Dissection of the 4D chromatin structure of the α-globin locus through *in vivo* erythroid differentiation with extreme spatial and temporal resolution

A. Marieke Oudelaar[1,2,5], Robert A. Beagrie[1,5], Matthew Gosden[1], Sara de Ornellas[1,3], Emily Georgiades[1,2], Jon Kerry[2], Daniel Hidalgo[4], Joana Carrelha[1], Arun Shivalingam[3], Afaf H. El-Sagheer[3], Jelena M. Telenius[1,2], Tom Brown[3], Veronica J. Buckle[1], Merav Socolovsky[4], Douglas R. Higgs[1*] and Jim R. Hughes[1,2*]

[1]MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK.

[2]MRC WIMM Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK.

[3]Department of Chemistry, University of Oxford, Oxford, UK.

[4]Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, Massachusetts, USA.

[5]These authors contributed equally: A. Marieke Oudelaar, Robert A. Beagrie.

*e-mail: jim.hughes@imm.ox.ac.uk, doug.higgs@imm.ox.ac.uk

Precise gene expression patterns during mammalian development are controlled by regulatory elements in the non-coding genome. Active enhancer elements interact with gene promoters within Topologically Associating Domains (TADs)[1-3]. However, the precise relationships between chromatin accessibility, nuclear architecture and gene activation are not completely understood. Here, we present Tiled-C, a new Chromosome Conformation Capture (3C) technology[4], which allows for the generation of high-resolution contact matrices of loci of interest at unprecedented depth, and which can be optimized for as few as 2,000 cells of input material. We have used this approach to study the chromatin architecture of the mouse α-globin locus through *in vivo* erythroid differentiation. Integrated analysis of matched chromatin accessibility and single-cell expression data shows that the α-globin locus lies within a pre-existing TAD, which is established prior to activation of the domain. During differentiation, this TAD undergoes further sub-compartmentalization as regulatory elements gradually become accessible and specific interactions between enhancers and promoters are formed. As these chromatin changes develop, gene expression is progressively upregulated. Our findings demonstrate that chromatin architecture and gene activation are tightly linked during development and provide insights into the distinct mechanisms contributing to the establishment of tissue-specific chromatin structures.

A major current goal in biology is to characterize the three-dimensional nuclear architecture of the genome, to determine how this structure changes during differentiation and development, and to understand how these changes relate to gene expression. It has been shown that the genome is extensively reorganized during differentiation[5,6] and that interactions between enhancers and promoters can be established prior to expression, which is suggested to prime gene loci for future activation[7,8]. However, it remains unclear precisely when specific interactions between enhancers and promoters are formed relative to activation of gene expression[9-11]. To better understand the relationship between chromatin architecture and gene expression, it is crucial to characterize chromatin structure at high resolution in pure, primary cell populations representing relevant developmental stages. This has been hampered by the lack of high-resolution 3C methods that are suitable for the analysis of limited numbers of primary cells.

To overcome these hurdles, we developed Tiled-C, a new 3C-based approach[4], which can generate high-resolution contact matrices of selected regions of interest. Tiled-C uses a panel of capture oligonucleotides tiled across all restriction fragments of specified genomic regions, combined with

an adapted Capture-C based protocol[12], to efficiently enrich for 3C contacts within this region. This allows for deep, targeted sequencing of chromatin interactions within regions of interest and thus for the generation of high-resolution, Hi-C-like data at unprecedented depth, across multiplexed samples and genomic regions (Methods). Tiled-C combines the ability of all vs all methods such as Hi-C[13] to map large-scale chromatin structures including TADs, as well as the ability of one vs all methods such as 4C[14,15] and Capture-C[12,16] to robustly identify enhancer-promoter interactions within TADs in detail (Supplementary Figure 1,2). To validate the Tiled-C approach, we compared Tiled-C data to the deepest currently available *in situ* Hi-C datasets (mouse ES cells[5]; Figure 1a). Tiled-C data at this region was ~28-fold higher in depth and required ~19-fold less sequencing (Supplementary Table 1). To further demonstrate the capabilities of Tiled-C, we performed experiments in which we multiplexed across several key gene loci in both primary mouse erythroid cells and ES cells, to characterize their tissue-specific configurations in depth (Figure 1b-c, Supplementary Figure 3). Importantly, the combination of the efficient tiled enrichment strategy with previous Capture-C library preparation optimizations[17] allows Tiled-C to generate high-resolution contact matrices from as few as 2,000 cells (Figure 2, Supplementary Figure 4, Supplementary Table 2) allowing for the analysis of previously intractable primary cell types. This is critical for the investigation of 4D (3D structure through developmental time) genome organization, as cell numbers become extremely limiting at early stages of development.

We used Tiled-C to study the changes in chromatin structure associated with gene activation in primary cells during erythroid differentiation. Using fluorescence-activated cell sorting (FACS), we isolated sequential stages of erythroid differentiation directly from mouse fetal livers[18,19] (Figure 3a,b). We focused our analysis on a ~3.3 Mb region containing the well-characterized α-globin genes and their associated regulatory elements. The α-globin genes are regulated by five erythroid-specific enhancer elements (R1-R4 and Rm), which classify as a super-enhancer[20], and interact with the gene promoters within a TAD flanked by multiple CTCF-binding elements[21-24] (Supplementary Figure 2). We generated a single-cell RNA-seq dataset[25], which is the first scRNA-seq dataset to include the full course of *in vivo* erythroid differentiation through to terminal differentiation in the mouse (Supplementary Figure 5). This dataset shows that α-globin is expressed at basal levels in the S0 populations. Expression of α-globin dramatically increases during the S2 stage and plateaus at S3, however the earliest cells showing elevated expression of α-globin are found in the S1 stage (Figure 3c). To validate that erythroid-specific α-globin upregulation begins at S1, we performed RNA-FISH to detect nascent transcription in FACS-sorted primary cells (Figure 3d). We detect a

small increase in nascent transcription from S0-low to S0-medium cells, and confirm a robust increase in expression from S0 to S1 cells ($P < 0.005$ by paired T-test; Figure 3e).

We used the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)[26] to profile chromatin accessibility in these stages. Interestingly, we find that both enhancer and promoter elements are accessible prior to the onset of erythroid-specific gene expression, and that the degree of accessibility gradually increases, concomitant with upregulation of α-globin expression (Figure 4, Supplementary Figure 6). Tiled-C shows that a TAD structure encompassing the α-globin locus is present at the earliest stage (S0-low), prior to the formation of weak enhancer-promoter interactions in the S0-medium stage (Figure 4, Supplementary Figure 7). These enhancer-promoter interactions further strengthen in the subsequent S1 and S2 stages, accompanied by increases in α-globin expression and accessibility. In the S3 stage, where chromatin accessibility and expression reach their maximum levels, we observe strong enhancer-promoter interactions in a sub-compartmentalized chromatin structure similar to that observed in primary erythroblasts derived from mature spleen tissue (Figure 1, Supplementary Figure 1,2). This smaller sub-compartmentalized structure, which forms within the pre-existing TAD, is delimited by convergent CTCF-binding elements that flank the α-globin enhancers and genes (Supplementary Figure 2). We have previously shown that these CTCF-binding elements are functionally important to restrict the interactions of the α-globin enhancers and prevent other genes within the TAD, but outside of the sub-compartmentalized structure, from being upregulated[22] (Figure 4, Supplementary Figure 2). This suggests that this smaller erythroid-specific domain is likely formed by similar CTCF-dependent mechanisms as TADs, although it is smaller in size (~70 kb) and has very high internal interaction frequencies compared to typical TADs.

Since both accessibility and the encompassing TAD structure are present prior to erythroid-specific α-globin expression, we purified early hematopoietic progenitor populations to investigate when in differentiation these features are established. Interestingly, we find that the pre-existing TAD containing the α-globin locus is already present in hematopoietic stem cells, despite four out of five enhancers and both promoters being inaccessible at this stage (Supplementary Figure 8).

To examine whether a similar order of events operates at other gene loci, we examined the chromatin architecture of Cpeb4, which is located ~420 kb upstream of the α-globin genes. Cpeb4 is an essential gene for terminal erythropoiesis[27] and is gradually upregulated during erythroid

differentiation (Supplementary Figure 9). Tiled-C shows that the TAD encompassing Cpeb4 is established prior to erythroid-specific gene expression. Enhancer accessibility and a sub-compartmentalized structure in which the enhancers and promoters interact are formed during differentiation, in parallel with activation of gene expression (Supplementary Figure 10), thus mirroring the behavior of the α-globin locus.

Our data therefore support a model in which an encompassing TAD structure is present prior to activation of the domain. During lineage commitment, tissue-specific open chromatin sites are established within this domain. Chromatin accessibility, sub-compartmentalization of the TAD, enhancer-promoter interactions and gene expression then increase progressively during subsequent differentiation stages (Figure 5).

The early and gradual establishment of accessible regulatory elements is consistent with previous analysis of transcription factor binding at the α-globin locus during erythroid differentiation[28]. Our findings are also consistent with previous chromatin conformation studies of the of the α- and β-globin loci in erythroid cell lines[29,30]. Furthermore, our results fit with our recent analysis of the conformation of the α-globin locus in two stages of *ex vivo* erythroid differentiation. Using super-resolution microscopy, we showed that the α-globin locus forms a self-interacting domain early in differentiation, prior to maximal gene activation. As erythropoiesis proceeds and α-globin transcription is upregulated, this domain decondenses, while the flanking CTCF sites come into more frequent proximity[24]. Here, we extend these observations *in vivo* at higher spatial and temporal resolution and provide new insights into the origins of these structures by analyzing earlier stages of erythroid differentiation.

Our findings are consistent with a model emerging from the work of several labs, in which pre-existing TADs are established prior to domain activation. During early differentiation, smaller sub-structures are formed, which facilitate interactions between enhancers and promoters to prime loci for gene activation. As cells differentiate further, these interactions are strengthened, concomitant with strong upregulation of gene activity. It has recently been shown for the globin loci that gene activation is associated with the formation of higher-order hub-like structures, in which multiple enhancers and promoters form simultaneous, specific interactions[23,31]. Our data suggest that these structures may only be formed in the final stage of differentiation, when chromatin accessibility and interactions between enhancers and promoters are strongest, and may be important to achieve maximal gene expression. This model is further supported by recent live imaging experiments in

Drosophila, in which gene activation only occurred upon the formation of tight associations between enhancers and a gene promoter, and not after induced enhancer-promoter proximity resulting from interactions between insulator elements[32]. Moreover, it is consistent with previous analyses of the conformations of the Hoxd cluster[33] and the Shh locus[34] across tissue types.

Interestingly, our model implies that there are multiple processes contributing to the formation of specific chromatin structures associated with gene activation. A pre-existing TAD encompassing the α-globin locus is formed prior to and thus independent of activation of the regulatory elements within the domain. This is likely driven by tissue-invariant loop extrusion mediated by cohesin and constitutive CTCF-binding elements[35]. During differentiation, chromatin accessibility increases, and a smaller sub-domain is formed within this TAD. We have previously shown that deletion of the CTCF-binding sites at the base of this sub-domain causes it to expand and leads to aberrant expression of the neighboring genes[22]. This indicates that sub-compartmentalization is dependent on these CTCF-binding sites and implies that its formation is mediated by loop extrusion. Since the CTCF-binding sites are constitutively occupied, erythroid-specific compartmentalization is likely driven by increased rate or processivity of loop extrusion in this region during differentiation. As we have previously observed erythroid-specific accumulation of cohesin at the α-globin enhancers[22], it is possible that this is mediated by increased cohesin recruitment at the activated regulatory elements. This is further supported by studies showing that cohesin co-localizes with transcription factors across the genome[36,37].

Within the sub-compartmentalized domain, specific interactions between the activated enhancers and promoters are formed gradually during differentiation. Somewhat surprisingly, the initial formation of accessible elements significantly precedes the onset of specific enhancer-promoter interactions. This indicates that chromatin opening can occur independently of chromatin reorganization, yet further increases in accessibility do occur alongside the establishment and progressive strengthening of enhancer-promoter interactions, suggesting only a partial decoupling. This is consistent with our previous work, which showed that deletion of the α-globin enhancers does not affect formation of the α-globin TAD, but does affect interactions between the enhancers and promoters within the domain[21,24]. This suggests that active regulatory elements play a role in the formation of tissue-specific chromatin structures, possibly mediated by interactions between the multi-protein complexes bound at these elements.

In conclusion, our dissection of the chromatin architecture of a well-understood gene locus during *in vivo* differentiation provides new insights into the mechanisms which regulate gene expression patterns during development. Importantly, Tiled-C provides an approach that enables such detailed analysis in cell types that were previously intractable.

Methods

Cells

Mature erythroid cells

Mature primary Ter 119+ erythroblasts were obtained from spleens of female C57BL/6 mice treated with phenylhydrazine as previously described[12].

Mouse ES cells

Mouse ES cells were cultured and harvested as previously described[12].

