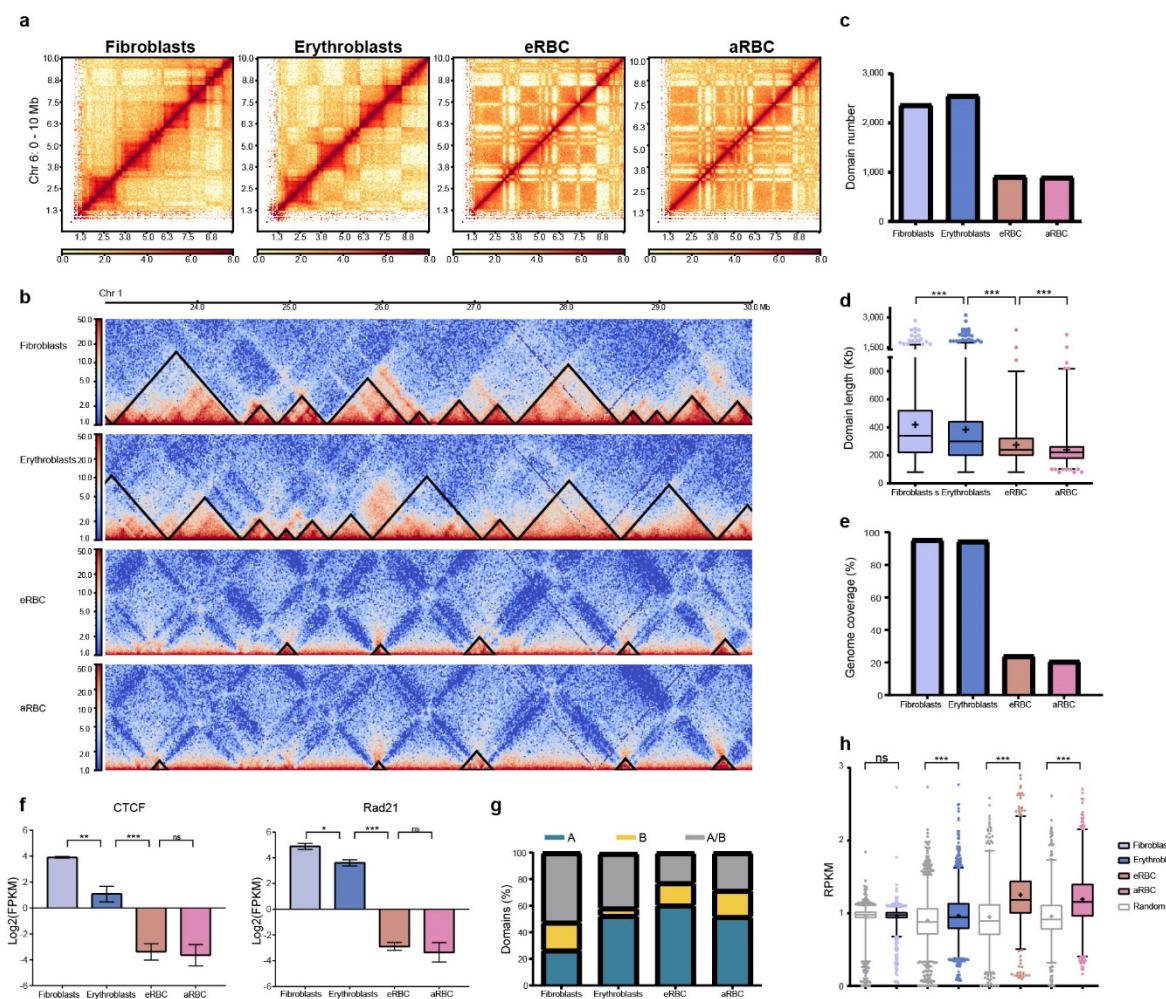
**a**, Number of ATAC-seq peaks detected in each cell type. **b**, Venn diagram of ATAC-seq peaks in each cell type. **c**, Annotation of ATAC-seq peaks at intergenic regions, promoters, exons and introns of transcriptionally active or inactive genes. **d**, ATAC-seq and RNA-seq tracks showing an example of inactive genes in eRBC and aRBC 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* snRNP involved in RNA polymerase pausing ( $n = 3$ ). Asterisk indicated significant differences based on a one-way ANOVA followed by Tukey's post hoc test. **g**, Imaging of the RNA pol II Ser5-P. In aRBC the

