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

Dominic D. G. Owens1,4, Giorgio Anselmi1, A. Marieke Oudelaar1,5, Damien J. Downes1, Alessandro Cavallo1, Joe R. Harman1, Ron Schwessinger1,2, Akin Bucakci1, Lucas Greder1, Sara de Ornellas1,3, Danuta Jeziorska1, Jelena Telenius2, Jim R. Hughes1,2✉ & Marella F. T. R. de Bruijn1✉

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.
Runx1/AML1 is a member of the RUNX family of transcription factors (TFs), which are key to many developmental processes\textsuperscript{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\textsuperscript{1,2,4}. Disruption of RUNX1 in humans leads to several hematopoietic disorders, including acute myeloid leukemia\textsuperscript{5} and familial platelet disorder with associated myeloid malignancy (FPD-AML)\textsuperscript{6}. All members of the RUNX family bind the same canonical DNA motif (GGYGGT) and their tissue-specific functions are thought to be governed largely by their specific expression patterns\textsuperscript{7}. Runx1 transcription is tightly regulated, with changes in gene dosage and expression level affecting both the spatiotemporal onset of hematopoiesis and hematopoietic homeostasis\textsuperscript{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\textsuperscript{12,13} (reviewed in de Bruijn and Dzierzak\textsuperscript{14}). 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\textsuperscript{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\textsuperscript{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 enhancer\textsuperscript{22}.

The 3D conformation of DNA in structures such as topologically associating domains (TADs) delimit the activities of enhancers in vivo\textsuperscript{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\textsuperscript{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 deserts\textsuperscript{29}. Indeed, studies of developmental regulators, such as Shh, have revealed exceptionally long-range enhancer–promoter interactions\textsuperscript{30}, 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)\textsuperscript{31}. CTCF, along with the loop extruding factor cohesin, mediates the establishment and maintenance of both E–P interactions and TAD structure\textsuperscript{32–34}. Interestingly, Runx1 was mis-regulated in zebrafish after perturbation of CTCF/cohesin\textsuperscript{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)\textsuperscript{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. \textsuperscript{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\textsuperscript{+} 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\textsuperscript{+} mesoderm, and emerging CD41\textsuperscript{+} CD45\textsuperscript{–} Runx1\textsuperscript{+} HPCs reflected the developmental trajectory as visualized in a Principal Component Analysis (PCA) plot (Fig. 1d). This was accompanied by silencing of pluripotency genes (Pof51, Sox2, Nanog), transient expression of mesodermal genes (Flk1, Eomes, T), and increasing levels of hematopoiesis-associated genes (Pecam1, Tall, Gfi1b, Meis1, Itga2b) (Fig. 1e). Runx1 expression was initiated in Flk1\textsuperscript{+} mesoderm and increased in HPCs (Fig. 1e\textsuperscript{17}).

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 loci\textsuperscript{38}, 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 mESCs\textsuperscript{40}, 31 binding sites were identified within the Runx1 TAD (Fig. 1g). MEME analysis\textsuperscript{41} 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 Clic6 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
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 × 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-P 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 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 × 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 × 10−43 and P2-3’ sub-TAD p = 6.6 × 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 <2kb upstream of the transcription start site (TSS; Fig. 3a, d), and CTCF is associated with sub-TAD and TAD boundaries31,48, 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 binding49–56, 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
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...
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).
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 to 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-
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, upper panel). Indeed, loss of P2-CTCF led to the region between Runx1 and Cic6 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.

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, the including 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

Runx1 P2 promoter-proximal CTCF site coordinates spatiotemporal 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 $p < 0.05$). 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 “gastulation” (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.
23 enhancer. As both leukemias with and without somatic or germline RUNX1 mutations were shown to depend on wild-type RUNX1 (reviewed in ref. 5) it is plausible that a similar dependency on the +23 enhancer may exist in other leukemic cells. 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 pathogenic. 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 as those lacked the required high-resolution, nor included a developmental time series.