Erythroid progenitors

Primary erythroid progenitor cells were isolated from fetal livers, which were freshly isolated at e12.5-e13.5 from C57BL/6 mouse embryos. 5-15 livers were pooled together for each experimental replicate, mechanically dissociated in staining buffer (PBS, 0.2% BSA, 5mM Glucose) and strained through a 30-µm strainer. Cells were immunostained at 4 °C in the presence of rabbit IgG (200 µg/ml, Jackson Laboratories 015-000-003) to block Fc receptors. To enrich for early erythroid progenitors, cells were first stained with 5 µg/ml biotin-conjugated anti-Ter119 (BD 553672) for 30 minutes, before magnetic depletion using streptavidin nanobeads (BioLegend Mojosort 480016) following the manufacturer's instructions. Cells were then incubated with 0.5 µg/ml APC-conjugated streptavidin (BD 553672), 0.33 µg/ml PE-Cy7-conjugated anti-CD71 (BD Biosciences) and a panel of 5 FITC-conjugated lineage antibodies (anti-CD41, anti-CD45R, anti-CD3e, anti-CD11b and anti-Ly-6G/6C, all at 1 µg/ml; BD 553848, 553087, 553061, 557396 and 553126, respectively) for 45 minutes. Cells were then resuspended in FACS running buffer (staining buffer plus 2 mM EDTA). 0.66 µg/ml Hoechst was added immediately prior to sorting in order to distinguish live cells. Cells were sorted into Eppendorf tubes containing 500 µl RPMI supplemented with 10% FCS using a BD FACSAria™ Fusion machine with a 100 µM nozzle size.

Hematopoietic progenitors

Hematopoietic stem and progenitor cells (HSPCs), lymphoid-primed multipotential progenitors (LMPPs) and multipotent progenitors (MPPs) were stained and FACS sorted from the bone marrow of 12-week old female C57BL/6 mice as previously described[50].

Replicates

The presented Tiled-C data derived from mature splenic erythroblasts and ES cells represent biological triplicates produced from separate mice or culture flasks, respectively. The presented Tiled-C data derived from hematopoietic and erythroid progenitor populations represent biological duplicates, with the exception of the S1 stage, for which we used a single biological replicate to generate technical duplicates. The presented ATAC data derived from hematopoietic and erythroid progenitor populations represent biological triplicates for the S0-low, S0-medium and S1 populations, biological duplicates for the S2 and S3 populations, and single replicates for the hematopoietic progenitor populations. The presented RNA-FISH data represent biological triplicates except for the brain and no-primary-antibody negative control, which represent biological duplicates.

Ethics

All protocols were approved through the Oxford University Local Ethical Review process and all experimental procedures were performed in accordance with European Union Directive 2010/63/EU and/or the UK Animals (Scientific Procedures) Act, 1986.

Tiled-C

Rationale

Tiled-C is an hybrid of the all vs all 3C methods, such as Hi-C[13], and the one vs all methods, such as 4C[14,15] and Capture-C[12,16]. Tiled-C generates all vs all contact matrices of specified genomic regions and thus combines an unbiased all vs all view with the ability to target regions of interest, without the need to sequence chromatin interactions genome-wide. Tiled-C has similarities to 5C[38], but overcomes shortcomings related to PCR primer design and duplicate filtering and is able to generate data at higher resolution and depth. Tiled-C also has similarities to Capture Hi-C[39]. The main differences are that Tiled-C enriches a 3C library rather than a biotinylated Hi-C library, thereby retaining optimal library complexity, which is critical for analysis of small cell numbers. In addition, Tiled-C uses an efficient capture oligonucleotide design targeted directly to the ends of all restriction fragments present in the region of interest. The combination of high library complexity and efficient enrichment in Tiled-C enables high-resolution data generation at great depth and makes the method suitable for the analysis of small cell numbers. Moreover, enriching for targeted regions of interest substantially decreases sequencing costs. We should note however that the synthesis of large amounts of capture oligonucleotides can also be expensive. We therefore

believe that Tiled-C is particularly useful for researchers who are interested in studying genomic regions of interest in multiple replicates and conditions and/or in primary cells with limited availability.

Oligonucleotide design

Tiled-C uses a panel of 70 bp oligonucleotides to enrich for regions of interest. The oligonucleotide sequences are designed complementary to both ends of each individual restriction fragment present in the region of interest. We use stringent BLAT-based filtering to ensure that the panel of oligonucleotides does not contain repetitive sequences that would decrease specific enrichment of the region of interest. To help users design capture oligonucleotides, we have developed a user-friendly python package (https://oligo.readthedocs.io/en/latest/). An overview of the oligonucleotide designs for the α-globin, Sox2 and Tal1 loci is shown in Supplementary Table 3.

Oligonucleotide synthesis

The enrichment step in the Tiled-C protocol can be performed with single-stranded or double-stranded biotinylated capture oligonucleotides (Supplementary Figure 4, Supplementary Table 4). We produced panels of single-stranded oligonucleotides on an in-house Combimatrix CustomArray B3 DNA synthesiser (B3Synth_v25.1 software) using CustomArray 12K Blank Slides (CustomArray Inc., PN: 2000100- Oligo pool Application). All the probe sequences were designed to be 70 bases in length and were placed at random positions on the microarray for synthesis using Layout Designer (v4.3.1). Synthesis of oligonucleotide probe sequences occurred on individual electrodes present on the semiconductor surface of the microarray by phosphoramidite chemistry in the 3' to 5' direction using standard software oligonucleotide pool synthesis settings and reagents prepared according to the manufacturer's protocols. Each sequence was synthesized in triplicate. After the synthesis of the unmodified oligonucleotide, 5'-biotin was added using a double coupling cycle with an extended 15 minutes coupling time. The final detritylation step was performed manually using the software by incubating the slides with TCA deblock (4 x 30s incubations) before washing the slide with acetonitrile four times and drying under argon. Oligonucleotides were then cleaved and deprotected on a stripping clamp system provided by the manufacturer using concentrated aqueous ammonia at 65 °C for 18 h. After cooling, the solution was recovered and the ammonia was removed by vacuum concentration. The oligonucleotide pool was purified using 2 x illustra NAP-5 Columns (GE Life Sciences, PN: 17085302). The resulting solution was evaporated to dryness, resuspended in water and quantified

by Nanodrop absorbance at 260 nm. We used ~0.1 fmol of each individual oligonucleotide per enrichment reaction.

We ordered panels of double-stranded capture oligonucleotides from Twist Bioscience (Custom probes for NGS target enrichment). As recommended by Twist, we used 13.67 fmol of each individual oligonucleotide per enrichment reaction.

Experimental procedure

We prepared 3C libraries as previously described[12]. Briefly, aliquots of ~10^7 cells were fixed with 2% formaldehyde for 10 minutes, after which the reaction was quenched with glycine. The fixed cells were washed in cold PBS and resuspended in cold lysis buffer (10 mM Tris pH 8, 10 mM NaCl, 0.2% Igepal CA-630, 1x cOmplete Protease Inhibitor Cocktail, water). After incubation on ice for 20 minutes, the cells were snap frozen. Prior to digestion, the cells were resuspended in restriction buffer, Dounce homogenized on ice, and treated with SDS and Triton X-100. The chromatin was digested with DpnII, using 3 aliquots of 1500 U DpnII restriction enzyme, which were added several hours apart over a total incubation time of 16–24 hours on a shaking thermomixer at 37 °C. The digestion reaction was heat-inactivated at 65 °C for 20 minutes and cooled on ice. Digested chromatin was ligated overnight with 720 U of T4 DNA ligase on a shaking thermomixer at 16 °C. The ligated DNA was reverse crosslinked and treated with proteinase K overnight at 65 °C. After RNase treatment, the DNA was purified using phenol-chloroform extraction and precipitation with ethanol and sodium acetate at −80 °C. The resulting 3C libraries were resuspended in PCR-grade water and stored at −20 °C. Aliquots of 5–6 µg of 3C library were sonicated to ~200 bp fragments using a Covaris S220 Focused Ultrasonicator (six cycles of 60s; duty cycle: 10%; intensity: 5; cycles per burst: 200). Illumina TruSeq adaptors were added using NEBNext DNA Library Prep reagents and the adapter-ligated libraries were indexed and amplified using Agilent Herculase II reagents in a 6-cycle PCR reaction, according to the manufacturer's protocols. All DNA clean-up steps were performed using AMPure XP beads in a 1.8:1 bead:sample ratio. Where possible, we processed 2 aliquots of each sample in parallel and ligated the DNA with the same index to generate maximum library complexity. All adaptations to the described 3C library preparation procedure for low-input samples are described in detail and are available on Bio-protocol[40].

For enrichment using single-stranded oligonucleotides, we used the Nimblegen SeqCap EZ reagents and followed the SeqCap EZ Library SR User's Guide (Chapters 5–7). We multiplexed up to 6 samples per enrichment reaction in a single tube, and multiplied the volumes described in the protocol by the number of multiplexed libraries. Briefly, 1–1.5 µg of indexed library (in exact

1:1 ratio if multiplexed reaction is performed) was mixed with 5 µg mouse Cot-1 DNA, 1 nmol of Nimblegen HE universal blocking oligonucleotides and 1 nmol of Nimblegen HE index-specific blocking oligonucleotides (corresponding to the Illumina TS index used) in a 1.5 µl microcentrifuge tube. This mixture was dried completely in a vacuum centrifuge at 50°C. The dried DNA was resuspended in 7.5 µl Nimblegen hybridization buffer and 3 µl Nimblegen hybridization component A and denatured at 95 °C for 10 minutes. Concurrently, 4.5 µl of the biotinylated capture oligonucleotides was placed in a 200 µl PCR tube and heated to 47 °C in a PCR thermocycler. After denaturation, the 3C library mixture was added to the biotinylated oligonucleotides without removing them from the heating block in the thermocycler. The hybridization reaction was incubated in the thermocycler at 47 °C for 64–72 h with a heated lid at 57 °C. After incubation was complete, we enriched for the captured DNA fragments using M270 streptavidin beads and the Nimblegen SeqCap EZ wash buffers, following the procedure described in the manufacturer's protocol. After the washing steps, the beads with captured material were resuspended in 40 µl PCR-grade water. Captured DNA was amplified directly of the beads using the KAPA master mix provided in the SeqCap EZ accessory kit v2, in 2 separate reactions of ~10 cycles, as described in the protocol. Ampure-XP beads were used in a 1.8:1 bead:sample ratio to clean up the amplification reaction and DNA was eluted in 30 µl PCR-grade water. To increase enrichment, a second round of oligonucleotide capture was performed following the same procedure, using up to 2 ug of enriched material in a single hybridization reaction (even if multiplexed in first round) of 20–24 hours.

For enrichment using double-stranded oligonucleotides, we used Twist Biosciences reagents and followed the Twist Custom Panel Protocol (Steps 4–7). To multiplex samples, we used 375–500 ng indexed library per sample, mixed up to 1.5 µg (in exact 1:1 ratio) in a single tube, and used single reaction volumes as described in the protocol. We processed multiple tubes simultaneously if required. Streptavidin C1 beads were used to enrich the hybridized DNA and the washed material was amplified using 10–12 cycles of PCR. Ampure-XP beads were used in a 1.8:1 bead:sample ratio to clean up the amplification reaction and DNA was eluted in 30 µl PCR-grade water. To increase enrichment, a second round of oligonucleotide capture was performed following the same procedure, using up to 1.5 µg of enriched material in a single hybridization reaction (even if multiplexed in first round) of 20–24 hours.

The enriched Tiled-C libraries were assessed using the Agilent Bioanalyzer or D1000 Tapestation and quantified using KAPA Library Quantification reagents, before sequencing using the Illumina NextSeq platform. In high-quality libraries, sequencing 3–5 million reads per enriched Mb per sample is sufficient for data at 5 kb resolution.

Analysis

The most straightforward way to analyze Tiled-C data is to use the HiCPro pipeline[41] with the options for Capture-Hi-C analysis. We have also adjusted our pipelines for Capture-C analysis[12] to be compatible with Tiled-C data. This pipeline is designed to analyze deep, targeted 3C data and provides very stringent filtering, especially regarding PCR-related artefacts. All data presented in the paper have been analyzed using a combination of this CCseqBasic pipeline (https://github.com/Hughes-Genome-Group/CCseqBasicF/releases), custom scripts and ICE normalization[42]. Detailed instructions and custom scripts are available on https://github.com/oudelaar/TiledC. Interaction profiles from virtual viewpoints (Supplementary Figure 2) can be generated using the CCseqBasic pipeline with the options for Capture-C analysis.

When Tiled-C data are compared between different cell types (Figures 1b, 1c, 4; Supplementary Figures 1, 3, 7, 8, 10), we have down-sampled the data in the different samples to an equal number of valid interactions within the tiled region.

To examine the reproducibility of Tiled-C in low-input samples, we used HiCRep to calculate stratum-adjusted correlation coefficients[43], considering a maximum distance of 100,000 bp.