RNA pol II Ser5-P accumulates at the nuclear periphery (arrow) Maximal projections are presented. Scale bar = 5  $\mu$ m. **h**, 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 and standard deviation from 4 replicates is presented. Statistical differences between against a mock control were calculated with a t-test.



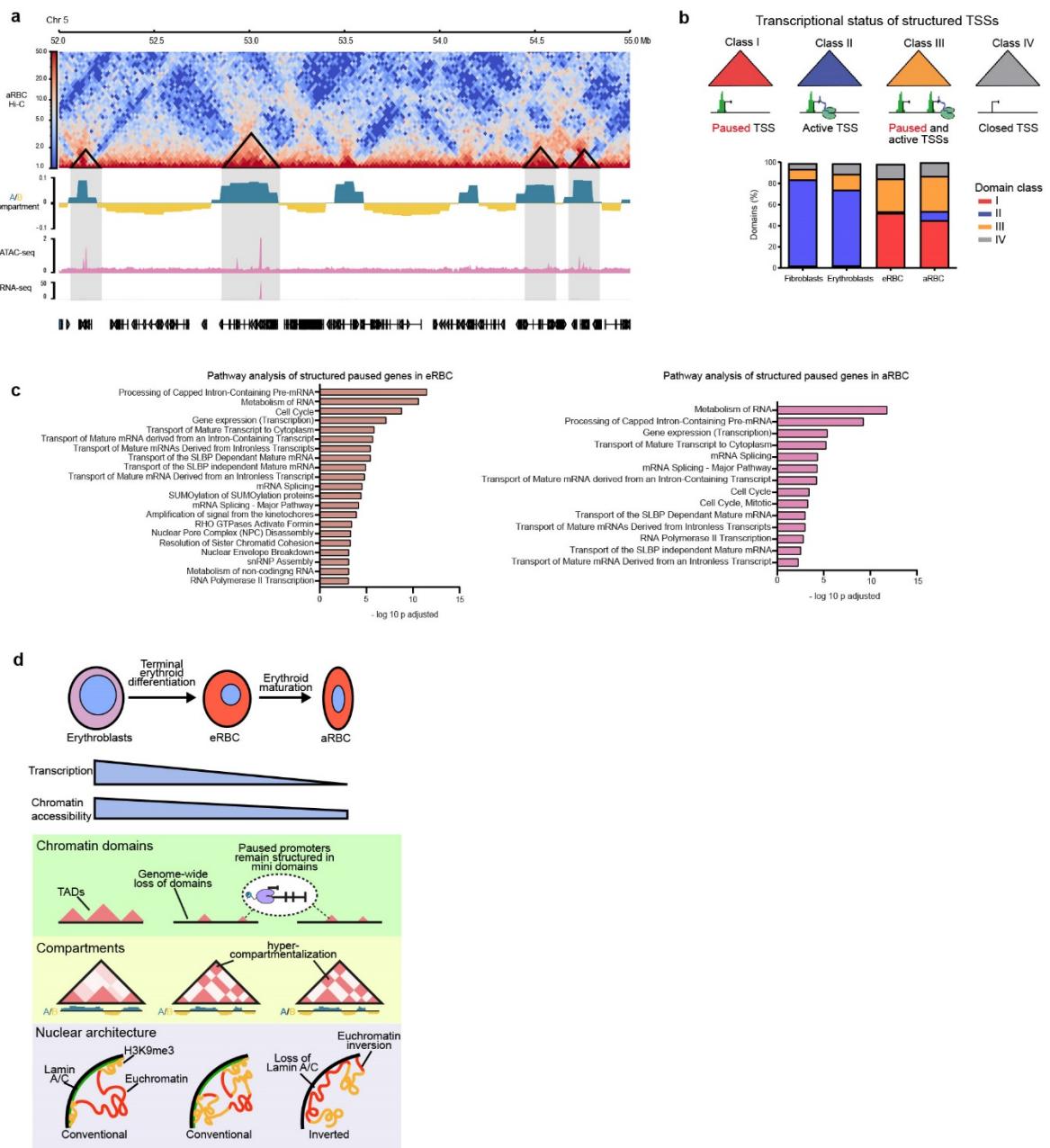
**Fig. 4. Chicken erythrocytes have hyper-compartmentalized genomes.**