**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 RNA-seq PCA of all poly(A)-minus RNA-seq replicates. d PCA of mesoderm RNA-seq samples. e PCA of mesoderm RNA-seq samples. 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 −log10 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).
In line with TADs at other developmentally regulated loci\textsuperscript{38,68–70}, the overall 1.1 Mb \textit{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\textsuperscript{71–74} as a predominant convergence of CTCF motifs was observed at the main \textit{Runx1} TAD boundaries, similar to what was found at other TADs\textsuperscript{75–77}. In addition to these preformed chromatin structures, increasing CTCF-CTCF interactions were observed upon \textit{Runx1} activation that might reflect a higher rate or processivity of loop extrusion in the \textit{Runx1} TAD, as was recently also observed for \textit{α-globin}\textsuperscript{38}. Alternatively, increased TF binding could be driving the specific increases in chromatin interactions that were observed over differentiation\textsuperscript{78,79}. Two tissue-specific sub-TADs formed over differentiation, leading to a sub-compartmentalization of the \textit{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 strengthening\textsuperscript{38}, 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 \textit{α-globin}, which reside entirely within one tissue-specific sub-TAD\textsuperscript{25,80}. This difference might simply reflect the smaller size of the \textit{α-globin} domain compared to the larger size of the \textit{Runx1} regulatory domain (97 kb compared to 1.1 Mb, respectively). The sub-TAD encompassing the entire \textit{α-globin} regulatory unit was shown to represent a discrete functional unit that delimits enhancer activity\textsuperscript{23}. We did not see evidence for this in \textit{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 place\textsuperscript{81}. Promoter-proximal CTCF site deletion facilitated new interactions with CTCF sites distal to the deleted CTCF sites, leading to expanded loop domains at \textit{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 \textit{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 \textit{Runx1} tissue-specific sub-TADs were strongest in HPCs, which express high levels of \textit{Runx1}, suggests they may be similar to the gene-body-associated domains (GADs) recently observed at genes highly expressed in hematopoietic cells\textsuperscript{82}. 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 processes\textsuperscript{82}, TF binding\textsuperscript{78,79}, and clustering of like histone modifications\textsuperscript{83}. Indeed, our data indicate that, as reported for other promoters\textsuperscript{38,84–86}, the \textit{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, interestingly, promoter-proximal CTCF sites have been suggested to play a role in E-P interactions and gene transcription. 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 commitment. 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), 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 P2. Finally, compared to in vitro differentiation, promoter-proximal CTCF-site loss may be more detrimental in vivo, where Runx1 levels are subject to spatiotemporal control and changes to Runx1 levels or dosage lead to knock-on effects on differentiation timing. 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, 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 mESCs were cultured in GMEM medium supplemented with 100 mM non-essential amino acids, 100 mM sodium pyruvate, 10% FCS, 2 mM L-glutamine, 100 mM 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 line was cultured at 2–8 x 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 by supplementing 40X defined serum replacement, 2 mM L-glutamine (Gibco), and 0.5 mM ascorbic acid (Sigma) with 45 mM monothioglycerol (Sigma). mESCs were seeded at a density of 5 x 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 Fli1+ mesodermal cells were cultured at 5 x 10^6 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(2BB)-2A-Puro-V2.0 containing two sgRNAs and confirmed by sequencing. mESCs were transfected with 5 μg plasmid using Lipofectamine 2000 (Invitrogen) and puromycin selected (1 μg/ml). Single colonies were achieved by limiting dilution. To detect larger on-target deletions, 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 genotypes, copy counting was performed across the targeted regions by droplet digital PCR (ddPCR). Reactions were performed in duplex, amplifying from an internal control region in every reaction. Internal control was located on mouse chromosome 4 (not targeted in these experiments and karyotypically stable in mESCs). 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). Reactions (22 μl) contained 11 μl Q2000 TaqMan Green Supermix (BioRad), 25–50 ng genomic DNA, 200 nM each of primers (Supplementary Table 2), 20 μM each of dNTPs, 150 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 QuantSeq 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 x 10^4 and 1 x 10^6 cells using a low-input protocol (Supplementary Table 5). 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 μg 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 (chr16:91,566,000–91,100,999). Probe sequences were stringently BLAT-filtered to exclude repetitive sequences, and synthesized in-house. A pooled capture reaction was performed on 1 μg of each indexed 3C library. Washing of captured material was done using NimbleSeq Cap 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-C capture-C data was processed at 2 kb resolution. Fastqs were analyzed using the CCSeqBasic CM5 pipeline. 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_C_Sam2RawMatrix.pl, https://github.com/TreeStar/Tiled-C). ICE normalization was done using Hic-Pro (2.111) and matrices were imported into R (3.6.0). Matrices were plotted (Tiled-C_matrix_visualisation.py) and plotted with a threshold between the 90th and 95th percentile. PCA was done on log normalized counts (Deseq2 1.30.1). Merged ICE normalized contact matrices were scaled to the mean number of total interactions (14,631,865) across samples using custom R scripts. Virtual Capture-C plots were generated by back-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. 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...
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. Cross-linking was done for 25 min. Fastq files were mapped to the mouse genome (mm9) (NGseqBasic VS2.0)\textsuperscript{125}. De novo CTCF motifs were identified in CTCF ChIP-seq data using meme (4.9.1)\textsuperscript{126}. CTCF peaks were called using MACS2 with parameters -p 0.02 using input track as a control. 2000 peaks were sampled using bedtools (2.25.0)\textsuperscript{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-merkov -m 0. A de novo motif file was generated using meme with options -revcomp -dna -nmotifs 1 -w 20 -maxsize 10000000 -mod zoos. De novo motifs were identified in CTCF peaks using fimo with options -motif 1 -thres 1e-3.

