Dynamics of the 4D genome during in vivo lineage specification and differentiation

A. Marieke Oudelaar

University of Oxford

Et al.

Let us know how access to this document benefits you.

Follow this and additional works at: https://escholarship.umassmed.edu/oapubs

Part of the Bioinformatics Commons, Biophysics Commons, Developmental Biology Commons, Genetics and Genomics Commons, Molecular Biology Commons, and the Structural Biology Commons

Repository Citation

Oudelaar AM, Hidalgo D, Socolovsky M, Higgs DR, Hughes JR. (2020). Dynamics of the 4D genome during in vivo lineage specification and differentiation. Open Access Publications by UMMS Authors. https://doi.org/10.1038/s41467-020-16598-7. Retrieved from https://escholarship.umassmed.edu/oapubs/4275

Creative Commons License

This work is licensed under a Creative Commons Attribution 4.0 License.

This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in Open Access Publications by UMMS Authors by an authorized administrator of eScholarship@UMassChan. For more information, please contact Lisa.Palmer@umassmed.edu.
Dynamics of the 4D genome during in vivo lineage specification and differentiation

A. Marieke Oudelaar1,2,6, Robert A. Beagrie1,6, Matthew Gosden1, Sara de Ornellas1,3, Emily Georgiades1,2, Jon Kerry2, Daniel Hidalgo4, Joana Carrelha1, Arun Shivalingam3, Afaf H. El-Sagheer3,5, Jelena M. Telenius1,2, Tom Brown3, Veronica J. Buckle1, Merav Socolovsky4, Douglas R. Higgs1,6✉ & Jim R. Hughes1,2✉

Mammalian gene expression patterns are controlled by regulatory elements, which interact within topologically associating domains (TADs). The relationship between activation of regulatory elements, formation of structural chromatin interactions and gene expression during development is unclear. Here, we present Tiled-C, a low-input chromosome conformation capture (3C) technique. We use this approach to study chromatin architecture at high spatial and temporal resolution through in vivo mouse erythroid differentiation. Integrated analysis of chromatin accessibility and single-cell expression data shows that regulatory elements gradually become accessible within pre-existing TADs during early differentiation. This is followed by structural re-organization within the TAD and formation of specific contacts between enhancers and promoters. Our high-resolution data show that these enhancer-promoter interactions are not established prior to gene expression, but formed gradually during differentiation, concomitant with progressive upregulation of gene activity. Together, these results provide new insight into the close, interdependent relationship between chromatin architecture and gene regulation during development.
Enhancers are non-coding regulatory elements required for precise control of gene expression during mammalian development. The interaction of active enhancer elements with their target genes occurs predominantly within Topologically Associating Domains (TADs), self-interacting regions of chromatin, which are usually ~0.1–1 Mb in size and delimited by convergent binding sites for the zinc-finger protein CTCF. The relationship between enhancer activation, formation of intra-TAD chromatin interactions and gene activation is not completely understood. To better understand how genome structure relates to function, it is important to characterize the three-dimensional nuclear architecture of the genome at higher resolution, and to determine how it changes during differentiation and development.

It has been shown that TAD boundaries are generally established early in development and remain relatively stable during differentiation. By contrast, interactions within TADs are extensively restructured in differentiating cells, which involves the formation of specific interactions between enhancers and promoters. It has been suggested that this reorganization precedes gene activity and that enhancer-promoter interactions are formed prior to gene expression. However, due to limitations in temporal and/or spatial resolution in previous studies, it is not known precisely when such interactions are formed during development. The detailed order of events and precise relationship between chromatin architecture and activation of regulatory elements and genes therefore remain unclear. For example, it is possible that enhancers and promoters form limited interactions prior to gene activation due to changes in TAD structure during early differentiation, but that strong upregulation of gene expression is associated with more specific, subtle changes in conformation that will only be detected in data with sufficient resolution and sensitivity.

To better understand the relationship between chromatin architecture and gene expression, it is crucial to characterize chromatin structures at high resolution in pure, primary cell populations representing relevant developmental stages. This has been hampered by the lack of high-resolution Chromosome Conformation Capture (3C) methods that are suitable for the analysis of limited numbers of primary cells. Therefore, we developed Tiled-C, a new 3C-based approach, which can generate high-resolution contact matrices of selected regions of interest from as few as 2000 cells and thereby allows for the analysis of cell populations that have previously been inaccessible.