**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 aRBC. **c**, Hi-C contacts counts as a function of genomic distance. An increase in long-range interaction in eRBC and aRBC is indicated with an arrow. **d**, A (blue) and B (yellow) compartment identification by PCA analysis in fibroblast, erythroblasts, eRBC, and aRBC in chromosome 3 (top) and 4 (bottom). **e**, Genome coverage of the A and B compartment. **f**, Average size of A and B compartment regions. Asterisks indicate P values based on one-way ANOVA followed by Bonferroni's post-hoc test. \*\*P < 0.01; \*\*\*P < 0.001.



**Fig. 5. Chicken erythrocytes present global TAD loss but retain structure at mini domains enriched in open chromatin.**

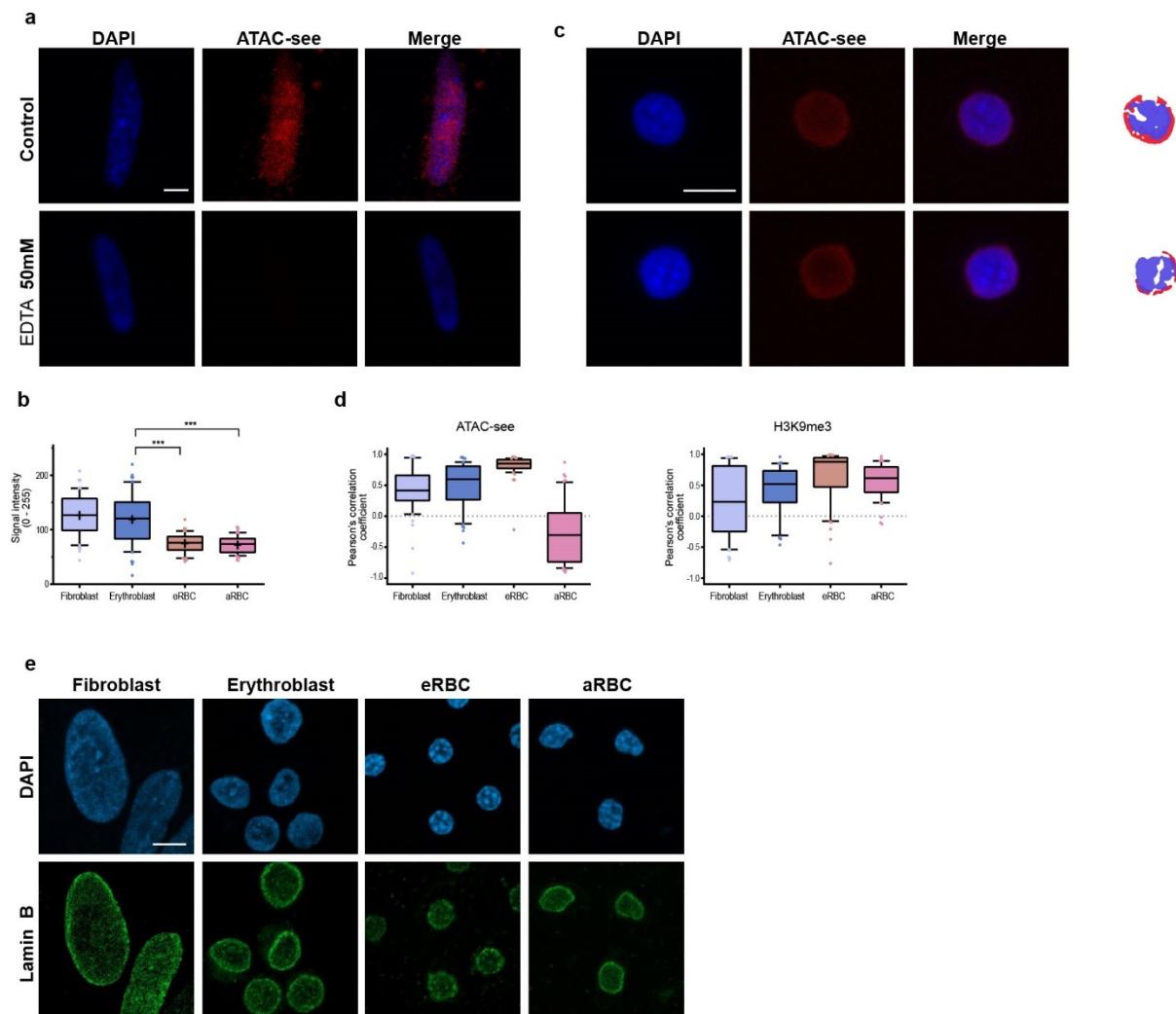
**a**, Hi-C contact matrices of the first 10 Mb of chromosome 6. TADs are clearly visible in fibroblast and erythroblasts while the most prominent signal in erythrocytes corresponds to the plaid pattern of compartments. **b**, Domains identified in each cell type indicated by black triangles. Number (**c**,) and average size of the domains is presented (**d**,). **e**, Percentage of the genome structured in domains. **f**, Expression levels of architectural proteins based on RNA-seq data. Asterisks indicate P values based on one-way ANOVA followed by Tukey's post hoc test. ns = P > 0.05; \* P < 0.05, \*\* P < 0.01, \*\*\*P < 0.001. **g**, Percentage of domains that contain only A compartment regions (blue), B compartment regions (yellow), or both (gray). **h**, ATAC-seq RPKM count inside domains or a random control. Asterisks indicate P values based on one-way ANOVA followed by Kruskal-Wallis post hoc test. ns = P > 0.05; \*\*\*P < 0.001.



