












Dynamic *Runx1* chromatin boundaries affect gene expression in hematopoietic development

Dominic D. G. Owens^{1,4}, Giorgio Anselmi ¹, A. Marieke Oudelaar ^{1,5}, Damien J. Downes ¹, Alessandro Cavallo ¹, Joe R. Harman ¹, Ron Schwessinger ^{1,2}, Akin Bucakci¹, Lucas Greder¹, Sara de Ornellas^{1,3}, Danuta Jeziorska¹, Jelena Telenius², Jim R. Hughes ^{1,2}  & Marella F. T. R. de Bruijn ¹ 

The transcription factor RUNX1 is a critical regulator of developmental hematopoiesis and is frequently disrupted in leukemia. *Runx1* is a large, complex gene that is expressed from two alternative promoters under the spatiotemporal control of multiple hematopoietic enhancers. To dissect the dynamic regulation of *Runx1* in hematopoietic development, we analyzed its three-dimensional chromatin conformation in mouse embryonic stem cell (ESC) differentiation cultures. *Runx1* resides in a 1.1 Mb topologically associating domain (TAD) demarcated by convergent CTCF motifs. As ESCs differentiate to mesoderm, chromatin accessibility, *Runx1* enhancer-promoter (E-P) interactions, and CTCF-CTCF interactions increase in the TAD, along with initiation of *Runx1* expression from the P2 promoter. Differentiation to hematopoietic progenitor cells is associated with the formation of tissue-specific sub-TADs over *Runx1*, a shift in E-P interactions, P1 promoter demethylation, and robust expression from both *Runx1* promoters. Deletion of promoter-proximal CTCF sites at the sub-TAD boundaries has no obvious effects on E-P interactions but leads to partial loss of domain structure, mildly affects gene expression, and delays hematopoietic development. Together, our analysis of gene regulation at a large multi-promoter developmental gene reveals that dynamic sub-TAD chromatin boundaries play a role in establishing TAD structure and coordinated gene expression.

¹MRC Molecular Hematology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. ²MRC WIMM Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. ³Physical and Theoretical Chemistry Building, Department of Chemistry, University of Oxford, Oxford, UK. ⁴Present address: Structural Genomics Consortium, University of Toronto, Toronto, Ontario, Canada. ⁵Present address: Max Planck Institute for Biophysical Chemistry, Göttingen, Germany. email: jim.hughes@imm.ox.ac.uk; marella.debruijn@imm.ox.ac.uk

Runx1/AML1 is a member of the RUNX family of transcription factors (TFs), which are key to many developmental processes^{1–3}. *Runx1* is best known for its critical role in the de novo generation of the hematopoietic system and maintenance of normal hematopoietic homeostasis^{1,2,4}. Disruption of *RUNX1* in humans leads to several hematopoietic disorders, including acute myeloid leukemia⁵ and familial platelet disorder with associated myeloid malignancy (FPD-AML)⁶. All members of the RUNX family bind the same canonical DNA motif (YGYGGT) and their tissue-specific functions are thought to be governed largely by their specific expression patterns⁷. *Runx1* transcription is tightly regulated, with changes in gene dosage and expression level affecting both the spatiotemporal onset of hematopoiesis and hematopoietic homeostasis^{8–11}. *Runx1* is transcribed from two alternative promoters, the distal P1 and proximal P2 that are differentially regulated and generate different transcripts and protein products^{12,13} (reviewed in de Bruijn and Dzierzak¹). During hematopoietic development, *Runx1* expression initiates from the P2 promoter and gradually switches to the P1 promoter with the majority of adult hematopoietic cells expressing P1-derived *Runx1*^{12,14–17}. The *Runx1* promoters do not confer tissue specificity and several distal *Runx1* cis-elements have been identified that mediate reporter gene expression in transient transgenic embryos in *Runx1*-specific spatiotemporal patterns^{16,18–21}. However, the combinatorial regulation of *Runx1* at different stages of hematopoiesis is currently unclear, as are the mechanisms through which the tight and dynamic spatiotemporal control of *Runx1* expression is achieved. A better understanding of *Runx1* regulation may yield insights into potential avenues for therapeutic targeting of *RUNX1* in a variety of hematological disorders, as recently highlighted by growth inhibition in a leukemia cell line upon loss of the *Runx1*+ 23 enhancer²².

The 3D conformation of DNA in structures such as topologically associating domains (TADs) delimit the activities of enhancers in vivo^{23–26}. Specific interactions achieved through chromatin folding, particularly enhancer–promoter (E–P) interactions, are thought to be a key component of spatiotemporal gene regulation^{27,28}. Many insights into principles of transcriptional regulation have come from studying a few relatively small gene loci, including the globin genes. However, genes encoding developmentally important TFs, including LIM-Homeobox, Hox, Eomes, Sox, and Sonic Hedgehog (Shh), often lie in larger regulatory domains and are frequently flanked by gene deserts²⁹. Indeed, studies of developmental regulators, such as Shh, have revealed exceptionally long-range enhancer–promoter interactions³⁰, suggesting that specific regulatory mechanisms may be at play at these large developmental loci in addition to the basic regulatory principles established at smaller genes. One aspect that remains unclear is whether large-scale chromatin conformation changes may be required to coordinate complex developmental expression patterns at larger genes. A known factor important for the regulation of chromatin conformation is CCCTC-binding factor (CTCF)³¹. CTCF, along with the loop extruding factor cohesin, mediates the establishment and maintenance of both E–P interactions and TAD structure^{32–34}. Interestingly, *Runx1* was mis-regulated in zebrafish after perturbation of CTCF/cohesin^{35–37}, suggesting that *Runx1* regulation may depend on chromatin structure. Elucidating *Runx1* transcriptional regulatory mechanisms is expected to contribute to a better understanding of the chromatin conformation changes employed by complex multi-promoter genes during development.

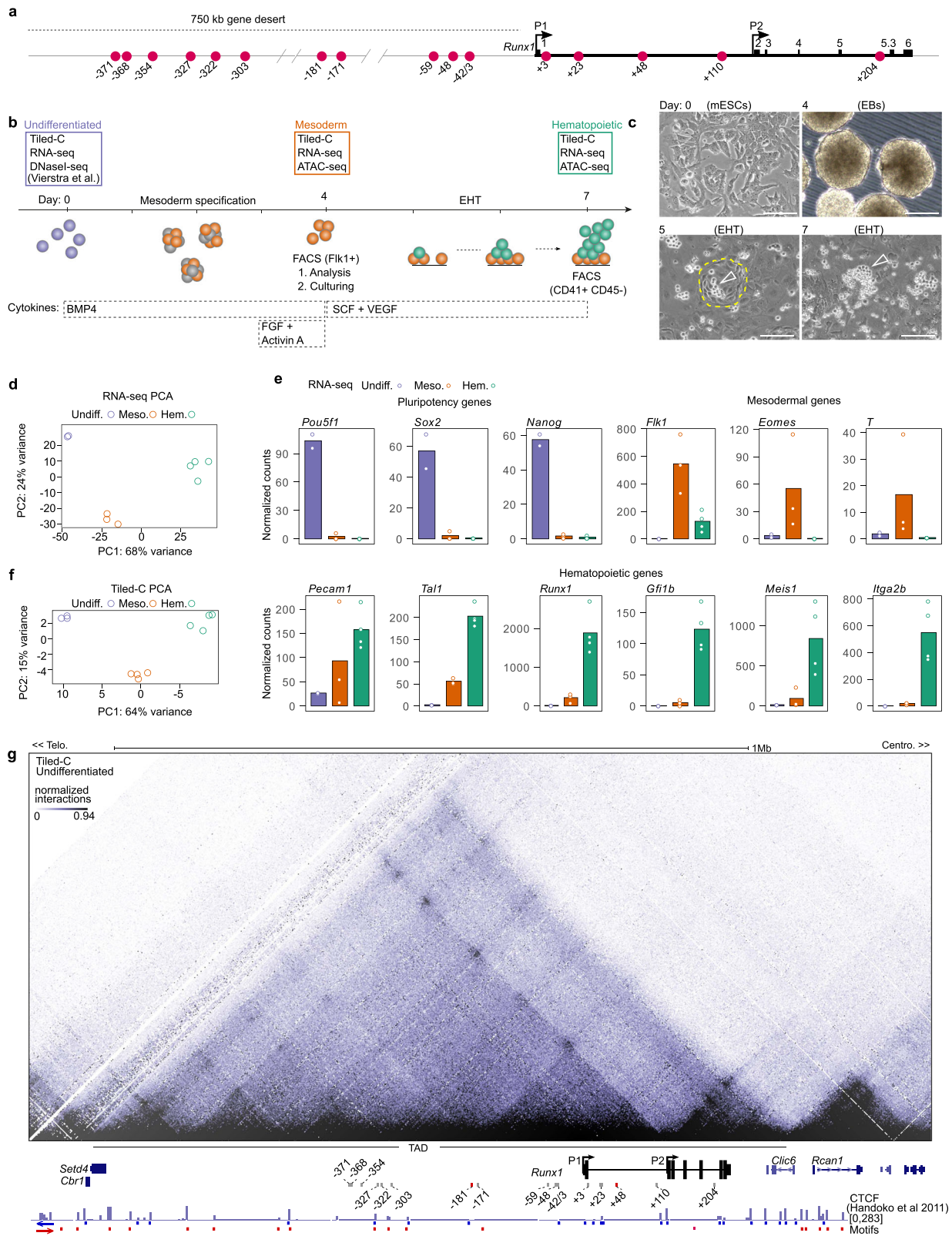
Here, we characterize the *Runx1* chromatin landscape in four-dimensions, i.e., in 3D space over time, in an in vitro mouse ESC differentiation model of developmental hematopoiesis. Using high-resolution chromosome conformation capture (Tiled-C)³⁸, we report the presence of a pre-formed 1.1 Mb TAD spanning the *Runx1* locus

in mouse ESCs that is conserved in human and forms prior to gene activation. Upon differentiation, accessible chromatin sites emerge within the TAD over known enhancers, CTCF sites, and candidate cis-regulatory elements. These regions interact with the *Runx1* promoters in a developmental stage-specific manner. Notably, an increased interaction of the P1 and P2 promoters within cell-type-specific *Runx1* sub-TADs is seen. These sub-TADs are bounded by highly conserved promoter-proximal CTCF sites, the role of which is poorly understood. Here, we use a machine learning approach and CRISPR/Cas9-mediated deletion to examine the importance of promoter-proximal CTCF binding for the *Runx1* chromatin landscape. Deletion of either the *Runx1* P1 or P2 promoter-proximal CTCF site partially disrupts TAD structure, while E–P interactions appear unaffected. *Runx1* levels show a decreased trend at the mesoderm stage, concomitant with significant changes in mesodermal gene expression indicative of a delay in hematopoietic differentiation. Together, we find that sub-TAD chromatin boundaries form dynamically within the large and complex *Runx1* regulatory domain during differentiation and are involved in coordinating gene expression and hematopoietic differentiation.

Results

***Runx1* lies in a conserved TAD which forms prior to gene activation.** To investigate dynamic changes in 3D chromatin conformation in the *Runx1* locus (schematically represented in Fig. 1a) during hematopoietic development, we used the in vitro mouse ESC (mESC) differentiation model that recapitulates the de novo generation of hematopoietic progenitor cells (HPCs) from mesoderm as it occurs in the embryo, including the endothelial-to-hematopoietic transition (EHT) specific to development (adapted from ref. 39). In this model, we assessed chromatin conformation (Tiled-C), along with gene expression (poly(A)-minus RNA-seq to capture nascent transcripts) and chromatin accessibility (ATAC-seq) over hematopoietic differentiation (Fig. 1b). Flk1⁺ mesodermal cells were isolated by flow cytometry from day 4 embryoid body (EB) cultures. Upon further differentiation in EHT media these gave rise to phenotypic HPCs with blood progenitor morphology and in vitro clonogenic potential (Fig. 1b, c, Supplementary Fig. 1). Gene expression analysis of mESCs, Flk1⁺ mesoderm, and emerging CD41⁺ CD45[−] Runx1⁺ HPCs reflected the developmental trajectory as visualized in a Principal Component Analysis (PCA) plot (Fig. 1d). This was accompanied by silencing of pluripotency genes (*Pof5f1*, *Sox2*, *Nanog*), transient expression of mesodermal genes (*Flk1*, *Eomes*, *T*), and increasing levels of hematopoiesis-associated genes (*Pecam1*, *Tal1*, *Gfi1b*, *Meis1*, *Itga2b*) (Fig. 1e). *Runx1* expression was initiated in Flk1⁺ mesoderm and increased in HPCs (Fig. 1e¹⁷).