We have used Tiled-C to study the chromatin architecture of key erythroid genes through sequential stages of in vivo erythroid differentiation in the mouse, including highly purified hematopoietic stem and progenitor cells. In addition, we have generated matched chromatin accessibility and single-cell expression data. We examined six loci, including the α-globin, Slc25a37, Tal1, Cd47, Cpeb4, and Btg2 genes, and focused our analyses on the α-globin genes, because the regulatory elements in this locus are extremely well-characterized. We find that the TAD encompassing the α-globin genes is already present in hematopoietic stem cells. We also find that the first steps in gene activation occur in early committed erythroid progenitors and involve opening of the α-globin enhancers, which become accessible prior to both chromatin reorganization and activation of α-globin RNA expression. Subsequent chromatin reorganization involves the appearance of smaller self-interacting domains (sub-TADs) within the larger TAD, in which specific interactions between enhancers and promoters are formed. In contrast to the current literature, we find that these enhancer-promoter interactions do not precede upregulation of gene activity, but are formed gradually and concomitantly with progressive activation of α-globin expression. Importantly, we find a similar order of events at other erythroid gene loci. Therefore, our data demonstrate that—at this improved level of resolution—chromatin architecture and gene activation are more tightly linked than previously appreciated. Together, these findings provide new insights into the mechanisms contributing to the establishment of tissue-specific chromatin structures during development.

Results
Tiled-C generates low-input and high-resolution 3C data. We developed Tiled-C, a 3C technique, which generates deep, high-resolution contact matrices of genomic regions of interest. Tiled-C maximizes library complexity by employing a single-tube protocol for 3C library preparation, which minimizes losses during the procedure. This is combined with an enrichment approach derived from the efficient Capture-C technology, which allows for up to a million-fold enrichment of interactions of interest. While Capture-C targets individual restriction fragments as viewpoints, Tiled-C uses a panel of capture oligonucleotides tiled across all restriction fragments of specified genomic regions to efficiently enrich for contacts within this region. This allows for deep, targeted sequencing of chromatin interactions within regions of interest and thus for the generation of high-resolution contact matrices at unprecedented depth, across multiplexed samples and genomic regions. Tiled-C therefore combines the ability of all vs all methods such as Hi-C to map large-scale chromatin structures including TADs, and the ability of one vs all methods such as 4C to robustly identify enhancer-promoter interactions within TADs in detail.

To validate the Tiled-C approach, we compared Tiled-C data to the deepest currently available in situ Hi-C datasets (mouse ES cells; Fig. 1a). Tiled-C data at this region was ~28-fold higher in depth and required ~19-fold less sequencing (Supplementary Table 1). To further demonstrate the capabilities of Tiled-C, we performed experiments in which we multiplexed across several key gene loci in both primary mouse erythroid cells and ES cells, to characterize their tissue-specific configurations in depth (Fig. 1b, c).

Current methods, such as Targeted Chromatin Capture (T2C), Capture Hi-C, and HYbrid Capture Hi-C (Hi-C2), also use an oligonucleotide capture procedure to enrich 3C or Hi-C libraries for regions of interest. These methods require millions of cells per sample. Tiled-C is specifically designed to maximize the efficiency of the experimental procedure and its increased sensitivity allows for the generation of reproducible, high-resolution contact matrices from as few as 2000 cells (Fig. 2, Supplementary Figs. 3 and 4, Supplementary Table 3), thereby enabling the analysis of previously intractable primary cell types. This is critical for the investigation of 4D (3D structure through developmental time) genome organization, as cell numbers become extremely limiting at early stages of development.

We used Tiled-C to study changes in chromatin structure associated with gene activation in primary cells during in vivo mouse erythropoiesis. We initially focused on the α-globin cluster because the α-globin genes and their regulatory elements are extremely well-characterized. The mouse α-globin cluster comprises the duplicated adult α-globin genes Hba-1 and Hba-2, as well as the embryonic gene Hba-x, and two genes of unknown function Hbq-1 and Hbq-2. These genes are located in a TAD, which also contains five additional genes upstream of the α-globin cluster: Nep13, Mpg, Rhdf1, Snurf25, and Il9r. The α-globin genes are regulated by five erythroid-specific enhancer elements (R1-R4 and Rm), which classify as a super-enhancer. In terminally differentiating erythroblasts these enhancers interact with the gene promoters in a sub-TAD flanked by multiple
CTCF-binding elements, which are predominantly in a convergent orientation (Supplementary Fig. 1).