Targeted bisulfite sequencing. DNA methylation analysis was performed as previously described\textsuperscript{126}. Genomic DNA (gDNA) was extracted from 1 to 5 × 10^6 cells using DNAeasy Blood and Tissue Kit (Qiagen) and 250 ng gDNA, or Universal Methylated Mouse DNA Standard (Zymo Research) was bisulfite converted using EZ DNA Methylation-GoldTM 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 HotStarTaq 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/TrimmGalore). Reads were mapped to an in silico bisulfite converted genome using bismark/0.20.0\textsuperscript{127}. Percentages of methylated CpG dinucleotides were determined using bismark methylation caller/0.20.0\textsuperscript{123}. The methylation data were mappable. BED files were generated using bedtools/2.25.0\textsuperscript{121} and filtered bam files were mapped to the mouse genome (mm9) (NGseqBasic VS2.0)\textsuperscript{123}. De novo CTCF motifs were identified in CTCF ChIP-seq data using meme (4.9.1)\textsuperscript{126}. CTCF peaks were called using MACS2 with parameters -p 0.02 using input track as a control. 2000 peaks were sampled using bedtools (2.25.0)\textsuperscript{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-merkov -m 0. A de novo motif file was generated using meme with options -revcomp -dna -nmotifs 1 -w 20 -maxsize 10000000 -mod zoos. De novo motifs were identified in CTCF peaks using fimo with options -motif 1 -thres 1e-3.

CTCF ChIP-seq analysis and de novo CTCF motif annotation. CTCF ChIP-seq was performed in 416B cells and publicly available E14 mESC data\textsuperscript{40} was downloaded from GEO (GSE82842). Fastq files were mapped to the mouse genome (mm9) (NGseqBasic VS2.0)\textsuperscript{125}. De novo CTCF motifs were identified in CTCF ChIP-seq data using meme (4.9.1)\textsuperscript{126}. CTCF peaks were called using MACS2 with parameters -p 0.02 using input track as a control. 2000 peaks were sampled using bedtools (2.25.0)\textsuperscript{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-merkov -m 0. A de novo motif file was generated using meme with options -revcomp -dna -nmotifs 1 -w 20 -maxsize 10000000 -mod zoos. De novo motifs were identified in CTCF peaks using fimo with options -motif 1 -thres 1e-3.