To generate high-resolution chromatin conformation maps of the *Runx1* regulatory domain, we performed Tiled-C, a targeted method that generates Hi-C-like data at specific loci³⁸, with probes against all *DpnII* fragments in a 2.5 Mb region centered on *Runx1*. PCA of individual Tiled-C replicates showed a clear developmental trajectory (Fig. 1f, Supplementary Fig. 2) similar to that seen based on gene expression analysis (Fig. 1d), demonstrating that *Runx1* exhibits reproducible dynamic chromatin conformation changes during differentiation. *Runx1* resides within a 1.1 Mb TAD in mESCs (Fig. 1g, mm9 chr16:92,496,000–93,617,999) that extends to encompass the upstream 750 kb gene desert, with the *Setd4* and *Cbr1* genes at its telomeric end, and *Clic6* at its centromeric end (Fig. 1g). Using previously published CTCF occupancy data from mESCs⁴⁰, 31 binding sites were identified within the *Runx1* TAD (Fig. 1g). MEME analysis⁴¹ identified core CTCF binding motif location and orientation underlying CTCF peaks and revealed a predominant convergence of CTCF motifs—



with primarily centromeric oriented motifs near the *Setd4* telomeric end of the *Runx1* TAD, and telomeric oriented motifs primarily at the *Cltc6* centromeric end (Fig. 1g). Together, this shows that the *Runx1* regulatory domain is established prior to gene expression, likely in a CTCF-dependent manner.

Mesodermal differentiation is accompanied by increased interactions between *Runx1* P2 and enhancers in the gene desert and *Runx1* gene body. Upon transition from mESC to Flk1⁺ mesoderm we observed increased chromatin accessibility in the *Runx1* TAD and low levels of transcription primarily from

Fig. 1 *Runx1* resides within a topologically associating domain (TAD) in undifferentiated cells. **a** Schematic of the *Runx1* locus on mouse chromosome 16, with *Runx1* proximal (P2) and distal (P1) promoters, exons, and adjacent gene desert labeled. Previously identified enhancers are indicated by red circles that are numbered according to the distance (in kb) from the *Runx1* start codon in exon 1^{13,15-17,19-21,128-129}. **b** Schematic of seven-day differentiation protocol with cytokines and markers used for isolation of cells by FACS indicated. EHT = endothelial-to-hematopoietic transition. DNaseI-seq data in mESC was previously published⁴². **c** Bright-field images of different stages of in vitro differentiation. Colonies of hemogenic endothelial (HE) cells are outlined with dashed yellow lines and clusters of emerging hematopoietic progenitors are indicated by hollow white arrowheads. Scale bars = 200 μ m. Representative images are shown. Experiments were performed more than ten times with similar results. **d** Principal component analysis (PCA) of individual poly(A) minus RNA-seq replicates colored by cell type. **e** Plot of normalized counts of lineage marker gene expression across differentiation. Undifferentiated $n = 2$, mesoderm $n = 3$, hematopoietic $n = 4$. **f** PCA of individual Tiled-C replicates colored by cell type. **g** Tiled-C matrix at 2 kb resolution for undifferentiated mESCs. Matrix is a merge of three independent replicates ($n = 3$). Interactions are visualized with a threshold at the 94th percentile. *Runx1* promoters (P1 and P2), neighboring genes, the adjacent gene desert, and approximate location of the 1.1 Mb *Runx1* TAD are labeled. Publicly available CTCF ChIP-seq in E14 mESCs⁴⁰ was reanalyzed and the orientation of CTCF motifs identified de novo under CTCF peaks is indicated. Previously published enhancer regions are indicated and numbered according to their distance from the *Runx1* start codon in exon 1. Enhancer regions that are accessible in undifferentiated cells are shown as red bars and enhancers that did not overlap DNaseI-seq⁴² peaks are identified by gray bars.

the *Runx1* P2 promoter (Fig. 2a, b). Compared to mESCs, at this mesodermal stage we identified 36 open chromatin sites that are lost and 33 sites that are gained (DNaseI-seq from ref. ⁴²) (Fig. 2a, Supplementary Table 1, adjusted p -value < 0.05). Ten accessible peaks unique to mesoderm corresponded to previously identified enhancers (-327 , -322 , -303 , -181 , -171 , -59 , $+3$, $+23$, $+48$, $+110$)^{16,18-21,43-47} (Supplementary Table 2), while 24 peaks did not overlap with any known regulatory elements (Supplementary Table 1). Interactions between CTCF sites also increased at the mesodermal stage, particularly between the two boundaries of the *Runx1* TAD (Fig. 2c, d, Kruskal–Wallis and Dunn’s test, adjusted $p = 0.003$), while insulation of the main TAD (a measure of intra-TAD interactions) decreased slightly (Fig. 2e, Kruskal–Wallis and Dunn’s test, adjusted $p = 1.4 \times 10^{-135}$). To determine promoter-specific enhancer interactions, we compared virtual Capture-C profiles across the *Runx1* locus using the P2 and P1 promoters as viewpoints (Fig. 2f). We observed an overall increase of E-P2 interactions in mesoderm compared to mESCs (Fig. 2f, Kruskal–Wallis and Dunn’s test, adjusted $p = 0.005$), specific increased interactions between the P2 promoter and the -327 , -322 , -303 , -181 , and -171 enhancer elements in the gene desert, and the $+3$, $+23$, and $+110$ enhancers in the *Runx1* gene body (Fig. 2g). In contrast to the P2, the P1 promoter did not show a significant overall increase in interactions with enhancers in mesodermal cells (Fig. 2f, Kruskal–Wallis and Dunn’s test, adjusted $p = 0.9$), in line with the absence of P1-derived *Runx1* expression (Fig. 2b). However, a slight specific increase was seen in interactions between P1 and elements -181 and -171 in the gene desert, and elements $+23$, $+48$, and $+110$ within *Runx1* intron 1, and with the P2 could be seen (Fig. 2g). Together, these data indicate that early spatiotemporal control of *Runx1* expression at the onset of hematopoiesis is associated with increased interactions between CTCF sites, reduced TAD insulation, and may be mediated by regulation of interactions between specific enhancer elements and both the P2 and P1 promoters.

Increased *Runx1* expression upon hematopoietic differentiation is associated with P1 activation, a shift in E-P interactions, and sub-TAD reinforcement. Differentiation of hematopoietic-fated mesoderm into HPCs is accompanied by increased expression from both *Runx1* promoters, with a three-fold higher expression from P2 than P1 (Fig. 3a, b). Compared to mesoderm, HPCs show dynamic shifts in chromatin accessibility in the *Runx1* TAD. ATAC-seq peaks were gained in the gene body at the $+204$ and -42 enhancers, lost at $+48$, -171 , -181 , -303 , -322 , and -328 and other sites in the gene desert, and maintained at the $+3$, $+23$, and $+110$ enhancers in HPCs (Fig. 3a, MACS2 adjusted $p < 0.05$, Supplementary Table 1). Five peaks are unique to HPCs (-42 , -48.6 , $+6.6$, $+38$, $+204$) while four peaks

are present in all three cell types (-778 [Setd4 promoter], -772 [Setd4 intronic element], -467 [gene desert CTCF site], $+128$ [*Runx1* P2 promoter]). Insulation of the main TAD slightly decreased further as cells differentiated from mesoderm to HPCs (Fig. 3c, Kruskal–Wallis and Dunn’s test, adjusted $p = 2.1 \times 10^{-7}$), while CTCF–CTCF interactions between the boundaries of the TAD were not different from mesoderm (Fig. 3c, Kruskal–Wallis and Dunn’s test, adjusted $p = 0.13$). Alongside these changes at the main TAD-level, two sub-TADs spanning the *Runx1* gene were strengthened significantly in HPCs (Fig. 3d, e, Kruskal–Wallis and Dunn’s test, P1–P2 sub-TAD $p = 3.1 \times 10^{-43}$ and P2–3’ sub-TAD $p = 6.6 \times 10^{-13}$). We next compared how specific E–P interactions changed between mesoderm and HPCs using virtual Capture-C plots of the Tiled-C data. Compared to mesoderm, total E–P interactions increased in HPCs for the P2 promoter and there was a trend of increased overall E–P interactions for P1 (Fig. 3f, Kruskal–Wallis and Dunn’s test, P2: adjusted $p = 0.005$, P1: adjusted $p = 0.06$). Both P1 and P2 showed specific increases in E–P interactions with -59 , -43 , -42 , $+23$, and $+110$ enhancers (Fig. 3g), while interactions with elements extending further in the gene desert were decreased generally but maintained at non-tissue-specific CTCF sites in the gene desert. Therefore, differentiation to HPCs is associated with specific E–P interactions primarily within HPC-specific sub-TADs that span the *Runx1* gene.

***Runx1* promoter-proximal CTCF sites play a role in establishing *Runx1* chromatin architecture but not E–P interactions.**

The sub-TAD spanning the first intron of *Runx1* had boundaries that correspond to the P1 and P2 promoters. As both promoters have telomeric orientated CTCF sites < 2 kb upstream of the transcription start site (TSS; Fig. 3a, d), and CTCF is associated with sub-TAD and TAD boundaries^{31,48}, we performed CTCF ChIP-seq in the 416B HPC cell line to determine if the observed changes in sub-TAD structure may be associated with differential CTCF binding in HPCs versus mESCs. Like mESC-derived HPCs, 416B HPCs express *Runx1* from both the P1 and P2 promoter (Fig. 3a, Supplementary Fig. 3a). Interestingly, while the majority of CTCF sites in the *Runx1* TAD were bound at similar levels in HPCs and mESCs, an increase in CTCF binding was seen in HPCs at the P1-proximal CTCF site, and at the $+23$ enhancer (Supplementary Fig. 3a). We next examined what the mechanism underlying the differential CTCF binding could be. DNA CpG dinucleotide methylation has been suggested to modulate dynamic CTCF binding^{49–56}, and is also known to be associated with promoter silencing⁵⁷. To investigate whether P1 promoter methylation could underlie differential activation and CTCF binding during hematopoietic differentiation, we performed targeted bisulfite sequencing of *Runx1* promoters in undifferentiated