Isolation of primary erythroid cell populations. Using fluorescence-activated cell sorting (FACS), we isolated cells at sequential stages of erythroid differentiation directly from mouse fetal livers (Fig. 3a, b). This allowed us to analyze highly purified primary cells through in vivo erythropoiesis. The S0-low cell population consists of early progenitors, predominantly Burst-Forming Unit-Erythroid (BFU-E) cells. S0-medium contains primarily of early Colony-Forming Unit-Erythroid (CFU-E) cells, while S1 and S2 contain the last CFU-E cell division before terminal differentiation (33,34). S3 through to S5 consist of terminally differentiating erythroblasts in progressively more mature states. Because erythroid cells enucleate in the final stages of differentiation, we have focused our analyses on stages S0 through to S3. In vitro, differentiation from S0 cells to S1 cells takes about 10 h, and differentiation to S3 cells takes an additional 10 h.

Gene expression through erythroid differentiation. We used Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-seq) (35), a variant of single-cell RNA-seq, to characterize gene expression in the isolated cell populations. We generated a dataset that covers the full course of in vivo erythroid differentiation through to terminal differentiation in the mouse (Supplementary Fig. 5), which shows that α-globin is expressed at basal levels in the S0 populations. Expression of α-globin dramatically increases during the S2 stage and plateaus at S3;

---

**Fig. 1 Tiled-C generates deep all vs all 3C data at regions of interest.** a Comparison of Tiled-C and Hi-C contact matrices at 2 kb resolution in mouse ES cells. Contact frequencies represent normalized, unique interactions in three and four replicates for Tiled-C and Hi-C data, respectively. Coordinates (mm9): chr11:29,902,000-33,228,000. b Tiled-C contact matrices of ~3.3 Mb spanning the mouse α-globin locus in primary mature erythroid cells (top) and ES cells (bottom) at 2 kb resolution. Contact frequencies represent normalized, unique interactions in three replicates. Gene annotation (α-globin genes highlighted in red), open chromatin (ATAC), and CTCF occupancy are shown below the matrices. Coordinates (mm9): chr11:29,902,000-33,228,000. c Tiled-C contact matrices of ~3.4 Mb spanning the mouse Sox2 locus in primary mature erythroid cells (top) and ES cells (bottom) at 5 kb resolution. Contact frequencies represent normalized, unique interactions in four replicates. Gene annotation (Sox2 gene highlighted in red), open chromatin (ATAC), and CTCF occupancy are shown below the matrices. Coordinates (mm9): chr3:33,200,000-36,565,000.
Fig. 2 Tiled-C generates high-resolution contact matrices from small numbers of cells. Tiled-C contact matrices of ~3.3 Mb spanning the mouse α-globin locus at 5 kb resolution, generated from small aliquots of primary mature erythroid cells. Contact frequencies represent normalized, unique interactions in three replicates. TADs are indicated below each matrix with a black bar. Gene annotation (α-globin genes highlighted in red), open chromatin (ATAC), and CTCF occupancy are shown below the matrices. Coordinates (mm9): chr11:29,900,000–33,230,000.
however, the earliest cells showing elevated expression of α-globin are found in the S1 stage (Fig. 3c). To validate that erythroid-specific α-globin upregulation begins at S1, we performed RNA-FISH to detect nascent transcription in FACS-sorted primary cells (Fig. 3d). We detect a small increase in nascent transcription from S0-low to S0-medium cells and confirm a robust increase in expression from S0 to S1 cells (P < 0.005 by paired t-test; Fig. 3e).