Hi-C

We compared Tiled-C data in mouse ES cells to the deepest currently available Hi-C data in mouse ES cells[5]. We explored the data using HiGlass[44] and downloaded and re-analyzed the Hi-C data using the HiC-Pro pipeline[41] with default options and ICE normalization[42].

ATAC-seq

Experimental procedure

For FACS-sorted erythroid progenitors from fetal liver, either 1 or 2 technical replicates were processed of ~50,000 cells each for each sorted population. ATAC-seq was performed as previously described[26].

For FACS-sorted hematopoietic stem and progenitor cells from adult bone marrow, 1 technical replicate of between 5,000 and 20,000 cells was processed for each population. Cells were spun at 500 g for 10 minutes at 4 °C. The supernatant was discarded and cells were resuspended directly in Dig-transposition buffer (25 µl 2x TD Buffer [Illumina], 2.5 µl Tn5 transposase, 0.5 µl 1%

digitonin and 22 µl H_2O) before incubating at 37 °C for 30 minutes with agitation at 600 rpm. After the transposition step, samples were processed as previously described[26].

Analysis

Reads were mapped to the mouse mm9 genome and PCR duplicates removed using NGseqBasic[45]. Technical replicates were merged and peaks called using MACS2[46]. Peaks were merged and the number of reads in each sample overlapping each peak was calculated using BEDTools merge and multicov[47]. For visualization, bedgraph files were generated using BEDTools genomecov with a scaling factor of 1e6 / (total number of reads in peaks). All analysis scripts are available at https://github.com/rbeagrie/alpha-tiledc.

Single-cell RNA-seq

Experimental procedure

Fetal livers were harvested and pooled from 7 e13.5 C57BL/6 mouse embryos and processed as above. Cells were first stained with 2.5 µg/ml biotin-conjugated anti-Ter119 (BD 553672) and 2.5 µg/ml purified anti-Ter119 (BioLegend 116241) conjugated to a DNA oligonucleotide (ADT2 GAGGCGATTGAT). After magnetic depletion, an equal number of Ter119+ cells were added back to the Ter119- fraction for further analysis, ensuring a balanced distribution of erythroid cells from all stages of differentiation. Cells were then stained with a panel of 9 further DNA-oligonucleotide conjugated antibodies (0.33 µg/ml anti-CD71: BioLegend 113802 conjugated with ADT9 [CGAAGAAGGAGT], 1 µg/ml anti-CD41: BioLegend 133919 conjugated with ADT3 [TGTCCGGCAATA], 1 µg/ml anti-CD45R: BioLegend 103249 conjugated with ADT5 [GATCGTAATACC], 1 µg/ml anti-CD3e: BioLegend 100345 conjugated with ADT7 [CATCGGTGTACA], 1 µg/ml anti CD11b: BioLegend 101249 conjugated with ADT1 [CATGATTGGCTC], 1 µg/ml anti-Ly6G/6C: BioLegend 108449 conjugated with ADT4 [TGGTGAACCTGG], 1 µg/ml anti-CD44: BioLegend 103051 conjugated with ADT8 [GTCTAGACTTCG], 1 µg/ml anti-cKit: BioLegend 105829 conjugated with ADT6 [AAGCGCTTGGCA], and 1 µg/ml non-specific IgG conjugated with ADT10 [CGGAGTAGTAAT]). Antibodies were conjugated to streptavidin as previously described[25] and mixed with biotinylated custom oligonucleotides. Cells were processed for scRNA-seq using the 10x Genomics Single Cell 3' v2 kit.

Analysis

cDNA reads were mapped to the mouse mm10 assembly and barcodes assigned to cells using Cell Ranger v2.1.1 (10x Genomics). ADT reads were mapped to cell/antibody barcodes using CITE-seq-count (https://github.com/Hoohm/CITE-seq-Count). Potential doublet cells were removed using Scrublet[48]. Further analysis was performed using Seurat v2. Low quality cells with less than 300 or more than 5,000 identified genes, or with more than 9% mitochondrial reads were also removed. Clusters were identified using the "FindClusters" function and UMAP projection was generated using the "RunUMAP" function, both with the first 16 principle components. Seurat clusters were annotated using marker genes and by reference to previously published data[19] – Seurat identified two clusters corresponding to committed erythroid progenitors (CEP) and four clusters corresponding to cells undergoing erythroid terminal differentiation (ETD). Average gene expression for populations matching those obtained by FACS sorting was generated by using the "SubsetData" function to select cells with low levels of cell-surface barcodes corresponding to lineage markers, and appropriate levels of barcodes corresponding to CD71 and Ter119. All analysis scripts are available at https://github.com/rbeagrie/alpha-tiledc.

RNA-FISH

Experimental procedure

Standard RNA-FISH was carried out as previously described[49]. Sorted cells from mouse fetal liver were placed back into culture for 6 hours to allow nascent transcription to be re-established. Samples were hybridized with digoxygenin-labelled oligonucleotide probes directed to α-globin introns (30 ng per slide) and visualized using FITC-conjugated antibodies (primary: sheep anti-DIG FITC [Roche] 1:50, secondary: rabbit anti-sheep FITC [Vector] 1:100). Two negative controls were also included: brain tissue from a male, adult CD1 mouse and Ter119+ (i.e. mature) fetal liver erythroid cells that were probed with secondary antibody but no primary antibody. Magnetically purified but not FACS-purified Ter119+ fetal liver erythroid cells were used as a positive control.

Imaging equipment and settings

Widefield fluorescence imaging was performed at 20 °C on a DeltaVision Elite system (Applied Precision) equipped with a 100 ×/1.40 NA UPLSAPO oil immersion objective (Olympus), a CoolSnap HQ2 CCD camera (Photometrics), DAPI (excitation 390/18; emission 435/40) and

FITC (excitation 475/28; emission 525/45) filters. 12-bit image stacks were acquired with a z-step of 150 nm giving a voxel size of 64.5 nm x 64.5 nm x 150 nm.

Image analysis

Image analysis was blinded by renaming image files from all experiments with random character strings and processing them together. Images were manually examined, and each cell was scored for the presence of active nascent-transcription foci. Analysis scripts are available at https://github.com/rbeagrie/alpha-tiledc.

Data availability

Tiled-C, ATAC-seq and scRNA-seq data generated in this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSEXXXX. All RNA-FISH image files are archived in Dryad. A UCSC hub for visualizing ATAC-seq and scRNA-seq mean expression data is available at http://sara.molbiol.ox.ac.uk/public/hugheslab/alpha-tiledc/hub.txt.

Acknowledgements

We thank J. Sloane-Stanley and J. Sharpe of the WIMM Transgenics Facility for mouse breeding and fetal liver provision, K. Clark, S-A. Clark and C. Waugh of the WIMM Flow Cytometry Facility for cell sorting services, N. Ashley of the WIMM Single Cell facility for antibody conjugation and 10X library preparation services, J. Brown for help with RNA-FISH, D. Downes for help with ATAC-seq and C. Garnett for advice on low-input ATAC-seq. We wish to acknowledge the Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine and University of Oxford for use of their services in this project. This work was supported by the Medical Research Council (MC_UU_00016 to D.R.H., J.R.H., V.J.B. and MR/N00969X/1 to J.R.H., T.B., V.J.B.), Wellcome Trust (209181/Z/17/Z to R.A.B.), National Institutes of Health (NIH NIDDK R01DK100915 to M.S.) and a Stevenson Junior Research Fellowship at University College, Oxford (to A.M.O.).

References

1. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361,** 1341–1345 (2018).
2. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nature Reviews Genetics* **20,** 437–455 (2019).
3. Stadhouders, R., Filion, G. J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569,** 345–354 (2019).
4. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295,** 1306–1311 (2002).
5. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171,** 557–572.e24 (2017).
6. Phillips-Cremins, J. E. *et al.* Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* **153,** 1281–1295 (2013).
7. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **513,** 89–100 (2014).
8. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503,** 290–294 (2013).
9. Andrey, G. & Mundlos, S. The three-dimensional genome: regulating gene expression during pluripotency and development. *Development* **144,** 3646–3658 (2017).
10. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502,** 499–506 (2013).
11. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **34,** 903–16 (2019).
12. Davies, J. O. J. *et al.* Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Meth* **86,** 1202–10 (2015).
13. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326,** 289–293 (2009).
14. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38,** 1348–1354 (2006).
15. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* **38,** 1341–1347 (2006).

16. Hughes, J. R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* **46,** 205–212 (2014).
17. Oudelaar, A. M., Davies, J. O. J., Downes, D. J., Higgs, D. R. & Hughes, J. R. Robust detection of chromosomal interactions from small numbers of cells using low-input Capture-C. *Nucleic Acids Research* (2017). doi:10.1093/nar/gkx1194
18. Pop, R. *et al.* A key commitment step in erythropoiesis is synchronized with the cell cycle clock through mutual inhibition between PU.1 and S-phase progression. *PLoS Biol.* **8,** e1000484 (2010).
19. Tusi, B. K. *et al.* Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature* **555,** 54–60 (2018).
20. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).
21. Hay, D. *et al.* Genetic dissection of the α-globin super-enhancer in vivo. *Nat Genet* **48,** 895–903 (2016).
22. Hanssen, L. L. P. *et al.* Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nature Cell Biology* **19,** 952–961 (2017).
23. Oudelaar, A. M. *et al.* Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat Genet* **50,** 1744–1751 (2018).
24. Brown, J. M. *et al.* A tissue-specific self-interacting chromatin domain forms independently of enhancer-promoter interactions. *Nature Communications* **9,** 376 (2018).
25. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Meth* **14,** 865–868 (2017).
26. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Meth* **10,** 1213–1218 (2013).
27. Hu, W., Yuan, B. & Lodish, H. F. Cpeb4-mediated translational regulatory circuitry controls terminal erythroid differentiation. *Dev. Cell* **30,** 660–672 (2014).
28. Anguita, E. *et al.* Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *EMBO J.* **23,** 2841–2852 (2004).
29. Palstra, R.-J. *et al.* The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* **35,** 190–194 (2003).

30. Vernimmen, D., De Gobbi, M., Sloane-Stanley, J. A., Wood, W. G. & Higgs, D. R. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J.* **26,** 2041–2051 (2007).
31. Allahyar, A. *et al.* Enhancer hubs and loop collisions identified from single-allele topologies. *Nat Genet* **295,** 1306 (2018).
32. Chen, H. *et al.* Dynamic interplay between enhancer-promoter topology and gene activity. *Nat Genet* **50,** 1296–1303 (2018).
33. Montavon, T. *et al.* A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147,** 1132–1145 (2011).
34. Williamson, I., Lettice, L. A., Hill, R. E. & Bickmore, W. A. Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* **143,** 2994–3001 (2016).
35. Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. A. Emerging Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harbor Symposia on Quantitative Biology* 034710 (2018). doi:10.1101/sqb.2017.82.034710
36. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467,** 430–435 (2010).
37. Schmidt, D. *et al.* A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Research* **20,** 578–588 (2010).
38. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Research* **16,** 1299–1309 (2006).
39. Jäger, R. *et al.* Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nature Communications* **6,** 6178 (2015).
40. Oudelaar, A. M., Downes, D., Davies, J. & Hughes, J. Low-input Capture-C: A Chromosome Conformation Capture Assay to Analyze Chromatin Architecture in Small Numbers of Cells. *Bio-protocol* **7,** (2017).
41. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16,** 259 (2015).
42. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Meth* **9,** 999–1003 (2012).
43. Yang, T. *et al.* HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Research* **27,** 1939–1949 (2017).

44. Kerpedjiev, P. *et al.* HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19,** 125–12 (2018).

45. Telenius, J. & Hughes, J. R. NGseqBasic - a single-command UNIX tool for ATAC-seq, DNaseI-seq, Cut-and-Run, and ChIP-seq data mapping, high-resolution visualisation, and quality control. *bioRxiv* (2018). doi:10.1101/393413

46. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).