**Fig. 6. Mini domains contain and organize paused promoters in chicken erythrocytes.**

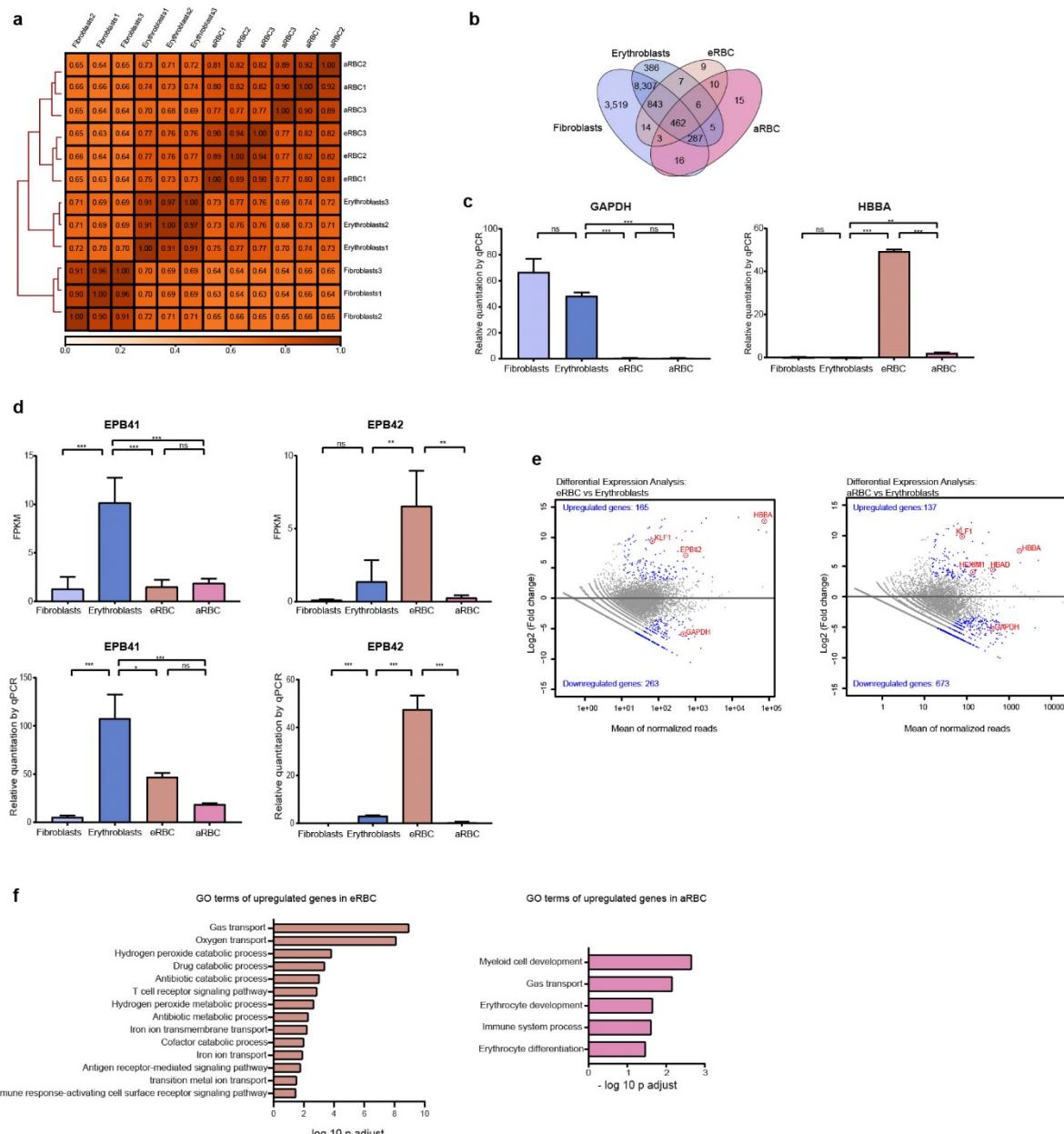
**a**, Hi-C contact matrix from aRBC (5: 52 – 55 Mb), the mini domains identified (black triangles) and the corresponding compartment signal, ATAC-seq, and RNA-seq tracks. aRBC mini domains contain ATAC-seq peaks independent from transcription (grey). **b**, Distribution of mini domains according to 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 mini domains. **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 mini domains 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 aRBC the absence of Lamin A/C causes chromatin inversion.