Chromatin architecture through erythroid differentiation. We used the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) to profile chromatin accessibility in these stages. Interestingly, we find that both enhancer and promoter elements are accessible prior to the onset of erythroid-specific gene expression, and that the degree of accessibility gradually increases, concomitant with upregulation of α-globin expression (Fig. 4, Supplementary Fig. 6). Tiled-C analysis shows that a TAD structure encompassing the α-globin locus is present at the earliest stage (S0-low), prior to the formation of weak enhancer-promoter interactions in the S0-medium stage (Fig. 4, Supplementary Figs. 7 and 8). These enhancer-promoter interactions strengthen in the subsequent S1 and S2 stages, accompanied by increases in α-globin expression and accessibility. In the S3 stage, where chromatin accessibility and expression reach their maximum levels, we observe a further increase in enhancer-promoter interactions.
promoter interactions in a sub-domain within the TAD, similar to the structure observed in primary erythroblasts derived from mature spleen tissue (Fig. 1, Supplementary Figs. 1 and 2). The smaller sub-TAD, which gradually forms within the pre-existing TAD, is delimited by convergent CTCF-binding elements that flank the α-globin enhancers and genes (Fig. 4, Supplementary Fig. 1). We have previously shown that these CTCF-binding elements are functionally important to restrict the interactions of the α-globin enhancers and prevent other genes within the TAD, but outside the sub-TAD, from being upregulated31,37. This suggests that this smaller erythroid-specific domain is likely formed by similar CTCF-dependent mechanisms as TADs, although it is smaller in size (~70 kb) and has very high internal interaction frequencies compared to typical TADs.
Discussion
The mouse α-globin cluster has a long history as model locus for studying gene regulation during differentiation and development. Previous analysis of transcription factor binding at the α-globin locus has shown that lineage commitment and differentiation are driven by sequential appearance of key transcription factors, consistent with the gradual increase in chromatin accessibility at the regulatory elements we describe here. In addition, previous 3C studies of both the α- and β-globin loci in erythroid cell lines have demonstrated that interactions between enhancers and promoters are tissue-specific. More recently, the tissue-specific conformation of the α-globin locus has also been described based on super-resolution imaging of two stages of ex vivo erythroid differentiation.

A limitation of the current body of work on chromatin organization during differentiation and development—both at the globin clusters and other gene loci—is that experiments have been performed at low spatial and temporal resolution and predominantly in vitro. Previous studies have therefore not been able to identify at what point in differentiation tissue-specific interactions between enhancers and promoters are established, nor how the formation of these interactions relates to changes in gene expression. Progress in this area has been limited by a lack of techniques capable of generating high-resolution interaction data from the small numbers of cells available in developmentally relevant primary cell populations.

To overcome this hurdle, we have developed Tiled-C, a new 3C-based approach, which can generate high-resolution chromatin interaction data from as few as 2000 cells and thus enables analysis of cell populations that have previously been inaccessible. We have used Tiled-C, in combination with ATAC and single-cell RNA-seq, to study the dynamic chromatin architecture and
expression of the α-globin cluster through in vivo erythropoiesis. Our data show that the boundaries of the TAD containing the α-globin cluster are already established in hematopoietic stem cells and maintained during further differentiation. This is consistent with previous reports, which have shown that TADs are relatively stable during differentiation and development. In contrast to the current literature, however, the high resolution of our data has enabled us to show that the formation of smaller domains within TADs and the subtle structural changes that strengthen specific enhancer-promoter interactions both occur gradually during erythroid differentiation, concomitant with progressive upregulation of gene activity (Fig. 6). Interestingly, initial chromatin accessibility is detectable at the regulatory elements of the α-globin locus prior to conformational change and activation of gene expression, although accessibility also increases gradually during differentiation. In addition to the α-globin cluster, we demonstrate the same order of events across five other erythroid gene loci.

It is likely that the formation of specific sub-domains within TADs early in differentiation facilitates interactions between enhancers and promoters to prime loci for gene activation. It has recently been shown for the globin loci that gene activation is associated with the formation of higher-order hub-like structures, in which multiple enhancers and promoters form simultaneous, specific interactions. Our data suggest that these structures may only be formed in the final stages of differentiation, when chromatin accessibility and interactions between enhancers and promoters are strongest, and may be important to achieve maximal gene expression. This is further supported by recent live imaging experiments in Drosophila, in which gene activation only occurred upon the formation of tight associations between enhancers and a gene promoter, and not after induced enhancer-promoter proximity resulting from interactions between insulator elements.