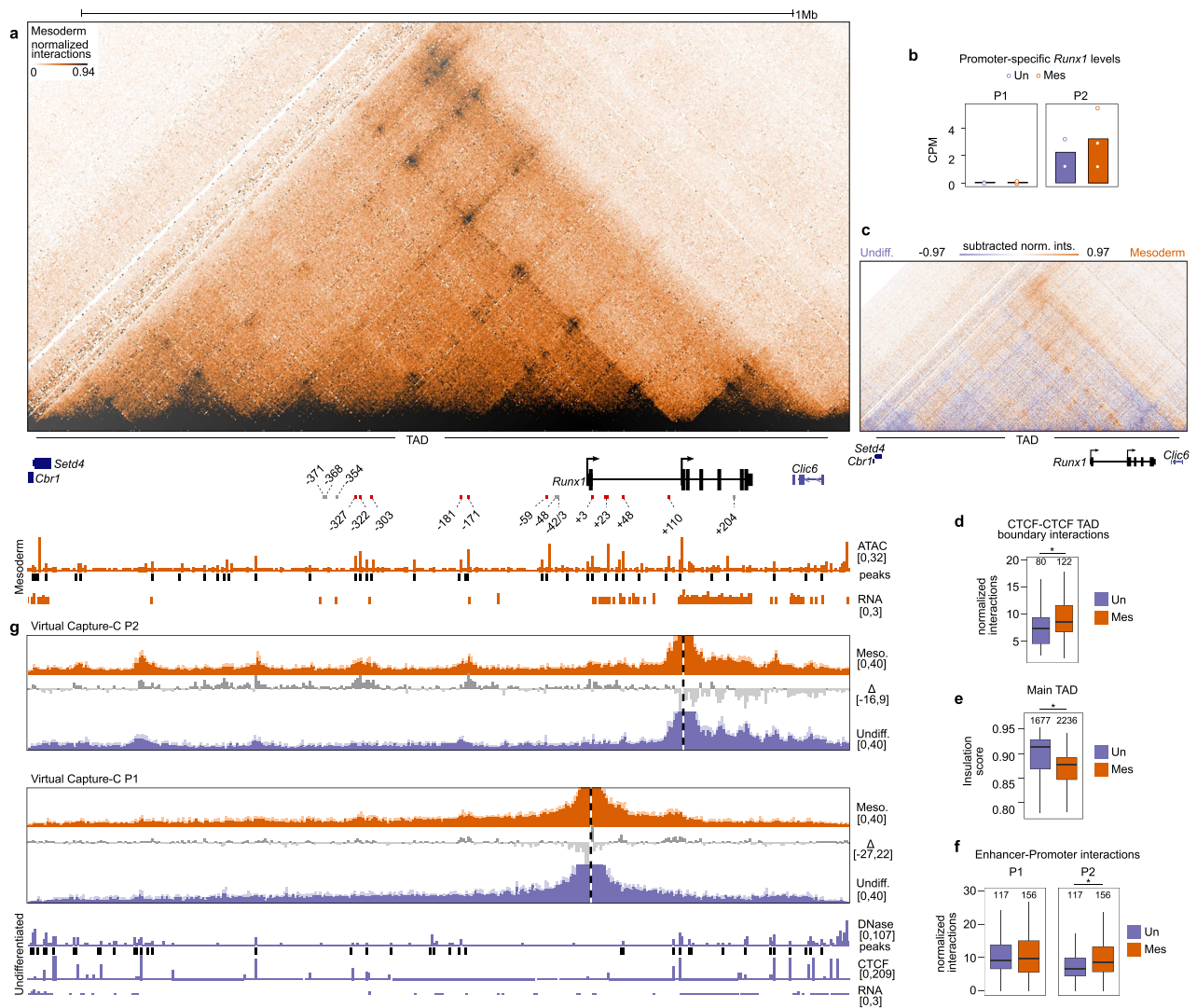


Fig. 2 Early hematopoietic differentiation leads to increased enhancer-*Runx1* P2 interactions. **a** Tiled-C matrix from mesoderm (2 kb resolution, threshold at the 94th percentile, $n = 4$). *Runx1* promoters and location of *Runx1* TAD are labeled below the matrix. RPKM-normalized ATAC-seq track is shown with called peaks (MACS2, adjusted $p < 0.05$, peaks called from one merged bam file). CPM-normalized poly(A)-minus RNA-seq ($n = 3$) is shown. Previously published enhancer regions are indicated. Enhancer regions that are accessible in mesoderm are shown as red bars and numbered according to their distance from the *Runx1* start codon in exon 1. Enhancers that did not overlap ATAC-seq peaks are identified by gray bars. **b** Promoter-specific *Runx1* levels in undifferentiated and mesoderm cells. Data were analyzed from two (undifferentiated) or three (mesoderm) biologically independent experiments. **c** Subtraction of normalized Tiled-C matrices between undifferentiated and mesoderm. The matrix is a subtraction of the signal between two merged matrices (undifferentiated $n = 3$, mesoderm $n = 4$, 2 kb resolution, threshold at +97th and -97th percentile). **d** Quantification of interactions between the four outermost CTCF peaks at the edges of the TAD (*, Kruskal-Wallis and Dunn's test, two-sided adjusted $p = 0.003$). **e** Insulation score (intra-TAD interaction ratio) of the main *Runx1* TAD (*, Kruskal-Wallis and Dunn's test, two-sided adjusted $p = 1.4 \times 10^{-135}$). **f** Quantification of total interactions from the viewpoint of each promoter with all previously published enhancers (Supplementary Table 2) (*, Kruskal-Wallis and Dunn's test, two-sided adjusted $p = 0.005$). **g** Virtual Capture-C profiles (obtained from Tiled-C data, see "Methods") from the viewpoint of both *Runx1* promoters in undifferentiated mESCs (blue tracks) and mesodermal cells (orange tracks). *Runx1* promoters (P1 and P2) are indicated by a vertical dashed line. Dark colors represent the mean reporter counts in 2 kb bins (undifferentiated $n = 3$, mesoderm $n = 4$) normalized to the total *cis*-interactions in each sample. Standard deviation is shown in the lighter color. Subtractions between two cell types as indicated are shown as gray tracks. DNase-seq⁴², DNase peaks, CTCF ChIP-seq⁴⁰, and RNA-seq from undifferentiated mESCs are indicated in blue below Capture-C tracks. **d-f** Boxplot centre shows median, bounds of the box indicate 25th and 75th percentiles, and maxima and minima show the largest point above or below 1.5 * interquartile range. Outlying points are not shown. Data were analyzed from the total number of bins indicated above each boxplot from three (undifferentiated) or four (mesoderm) biologically independent experiments.

mESCs and 416B HPCs. While the *Runx1* P2 promoter harbors a 2.0 kb CpG island that was hypomethylated in both cell types (Supplementary Fig. 3b), in contrast, *Runx1* P1 was near-completely methylated in mESCs, and became demethylated in hematopoietic cells (Supplementary Fig. 3b). Together, this shows that *Runx1* sub-TAD strengthening over hematopoietic

differentiation is associated with *Runx1* P1 promoter demethylation and increased CTCF binding at promoter-proximal sites.

Genome-wide, we observed a significant enrichment of CTCF binding close to active promoters in both mESCs and HPCs (<5 kb from TSS; Supplementary Fig. 3c, d, Chi-square test, $p < 1 \times 10^{-10}$). The role of promoter-proximal CTCF has not

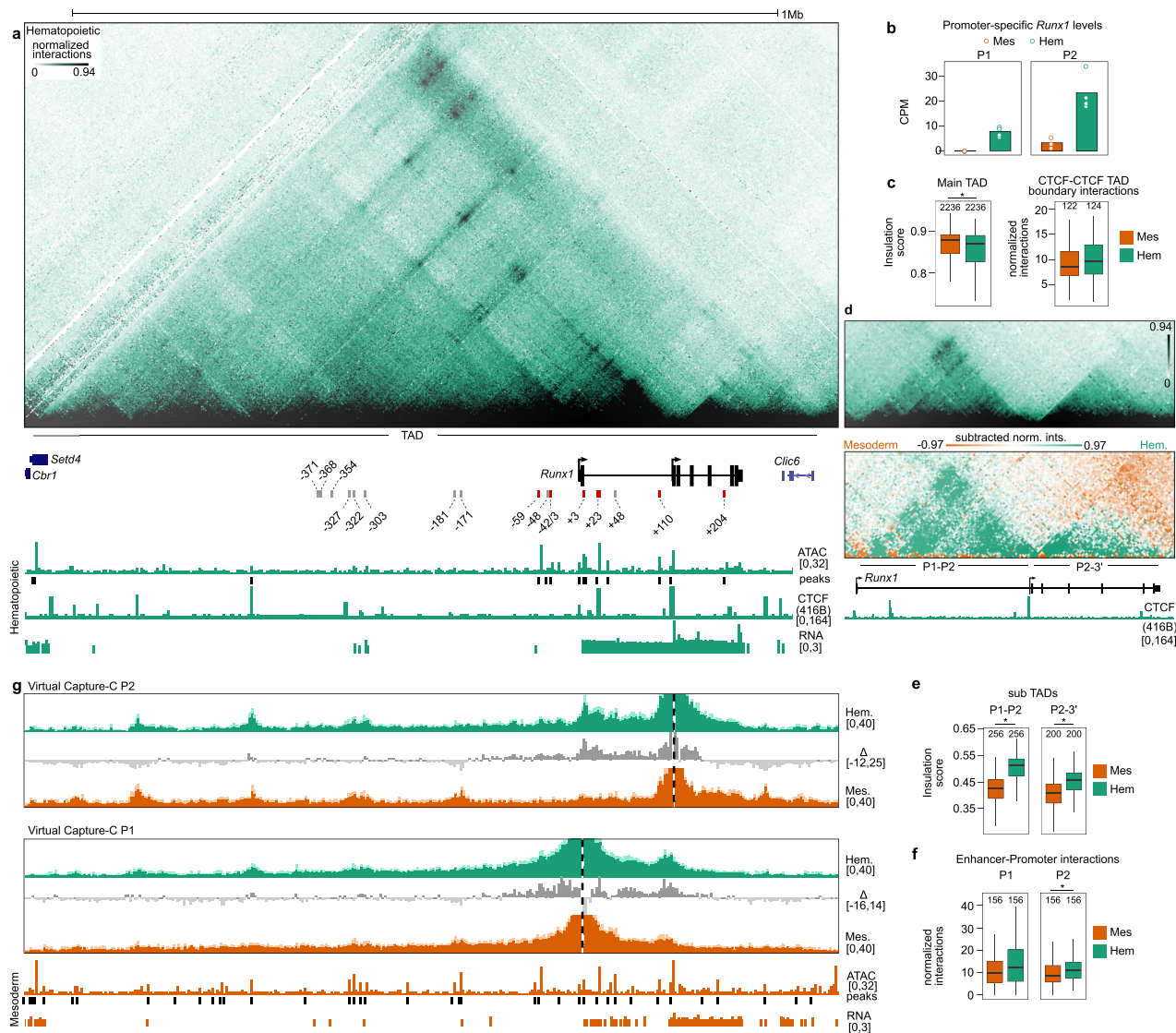
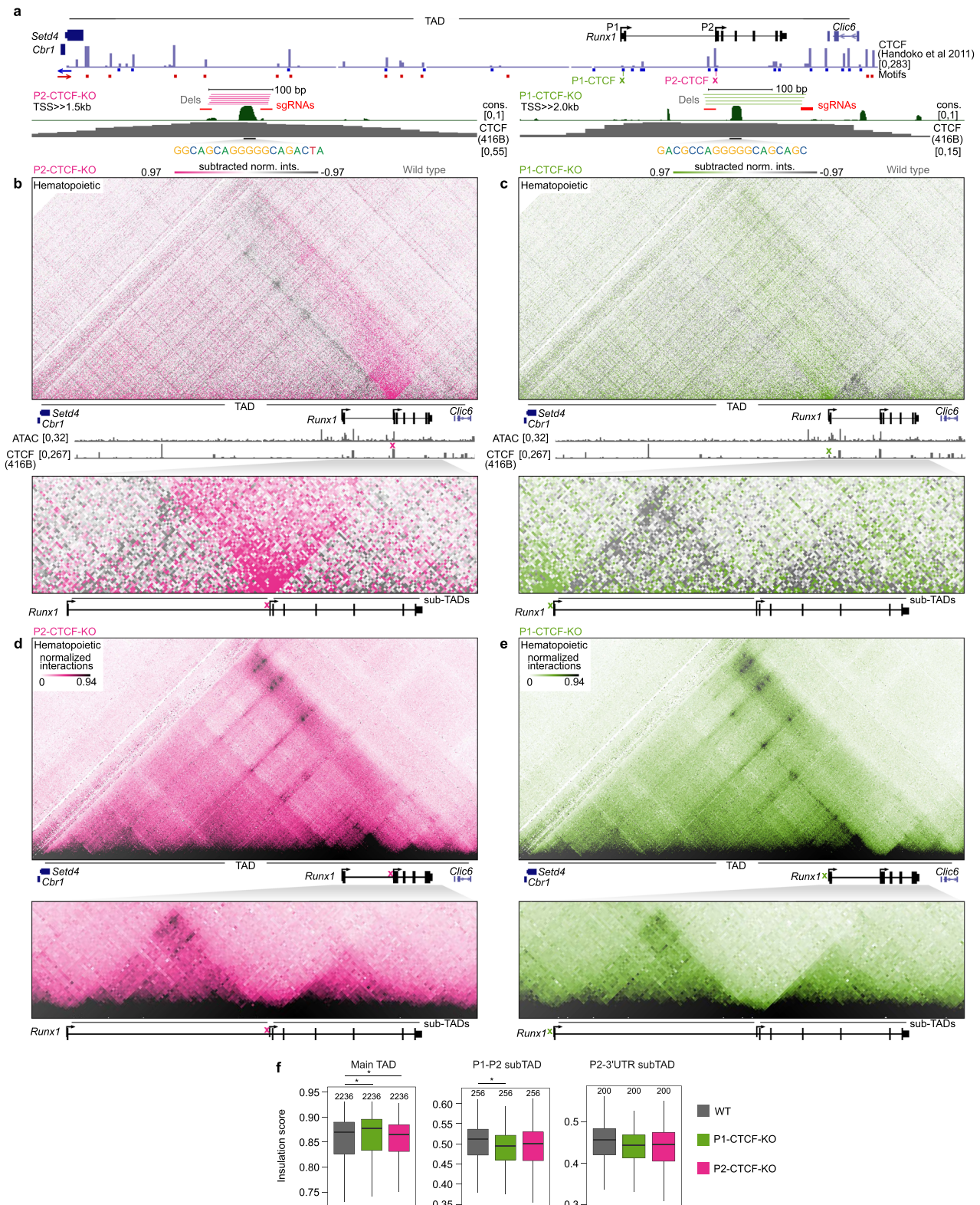