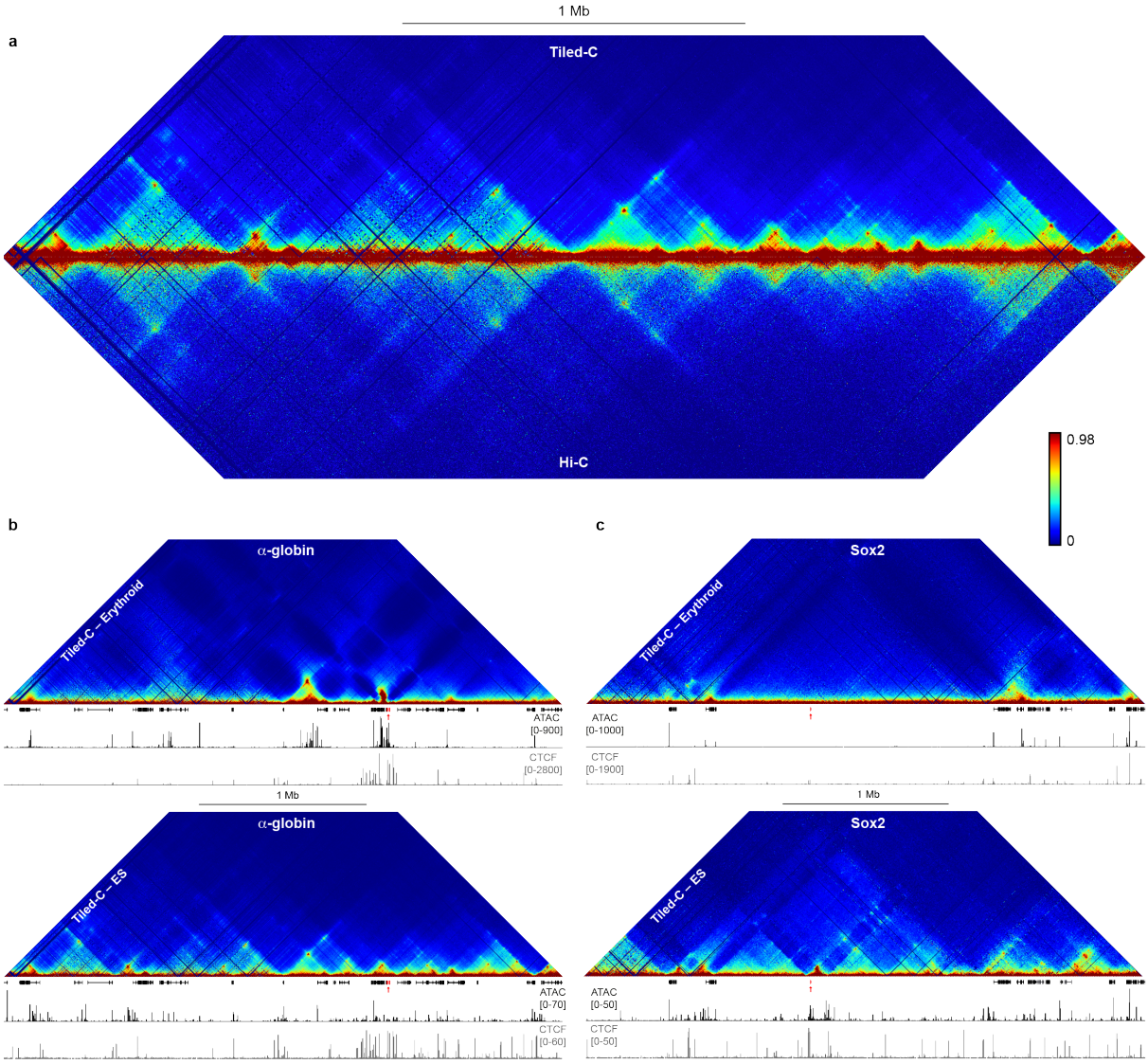
47. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

48. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8,** 281–291.e9 (2019).

49. Brown, J. M. & Buckle, V. J. Detection of nascent RNA transcripts by fluorescence in situ hybridization. *Methods Mol Biol* **659,** 33–50 (2010).

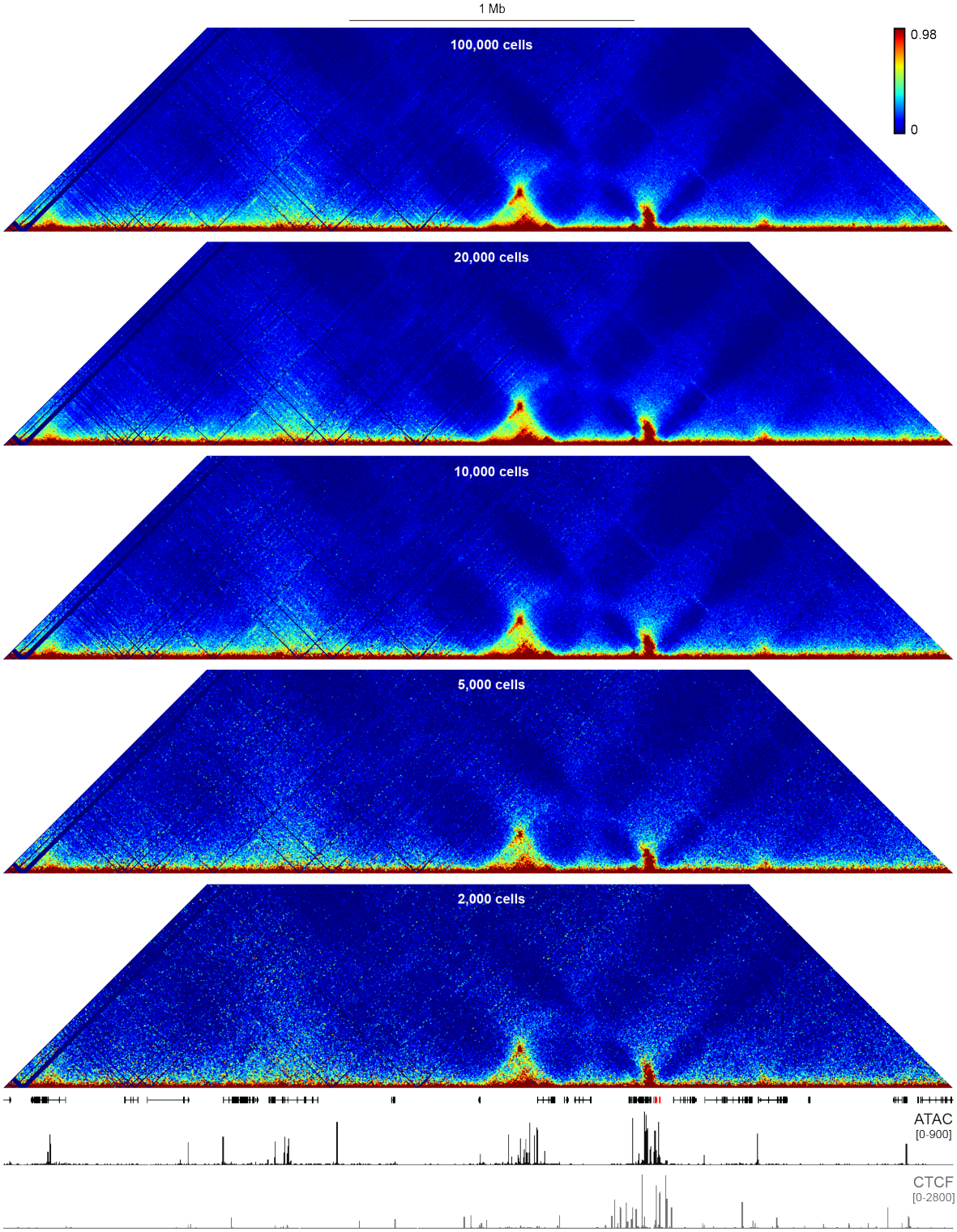
50. Pietras, E. M. *et al.* Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell* **17,** 35–46 (2015).

Figure 1: Tiled-C generates deep all vs all 3C data at regions of interest.



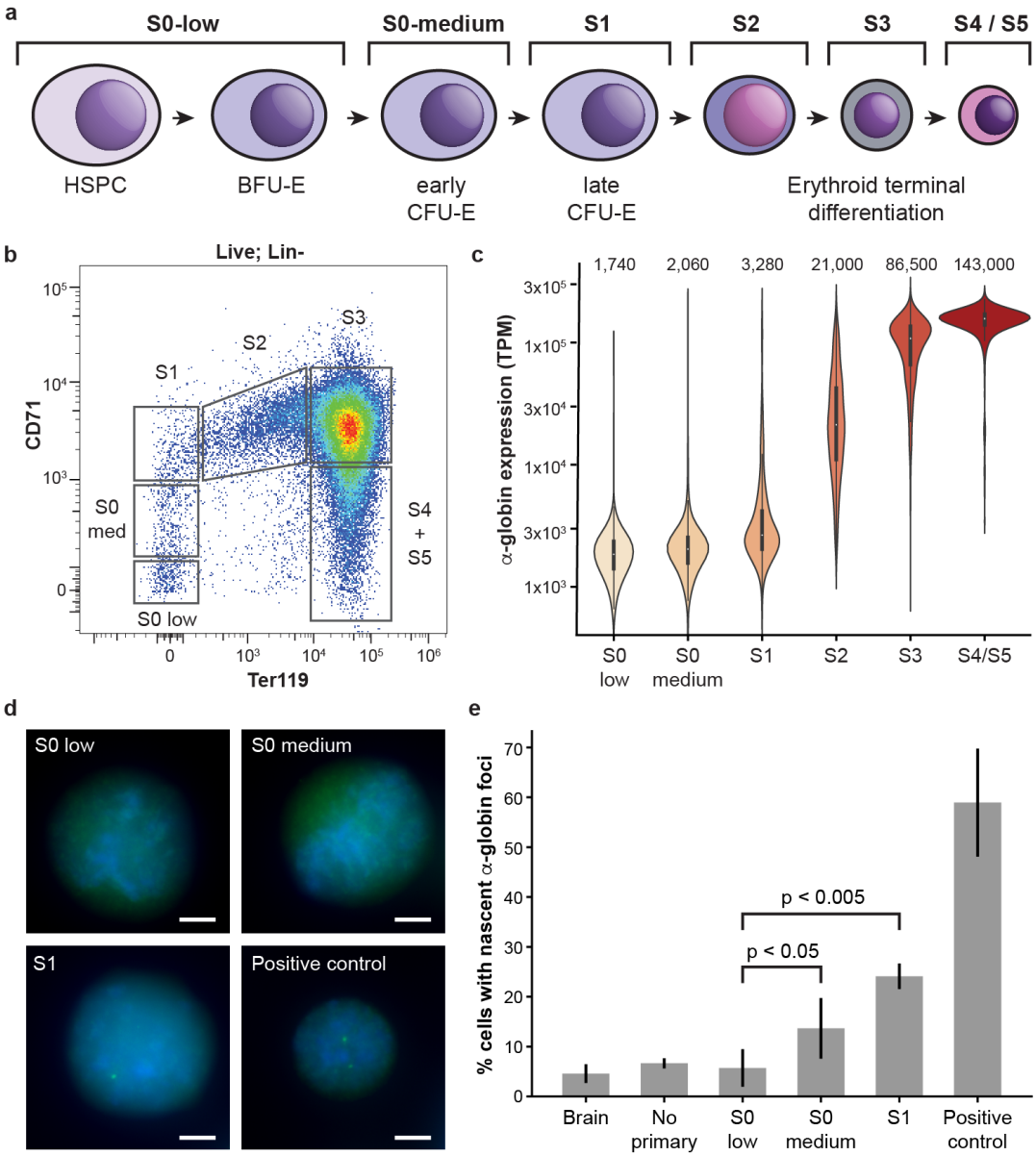
(a) Comparison of Tiled-C and Hi-C contact matrices at 2 kb resolution in mouse ES cells. Contact frequencies represent normalized, unique interactions in 3 and 4 replicates for Tiled-C and Hi-C data, respectively. Coordinates (mm9): chr11:29,902,000-33,228,000. **(b)** Tiled-C contact matrices of ~3.3 Mb spanning the α-globin locus in primary mouse erythroid cells (top) and ES cells (bottom) at 2 kb resolution. Contact frequencies represent normalized, unique interactions in 3 replicates. Gene annotation (α-globin genes highlighted in red), open chromatin (ATAC) and CTCF occupancy are shown below the matrices. Coordinates (mm9): chr11:29,902,000-33,228,000. **(c)** Tiled-C contact matrices of ~3.4 Mb spanning the Sox2 locus in primary mouse erythroid cells (top) and ES cells (bottom) at 5 kb resolution. Contact frequencies represent normalized, unique interactions in 4 replicates. Gene annotation (Sox2 gene highlighted in red), open chromatin (ATAC) and CTCF occupancy are shown below the matrices. Coordinates (mm9): chr3:33,200,000-36,565,000.

Figure 2: Tiled-C generates high-resolution contact matrices from small numbers of cells.



Tiled-C contact matrices of ~3.3 Mb spanning the α-globin locus generated from small aliquots of primary mouse erythroid cells at 5 kb resolution. Contact frequencies represent normalized, unique interactions in 3 replicates. Gene annotation (α-globin genes highlighted in red), open chromatin (ATAC) and CTCF occupancy are shown below the matrices. Coordinates (mm9): chr11:29,900,000-33,230,000.

Figure 3: Expression of α-globin is gradually upregulated during *in vivo* erythroid differentiation.



(a) Scheme of erythroid differentiation showing the various populations analyzed. **(b)** Example FACS plot showing the gating strategy used to isolate erythroid progenitors from mouse fetal liver. **(c)** Expression (in tags per million) of α-globin transcripts in each population as determined by single-cell RNA-seq. Mean expression for each population is given at the top of each bar. **(d)** Representative RNA-FISH images showing detection of nascent α-globin transcripts in sorted early erythroid progenitors. Scale bar is 3 μM for each image. **(e)** RNA-FISH quantification, showing the mean ± s.d. of 3 independent experiments (except for brain and "no primary" negative controls, which have n=2). P values were calculated by two-tailed paired T-tests.

Figure 4: Upregulation of α-globin expression correlates with increased chromatin accessibility and enhancer-promoter interactions.