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

The initial appearance of chromatin accessibility at the regulatory elements of α-globin occurs early in differentiation and significantly precedes the onset of specific enhancer-promoter interactions. This suggests that chromatin opening can occur independently of large scale chromatin reorganization, yet further increases in accessibility do occur alongside the establishment and progressive strengthening of enhancer-promoter interactions, suggesting only a partial decoupling. Our observation that the TAD at the α-globin locus is present prior to accessibility of the regulatory elements indicates that activation of regulatory elements is not required for the establishment of the TAD, which is consistent with our previous work showing that deletion of the α-globin enhancers has no impact on the formation of the α-globin TAD. In contrast, the α-globin enhancer deletions do affect specific interactions between the α-globin promoters and the remaining enhancers, reinforcing that regulatory elements play a role in the formation of tissue-specific chromatin structures, possibly mediated by interactions between the multi-protein complexes bound at these elements.

In conclusion, our dissection of the chromatin architecture of a well-understood gene locus through in vivo erythroid differentiation demonstrates that chromatin architecture and gene activation are tightly linked during development and provides new insights into the distinct mechanisms contributing to the establishment of tissue-specific chromatin structures. Importantly, Tiled-C provides an approach that enables such detailed analysis in cell types that were previously intractable.

**Methods**

**Cells.** Mature primary Ter 119+ erythroblasts were obtained from spleens of female C57BL/6 mice treated with phenylhydrazine as previously described. Mouse ES cells (E14 and A9 lines) were cultured and harvested as previously described.

**Primary erythroid progenitor cells** were isolated from fetal livers, which were freshly isolated at e12.5-e13.5 from C57BL/6 mouse embryos. 5-15 livers were pooled together for each experimental replicate, mechanically dissociated in staining buffer (PBS, 0,2% BSA, 5 mM Glucose) and strained through a 30-μm
strainer. Cells were immunostained at 4 °C in the presence of rabbit IgG (200 μg/ml Jackson Laboratories 015-009-003) to block Fc receptors. To enrich for early erythroid cells, they were first stained with 5 μg/ml biotinylated anti-Ter119 (BD 553672) for 30 min, before magnetic depletion using streptavidin nanobeads (BioLegend Mojosort 480016) following the manufacturer’s instructions. Cells were then incubated with 0.5 μg/ml APC-conjugated streptavidin (BD 553672), 0.33 μg/ml PE-Cy7-conjugated anti-D71 (BioLegend 113811) and a panel of 2 FITC-conjugated antibodies (CustomBio Blocking Antibodies) against CD45R anti-CD3e, anti-CD11b and anti-Ly-6G/6C, all at 1 μg/ml; BD 553848, 553087, 553061, 557396, and 553126, respectively) for 45 min. Cells were then resuspended in FACs running buffer (staining buffer plus 2 μM EDTA). 0.66 μg/ml Hoechst 33342 was added immediately prior to the solution to distinguish live cells. Cells were sorted into Eppendorf tubes containing 500 μl RPMI supplemented with 10% FCS using a BD FACSAria® Fusion machine with a 100 μM nozzle size. Hematopoietic stem and progenitor cells (HSPCs), lymphoid–primed multipotent progenitors (LMPPs) and multipotent progenitors (MPPs) were stained and FACs sorted from the bone marrow of 12-week-old female C57BL/6 mice as previously described.48 All protocols were approved through the Oxford University Local Ethical Review process and all experimental procedures were performed in accordance with European Union Directive 2010/63/EU and/or the UK Animals (Scientific Procedures) Act, 1986.

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

Tiled-C—reagents. Tiled-C is a hybrid of the all vs all 3C methods, such as Hi-C,49 and the one vs all methods, such as 4C2,50 and Capture-C2,22 Tiled-C generates all vs all contact matrices of specified genomic regions and thus combines an unbiased all vs all view with the ability to target regions of interest, without the need to sequence chromatin interactions genome-wide. Tiled-C has similarities to SC20,51 and Targeted Chromatin Capture (T2C)52. However, Tiled-C allows for reliable PCR dilution filtering based on random sonication ends and uses more efficient capture enrichment, and is therefore able to generate data at higher resolution and depth. Tiled-C also has similarities to Capture-Hi-C (Hi-C2)28, which use oligonucleotide capture to enrich Hi-C libraries for regions of interest (Supplementary Table 2). The main differences are that these methods enrich a biotinylated Hi-C library, while Tiled-C enriches a 3C library generated with an optimized procedure to retain maximal library complexity, which is critical for the analysis of small cell numbers. In addition, Tiled-C uses an efficient capture oligonucleotide design targeted directly to the ends of all restriction fragments present in the region of interest and an efficient capture enrichment procedure enabling coupling up to a million-fold enrichment of restriction fragments of interest. The combination of high library complexity and efficient enrichment in Tiled-C enables high-resolution data generation at great depth and makes the method suitable for the analysis of small cell numbers. Moreover, enriching for targeted regions of interest substantially decreases sequencing costs. We should note, however, that the synthesis of large amounts of capture oligonucleotides can also be expensive. We therefore believe that Tiled-C is particularly useful for researchers who are interested in studying genomic regions of interest in multiple replicates and conditions and/or in primary cells with limited availability.