Fig. 3 EHT progression is associated with sub-TAD reinforcement, increased *Runx1* expression and P1 activation. **a** Tiled-C matrix from HPCs (2 kb resolution, threshold at the 94th percentile, $n = 4$). *Runx1* promoters and location of *Runx1* TAD are labeled below the matrix. RPKM-normalized ATAC-seq track is shown with called peaks (MACS2 adjusted $p < 0.05$, peaks called from one merged bam file). CPM-normalized poly(A)-minus RNA-seq ($n = 4$) is shown. CTCF occupancy in 416B hematopoietic progenitor cells is shown. Previously published enhancer regions are indicated. Enhancer regions that are accessible in HPCs are shown as red bars and numbered according to their distance from the *Runx1* start codon in exon 1. Enhancers that did not overlap ATAC-seq peaks are identified by gray bars. **b** Promoter-specific *Runx1* levels in mesoderm and HPCs. Data were analyzed from three biologically independent experiments. **c** Left, insulation score (intra-TAD interaction ratio) of the main *Runx1* TAD (*, Kruskal-Wallis and Dunn's test, two-sided adjusted $p = 2.1 \times 10^{-7}$). Right, quantification of interactions between the four outermost CTCF peaks at the edges of the TAD. **d** Top, zoom of Tiled-C data at 2 kb resolution with a threshold at 94th percentile. Below, subtraction of normalized Tiled-C matrices between mesoderm and HPCs. The matrix is a subtraction of the signal between two merged matrices ($n = 4$, 2 kb resolution, threshold at +97th and -97th percentile). **e** Insulation scores (intra-TAD interaction ratio) of the two *Runx1* sub-TADs (*, Kruskal-Wallis and Dunn's test, P1-P2 TAD two-sided adjusted $p = 3.1 \times 10^{-43}$ and P2-3' TAD two-sided adjusted $p = 6.6 \times 10^{-13}$). **f** Quantification of total interactions from the viewpoint of each promoter with all previously published enhancers (Supplementary Table 2) (*, Kruskal-Wallis and Dunn's test, two-sided adjusted $p = 0.004$). **g** Virtual Capture-C profiles (obtained from Tiled-C data, see "Methods") from the viewpoint of both *Runx1* promoters in mesoderm (orange tracks) and HPCs (green tracks). *Runx1* promoters (P1 and P2) are indicated by a vertical dashed line. Dark colors represent the mean reporter counts in 2 kb bins ($n = 4$) normalized to the total *cis*-interactions in each sample. Standard deviation is shown in the lighter color. Subtractions of the signal between two cell types as indicated are shown as gray tracks. ATAC-seq and peaks and RNA-seq from mesoderm are indicated in orange below Capture-C tracks. **c**, **e**, **f** Boxplot centre shows median, bounds of the box indicate 25th and 75th percentiles, and maxima and minima show the largest point above or below 1.5 * interquartile range. Outlying points are not shown. Data were analyzed from the total number of bins indicated above each boxplot from four biologically independent experiments.

been widely explored. To examine whether the deeply conserved P1 and/or P2 promoter-proximal CTCF sites (Fig. 4a), hereon referred to as P1-CTCF and P2-CTCF, may play a role in establishing the dynamic *Runx1* sub-TADs in HPCs we first utilized a deep learning approach (deepC; ref. 58) to predict

chromatin interactions at the *Runx1* locus in mESCs. The deepC model was trained on Hi-C data from mESCs⁵⁹, withholding mouse chromosome 16 containing *Runx1*. The overall *Runx1* TAD predicted by deepC agreed well with the TAD observed in Tiled-C data from mESCs (Supplementary Fig. 4a). In silico



deletion of the CTCF site proximal to the P2 promoter was predicted to reduce the stripe of interactions emanating from this site into the gene desert, and to increase interactions across the boundary in mESCs (Supplementary Fig. 4b). In contrast, deletion of the P1-proximal CTCF was predicted have little effect on chromatin interactions in mESCs (Supplementary Fig. 4c).

Next, we determined the impact of promoter-proximal CTCF site deletion on chromatin conformation experimentally by Tiled-C. We

generated P1-CTCF-KO and P2-CTCF-KO mESC clones using CRISPR-Cas9; these lacked the entire CTCF site, including the core motif, but retained nearby conserved sequences (Fig. 4a, Supplementary Figs. 5–7). Hematopoietic differentiation in the KO clones was unaffected (three independent mESCs clones analyzed each for the P1-CTCF-KO and P2-CTCF-KO; Supplementary Fig. 8) and Tiled-C was performed on undifferentiated mESCs, Flk1⁺ mesoderm, and HPCs. PCA showed that P1-CTCF-KO and P2-CTCF-

Fig. 4 *Runx1* promoter-proximal CTCF sites play a role in establishing *Runx1* chromatin architecture. **a** Schematic of *Runx1* TAD showing CTCF binding in mESCs⁴² and the orientation of CTCF motifs underlying peaks. P1 and P2 promoter-proximal CTCF sites are indicated with CRISPR/Cas9 strategies to delete them. Distance to *Runx1* transcription start sites is indicated. Vertebrate conservation (phastCons, cons), CTCF occupancy in 416B HPCs, core motif sequence, single guide (sg)RNA, and deletion alleles (dels) are indicated. **b, c** Subtraction of Tiled-C matrices between P2-CTCF-KO (**b**) P1-CTCF-KO (**c**) and wild-type hematopoietic cells is shown at 2 kb resolution with threshold at \pm 97th percentile of subtracted normalized interactions (subtracted norm. ints.) ($n = 4$). Locations of CTCF site deletions are indicated by a pink and green cross. RPKM-normalized ATAC-seq in wild-type HPCs and CTCF occupancy in 416B cells is shown. The locations of the main *Runx1* TAD and sub-TADs are indicated. **d, e** Tiled-C matrix from P2-CTCF-KO (**d**) and P2-CTCF-KO (**e**) (2 kb resolution, threshold at 94th percentile, $n = 4$). **f** Insulation scores (intra-TAD interaction ratio) for main *Runx1* TAD and sub-TADs in wild type, P1-CTCF-KO, and P2-CTCF-KO HPCs (*, Kruskal–Wallis and Dunn’s test, two-sided adjusted p -values: main TAD WT and P1-CTCF-KO $p = 4.8 \times 10^{-4}$, WT and P2-CTCF-KO $p = 0.03$, P1-P2 sub-TAD WT and P1-CTCF-KO $p = 0.003$). Boxplot centre shows median, bounds of the box indicate 25th and 75th percentiles, and maxima and minima show the largest point above or below 1.5 * interquartile range. Outlying points are not shown. Data were analyzed from the total number of bins indicated above each boxplot from four biologically independent experiments.

KO cells clustered along the same developmental trajectory as wild-type mESC differentiation cultures (Supplementary Figs. 9 and 10). Strikingly, CTCF-CTCF interactions were reduced across the entire TAD in P2-CTCF-KO HPCs, in agreement with the deepC prediction and consistent with P2-CTCF forming an insulated boundary (Fig. 4b, upper panel). Tiled-C in P1-CTCF-KO mESCs also agreed with the deepC predictions, with loss of the P1-CTCF site showing no effect compared to wild-type mESCs (Supplementary Fig. 4c). P1-CTCF-KO HPCs, however, exhibited a subtle decrease in interactions with CTCF sites in the gene desert (Fig. 4c, upper panel), highlighting the tissue-specific binding of CTCF to this site. Deletion of either P1-CTCF or P2-CTCF increased interaction frequencies between regions upstream and downstream of these sites, indicating that both CTCF sites act as boundaries (Fig. 4b and c, lower panels). Indeed, loss of P2-CTCF led to the region between *Runx1* and *Clic6* interacting ectopically with upstream regions (Fig. 4b, upper panel). The main TAD and tissue-specific sub-TADs were still present in P1-CTCF-KO or P2-CTCF-KO HPCs (Fig. 4d, e), though insulation scores were affected, indicating that sub-TAD boundary strengths were reduced (Fig. 4f).

In HPCs, both P1 and P2 promoters primarily interacted with enhancers lying within the tissue-specific sub-TADs (Fig. 3g). As these sub-TADs were altered upon deletion of promoter-proximal CTCF motifs, E-P interactions in P1- and P2-CTCF-KO HPCs were compared to wild-type cells. Surprisingly, despite generally weaker sub-TAD interactions in P1- and P2-CTCF-KO HPCs (Fig. 4a, b, and e), total E-P interactions were not significantly different compared to wild type for either promoter at any stage of differentiation (Supplementary Fig. 11a, b, Kruskal–Wallis test, adjusted $p > 0.4$) nor was any individual E-P interaction (Supplementary Fig. 12, Kruskal–Wallis test, adjusted $p = 1.0$). Together, our results show that despite perturbed chromatin architecture resulting from the absence of conserved promoter-proximal CTCF sites, specific *Runx1* E-P interactions are maintained.

***Runx1* P2 promoter-proximal CTCF site coordinates spatio-temporal gene expression and differentiation.** Since CTCF binding close to promoters genome-wide, including at *Runx1* P1, is associated with promoter activity (Supplementary Fig. 3c, d), and since loss of promoter-proximal CTCF sites disrupted *Runx1* chromatin architecture, the effect of promoter-proximal CTCF loss on *Runx1* expression was examined during hematopoietic differentiation of KO mESC clones. We observed a non-significant trend for reduced total *Runx1* expression in P2-CTCF-KO mesoderm compared to both wild-type and P1-CTCF-KO (Fig. 5a, zoomed in graph with dashed outline, DESeq2 adjusted $p = 0.6$). No changes were observed in alternative P1 or P2 promoter usage after deletion of promoter-proximal CTCF sites (Fig. 5b). PCA of global RNA-seq profiles across all stages and genotypes showed clustering based on cell type rather than genotype (Fig. 5c). However, when considering

mesoderm samples alone, all three P2-CTCF-KO samples were located at the far end of the distribution of samples (Fig. 5d). Indeed, differential expression analysis revealed that, globally, 168 genes were differentially expressed between P2-CTCF-KO and wild-type mesoderm (Fig. 5e, DESeq2 adjusted $p < 0.05$, fold change > 1). Notably, expression of several mesodermal markers (including *T* and *Eomes*) was higher in P2-CTCF-KO mesoderm compared to wild type, while several hematopoietic markers and *Runx1* target genes were downregulated (Fig. 5f, and Supplementary Fig. 13, adjusted $p < 0.05$)^{60–62}. GO analysis of genes downregulated by P2-CTCF-KO were associated with biological processes including “response to growth factor” and “blood vessel remodeling”, while upregulated genes were associated with terms including “mesoderm development” and “gastrulation” (Fig. 5g, adjusted $p < 0.05$). Collectively, this indicates that loss of the P2-proximal CTCF binding site caused a delay in *in vitro* hematopoietic differentiation, providing functional support for a mild decrease in P2-derived *Runx1* transcription at or prior to the mesoderm stage.

Discussion

In the present study, we used a cutting-edge 3C-based method to reveal dynamic changes in the *Runx1* chromatin architecture in four-dimensions, i.e., the three-dimensional folding of chromatin over developmental time. Tiled-C³⁸ analysis of the *Runx1* regulatory domain provided an unprecedented high-resolution view of *Runx1* chromatin architecture during *in vitro* differentiation from mESC through Flk1⁺ mesoderm to differentiating HPCs. Our detailed dissection of *Runx1* transcriptional regulation during developmental hematopoiesis sheds light on regulatory mechanisms of complex large developmental genes. We found that *Runx1* resides in a preformed, transcription-independent, and evolutionarily conserved main TAD that is present throughout differentiation (Fig. 6). Within this TAD dynamic sub-structures formed over development, namely sub-TADs spanning the *Runx1* gene that appeared specifically in HPCs. We showed that promoter-proximal CTCF sites played a role in the maintenance of *Runx1* sub-TADs, but, interestingly, not in mediating the dynamic changes in E-P interactions associated with hematopoietic differentiation. Yet, loss of the P2-proximal CTCF site led to delayed hematopoietic differentiation and disrupted gene expression specifically at the Flk1⁺ mesoderm stage, possibly by slight reductions in *Runx1* levels. Finally, we found that during hematopoietic development from mesoderm to HPCs, the *Runx1* promoters switched from interacting with enhancers located throughout the TAD, including the gene desert, to primarily interacting with cis-elements closer to the gene and within the tissue-specific sub-TADs. This refines the region within which functional enhancers important for driving hematopoietic-specific *Runx1* expression are likely to be found. These hematopoietic enhancers may represent therapeutic targets in leukemia, similar to what was recently shown for the *RUNX1*