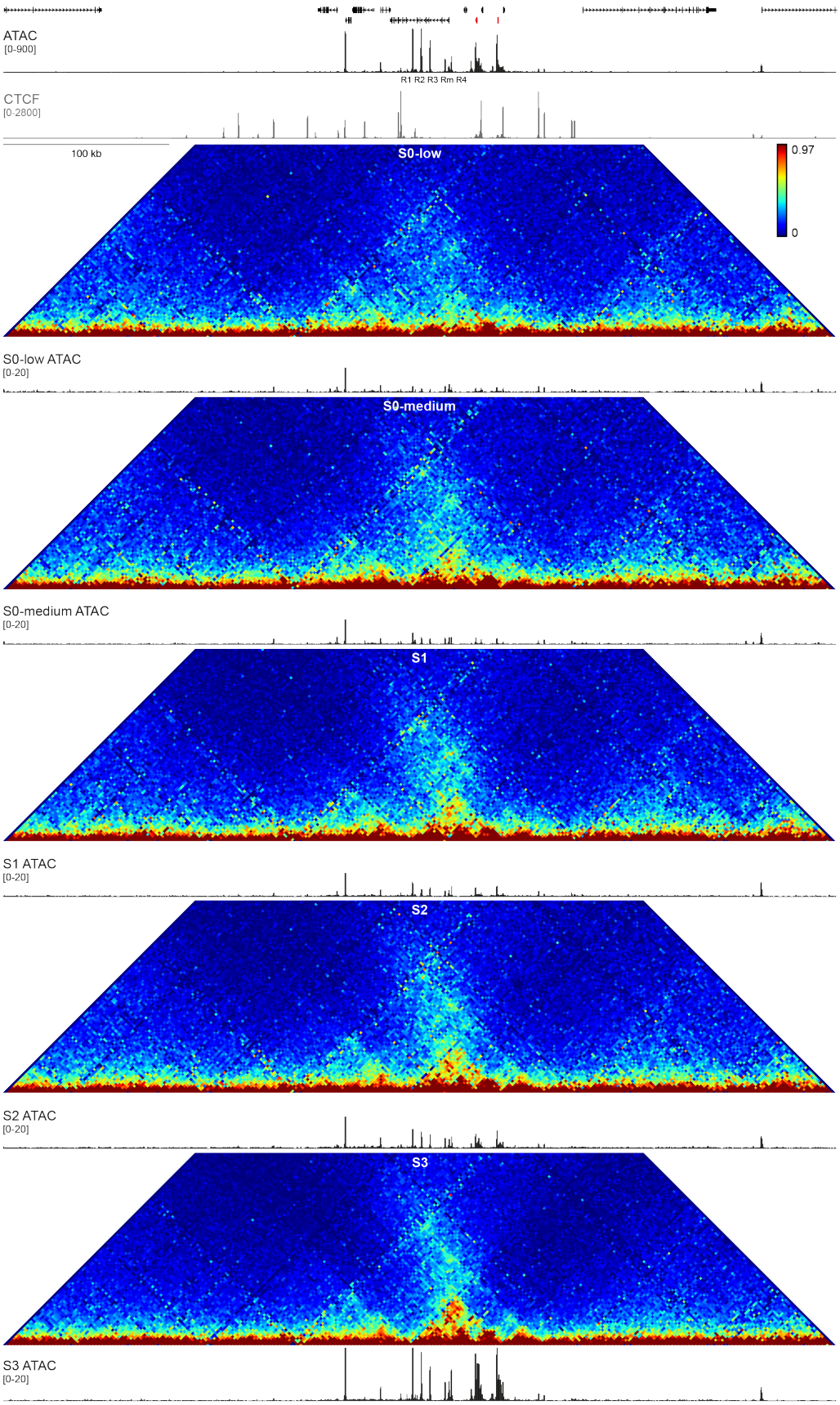
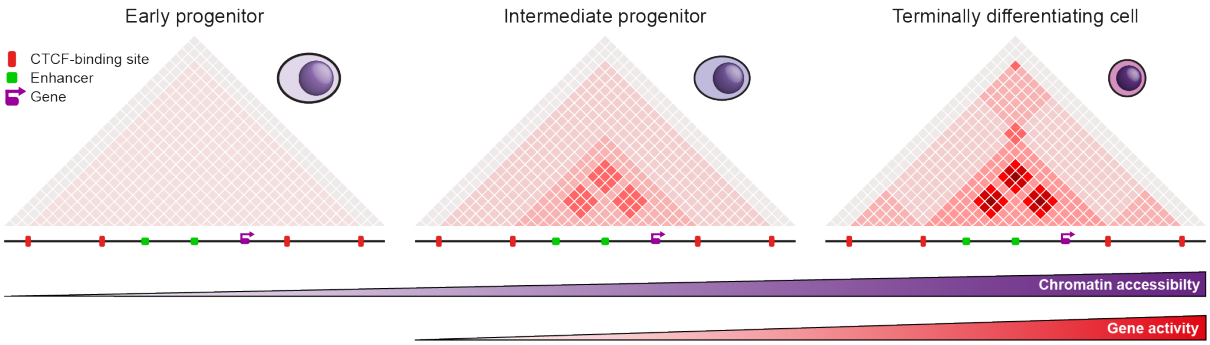


Figure 4: Upregulation of α-globin expression correlates with increased chromatin accessibility and enhancer-promoter interactions. [continued]

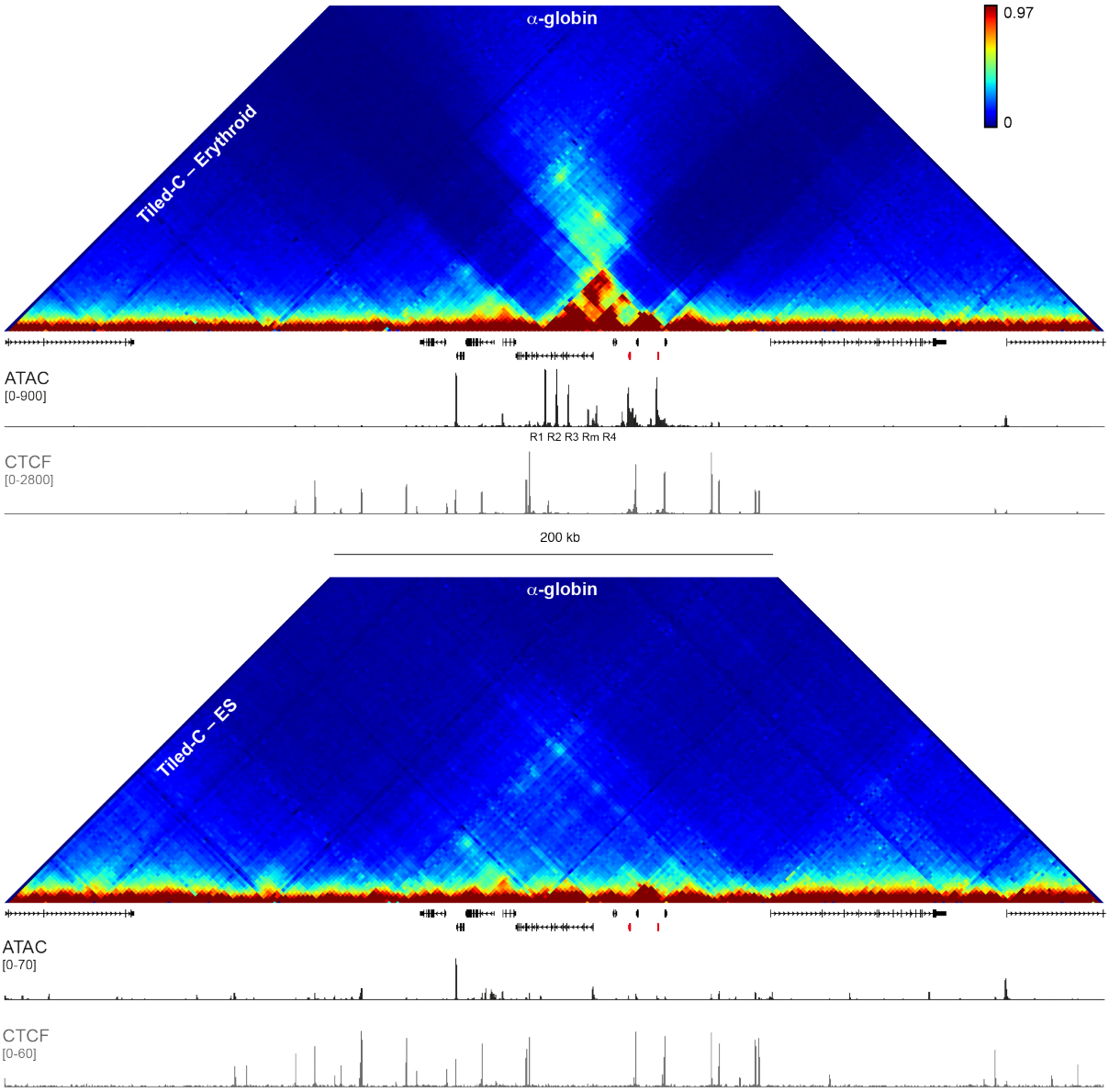
Tiled-C contact matrices of 500 kb spanning the mouse α-globin locus in sequential stages of *in vivo* erythroid differentiation at 2 kb resolution. Contact frequencies represent normalized, unique interactions in 2 replicates. Matched open chromatin (ATAC) profiles are shown underneath each matrix and represent normalized data from 3 S0-low, S0-medium and S1 replicates and 2 S2 and S3 replicates. Gene annotation (α-globin genes highlighted in red), open chromatin (ATAC) and CTCF occupancy in mature mouse erythroblast cells are shown at the top. Coordinates (mm9): chr11:31,900,000-32,400,000.

Figure 5: Graphical summary.



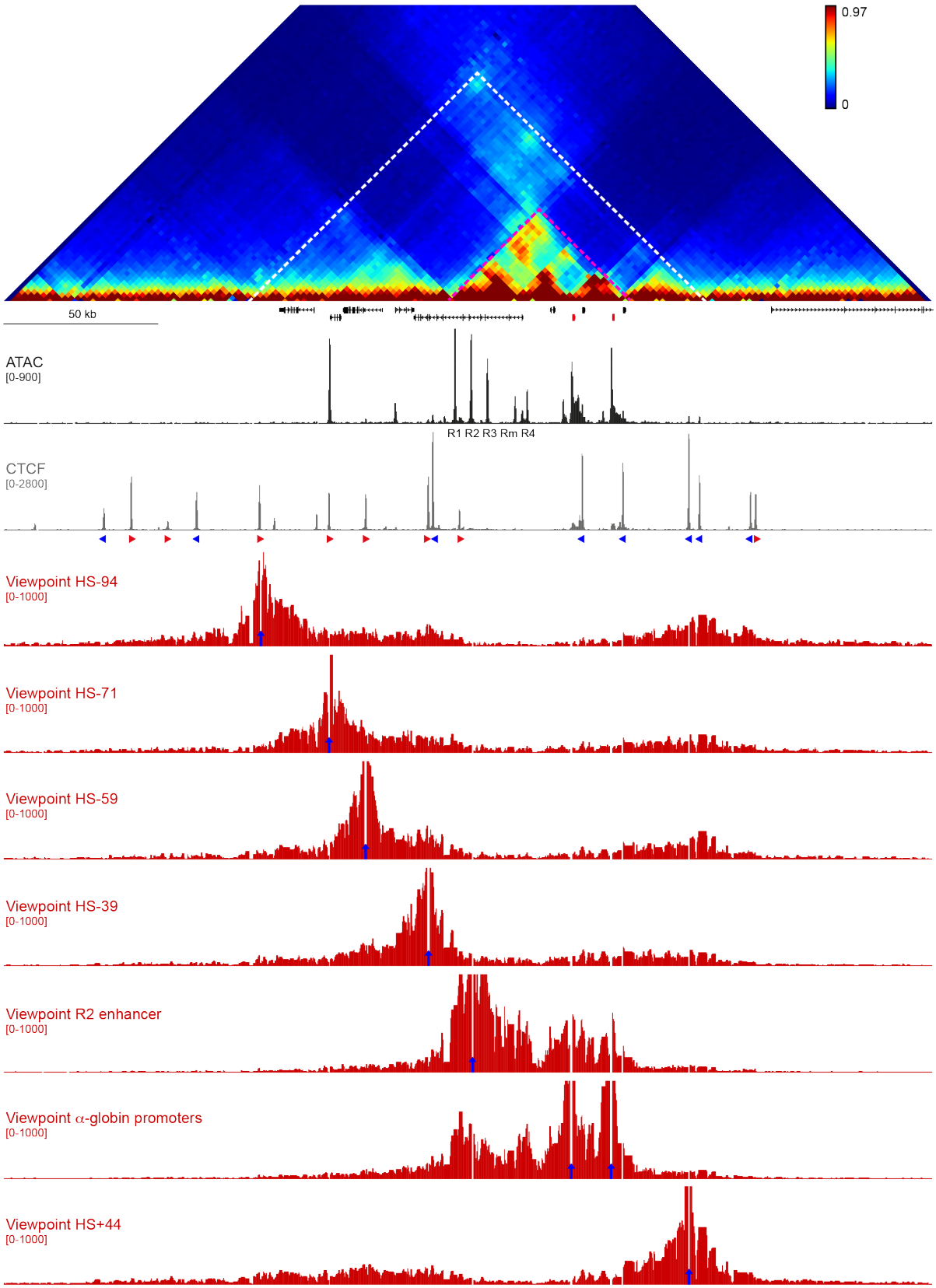
Our data support a model in which TADs are established early in differentiation, prior to activation of the domain. During differentiation, accessible regulatory elements are formed. This is followed by sub-compartmentalization of the TAD into smaller domains, in which enhancers and promoters form specific interactions. Through differentiation, accessibility and interactions between enhancers and promoters are gradually increased, concomitant with upregulation of gene expression.