Tiled-C—oligonucleotide design. Tiled-C uses a panel of 70 bp oligonucleotides to enrich for regions of interest. The oligonucleotide sequences are designed complementary to both ends of each individual restriction fragment present in the region of interest. For efficient capture enrichment, we are using 80Extend BLAT-based primer design of oligonucleotides does not contain repetitive sequences that would decrease specific enrichment of the region of interest. To help users design capture oligonucleotides, we have developed a user-friendly python package (https://oligo.readthedocs.io/en/latest/). An overview of the oligonucleotide designs used in our study is shown in Supplementary Table 4.

Tiled-C—oligonucleotide synthesis. The enrichment step in the Tiled-C protocol can be performed with single-stranded or double-stranded biotinylated oligonucleotides (Supplementary Fig. 4, Supplementary Table 5). We produced panels of single-stranded oligonucleotides on an in-house Combinatrix CustomArray B3 DNA synthesiser (B3Synth, v2.5 software) using CustomBio Blocking Antibodies (CustomBio Blocking Antibody Application). All the probe sequences were designed to be 70 bases in length and were placed at random positions on the microarray for synthesis using Layout Designer (v4.3.1). Synthesis of oligonucleotide probe sequences occurred on individual electrodes present on the semiconductor surface of the microarray by photolithography. The photolithography and the conformation of the oligo do not affect the final probe. The oligonucleotide pool synthesis settings and reagents prepared according to the manufacturer’s protocols. Each sequence was synthesized in triplicate. After the synthesis of the unmodified oligonucleotide, 5’-biotin was added using a double coupling cycle with an extended 15 min coupling time. The final deprotection step was performed manually using the software by incubating the slides with TCA deblock (4 x 30 s incubations) before washing the slide with acetone four times and drying under argon. Oligonucleotides were then cleaved and deprotected on a stripping clamp system provided by the manufacturer using concentrated aqueous H2SO4 at 90 °C for 18 h. After the solution temperature was cooled, cells were pelleted and all lysis buffer was removed by careful pipetting. Chromatin was subsequently digested with the DpnII restriction enzyme in a 200 μl reaction, to which three doses of 150 U DpnII enzyme were added several hours apart and which was incubated 16–24 h at 37 °C. After heat-inactivation of DpnII, the ligation reaction was performed in the same tube, using 120 U T4 ligase in an overnight incubation at 16 °C. Ligated DNA was reverse crosslinked and treated with proteinase K at 65 °C in an overnight incubation. After RNase treatment, DNA was extracted using phenol-chloroform and transferred to a light phase lock tube for separation. To maximize yield, DNA was precipitated overnight in 70% ethanol at -20 °C after which DNA was pelleted and resuspended in PCR-grade water. The resulting 3 C libraries were sonicated to 200 bp fragments using a Covaris S220 Focused Ultrasonicator (six cycles of 60 s; duty cycle: 10%; cycles per burst: 200). Illumina TruSeq adapters were subsequently added using the NEBNext Ultra II DNA Library Prep kit, according to the manufacturer’s protocol. DNA clean-up steps were performed with Ampure XP beads at a 1:1.8 ratio to minimize loss of material. The libraries were indexed and amplified in 8–12 rounds of PCR amplification using the Hercule II PCR kit. 700 ng–1 μg of indexed material was used during subsequent capture-based enrichment.