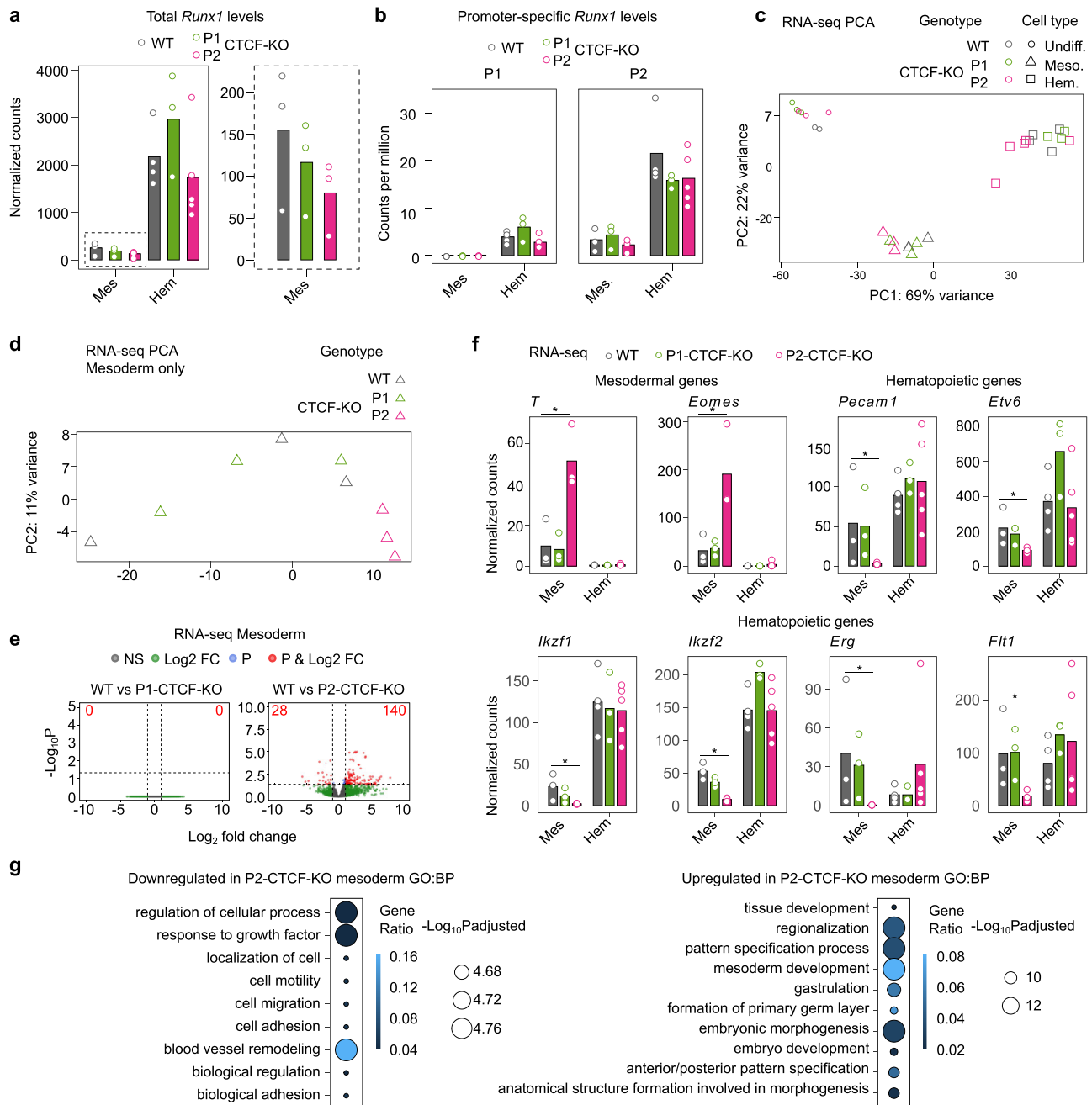


Fig. 5 *Runx1* spatiotemporal expression is slightly altered after loss of P2-proximal CTCF. **a** Total *Runx1* levels in poly(A)-minus RNA-seq in the cell types and genotypes indicated. The expanded graph with a dashed outline shows data just for mesodermal cells on a different axis. **b** Promoter-specific *Runx1* levels for each promoter in the cell types and genotypes indicated. **c** PCA of all poly(A)-minus RNA-seq replicates. **d** PCA of mesoderm RNA-seq samples. **e** Volcano plots showing differentially expressed genes (DEGs, DESeq2 adjusted two-sided $p < 0.05$, fold change > 1) in P2-CTCF-KO compared to wild-type mesoderm. **f** Expression of lineage marker genes across differentiation in the genotypes indicated (*, DESeq2 adjusted two-sided p -values: *T* (0.048), *Eomes* (0.0021), *Pecam1* (0.037), *Etv6* (0.043), *Ikzf1* (0.016), *Ikzf2* (1.1×10^{-5}), *Erg* (0.0079), *Flt1* (0.0078), fold change > 1). **g** GO term biological processes associated with the DEG list between wild-type and P2-CTCF-KO mesoderm. Gene ratios and $-\log_{10}$ p -values adjusted using the Benjamini-Hochberg method are indicated for significantly enriched (goseq p -values adjusted with Benjamini-Hochberg procedure $p < 0.05$) GO terms. **a-g** Data were analyzed from $n = 3$ independent experiments for Wild type, P1-CTCF-KO, P2-CTCF-KO mesoderm, $n = 4$ independent experiments wild-type hematopoietic, $n = 3$ independent experiments P1-CTCF-KO hematopoietic, $n = 5$ independent experiments P2-CTCF-KO hematopoietic).

+23 enhancer²². As both leukemias with and without somatic or germline *RUNX1* mutations were shown to depend on wild-type *RUNX1*⁶³ (reviewed in ref. 5) it is plausible that a similar dependency on the +23 enhancer may exist in other leukemic cells⁶⁴. Of note, a recent report exploring GWAS studies for SNPs in *RUNX1* regulatory elements found just one mutation in the +23 enhancer and this was predicted to be pathogenic⁶⁵. This

suggested that there is strong selection pressure to conserve the +23 enhancer in normal hematopoiesis. However, further studies are required to elucidate the dependency on *Runx1* enhancers in normal hematopoiesis and leukemia. The 4D regulatory interactions at *Runx1* described here were missed in previous reports^{21,66,67} as those lacked the required high-resolution, nor included a developmental time series.

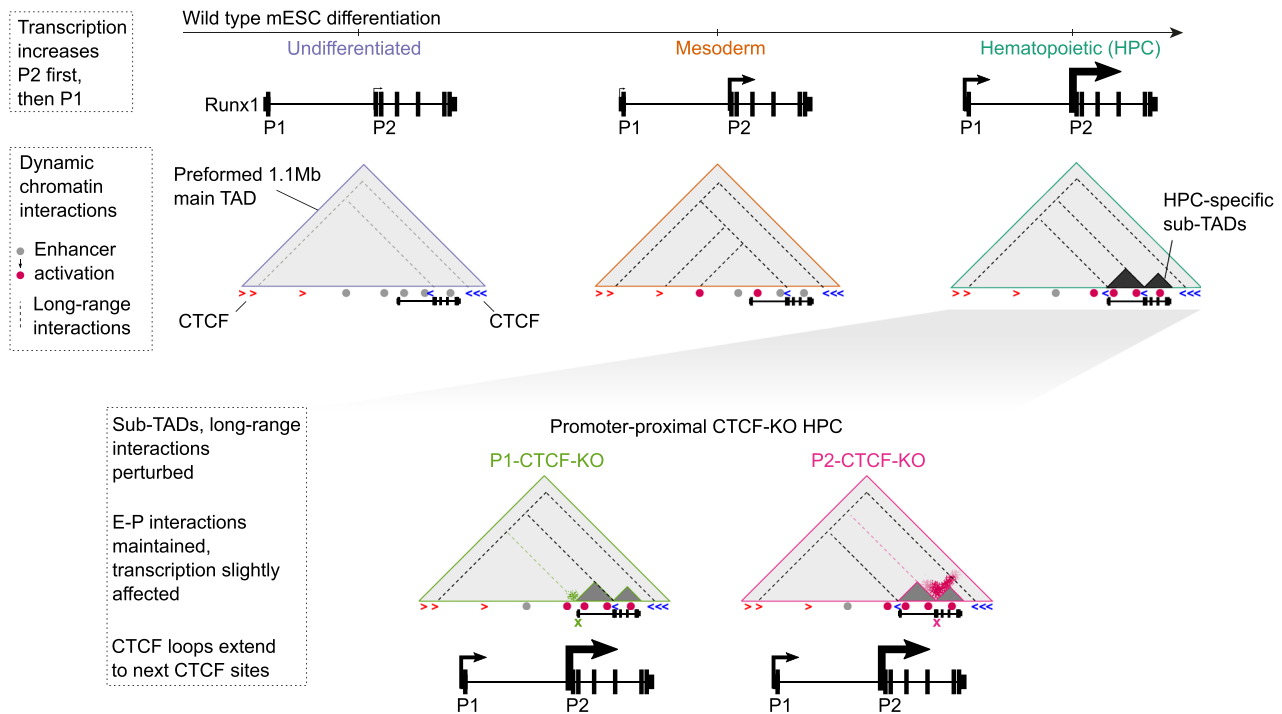


Fig. 6 Schematic model of dynamic chromatin changes at *Runx1* during hematopoietic development and after promoter-proximal CTCF site deletions.

Large light gray triangles represent the preformed main 1.1 Mb *Runx1* TAD. The orientation of selected CTCF motifs is indicated below the TAD. Selected previously identified enhancer elements in each cell type are represented by gray circles (inaccessible) or red circles (accessible). Dashed lines throughout the TAD represent long-range chromatin interactions, with the darker color indicating a stronger interaction. Hematopoietic progenitor cell (HPC)-specific sub-TADs over the *Runx1* gene are indicated by dark gray smaller triangles within the larger TAD. A larger schematic of the *Runx1* gene is shown in each cell type with larger arrows at each promoter representing more transcription from that promoter. P1 and P2-CTCF-KO interactions in HPCs are indicated in the bottom two triangles, with the location of the deleted sites indicated by green and pink crosses. The lighter sub-TAD triangles over the *Runx1* gene indicate reduced sub-TAD insulation after promoter-proximal CTCF site deletion. The green and pink colored areas in the TAD represent increased interactions compared to wild type after deletion of P1-CTCF and P2-CTCF, respectively. The lighter green and pink dashed lines represent reduced long-range interactions from the promoters after P1 and P2-CTCF site deletion.

In line with TADs at other developmentally regulated loci^{38,68–70}, the overall 1.1 Mb *Runx1* TAD formed prior to differentiation and in the absence of virtually any gene transcription. The mechanism behind its establishment is likely CTCF/cohesin-mediated loop extrusion^{71–74} as a predominant convergence of CTCF motifs was observed at the main *Runx1* TAD boundaries, similar to what was found at other TADs^{75–77}. In addition to these preformed chromatin structures, increasing CTCF-CTCF interactions were observed upon *Runx1* activation that might reflect a higher rate or processivity of loop extrusion in the *Runx1* TAD, as was recently also observed for α -globin³⁸. Alternatively, increased TF binding could be driving the specific increases in chromatin interactions that were observed over differentiation^{78,79}. Two tissue-specific sub-TADs formed over differentiation, leading to a sub-compartmentalization of the *Runx1* gene itself that was correlated with promoter activation. This agrees with recent findings at alpha-globin, where gene activation is correlated with sub-TAD strengthening³⁸, indicating that common mechanisms underlie chromatin architecture changes during transcriptional activation at different sized gene loci. However, we also observed differences in chromatin structures between smaller and larger genes, in that sub-compartmentalization within a gene is not generally seen at smaller genes, such as α -globin, which reside entirely within one tissue-specific sub-TAD^{23,80}. This difference might simply reflect the smaller size of the α -globin domain compared to the larger size of the *Runx1* regulatory domain (97 kb compared to 1.1 Mb, respectively). The sub-TAD encompassing the entire α -globin regulatory unit was shown to represent a discrete functional unit

that delimits enhancer activity²³. We did not see evidence for this in *Runx1* as E-P interactions were not significantly changed upon sub-TAD perturbation by deleting promoter-proximal CTCF sites. However, residual sub-TAD structures may still have provided a framework within which specific regulatory interactions could take place⁸¹. Promoter-proximal CTCF site deletion facilitated new interactions with CTCF sites distal to the deleted CTCF sites, leading to expanded loop domains at *Runx1*. Namely, deletion of P2-CTCF led to increased interactions between upstream regions and the centromeric end of the TAD up to a cluster of CTCF sites close to *Clic6* (Fig. 6). Moreover, P1-CTCF-KO led to increased interactions upstream of the P1 promoter at the expense of downstream contacts in the P1-P2 sub-TAD which was weakened. The finding that the *Runx1* tissue-specific sub-TADs were strongest in HPCs, which express high levels of *Runx1*, suggests they may be similar to the gene-body-associated domains (GADs) recently observed at genes highly expressed in hematopoietic cells⁸². The fact that P1 and P2-CTCF sites are not in a convergent orientation, and the fact that sub-TADs/GADs form within the gene-body of actively transcribed genes, suggests that instead of being caused by a CTCF-dependent mechanism like loop extrusion, these structures may be dependent on a combination of factors including transcriptional processes⁸², TF binding^{78,79}, and clustering of like histone modifications⁸³. Indeed, our data indicate that, as reported for other promoters^{58,84–86}, the *Runx1* promoters may function as chromatin boundaries in a CTCF-independent manner, which would explain the residual sub-TADs and E-P interactions, as well as the expanded chromatin interactions observed in P1/P2-CTCF-KO

cells. Together, our findings suggest that loop extrusion and transcription-related mechanisms may act in concert to produce dynamic chromatin structures during differentiation.