Supplementary Figure 1: Tiled-C contact matrices of the α-globin locus.



Tiled-C contact matrices of 500 kb spanning the α-globin locus in primary mouse erythroid cells (top) and ES cells (bottom) at 2 kb resolution. Contact frequencies represent normalized, unique interactions in 3 replicates. Gene annotation (α-globin genes highlighted in red), open chromatin (ATAC) and CTCF occupancy are shown below the matrices. Coordinates (mm9): chr11:31,900,000-32,400,000.

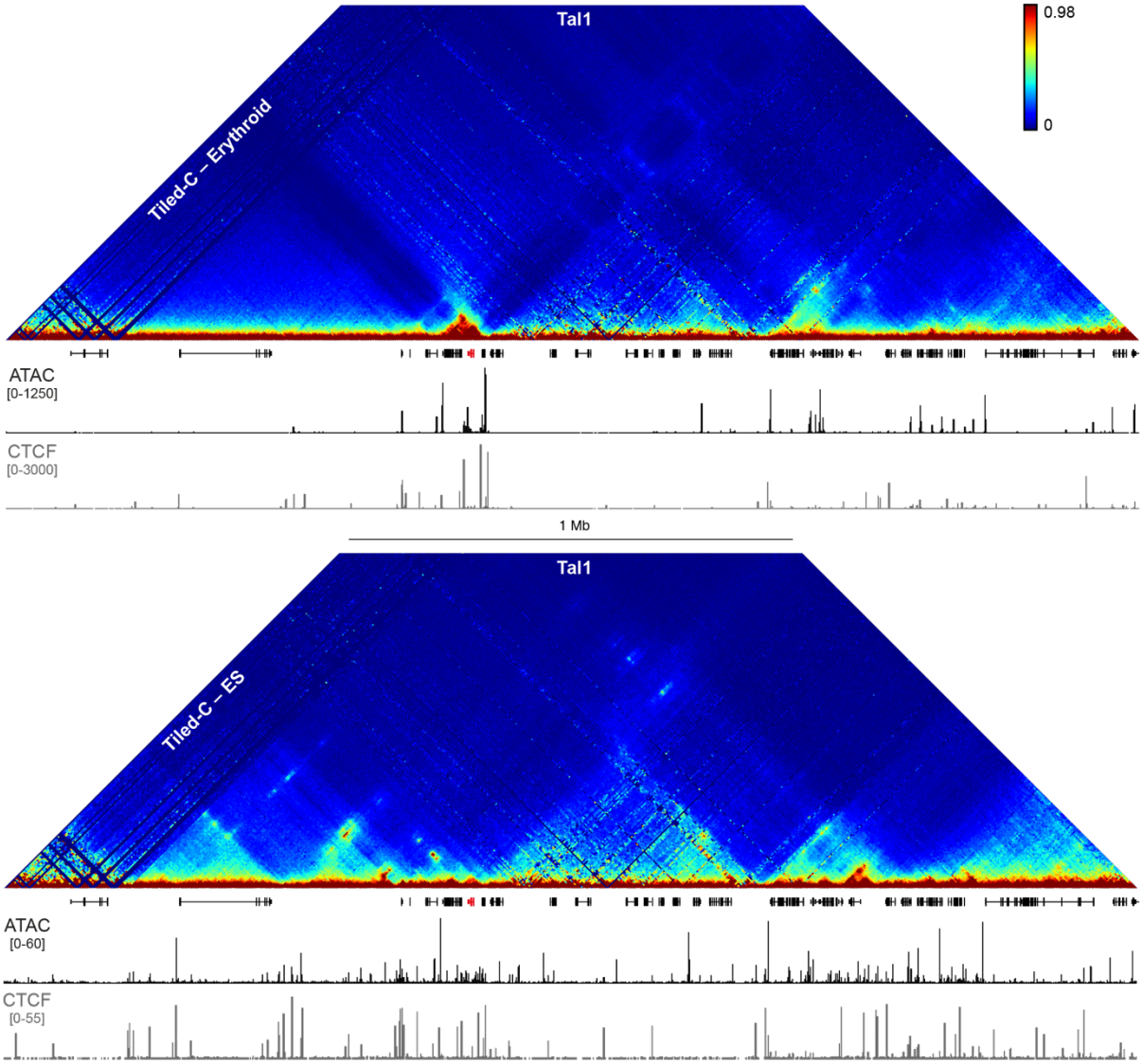
Supplementary Figure 2: Tiled-C contact matrix and virtual viewpoints of the α-globin locus.



Supplementary Figure 2: Tiled-C contact matrix and virtual viewpoints of the α-globin locus. [continued]

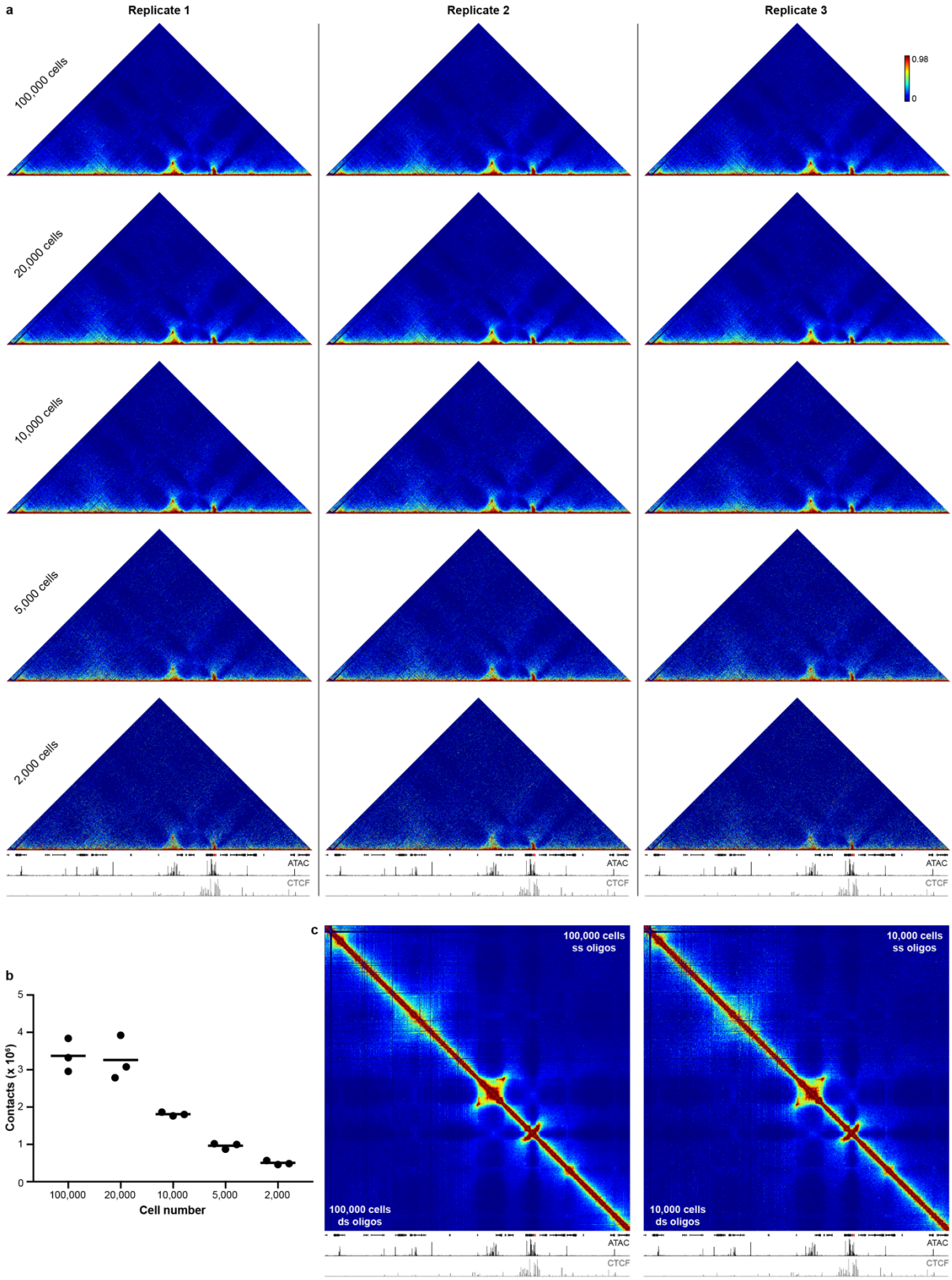
Tiled-C contact matrix of 300 kb spanning the α-globin locus in primary mouse erythroid cells at 2 kb resolution. Gene annotation (α-globin genes highlighted in red), open chromatin (ATAC), CTCF occupancy and interaction profiles from virtual viewpoints (highlighted by blue arrows) across the α-globin locus are shown underneath. Contact frequencies represent normalized, unique interactions in 3 replicates. The pre-existing TAD and sub-compartmentalized erythroid-specific domain are highlighted in white and magenta dashed lines, respectively. Coordinates (mm9): chr11:32,000,000-32,300,000.

Supplementary Figure 3: Tiled-C contact matrices of the Tal1 locus.



Tiled-C contact matrices of ~2.6 Mb spanning the Tal1 locus in primary mouse erythroid cells (top) and ES cells (bottom) at 5 kb resolution. Contact frequencies represent normalized, unique interactions in 4 replicates. Gene annotation (Tal1 gene highlighted in red), open chromatin (ATAC) and CTCF occupancy are shown below the matrices. Coordinates (mm9): chr4:113,690,000-116,240,000.

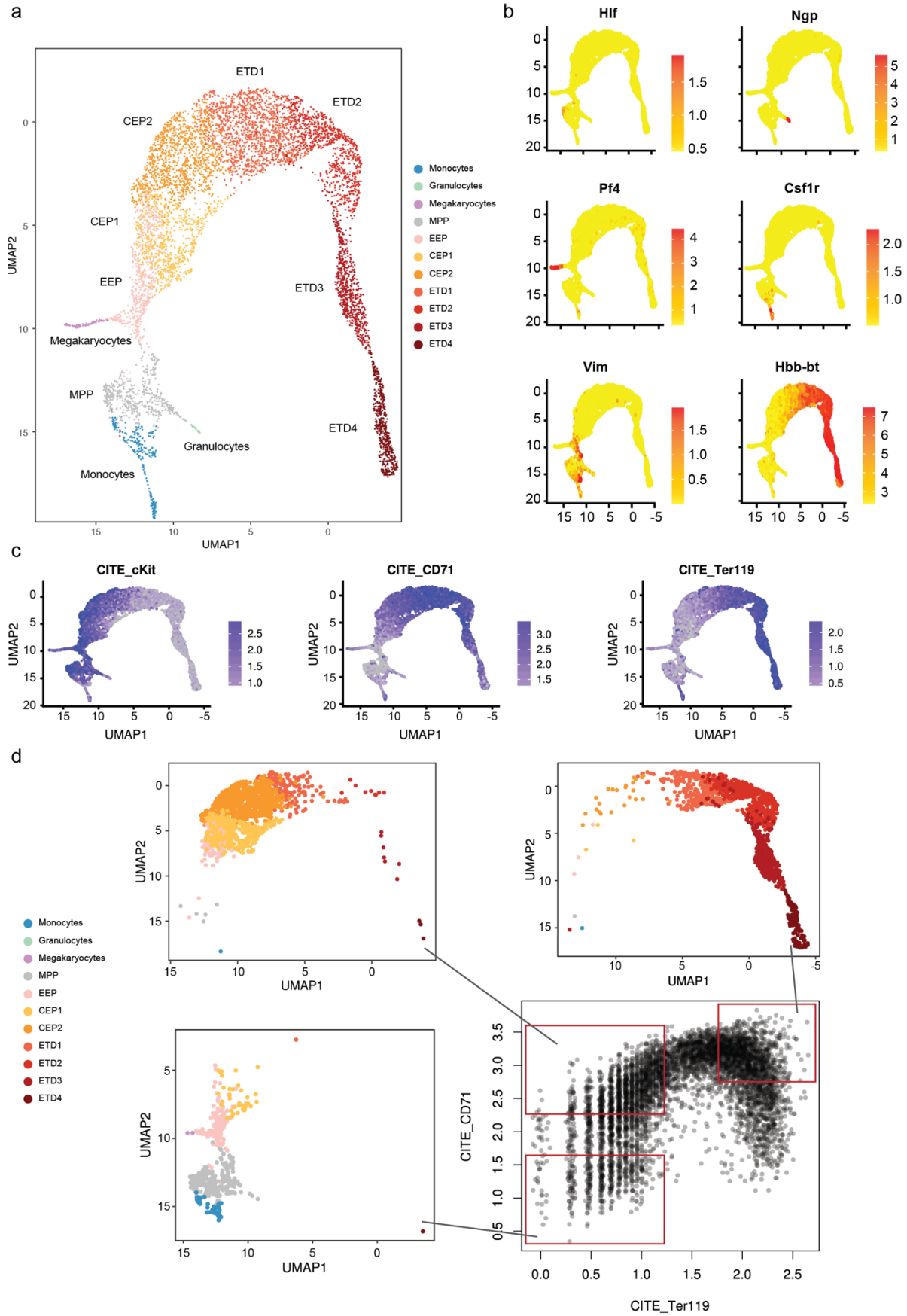
Supplementary Figure 4: Reproducibility of Tiled-C data derived from low-input samples.