**Statistical analysis.** All statistical tests were performed in R and were two-tailed. Tiled-C contact data were non-normal (Shapiro–Wilk test, p < 2 × 10\textsuperscript{-10}) 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 GSE114490. Processed Tiled-C matrices are available on github (https://github.com/dlitominic/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 GSE98247. The 416B HEK293c ChIP-seq data used in this study are available in the GEO database under accession code GSE69776. The 416B and E14 mESC Dnamaseq 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/dlitominic/Owens_et_al_Tiled-C and Zenodo at https://doi.org/10.5281/zenodo.5781832\textsuperscript{113}. 

Received: 8 June 2021; Accepted: 12 January 2022; 
Published online: 09 February 2022

References

1. de Brujin, M. & Dräzerk, 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. Guo, L. et al. RUNX1 and the endothelial origin of blood. Exp. Hematol. 68, 2–9 (2018).

4. Yazguirre, A. D., de Brujin, 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 hematopoiesis. Blood 129, 2070–2082 (2017).

6. Bellissimo, D. C. & Speck, N. A. RUNX1 mutations in inherited and sporadic leukemia. Front. Cell Dev. 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 (2000).

8. Caiz, 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 Cbfα2 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, dev14919 (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. & Speck, N. A. The role of Runx1 in murine CBFα2 gene: distinct roles and regulation in normal killer and T cell development. Developmental Biol. 229, 363–382 (2001).

13. Ghozi, 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 in Proc. Natl. Acad. Sci. USA 93, 1935–1940 (1996).

14. North, T. et al. Cbfα2 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. Soczyńska, P., Cancrin, 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. Schütte, J. et al. An experimentally validated network of nine hematopoietic transcription factors reveals mechanisms of cell state stabilization. eLife 5, e11469 (2016).

21. Marsman, J., Thomas, A., Osato, M., O’Reilly, S. & Speck, N. A. A chromatin immunoprecipitation screen in mouse keratinocytes reveals Runx1 as a direct transcriptional target of DeltaNp63. J. Cell. Biochem. 104, 1204–1219 (2008).

22. 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).

23. Vierstra, J. et al. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. Science 346, 1007–1012 (2014).

24. Cheng, C. K. et al. RUNX1 upregulation via disruption of long-range transcriptional control by a novel t(5;21)(q31;q22) translocation in acute myeloid leukemia. Mol. Cancer 17, 133 (2018).

25. Harland, L. T. G. et al. The T-box transcription factor Eomesoderm governs haemogenic competence of yolk sac mesodermal progenitors. Nat. Cell Biol. 23, 61–74 (2021).

26. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-bound transcriptional boundary controls imprinted expression of the Igf2 gene. Nature 405, 482–485 (2000).

27. Cheng, C. K. et al. RUNX1 upregulation via disruption of long-range transcriptional control by a novel t(5;21)(q31;q22) translocation in acute myeloid leukemia. Mol. Cancer 17, 133 (2018).

28. Fox, N. F. et al. The t(15;17) translocation in acute promyelocytic leukemia defines a genetic basis for specific transcriptional deregulation. Cell 85, 833–842 (1996).

29. Lie, A. L. M. et al. Regulation of RUNX1 dosage is crucial for efficient blood formation from hemogenic endothelium. Development 145, dev14919 (2018).

30. Wang, H. et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res. 22, 1680–1688 (2012).

31. Flavahan, W. A. et al. Insulator dysfunction and oncogenic activation in IDH mutant gliomas. Nature 529, 110–114 (2016).

32. Harimoto, H. et al. Structural dissection of the versatile and methylation dependent binding of CTCF to DNA. Mol. Cell 66, 711–720 e713 (2017).

33. Xu, C. & Corces, V. G. Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. Science 359, 1166–1170 (2017).

34. Engel, J. H. et al. Transcriptional dysregulation of MYC reveals common enhancer-docking mechanism. Cell Rep. 23, 349–360 (2018).