Interestingly, deletion of the *Runx1* P2 promoter-proximal CTCF binding site resulted in a developmental delay in hematopoietic mesoderm as shown by increased expression of early mesodermal markers at the expense of later hematopoietic ones. Although studies in cell lines indicate that CTCF sites are required for TAD formation^{32–34,87}, interestingly, promoter-proximal CTCF sites have been suggested to play a role in E-P interactions and gene transcription^{55,88–92}. Therefore, a plausible explanation for the observed differentiation delay could be that promoter-proximal CTCF loss leads to a later or perturbed onset of *Runx1* expression, as *Runx1* is well known to promote hematopoietic commitment⁹³. Although *Runx1* expression in the total mesodermal cell population was not significantly altered, a decreased trend was seen. This surprisingly robust *Runx1* expression and alternative promoter usage upon promoter-proximal CTCF-site loss may reflect the residual sub-TADs still found which could be attributed to redundant CTCF sites not targeted in this study (redundancy between CTCF sites has been observed before^{23,58,94}), or the *Runx1* promoters themselves. An alternative explanation could be the relatively asynchronous development of cells in culture, where some cells may have had enough time to restore *Runx1* expression levels. A more substantial phenotype may be seen in a *Runx1*-heterozygous background, similar to what was observed in mouse embryos carrying an attenuated *Runx1* P2 allele and a non-functional *Runx1* allele, where the compound phenotype was more severe than that of a homozygous attenuated P2¹⁶. Finally, compared to in vitro differentiation, promoter-proximal CTCF-site loss may be more detrimental in vivo, where *Runx1* levels are subject to tight spatiotemporal control and changes to *Runx1* levels or dosage lead to knock-on effects on differentiation timing^{8–11,16,95}. Together this indicates that even subtle changes in *Runx1* levels, such as the trend seen in P2-CTCF-KO mesoderm, have the potential to alter hematopoietic developmental dynamics. This underlines that *Runx1* requires an exceptionally fine-scale spatiotemporal transcriptional control and isolation from neighboring regulatory domains to support its pivotal role in development. Given that *Runx1* has important functions in development and human disease^{1,2,96}, an increased understanding of dynamic cis-regulatory mechanisms underpinning its regulation will be vital to future efforts to develop potential therapeutic approaches to manipulate *RUNX1* expression in human blood disorders.

Methods

Cell culture. E14-TG2a mESCs⁹⁷ were cultured in GMEM medium supplemented with 100 mM non-essential amino acids, 100 mM sodium pyruvate, 10% FCS, 2 mM L-glutamine, 100 μ M β -mercaptoethanol (all Gibco), and 1% Leukemia Inhibitory Factor (prepared in house). Cells were passaged using 0.05% trypsin (Gibco) every 2–3 days. The 416B mouse immortalized myeloid progenitor cell line⁹⁸ was cultured at $2–8 \times 10^5$ cells/ml in Fischer's medium with 20% horse serum and 2 mM L-glutamine (all Gibco).

Hematopoietic differentiation of mESCs. Differentiation of mESCs was performed using a modified serum-free protocol^{39,99} in StemPro-34 (SP34, Gibco) supplemented with 40X defined serum replacement, 2 mM L-glutamine (Gibco), and 0.5 mM ascorbic acid, 0.45 mM monothioglycerol (Sigma). mESCs were seeded at a density of 5×10^4 cells/ml into SP34 medium plus BMP-4 (R&D, 5 ng/ml). At day 3, bFGF and Activin A were added (R&D, 5 ng/ml). At day 4, single-cell suspension was generated from embryoid bodies using 0.05% trypsin. FACS-isolated Flk1+ mesodermal cells were cultured at 5×10^4 cells/cm² in SP34 plus SCF (Peprotech) and VEGF (R&D), 10 ng/ml each. After a further 3 days of culture, adherent and suspension cells were treated with 0.05% trypsin and analyzed.

Flow cytometry and cell sorting. Cells were stained in PBS plus 10% FCS with the antibodies listed in Supplementary Table 3. Dead cells were identified with Hoechst 33258. Cells were analyzed using a Fusion 2 flow cytometer (BD Biosciences, BD

FACSDiva Software version 8.0.1) and data analysis was performed using FlowJo (TreeStar, version 10.8).

Colony forming unit assays. Unsorted cells at day 6 (4 + 2) of differentiation were cultured in MethoCult 04434 (Stem Cell Technologies) in 35 mm dishes. Colonies were counted after 10 days.

Immunocytochemistry and confocal microscopy. Cells were plated into glass-bottom 24-well plates (ibidi) and fixed after culture for 10 min with 4% paraformaldehyde (Sigma), permeabilized, and labeled using antibodies (Supplementary Table 3) for 1 h at room temperature. Imaging was performed using a Zeiss 880 laser scanning confocal microscope.

Deletion of CTCF sites using CRISPR/Cas9. Single guide RNAs (sgRNAs, Supplementary Table 4) were designed (crispr.mit.edu) to flank conserved CTCF motifs. sgrNAs were cloned into pSpCas9(BB)-2A-Puro V2.0¹⁰⁰ containing two sgRNAs and confirmed by sequencing. mESCs were transfected with 5 μ g plasmid using Lipofectamine 2000 (Invitrogen) and puromycin selected (1 μ g/ml). Single colonies were isolated by limiting dilution¹⁰¹. To detect larger on-target deletions^{102–104}, PCR amplification was done (500 bp to 5 kb, Supplementary Table 5). Sequencing confirmed the presence of two distinct deletion alleles and the retention of DpnII restriction sites.

Copy counting by droplet digital PCR (ddPCR). To further rule out the presence of undesired complex genotypes¹⁰⁵, copy counting was performed across the targeted regions by droplet digital PCR (ddPCR). Reactions were performed in duplex, amplifying from an internal control and a test region in every reaction. Internal control was located on mouse chromosome 4 (not targeted in these experiments and karyotypically stable in mESCs¹⁰⁶). Test regions were amplified directly over the targeted region to detect loss of allele (LOA) and at 100 bp and 1 kb up- and down-stream from sgRNA target sites (Supplementary Table 5)¹⁰². Reactions (22 μ l) contained 11 μ l QX200 ddPCR EvaGreen Supermix (Bio-Rad), 25–50 ng genomic DNA purified using DNeasy Blood and Tissue Kit (Qiagen), 250 nM each of internal control primer, and 125 nM each of test primer. Standard reagents and consumables supplied by Bio-Rad were used. Ratios between test and internal control amplicons was determined in QuantaSoft software (Bio-Rad). Ratios were normalized to the mean ratio of three test amplicons located on different non-targeted chromosomes (1, 6, and 7) to determine relative copy numbers of the test amplicons.

Chromatin interaction analysis (Tiled-C). Tiled-C was performed on between 7.7×10^4 and 1×10^6 cells using a low-input protocol^{38,107,108}. Cells were cross-linked using 2% formaldehyde for 10 min. DpnII (NEB) digestion was performed shaking overnight at 37 °C. Ligation was performed overnight at 16 °C. Samples were treated with RNase A (Roche) for 30 min at 37 °C and decross-linked using Proteinase K (Thermo Fisher) overnight at 65 °C. Digestion efficiency was quantified by qPCR (Supplementary Table 5). Libraries with >70% digestion efficiency were used (mean 83%). Up to 1 μ g DNA was sonicated using a Covaris ultrasonicator. End-repair, adapter ligation, and PCR addition of indices (7–11 cycles) was done using NEBNext Ultra II DNA library prep kit (NEB). Biotinylated capture probes 70 nt in length were designed against every DpnII restriction fragment in a 2.5 Mb window centered on *Runx1* (chr16:91,566,000–94,101,999). Probe sequences were stringently BLAT-filtered to exclude repetitive sequences, and synthesized in-house³⁸. A pooled capture reaction was performed on 1 μ g of each indexed 3C library. Washing of captured material was done using Nimblegen SeqCap EZ hybridisation and wash kit (Roche) and captured sequences were isolated using M-270 Streptavidin Dynabeads (Invitrogen). PCR amplification was performed for 12 cycles. Amplified DNA was purified and a second capture and PCR amplification step were performed. Libraries were sequenced on two Illumina NextSeq high-output 150 cycle runs (paired-end).

Analysis of Tiled-C data. Tiled Capture-C data was processed at 2 kb resolution³⁸. Fastqs were analyzed using the CCSeqBasic CM5 pipeline¹⁰⁹ (<https://github.com/Hughes-Genome-Group/CCseqBasicF/releases>). Individual samples were analyzed before merging biological replicates. PCR duplicate-filtered bam files containing uniquely mapping reads were converted to sam files (samtools) and then into sparse raw contact matrices (Tiled_sam2rawmatrix.pl, <https://github.com/oudelaar/TiledC>). ICE normalization was done using HiC-Pro (2.11.1)^{110,111} and matrices were imported into R (3.6.0). Matrices were plotted (TiledC_matrix_visualisation.py, <https://github.com/oudelaar/TiledC>) with a threshold between the 90th and 95th percentile. PCA was done on log normalized counts (DESeq2, 1.30.1)¹¹². Merged ICE normalized contact matrices were scaled to the mean number of total interactions (14,631,865) across samples using custom R scripts¹¹³. Virtual Capture-C plots were generated by sub-setting the matrices on individual viewpoints of interest. E-P contacts were quantified in count and ICE-normalized matrices from the viewpoint of the bin containing each promoter and bins overlapping previously published *Runx1* enhancers (Supplementary Table 2). TADs were detected by visual inspection. Intra-TAD interactions were calculated by quantifying the ratio between intra- and extra-TAD interactions for each bin within the TAD in each sample. TAD boundary contacts were quantified between

bins overlapping the four outermost CTCF sites at each of the centromeric and telomeric ends of the main TAD.

Analysis of Hi-C data. Publicly available Hi-C data in mESCs⁵⁹ was analyzed as previously described³⁸. Data were analyzed using HiC-Pro¹¹⁰ with ICE normalization¹¹¹, and plotted using python as described above.

DeepC prediction of chromatin architecture. Predictions of chromatin architecture were performed using deepC⁵⁸. Briefly, deepC was trained using a transfer learning approach on distance stratified and percentile binned Hi-C data from mESCs⁵⁹, withholding mouse chromosome 16 (that contains *Runx1*) and chromosome 17 from training. Training was done in two stages. First, a convolutional neural network was trained to predict chromatin features given a 1 kb DNA sequence input. The chromatin features cover open chromatin, transcription factor binding, including CTCF, and histone modifications using publicly available DNase-seq, ATAC-seq, and ChIP-seq peaks across a range of cell types. Second, a neural network using a convolutional module followed by a dilated convolutional module was trained to predict Hi-C data given 1 Mb of DNA sequence input. The convolutional filters of the first network are used in the transfer learning to seed the filters of the convolutional module. Promoter-proximal CTCF site deletions were modeled by mutating the region spanning CRISPR/Cas9 deletion alleles that were confirmed by sequencing and predicting the chromatin interactions of the reference and deletion alleles.

Gene expression analysis (RNA-seq). RNA was isolated from 1×10^3 to 2.5×10^6 cells using QIAzol (Qiagen). Total RNA was extracted using miRNeasy Mini kit (Qiagen). RNA integrity was determined using a 4200 TapeStation RNA Screen-Tape (Agilent). Ribosomal RNA was depleted from 2.5 µg total RNA per sample of undifferentiated mESC and 416B cells using the RiboMinus™ Eukaryote System v2 (Invitrogen). A poly A selection module (NEB) was used to extract poly A minus RNA and was eluted directly in First Strand Synthesis Reaction Buffer. cDNA was synthesized using the NEBNext Ultra directional library prep kit. Adapter ligation and 8–15 cycles of PCR were performed. Libraries were sequenced on Illumina NextSeq high-output 75 cycle kit (paired-end).

RNA-seq analysis. Fastq files were mapped to the mouse genome (mm9) using STAR (2.6.1d)¹¹⁴. PCR duplicates were removed using picard-tools (2.3.0) MarkDuplicates. Counts per million (CPM)-normalized bigwig files were generated using deeptools (2.2.2 and 3.0.1)¹¹⁵. A blacklist file was used to exclude mapping artifacts. Bigwig files were converted to bedGraph (ucscutils (373)) and imported into R. Mean CPM was calculated for each merged sample. Reads were assigned using subread (2.0.0) featureCounts¹¹⁶. For poly(A)-minus RNA-seq data reads were assigned to both exons and introns. Assigned counts were imported into R and analyzed using DESeq2 (1.24.0)¹¹². Sample clustering was performed on log normalized counts. Differential expression analysis was done using DESeq2 (adjusted two-sided p -value < 0.05 , fold change > 1). Volcano plots were made using EnhancedVolcano (1.2.0)¹¹⁷. GO terms were calculated using goseq (1.36.0)¹¹⁸ and KEGG.db (3.2.3)¹¹⁹. Gene expression was visualized using plotCounts and ggplot2 (3.3.0)¹²⁰. Promoter-specific counts were quantified from over a 5 kb window downstream of each TSS (*Runx1*-P1 chr16:92,823,811–92,828,811; *Runx1*-P2 chr16:92,695,073–92,700,073) using bedtools (2.25.0)¹²¹.