When cell numbers were not limiting, we used aliquots of ~107 cells to prepare 3C libraries, following the efficient capture enrichment protocol. When cell numbers were limiting, we used aliquots of 1500 U DpnII restriction enzyme, which were added several hours apart over a total incubation time of 16–24 h at 37 °C. The digestion reaction was heat-inactivated and digested chromatin was ligated overnight with 720 U of T4 DNA ligase at 16 °C. The ligated DNA was reverse crosslinked and treated with proteinase K overnight at 65 °C. After RNase treatment, DNA was cleaned up using phenol-chloroform extraction and precipitation with ethanol and sodium acetate at -80 °C. The resulting 3 C libraries were resuspended in PCR-grade water. Aliquots of 5–6 μg of 3C library were sonicated to ~200 bp fragments using a Covaris S220 Focused Ultrasonicator (six cycles of 60 s; duty cycle: 10%; cycles per burst: 200). Illumina TruSeq adapters were added using the NEBNext Ultra II DNA Library Prep kit, according to the manufacturer’s protocol. DNA clean-up steps were performed with Ampure XP beads in a 1:1.8 beads/sample ratio. Where possible, we processed two aliquots of each sample in parallel and ligated the DNA with the same index to generate maximum library complexity. 1–1.5 μg of indexed material was used during subsequent capture-based enrichment.

For enrichment using single-stranded oligonucleotides, we used the Nimblegen SeqCap EZ reagents and followed the SeqCap EZ Library SR User’s Guide (Chapters 5–7). We multiplexed six samples per protocol in a single tube, and multiplied the volumes described in the protocol by the number of multiplexed libraries. Briefly, 700 ng–1.5 μg of indexed library (in exact 1:1 ratio if multiplexed reaction is performed) was mixed with 5 μg mouse Cot-1 DNA, 1 mmol of Nimblegen HE universal blocking oligonucleotides, and 1 mmol of Nimblegen HE hybridization blocking oligonucleotides (BSO). Only one of the indexed libraries was used in each 1:5 μl microcentrifuge tube. This mixture was dried completely in a vacuum centrifuge at 50 °C. The dried DNA was resuspended in 7.5 μl Nimblegen
hybridization buffer and 3 μl Nimblegen hybridization component A and denatured at 95 °C for 10 min. Concurrently, 4.5 μl of the biotinylated capture oligonucleotides was placed in a 200 μl PCR tube and heated to 47 °C in a PCR thermocycler. After denaturation, the 3C library mixture was added to the biotinylated oligonucleotides without removing them from the heating block in the thermocycler. The hybridization reaction was incubated in the thermocycler at 47 °C for 64–72 h with a heated lid at 57 °C. After incubation was complete, we transferred the samples to a 1.5 ml Eppendorf tube in 1 N ICl, 2.5 μl of 10× TAE, and the Nimblegen SeqCap EZ wash buffers, following the procedure described in the manufacturer’s protocol. After the washing steps, the beads with captured material were resuspended in 40 μl PCR-grade water. Captured DNA was amplified directly of the Tiled-C mastermix provided in the SeqCap EZ discovery kit v2 in two separate reactions of ~10 cycles, as described in the protocol. Ampure-XP beads were used in a 1:8:1 bead/sample ratio to clean-up the amplification reaction and DNA was eluted in 30 μl PCR-grade water. To increase enrichment, a second round of oligonucleotide capture was performed following the same procedure, using up to 2 μg of enriched material in a single hybridization reaction (even if multiplexed in first round) of 20–24 h.

For enrichment using double-stranded oligonucleotides, we used Twist Biosciences reagents and followed the Twist Custom Panel Protocol (Steps 4–7). To multiplex samples, we used 375–500 ng indexed library per sample (in exact 1:1 ratio), mixed up to 1.5 μg in a single tube, and used single reaction volumes as described in the protocol. We processed multiple tubes simultaneously if required. Streptavidin C1 beads were used to enrich the hybridized DNA and the washed material was amplified using 10–12 cycles of PCR. Ampure-XP beads were used in a 1:8:1 beaddisposal ratio to clean-up the amplification reaction and DNA was eluted in 30 μl PCR-grade water. To increase enrichment, a second round of oligonucleotide capture was performed following the same procedure, using up to 1.5 μg of enriched material in a single hybridization reaction (even if multiplexed in first round) of 20–24 h.