35. Canzio, D. & Maniatis, T. The generation of a protocadherin cell-surface recognition code for neural circuit assembly. Curr. Opin. Neurobiol. 59, 213–220 (2019).

36. Kurradi, M., Izzo, A., Badaracco, G. & Landesberger, N. Molecular mechanisms of gene silencing mediated by DNA methylation. Mol. Cell Biol. 22, 3173–3177 (2002).

37. Schwessinger, R. et al. DeepC: predicting 3D genome folding using megabase-scale transfer learning. Nat. Methods 17, 1118–1124 (2020).

38. Vierstra, J. et al. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. Science 346, 1007–1012 (2014).

39. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. Cell 171, 1012–1025 (2017).

40. McLachlan, J. & Maniatis, T. The generation of a protocadherin cell-surface recognition code for neural circuit assembly. Curr. Opin. Neurobiol. 59, 213–220 (2019).

41. Iacovino, M. et al. HoxA1 is an apical regulator of haemogenic endothelium. Nat. Cell Biol. 13, 72–78 (2011).

42. Ben-Ami, O. et al. Addition of 18(21) and inv(16) acute myeloid leukemia to native RUNX1. Cell Rep. 4, 1131–1143 (2013).

43. Antony, J. et al. BET inhibition prevents aberrant RUNX1 and ERG transcription in STAG2 mutant leukemia cells. J. Mol. Cell Biol. 12, 397–399 (2020).

44. 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 Shh during limb development. *Proc. Natl Acad. Sci. USA* **116**, 2180–2189 (2019).
70. Haploinsuf ciency of Runx1 results in the acceleration of mesodermal development and hemangioblast specification upon in vitro differentiation of ES cells. *Blood* **103**, 868–889 (2004).
71. Mevel, R., Draper, J. E., Lie, A. L. M., Kouskoff, V. & Lacaud, G. RUNX transcription factors: orchestrators of development. *Development* **146**, 1482–148296 (2019).
72. Handside, 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. Dev. Biol.* **198**, 845–85 (1989).
73. Stead, T. M., Allen, T. D., Scott, D. & Teich, N. M. Isolation and characterisation of a bipotential haematopoietic cell line. *Nature* **277**, 471–474 (1979).
74. Pearson, S., Cuverino, 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).
75. Ren, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
76. Gruzdiev, 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).
77. Voss, D. D. G. et al. Microphobinomeres are prevalent at Cas9-induced larger deletions. *Nucleic Acids Res.* **47**, 7402–7417 (2019).
78. Tsuboi, I., Herrault, Y., Walls, S., Qasim, W. & Piatovich, G. Variability in genome editing outcomes: challenges for research reproducibility and clinical safety. *Mol. Ther.* **28**, 1422–1431 (2020).
79. 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).
80. Hay, D. et al. Genetic dissection of the alpha-globin super-enhancer in vivo. *Cell Rep.* **35**, 108206 (2020).
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. Blighe, K. EnhancedVolcano: Publication-ready volcano plots with enhanced annotations by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **17**, 30 (2016).
84. Blighe, K. Transposition of native chromatin for fast and sensitive epigenomic profiling of mammalian chromatin folding. *Mol. Cell Biol.* **38**, 3599–3612 (2018).
85. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **23**, 4247–4261 (2018).
86. Haploinsuf ciency of Runx1 results in the acceleration of mesodermal development and hemangioblast specification upon in vitro differentiation of ES cells. *Blood* **103**, 868–889 (2004).
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 Bisulfi-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_001614), 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 Clayden 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 HU, MRC MHU [MC_UU_12009], NIHR Oxford BRC and John Fell Fund [131/030 and 101/517], EPA fund [CFI82 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/EPSRC grant (MR/K015777/1) to MICa—NanoscropyOxford (NanoO): Novel Super-resolution Imaging Applied to Biomedical Sciences, Micron (107457/Z/15Z).

Author contributions
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