Chromatin accessibility analysis (ATAC-seq and DNaseI-seq). ATAC-seq libraries were generated in differentiated E14-TG2a-RV mESCs (stably transfected with a Venus reporter at the 3' end of *Runx1*⁴⁷ and a hsp68-mCherry-*Runx1* +23 enhancer-reporter transgene in the *Col1a1* locus). Libraries were generated as previously described¹²². $2\text{--}5 \times 10^4$ differentiated cells were FACS-isolated, resuspended in cold lysis buffer, and incubated for 10 min on ice. Cells were centrifuged, supernatant discarded, and resuspended in 10 µl transposition mix. Samples were incubated for 30 min at 37 °C and quenched using 1.1 µl 500 mM EDTA. Reactions were centrifuged and incubated at 50 °C for 10 min. A total of 13 cycles of PCR were performed as in Buenrostro, Giresi, Zaba, Chang, and Greenleaf¹²² with transposition reaction as a template. PCR reactions were purified using MinElute PCR purification kit (Qiagen). Libraries were sequenced using Illumina NextSeq 75 cycle kit with paired-end reads. Fastq files were mapped to the mouse genome (mm9) (NGseqBasic VS2.0)¹²³. PCR duplicate-filtered bam files from individual samples (four mesoderm replicates from two experiments and two hematopoietic replicates from one experiment) were merged and filtered to remove reads mapping to chrM, ploidy regions, or the *Runx1*-Venus targeting construct (chr16:92,602,138–92,605,899, chr16:92,606,403–92,609,879), and only reads with short (<100 bp) insert sizes were retained. RPKM-normalized bigwig files were generated using deeptools (2.2.2 and 3.0.1)¹¹⁵. DNaseI-seq data in undifferentiated mESCs were downloaded from GEO (GSM1014154)⁴² and analyzed as ATAC-seq data were. Peaks were called from a single merged bam file for each sample using MACS2¹²⁴ (adjusted p -value < 0.05).

CTCF binding (ChIP-seq). CTCF ChIP was conducted using Millipore ChIP agarose kit (Millipore). 1×10^6 cross-linked 416B cells were lysed and sonicated using a Covaris ultrasonicator. Sonicated chromatin was diluted using dilution

buffer and 50 µL was removed as the 5% input control. 2 µL CTCF antibody (Supplementary Table 3) was added to 1 mL chromatin and incubated overnight at 4 °C. Decross-linking was done at 65 °C overnight. DNA was purified using phenol-chloroform-isoamylalcohol (25:24:1, Sigma) and enrichment was determined using qPCR (Supplementary Table 5). NEBNext Ultra II DNA Library Prep Kit (NEB) with 11 cycles of PCR was used to prepare sequencing libraries. CTCF ChIP libraries were sequenced using Illumina NextSeq high-output 75 cycle kit (paired-end).

CTCF ChIP-seq analysis and de novo CTCF motif annotation. CTCF ChIP-seq was performed in 416B cells and publicly available E14 mESC data⁴⁰ was downloaded from GEO (GSE28247). Fastq files were mapped to the mouse genome (mm9) (NGseqBasic VS2.0)¹²³. De novo CTCF motifs were identified in CTCF ChIP-seq data using meme (4.9.1_1)^{41,23}. CTCF peaks were called using MACS2¹²⁵ with parameters $-p$ 0.02 using input track as a control. 2000 peaks were sampled using bedtools (2.25.0)¹²¹ and flanking regions were extracted from the sampled peaks. Sequences of sampled peaks and flanking regions were retrieved and a background file was generated using fasta-get-markov -m 0. A de novo motif file was generated using meme with options $-revcomp$ $-dna$ $-nmotifs$ 1 $-w$ 20 $-maxsize$ 1000000 $-mod$ zoops. De novo motifs were identified in CTCF peaks using fimo with options $-motif$ 1 $-thresh$ 1e-3.

Targeted bisulfite sequencing. DNA methylation analysis was performed as previously described¹²⁶. Genomic DNA (gDNA) was extracted from 1 to 5×10^6 cells using DNeasy Blood and Tissue Kit (Qiagen) and 250 ng gDNA, or Universal Methylated Mouse DNA Standard (Zymo Research) was bisulfite converted using EZ DNA Methylation-Gold™ Kit (Zymo Research). Nested PCR primer sets (Supplementary Table 5) were designed to amplify 281–379 bp overlapping target regions. External PCR reactions were performed on 1 µL bisulfite converted DNA using Hot-StarTaq DNA Polymerase (Qiagen). Internal nested PCR reactions were performed using 1 µL of the external PCR reaction. Amplicons were size selected by gel and purified. 250 ng equimolar PCR amplicons were combined for each biological sample and indexed using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) with 6 PCR cycles. Quality of reads was assessed using fastqc/0.10.1. Reads were quality and adapter trimmed using trim galore/0.3.1 (<https://github.com/FelixKrueger/TrimGalore>). Reads were mapped to an in silico bisulfite converted genome using bismark/0.20.0¹²⁷. Percentages of methylated CpG dinucleotides were determined using bismark methylation extractor. Bedgraph output files were filtered on CpG dinucleotides with coverage greater than 100 reads and imported into R. Average methylated CpG dinucleotide percentages were plotted over each region using ggplot2.

Statistics. All statistical tests were performed in R and were two-tailed. Tiled-C contact data were non-normal (Shapiro–Wilks test, $p < 2 \times 10^{-16}$) and so non-parametric two-sided Kruskal–Wallis test with Dunn's post hoc comparisons test was applied. Post hoc testing with Dunn's test was applied when Kruskal–Wallis test was significant and p -values were adjusted using the Holm method or the Benjamini–Hochberg method. A significance threshold of $p < 0.05$ was used for all statistical tests.

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

Data availability

The data that support this study are available from the corresponding authors upon reasonable request. The sequencing data generated in this study have been deposited in the GEO database under accession code GSE184490. Processed Tiled-C matrices are available on github (https://github.com/d0minicO/Owens_et_al_Tiled-C). The E14 mESC CTCF ChIP-seq data used in this study are available in the GEO database under accession code GSE28247. The 416B H3K27ac ChIP-seq data used in this study are available in the GEO database under accession code GSE69776. The 416B and E14 mESC DNaseI-seq data used in this study are available in the GEO database under accession code GSE37074. Source data are provided with this paper.

Code availability

Code is freely available on github at https://github.com/d0minicO/Owens_et_al_Tiled-C and Zenodo at <https://doi.org/10.5281/zenodo.5781832>¹¹³.

Received: 8 June 2021; Accepted: 12 January 2022;

Published online: 09 February 2022

References

- de Bruijn, M. & Dzierzak, E. Runx transcription factors in the development and function of the definitive hematopoietic system. *Blood* **129**, 2061–2069 (2017).

2. Levanon, D. & Groner, Y. Structure and regulated expression of mammalian RUNX genes. *Oncogene* **23**, 4211–4219 (2004).
3. Gao, L. et al. RUNX1 and the endothelial origin of blood. *Exp. Hematol.* **68**, 2–9 (2018).
4. Yzaguirre, A. D., de Bruijn, M. F. & Speck, N. A. The role of Runx1 in embryonic blood cell formation. *Adv. Exp. Med. Biol.* **962**, 47–64 (2017).
5. Sood, R., Kamikubo, Y. & Liu, P. Role of RUNX1 in hematological malignancies. *Blood* **129**, 2070–2082 (2017).
6. Bellissimo, D. C. & Speck, N. A. RUNX1 mutations in inherited and sporadic leukemia. *Front. Cell Develop. Biol.* **5**, 111 (2017).
7. Levanon, D. et al. Architecture and anatomy of the genomic locus encoding the human leukemia-associated transcription factor RUNX1/AML1. *Gene* **262**, 23–33 (2001).
8. Cai, Z. et al. Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. *Immunity* **13**, 423–431 (2000).
9. Wang, Q. et al. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc. Natl Acad. Sci. USA* **93**, 3444–3449 (1996).
10. Lie, A. L. M. et al. Regulation of RUNX1 dosage is crucial for efficient blood formation from hemogenic endothelium. *Development* **145**, dev149419 (2018).
11. Song, W. J. et al. Haploinsufficiency of Cbfa2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat. Genet.* **23**, 166–175 (1999).
12. Telfer, J. C. & Rothenberg, E. V. Expression and function of a stem cell promoter for the murine Cbfa2 gene: distinct roles and regulation in natural killer and T cell development. *Developmental Biol.* **229**, 363–382 (2001).
13. Ghози, M. C., Bernstein, Y., Negreanu, V., Levanon, D. & Groner, Y. Expression of the human acute myeloid leukemia gene AML1 is regulated by two promoter regions. *Proc. Natl Acad. Sci. USA* **93**, 1935–1940 (1996).
14. North, T. et al. Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development* **126**, 2563–2575 (1999).
15. Bee, T. et al. Alternative Runx1 promoter usage in mouse developmental hematopoiesis. *Blood Cells Mol. Dis.* **43**, 35–42 (2009).
16. Bee, T. et al. Nonredundant roles for Runx1 alternative promoters reflect their activity at discrete stages of developmental hematopoiesis. *Blood* **115**, 3042–3050 (2010).
17. Sroczyńska, P., Lancrin, C., Kouskoff, V. & Lacaud, G. The differential activities of Runx1 promoters define milestones during embryonic hematopoiesis. *Blood* **114**, 5279–5289 (2009).
18. Nottingham, W. T. et al. Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. *Blood* **110**, 4188–4197 (2007).
19. Ng, C. E. et al. A Runx1 intronic enhancer marks hemogenic endothelial cells and hematopoietic stem cells. *Stem Cells* **28**, 1869–1881 (2010).
20. Schutte, J. et al. An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. *eLife* **5**, e11469 (2016).
21. Marsman, J., Thomas, A., Osato, M., O'Sullivan, J. M. & Horsfield, J. A. A DNA contact map for the mouse Runx1 gene identifies novel haematopoietic enhancers. *Sci. Rep.* **7**, 13347 (2017).
22. Mill, C. P. et al. RUNX1-targeted therapy for AML expressing somatic or germline mutation in RUNX1. *Blood* **134**, 59–73 (2019).
23. Hanssen, L. L. P. et al. Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat. Cell Biol.* **19**, 952–961 (2017).
24. Lettice, L. A. et al. Enhancer-adoption as a mechanism of human developmental disease. *Hum. Mutat.* **32**, 1492–1499 (2011).
25. Symmons, O. et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
26. Lupianez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
27. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
28. Oudelaar, A. M. & Higgs, D. R. The relationship between genome structure and function. *Nat. Rev. Genet.* **22**, 154–168 (2021).
29. Ovcharenko, I. et al. Evolution and functional classification of vertebrate gene deserts. *Genome Res.* **15**, 137–145 (2005).
30. Lettice, L. A. et al. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc. Natl Acad. Sci. USA* **99**, 7548–7553 (2002).
31. Braccioli, L. & de Wit, E. CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays Biochem.* **63**, 157–165 (2019).
32. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 e922 (2017).
33. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 e324 (2017).
34. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
35. Horsfield, J. A. et al. Cohesin-dependent regulation of Runx genes. *Development* **134**, 2639–2649 (2007).
36. Marsman, J. et al. Cohesin and CTCF differentially regulate spatiotemporal runx1 expression during zebrafish development. *Biochimica et. biophysica acta* **1839**, 50–61 (2014).
37. Mazzola, M. et al. Dysregulation of NIPBL leads to impaired RUNX1 expression and haematopoietic defects. *J. Cell. Mol. Med.* **24**, 6272–6282 (2020).
38. Oudelaar, A. M. et al. Dynamics of the 4D genome during in vivo lineage specification and differentiation. *Nat. Commun.* **11**, 2722 (2020).
39. Sroczyńska, P., Lancrin, C., Pearson, S., Kouskoff, V. & Lacaud, G. In vitro differentiation of mouse embryonic stem cells as a model of early hematopoietic development. *Methods Mol. Biol.* **538**, 317–334 (2009).
40. Handoko, L. et al. CTCF-mediated functional chromatin interactions in pluripotent cells. *Nat. Genet.* **43**, 630–638 (2011).
41. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
42. Vierstra, J. et al. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **346**, 1007–1012 (2014).
43. Ortt, K., Raveh, E., Gat, U. & Sinha, S. A chromatin immunoprecipitation screen in mouse keratinocytes reveals Runx1 as a direct transcriptional target of DeltaNp63. *J. Cell. Biochem.* **104**, 1204–1219 (2008).
44. Fitch, S. R. et al. Gata3 targets Runx1 in the embryonic haematopoietic stem cell niche. *IUBMB Life* **72**, 45–52 (2020).
45. Cauchy, P. et al. Chronic FLT3-ITD signaling in acute myeloid leukemia is connected to a specific chromatin signature. *Cell Rep.* **12**, 821–836 (2015).
46. Cheng, C. K. et al. RUNX1 upregulation via disruption of long-range transcriptional control by a novel t(5;21)(q13;q22) translocation in acute myeloid leukemia. *Mol. Cancer* **17**, 133 (2018).
47. Harland, L. T. G. et al. The T-box transcription factor Eomesodermin governs haemogenic competence of yolk sac mesodermal progenitors. *Nat. Cell Biol.* **23**, 61–74 (2021).
48. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
49. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482–485 (2000).
50. Shukla, S. et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011).
51. Wang, H. et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–1688 (2012).
52. Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).
53. Hashimoto, H. et al. Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. *Mol. Cell* **66**, 711–720 e713 (2017).
54. Xu, C. & Corces, V. G. Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. *Science* **359**, 1166–1170 (2018).
55. Schuijers, J. et al. Transcriptional dysregulation of MYC reveals common enhancer-docking mechanism. *Cell Rep.* **23**, 349–360 (2018).
56. Canzio, D. & Maniatis, T. The generation of a protocadherin cell-surface recognition code for neural circuit assembly. *Curr. Opin. Neurobiol.* **59**, 213–220 (2019).
57. Curradi, M., Izzo, A., Badaracco, G. & Landsberger, N. Molecular mechanisms of gene silencing mediated by DNA methylation. *Mol. Cell Biol.* **22**, 3157–3173 (2002).
58. Schwessinger, R. et al. DeepC: predicting 3D genome folding using megabase-scale transfer learning. *Nat. Methods* **17**, 1118–1124 (2020).
59. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 e524 (2017).
60. Matys, V. et al. TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108–D110 (2006).
61. Harman, J. R. et al. A KMT2A-AFF1 gene regulatory network highlights the role of core transcription factors and reveals the regulatory logic of key downstream target genes. *Genome Res.* **35**, 1159–1173 (2021).
62. Iacovino, M. et al. HoxA3 is an apical regulator of haemogenic endothelium. *Nat. Cell Biol.* **13**, 72–78 (2011).
63. Ben-Ami, O. et al. Addiction of t(8;21) and inv(16) acute myeloid leukemia to native RUNX1. *Cell Rep.* **4**, 1131–1143 (2013).
64. Antony, J. et al. BET inhibition prevents aberrant RUNX1 and ERG transcription in STAG2 mutant leukaemia cells. *J. Mol. Cell Biol.* **12**, 397–399 (2020).
65. Thomas, A. L. et al. Transcriptional regulation of RUNX1: an informatics analysis. *Genes* **12**, 1175 (2021).