Supplementary Figure 4: Reproducibility of Tiled-C data derived from low-input samples. [continued]

(a) Tiled-C contact matrices of ~3.3 Mb spanning the α-globin locus in individual biological replicates of small aliquots of primary mouse erythroid cells at 5 kb resolution. Contact frequencies represent normalized, unique interactions. Gene annotation (α-globin genes highlighted in red), open chromatin (ATAC) and CTCF occupancy are shown below the matrices. Coordinates (mm9): chr11:29,900,000-33,230,000. Stratum-adjusted correlation coefficients between individual replicates are 0.979-0.983 for 100,000 cells, 0.987-0.990 for 20,000 cells, 0.972-0.980 for 10,000 cells, 0.967-0.971 for 5,000 cells, and 0.953-0.962 for 2,000 cells. **(b)** Overview of the number of unique contacts in the α-globin locus (chr11:29,900,000-33,230,000) identified in individual replicates of small aliquots of primary erythroid cells. At this sequencing depth (Supplementary Table 2), library complexity is limiting when fewer than 20,000 cells are used. **(c)** Tiled-C contact matrices of 3.3 Mb spanning the α-globin locus at 5 kb resolution generated using single-stranded (ss; top) and double-stranded (ds; bottom) oligonucleotides in aliquots of 100,000 (left) and 10,000 (right) primary erythroid cells. Contact frequencies represent normalized, unique interactions in 3 replicates. Gene annotation (α-globin genes highlighted in red), open chromatin (ATAC) and CTCF occupancy are shown below the matrices. Coordinates (mm9): chr11:29,900,000-33,230,000.

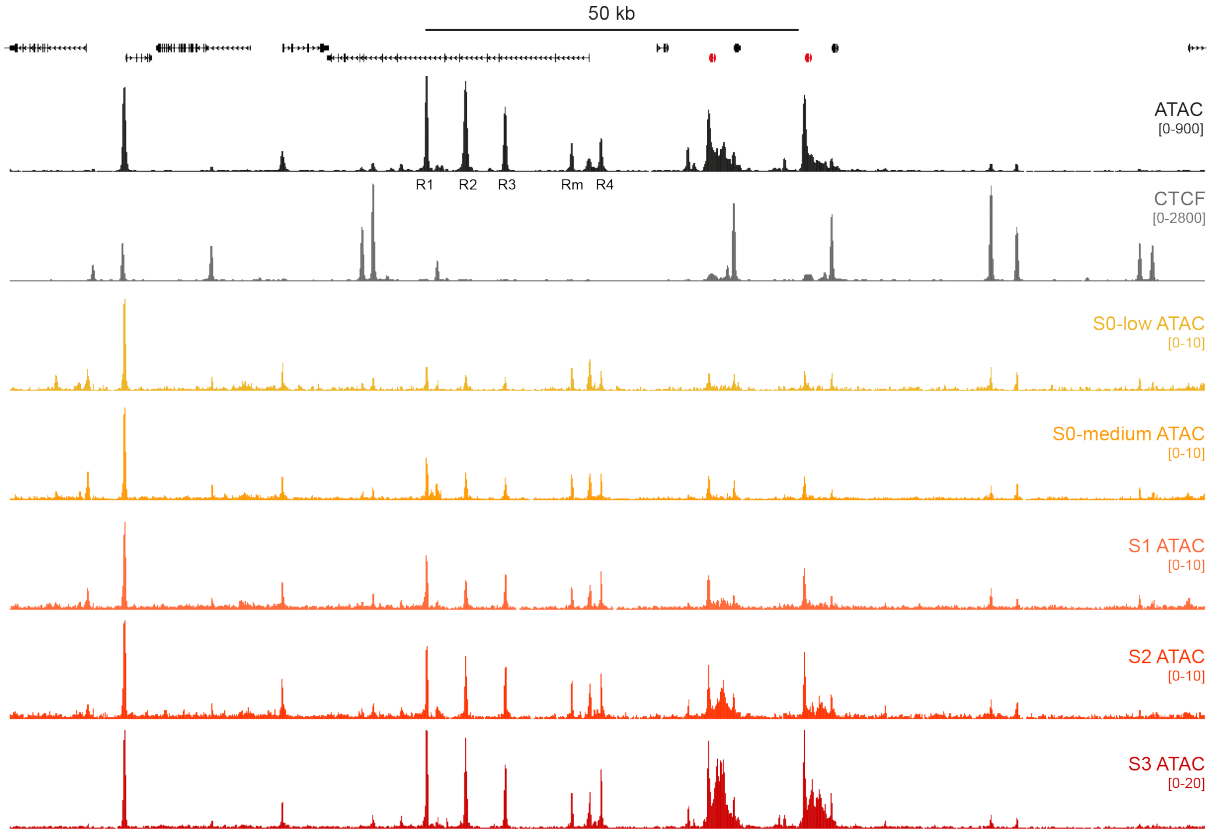
Supplementary Figure 5: Single-cell RNA-seq.



Supplementary Figure 5: Single-cell RNA-seq. [continued]

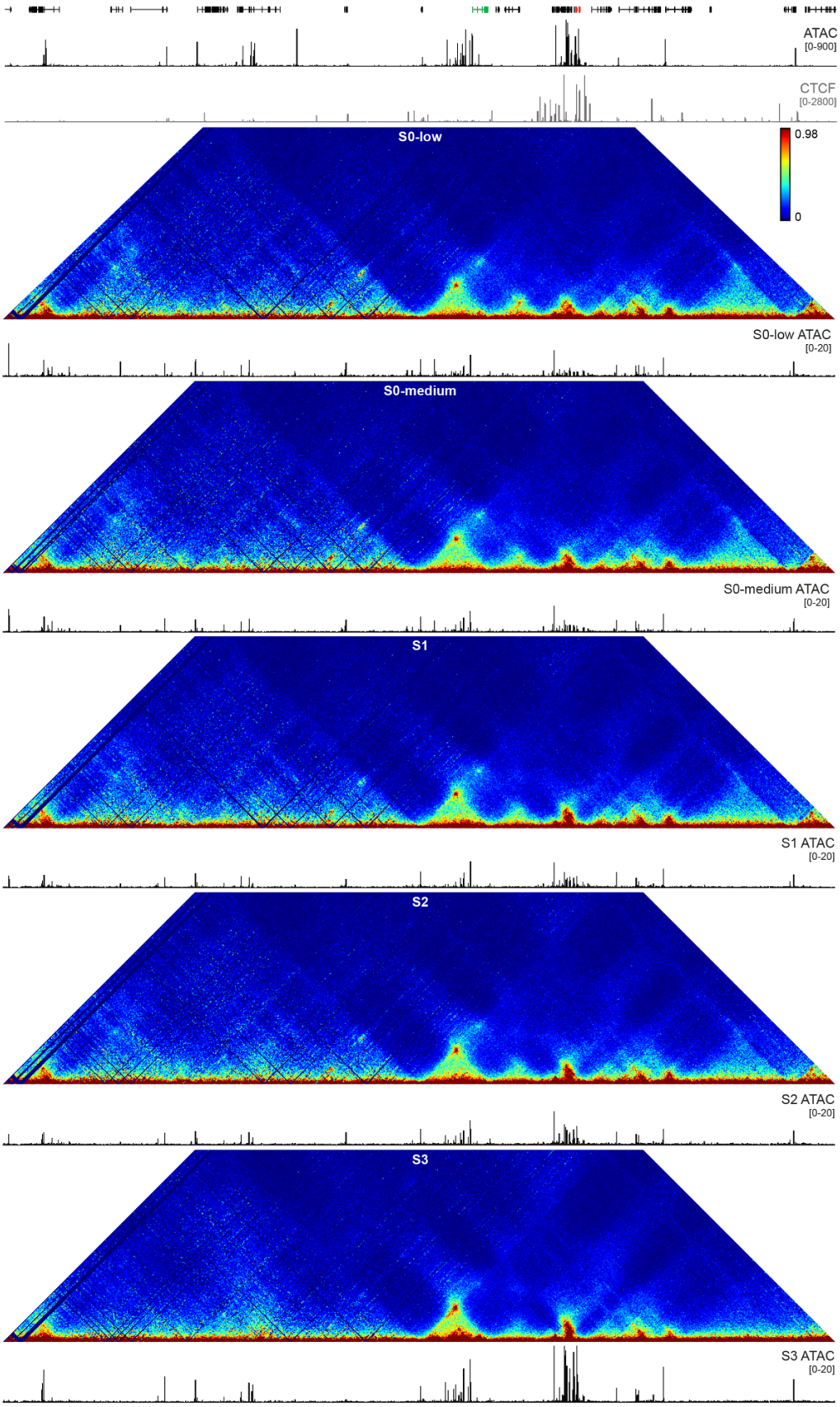
(a) UMAP projection of scRNA-seq data showing cell clusters detected by Seurat. Based on marker gene expression, three smaller trajectories were assigned as Monocytes, Granulocytes and Megakaryocytes, whilst cells in the center of these trajectories were identified as multi-potent progenitors (MPP). Based on previously published scRNA-seq data (Tusi et al. Nature 2018), the erythroid clusters were identified as early erythroid progenitors (EEP), committed erythroid progenitors (CEP1-2) or cells undergoing erythroid terminal differentiation (ETD1-4). **(b)** Example marker genes used to annotate scRNA-seq clusters, displayed with Seurat normalized expression values. **(c)** Cells were also labelled with barcoded antibodies against the same surface markers used for FACS purification. Plots show three example surface markers highlighting early populations (cKit), mid-differentiation (CD71) and late-differentiation (Ter119) erythroid cells. **(d)** *In silico* gating of the scRNA-seq data was used to match cell populations from scRNA-seq to bulk datasets obtained through FACS sorting. *Bottom right:* Gating of Lin- cells based on barcode counts for CD71 and Ter119. *Bottom left, top left, top right:* For three example gates (S0-low, S1 and S3, respectively) the cluster composition of the included cells is shown. S0-low contains predominantly MMP and EEP, S1 contains mostly CEP and S3 contains ETD cells.

Supplementary Figure 6: Chromatin accessibility in the α-globin locus through erythroid differentiation.



Chromatin accessibility (ATAC) in a region of 160 kb spanning the mouse α-globin locus in sequential stages of *in vivo* erythroid differentiation. ATAC profiles represent normalized data from 3 S0-low, S0-medium and S1 replicates and 2 S2 and S3 replicates. The profiles are shown at different scales to highlight changes in accessibility in early stages of differentiation. Gene annotation (α-globin genes highlighted in red), open chromatin (ATAC) and CTCF occupancy in mature mouse erythroblast cells are shown at the top. Coordinates (mm9): chr11:32,090,000-32,250,000.