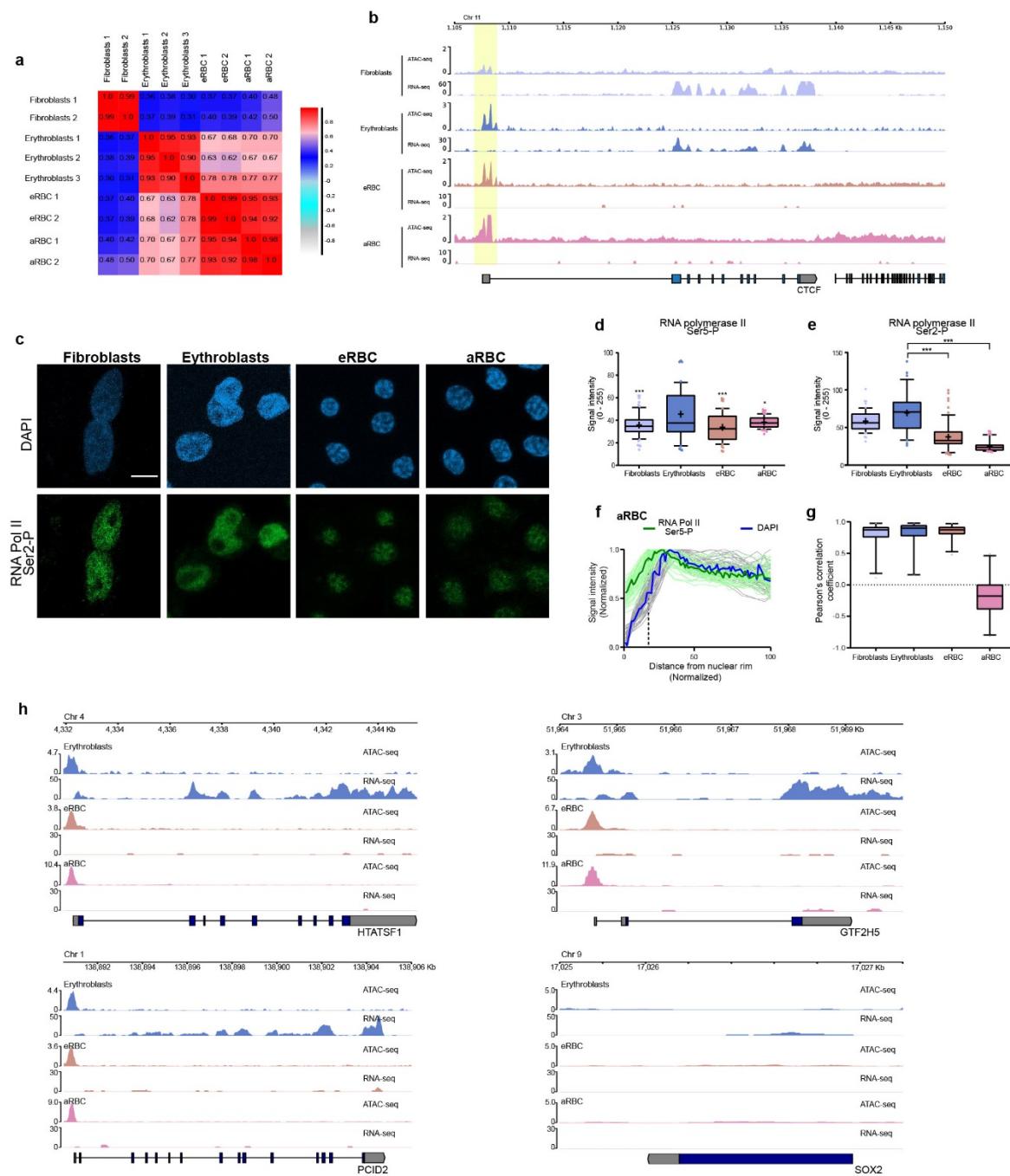
**Extended data 1. Validation of ATAC-see and chromatin inversion in aRBC.**

**a**, Imaging of open chromatin by ATAC-see (red) and DAPI (blue, nuclei) in chicken fibroblasts. A negative control of ATAC-see in presence of EDTA 50 mM shows no incorporation of the fluorophore. **b**, ATAC-see signal average intensity per nucleus in chicken erythrocytes. Boxplots and mean of measurements from 50 nuclei are shown. Asterisks indicate P values based on one-way ANOVA followed by Tukey's post-hoc test. \*\*\*P < 0.001. **c**, Representative images of ATAC-see in aRBC. Brightest pixels enriched in ATAC-see (red) or DAPI (blue) are shown on the right. **d**, Pearson's correlation score between ATAC-see or H3K9me3 signal intensity and the distance to the nuclear rim (n = 50) **e**, Fluorescence imaging of Lamin B1 (green) and DAPI. In every case, maximal projections of three-dimensional (3D) stacks are presented. Scale bar = 5  $\mu$ m.