66. Wilson, N. K. et al. Integrated genome-scale analysis of the transcriptional regulatory landscape in a blood stem/progenitor cell model. *Blood* **127**, e12–e23 (2016).
67. Chen, C. et al. Spatial genome re-organization between fetal and adult hematopoietic stem cells. *Cell Rep.* **29**, 4200–4211 e4207 (2019).
68. Brown, J. M. et al. A tissue-specific self-interacting chromatin domain forms independently of enhancer-promoter interactions. *Nat. Commun.* **9**, 3849 (2018).
69. 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 e219 (2017).
70. Paliou, C. et al. Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proc. Natl Acad. Sci. USA* **116**, 12390–12399 (2019).
71. Nasmyth, K. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu. Rev. Genet.* **35**, 673–745 (2001).
72. Alipour, E. & Marko, J. F. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.* **40**, 11202–11212 (2012).
73. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, E6456–E6465 (2015).
74. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
75. de Wit, E. et al. CTCF Binding Polarity Determines Chromatin Looping. *Mol. Cell.* **60** 676–684 (2015).
76. Rao, S. S. P. et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **162**, 687–688 (2015).
77. Vietri Rudan, M. et al. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Rep.* **10**, 1297–1309 (2015).
78. Hsieh, T. S. et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* **78**, 539–553 e538 (2020).
79. Hua, P. et al. Defining genome architecture at base-pair resolution. *Nature* **595**, 125–129 (2021).
80. Hay, D. et al. Genetic dissection of the alpha-globin super-enhancer in vivo. *Nat. Genet.* **48**, 895–903 (2016).
81. Ing-Simmons, E. et al. Independence of chromatin conformation and gene regulation during *Drosophila* dorsoventral patterning. *Nat. Genet.* **53**, 487–499 (2021).
82. Zhang, C. et al. tagHi-C reveals 3D chromatin architecture dynamics during mouse hematopoiesis. *Cell Rep.* **32**, 108206 (2020).
83. Ruthenburg, A. J., Li, H., Patel, D. J. & Allis, C. D. Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **8**, 983–994 (2007).
84. Cho, S. W. et al. Promoter of lncRNA gene PVT1 is a tumor-suppressor dna boundary element. *Cell* **173**, 1398–1412 e1322 (2018).
85. Harrold, C. L. et al. A functional overlap between actively transcribed genes and chromatin boundary elements. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.07.01.182089> (2020).
86. Bozhilov, Y. K. et al. A gain-of-function single nucleotide variant creates a new promoter which acts as an orientation-dependent enhancer-blocker. *Nat. Commun.* **12**, 3806 (2021).
87. Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).
88. Ren, G. et al. CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. *Mol. Cell* **67**, 1049–1058 e1046 (2017).
89. Lee, J., Krivega, I., Dale, R. K. & Dean, A. The LDB1 complex Co-opts CTCF for erythroid lineage-specific long-range enhancer interactions. *Cell Rep.* **19**, 2490–2502 (2017).
90. Canzio, D. et al. Antisense lncRNA transcription mediates DNA demethylation to drive stochastic protocadherin alpha promoter choice. *Cell* **177**, 639–653 e615 (2019).
91. Kubo, N. et al. Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat. Struct. Mol. Biol.* **28**, 152–161 (2021).
92. Zhou, Q. et al. ZNF143 mediates CTCF-bound promoter-enhancer loops required for murine hematopoietic stem and progenitor cell function. *Nat. Commun.* **12**, 43 (2021).
93. Swiers, G. et al. Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat. Commun.* **4**, 2924 (2013).
94. Kentepozidou, E. et al. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol.* **21**, 5 (2020).
95. Lacaud, G., Kouskoff, V., Trumble, A., Schwantz, S. & Keller, G. Haploinsufficiency of Runx1 results in the acceleration of mesodermal development and hemangioblast specification upon in vitro differentiation of ES cells. *Blood* **103**, 886–889 (2004).
96. Mevel, R., Draper, J. E., Lie, A. L. M., Kouskoff, V. & Lacaud, G. RUNX transcription factors: orchestrators of development. *Development* **146**, dev148296 (2019).
97. Handyside, A. H., O'Neill, G. T., Jones, M. & Hooper, M. L. Use of BRL-conditioned medium in combination with feeder layers to isolate a diploid embryonal stem cell line. *Roux's Arch. Develop. Biol.* **198**, 48–56 (1989).
98. Dexter, T. M., Allen, T. D., Scott, D. & Teich, N. M. Isolation and characterisation of a bipotential haematopoietic cell line. *Nature* **277**, 471–474 (1979).
99. Pearson, S., Cuvertino, S., Fleury, M., Lacaud, G. & Kouskoff, V. In vivo repopulating activity emerges at the onset of hematopoietic specification during embryonic stem cell differentiation. *Stem Cell Rep.* **4**, 431–444 (2015).
100. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
101. Gruzdev, A., Scott, G. J., Hagler, T. B. & Ray, M. K. CRISPR/Cas9-assisted genome editing in murine embryonic stem cells. *Methods Mol. Biol.* **1960**, 1–21 (2019).
102. Owens, D. D. G. et al. Microhomologies are prevalent at Cas9-induced larger deletions. *Nucleic Acids Res.* **47**, 7402–7417 (2019).
103. Teboul, L., Herault, Y., Wells, S., Qasim, W. & Pavlovic, G. Variability in genome editing outcomes: challenges for research reproducibility and clinical safety. *Mol. Ther.* **28**, 1422–1431 (2020).
104. Mianne, J. et al. Analysing the outcome of CRISPR-aided genome editing in embryos: Screening, genotyping and quality control. *Methods* **121–122**, 68–76 (2017).
105. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
106. Codner, G. F. et al. Aneuploidy screening of embryonic stem cell clones by metaphase karyotyping and droplet digital polymerase chain reaction. *BMC Cell Biol.* **17**, 30 (2016).
107. Oudelaar, A. M., Downes, D. J., Davies, J. O. J. & Hughes, J. R. Low-input capture-C: a chromosome conformation capture assay to analyze chromatin architecture in small numbers of cells. *Bio-Protocol* **7**, e2645 (2017).
108. 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 Res.* **45**, e184 (2017).
109. Telenius, J. M. et al. CaptureCompendium: a comprehensive toolkit for 3C analysis. Preprint at *bioRxiv*. <https://doi.org/10.1101/2020.02.17.952572> (2020).
110. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
111. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
112. 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).
113. Owens, D. D. G. dOminicO/Owens_et_al_Tiled-C: Owens_et_al_Runx1_Tiled-C_scripts_v1. <https://doi.org/10.5281/zenodo.5781832> (2021).
114. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
115. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
116. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
117. Blighe, K. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.2.0. <https://github.com/kevinblighe/EnhancedVolcano>. (2019).
118. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
119. Carlson, M., Vol. R package version 3.2.3 (2016).
120. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).
121. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
122. 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. Methods* **10**, 1213–1218 (2013).
123. 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. Preprint at *bioRxiv*. <https://doi.org/10.1101/393413> (2018).
124. Liu, T. Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. *Methods Mol. Biol.* **1150**, 81–95 (2014).

125. Zhang, Y. et al. Model-based analysis of CHIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
126. Jeziorska, D. M. et al. DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease. *Proc. Natl Acad. Sci. USA* **114**, E7526–E7535 (2017).
127. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
128. Miyoshi, H. et al. Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia. *Nucleic Acids Res.* **23**, 2762–2769 (1995).
129. Bee, T. et al. The mouse Runx1 +23 hematopoietic stem cell enhancer confers hematopoietic specificity to both Runx1 promoters. *Blood* **113**, 5121–5124 (2009).

Acknowledgements

We thank Christina Rode, Emanuele Azzoni, Vincent Frontera, Ruth Williams, and Tatjana Sauka-Spengler for helpful discussions and advice. We thank Nick Crump and Tom Milne for expert technical assistance and advice. We thank Kevin Clark, Sally Clark, and Paul Sopp from the WIMM FACS facility for assistance with cell sorting and technical expertise. This work was supported by programs in the MRC Molecular Hematology Unit Core award to M.F.T.R.d.B. (MC_UU_12009/2) and J.R.Hu (MC_UU_00016/14), MRC MR/N00969X/1 and Wellcome Trust (106130/Z/14/Z) to J.R.Hu, Wellcome Trust Doctoral Programmes supported A.C. (108870/Z/15/Z), R.S. (203728/Z/16/Z), and A.M.O. (105281/Z/14/Z), who was also supported by the Stevenson Junior Research Fellowship (University College, Oxford), an EPSRC studentship supported J.R.Ha, a Clarendon Scholarship L.G., and S.d.O. was supported by an MRC Project Award (MR/N00969X/1) to J.R.Hu. The WIMM Flow Cytometry facility is supported by the MRC HIU, MRC MHU [MC_UU_12009], NIHR Oxford BRC and John Fell Fund [131/030 and 101/517], EPA fund [CF182 and CF170], WIMM Strategic Alliance awards [G0902418 and MC_UU_12025]. The Wolfson Imaging Centre Oxford is supported by the Medical Research Council via the WIMM Strategic Alliance (G0902418), the Molecular Hematology Unit (MC_UU_12009), the Human Immunology Unit (MC_UU_12010), the Wolfson Foundation (Grant 18272), and by an MRC/BBSRC/EPSC grant (MR/K015777X/1) to MICA—NanoscopyOxford (NanO): Novel Super-resolution Imaging Applied to Biomedical Sciences, Micron (107457/Z/15Z).

Author contributions

D.D.G.O., A.M.O., D.J.D., J.R.Hu. and M.F.T.R.d.B. designed and conceptualized the study. D.D.G.O., G.A., A.M.O., D.J.D., A.C., J.R.Ha. and A.B. performed experiments.

D.D.G.O., A.M.O., J.R.Ha., R.S. and D.J. performed data analysis and interpretation. J.T. generated scripts used in bioinformatics analysis. L.G., S.d.O. and D.J. generated vital reagents. D.D.G.O., G.A., A.M.O., D.J.D., A.C., J.R.Hu. and M.F.T.R.d.B. wrote the paper. All authors approved of the final version to be published.

Competing interests

J.R.Hu and D.J. are founders and shareholders of, and D.J.D. and R.S. are paid consultants for Nucleome Therapeutics. Other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-28376-8>.

Correspondence and requests for materials should be addressed to Jim R. Hughes or Marella F. T. R.de Bruijn.

Peer review information *Nature Communications* thanks Hui Cheng, Julia Horsfield and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022