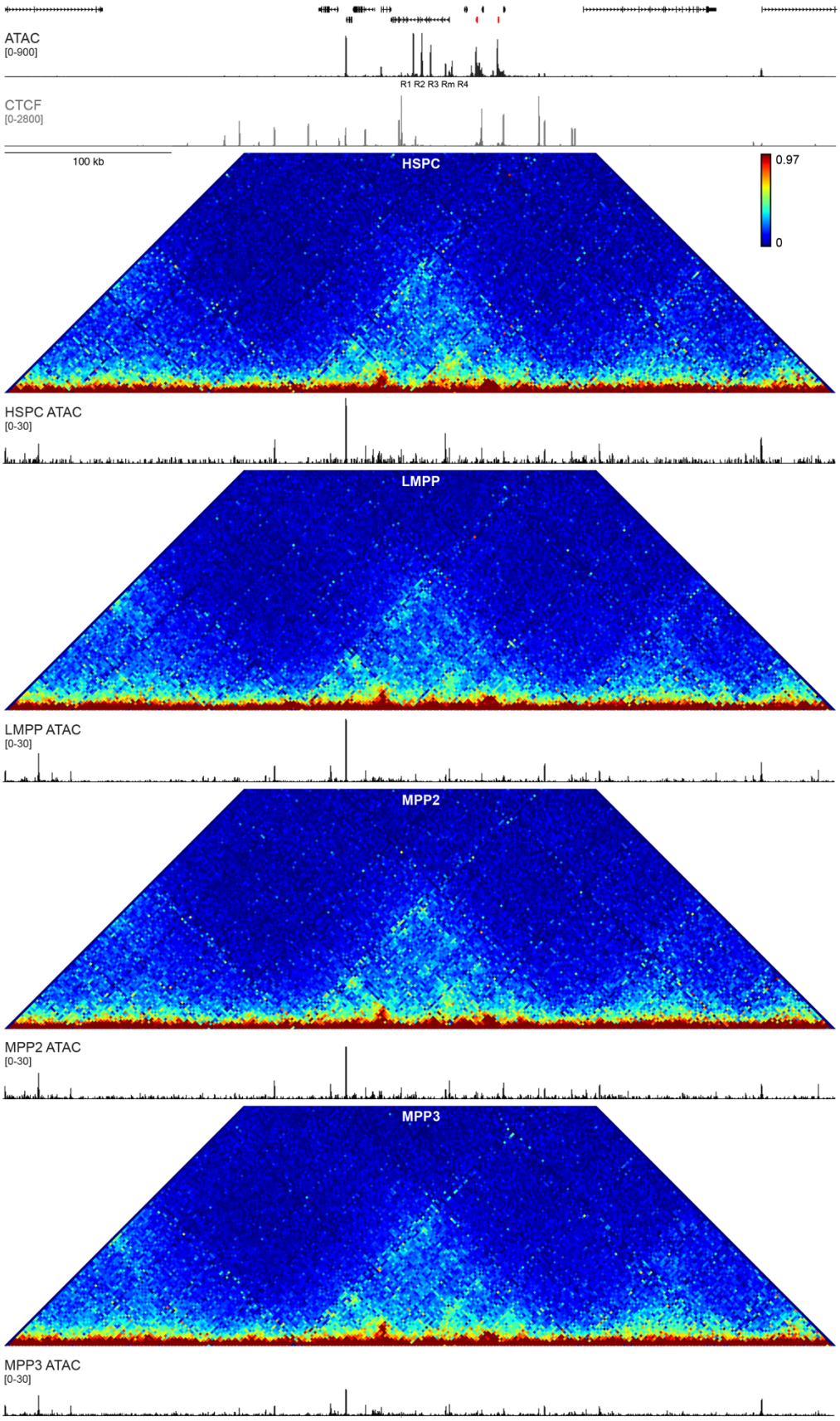
Supplementary Figure 7: Chromatin interactions and accessibility in the extended α-globin locus through erythroid differentiation.



Supplementary Figure 7: Chromatin interactions and accessibility in the extended α-globin locus through erythroid differentiation. [continued]

Tiled-C contact matrices of ~3.3 Mb spanning the mouse α-globin locus in sequential stages of *in vivo* erythroid differentiation at 5 kb resolution. Contact frequencies represent normalized, unique interactions in 2 replicates. Matched open chromatin (ATAC) profiles are shown underneath each matrix and represent normalized data from 3 S0-low, S0-medium and S1 replicates and 2 S2 and S3 replicates. Gene annotation (α-globin and Cpeb4 genes highlighted in red and green, respectively), open chromatin (ATAC) and CTCF occupancy in mature mouse erythroblast cells are shown at the top. Coordinates (mm9): chr11:29,900,000-32,230,000.

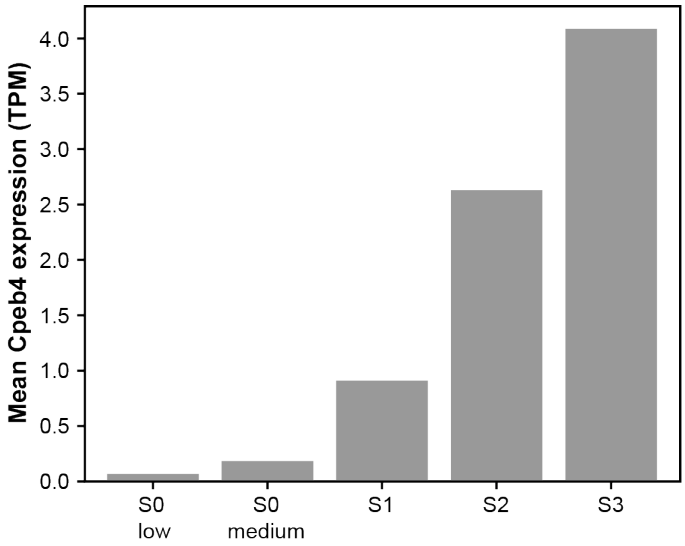
Supplementary Figure 8: Chromatin interactions and accessibility in the α-globin locus in early hematopoietic progenitor populations.



Supplementary Figure 8: Chromatin interactions and accessibility in the α-globin locus in early hematopoietic progenitor populations. [continued]

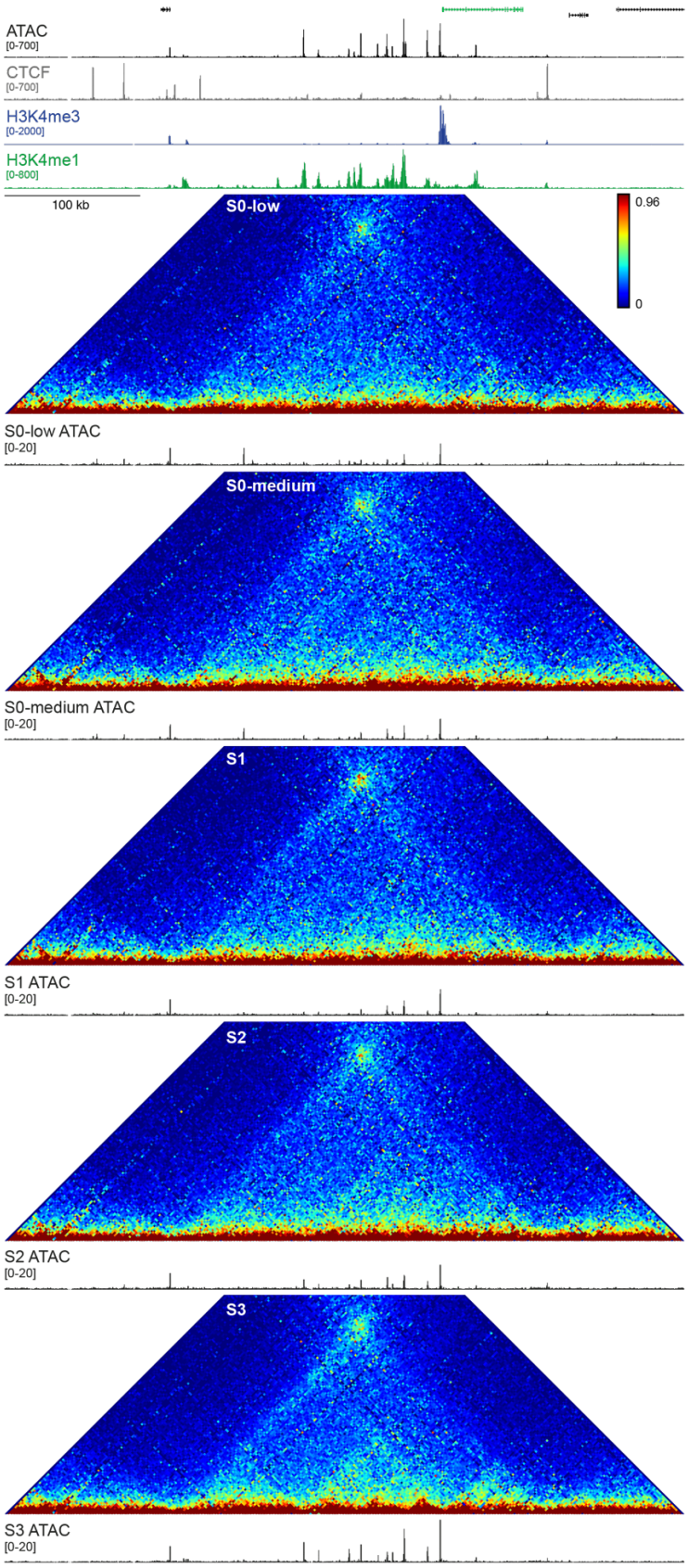
Tiled-C contact matrices of 500 kb spanning the mouse α-globin locus in early hematopoietic progenitor populations at 2 kb resolution. Contact frequencies represent normalized, unique interactions from 2 replicates. Matched open chromatin (ATAC) profiles are shown underneath each matrix. Gene annotation (α-globin genes highlighted in red), open chromatin (ATAC) and CTCF occupancy in mature mouse erythroblast cells are shown at the top. Coordinates (mm9): chr11:31,900,000-32,400,000. HSPC = hematopoietic stem and progenitor cell, LMPP = lymphoid-primed multipotential progenitor, MPP = multipotent progenitor.

Supplementary Figure 9: Expression of Cpeb4 through erythroid differentiation.



Expression levels of Cpeb4 through erythroid differentiation. Values represent mean tags per million calculated from single-cell RNA-seq data.

Supplementary Figure 10: Chromatin interactions and accessibility in the Cpeb4 locus through erythroid differentiation.



Supplementary Figure 10: Chromatin interactions and accessibility in the Cpeb4 locus through erythroid differentiation. [continued]

Tiled-C contact matrices of 500 kb spanning the mouse Cpeb4 locus in sequential stages of *in vivo* erythroid differentiation at 2 kb resolution. Contact frequencies represent normalized, unique interactions from 2 replicates. Matched open chromatin (ATAC) profiles are shown underneath each matrix and represent normalized data from 3 S0-low, S0-medium and S1 replicates and 2 S2 and S3 replicates. Gene annotation (Cpeb4 gene highlighted in green), open chromatin (ATAC) and CTCF, H3K4me3 and H3K4me1 occupancy in mature mouse erythroblast cells are shown at the top. Coordinates (mm9): chr11:31,450,000-31,950,000. The matrices show that the Cpeb4 TAD is established early in differentiation and that a smaller domain, in which the Cpeb4 promoter interacts with accessible elements marked with H3K4me1, is formed later in differentiation, concomitant with upregulation of Cpeb4 expression (Supplementary Figure 9).

Supplementary Table 1: Comparison of Hi-C and Tiled-C data.

	Read pairs	Contacts (α-globin region)	Contacts per restriction fragment
Hi-C ES	7,260,480,082	3,518,538	449
Tiled-C ES	379,033,076	97,807,262	12,490
Tiled-C Erythroid	293,410,431	60,515,876	7,728

Overview of the number of sequencing reads and detected contacts in the α-globin region (chr11:29,902,000-33,228,000) in Hi-C data in mouse ES cells (Bonev et al. Cell 2017) and Tiled-C data in mouse ES and primary erythroid cells. Numbers are derived from pooled data of 4 and 3 replicates for the Hi-C and Tiled-C data, respectively, and contacts represent filtered, non-duplicated contact counts.

Supplementary Table 2: Comparison of Tiled-C data across low-input primary erythroid samples.

Cells	Read pairs	Contacts (α-globin region)	Contacts per restriction fragment
100,000	58,362,991	6,108,837	780
20,000	57,242,048	5,782,472	738
10,000	55,480,867	3,251,628	415
5,000	56,476,873	1,769,051	226
2,000	33,898,548	926,504	118

Overview of the number of sequencing reads and detected contacts in the α-globin region (chr11:29,902,000-33,228,000) in Tiled-C data generated from low-input aliquots of primary erythroid cells. Numbers are derived from pooled data of 3 replicates and contacts represent filtered, non-duplicated contact counts.

Supplementary Table 3: Tiled-C capture oligonucleotides.

Region	Coordinates	Oligo number
α-globin	chr11:29,902,950-33,440,011	8,571
Sox2	chr3:33,200,154-36,559,869	8,266
Tal1	chr4: 113,693,194-116,236,558	6,200

Overview of the numbers of oligonucleotides used in Tiled-C experiments to target genomic regions of interest.

Supplementary Table 4: Comparison of Tiled-C data generated using single- and double-stranded oligonucleotides.

Oligo	Cell number	Read pairs	Contacts (α-globin region)	Contacts per restriction fragment
Single-stranded	100,000	58,362,991	6,108,837	780
Double-stranded	100,000	58,636,489	5,400,851	690
Single-stranded	10,000	55,480,867	3,251,628	415
Double-stranded	10,000	56,129,518	3,190,934	407

Overview of the number of sequencing reads and detected contacts in the α-globin region (chr11:29,902,000-33,228,000) in Tiled-C data generated in erythroid cells using single- and double-stranded oligonucleotides. Numbers are derived from pooled data of 3 replicates and contacts represent filtered, non-duplicated contact counts.