## Extended data 2. Characterization of the erythroid transcriptome.

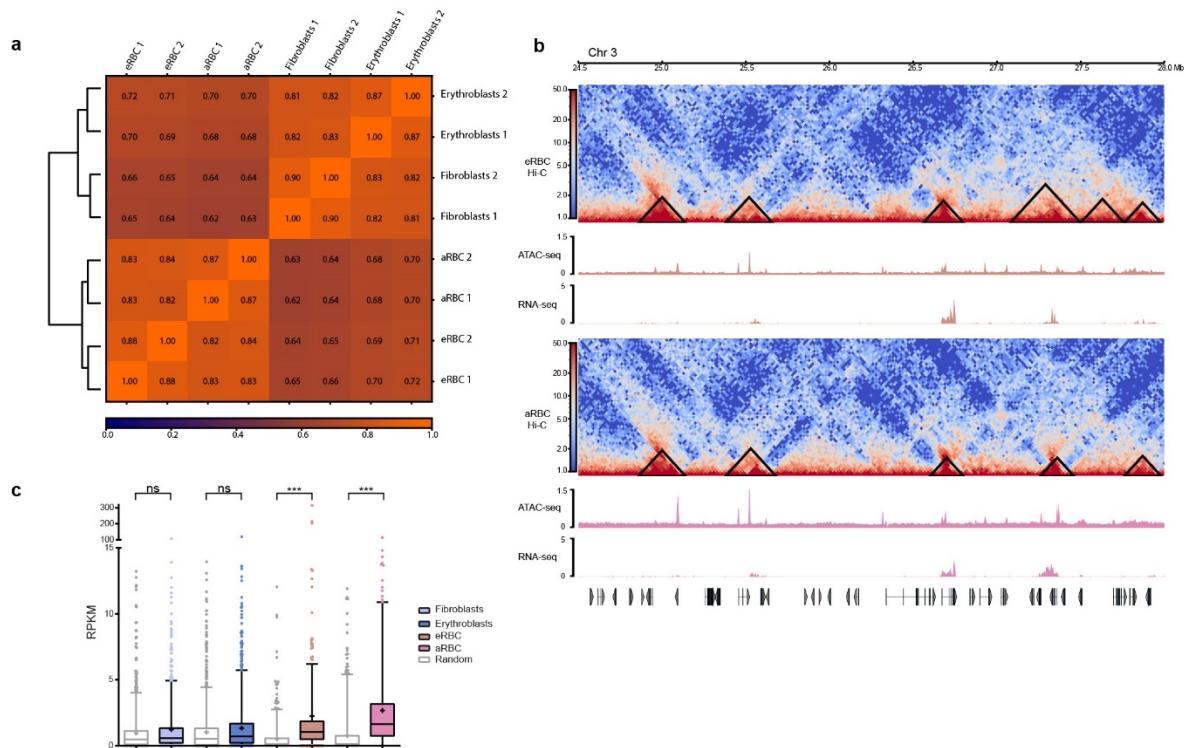
**a**, Spearman correlation heatmap of RNA-seq replicates. **b**, Venn diagram of expressed genes in fibroblasts and erythroid cells. **c**, Relative expression of GAPDH and HBBA measured by RT-qPCR. **d**, Expression levels of the erythroid genes EPB41 and EPB42 measured by RNA-seq (up) and RT-qPCR (down). Asterisks indicate P values based on one-way ANOVA followed by Tukey's post-hoc test. **e**, MA plots of differentially expressed genes between eRBC and erythroblasts, and aRBC and erythroblasts. **f**, GO terms enriched in upregulated genes from **e**.



### Extended data 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. Maximal projections of three-dimensional (3D) stacks are presented. Scale bar = 5  $\mu$ m. Average signal intensity of RNA pol II Ser5-P (**d**,**e**) and Ser2-P (**e**,**f**) signal average intensity per nucleus in chicken erythrocytes. Boxplots and mean of measurements from 50 nuclei are shown. Asterisks indicate P values based on one-way ANOVA followed by Tukey's post-hoc test. \*P < 0.05; \*\*\*P < 0.001; ns not significant differences. **f**, 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. **g**, Pearson's correlation score between RNA pol II Ser5-P signal intensity and the distance to the nuclear rim (n = 50). **d**, ATAC-seq and RNA-seq tracks of HTATSF1, GTF2H5, PCID2, and SOX2.



#### Extended data 4. 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**, RNA-seq RPKM count inside domains or a random control. Asterisks indicate P values based on one-way ANOVA followed by Kruskal-Wallis post hoc test. ns = P > 0.05; \*\*\*P < 0.001.