All Tiled-C data in this study were generated using a double capture enrichment procedure, since we have previously shown that this substantially increased enrichment in the Capture-C approach. However, we have recently noticed that a single Tiled-C capture step results in very efficient enrichment (~85% capture efficiency), which is only marginally improved after a second capture step (~88% capture efficiency). It might therefore be preferable to perform a single enrichment step as this is less labor-intensive, cheaper and results in fewer PCR duplicates.

The enriched Tiled-C libraries were assessed using the Agilent Bioanalyzer or DI1000 Tapestation and quantified using KAPA Library Quantification reagents, before sequencing using the Illumina NextSeq platform. In high-quality libraries, sequencing 3–5 million reads per enriched Mb per sample is sufficient for data at 5 kb resolution.

Tiled-C—analysis. The most straightforward way to analyze Tiled-C data is to use the HiCpro pipeline with the options for Capture-Hi-C analysis. We have also adjusted our pipelines for Capture-C analysis to be compatible with Tiled-C data. This pipeline is designed to analyze deep, targeted 3C data and provides very stringent filtering of 3C artefacts, including PCR duplicates. All data presented in the paper have been analysed using a combination of this CCseqBasic pipeline (https://github.com/Hughes-Genome-Group/CCseqBasic), custom analysis scripts and ICE normalization. Detailed instructions and custom scripts are available on https://github.com/oudelaar/TiledC. Interaction profiling and peak calling were performed using the MACS257. Peaks were merged and the number of reads in peaks called using MACS257. For visualization, bedgraph files were generated using the BEDTools genomecov with a scaling factor of 1/g (total number of reads in peaks). All analysis scripts are available at https://github.com/rbeagrie/alpha-tiledc.

Single-cell RNA-seq—experimental procedure. Fetal livers were harvested and pooled from 7 e13.5 C57Bl/6 mouse embryos and processed as above. Cells were first stained with 2.5 μg/ml biotin-conjugated anti-Ter119 (BD 553567) and 2.5 μg/ml purified anti-Ter119 (BioLegend 116241) conjugated to a DNA oligonucleotide (ADT2 GAGGCGATTGAT). After magnetic depletion, an equal number of Ter119+ cells were added back to the Ter119 fraction for further analysis, ensuring a balanced distribution of erythroid cells from all stages of differentiation. Cells were then stained with a panel of 9 further DNA-oligonucleotide-conjugated antibodies (0.33 μg/ml anti-CD71: BioLegend 113802 conjugated with ADT9 [CGAAGAGGGAGT], 1 μg/ml anti-CD41: BioLegend 133919 conjugated with ADT3 [TGTCGGCAATA], 1 μg/ml anti-CD45R: BioLegend 103049 conjugated with ADT5 [GATCGTAATAC], 1 μg/ml anti-CD3e: BioLegend 100345 conjugated with ADT7 [CATCGTGATTCA], 1.5 μg/ml anti-CD11b: BioLegend 101249 conjugated with ADT1 [CATGATTG GCC], 1.5 μg/ml anti-Ly6G/6C: BioLegend 1084249 conjugated with ADT4 [TGTGGAACCTTG], 1 μg/ml anti-CD44: BioLegend 103051 conjugated with ADT8 [GTCTGACTTGGC], 1 μg/ml anti-cKit: BioLegend 105829 conjugated with ADT6 [AAGGCCTGTCGA], and 1 μg/ml non-specific conjugated with ADTI06 [GGGAGT] and ATAC-seq—experimental procedure. For FACS-sorted erythroid progenitors from fetal liver, either two or three replicates were processed of ~50,000 cells each for each sorted population. ATAC-seq was performed as previously described. For FACS-sorted hematopoietic stem and progenitor cells from adult bone marrow, one replicate of between 5000 and 20,000 cells was processed for each population. Cells were spun at 500 x g for 10 min at 4 °C. The supernatant was discarded and cells were resuspended directly in Dg-transposition buffer (25 μl 2x TATase buffer, 2 μg each of TATase (NEB), 1 mM MgCl2, 5 μM ATP, 0.5 μM C2H2 OAc, 5 μl 20X T4 ligase buffer and 5 μl H2O) before incubating at 37 °C for 30 min with agitation at 600 rpm. After the transposition step, samples were processed as previously described.

ATAC-seq—analysis. Reads were mapped to the mouse mm9 genome and PCR duplicates removed using NGEseqBasic. Technical replicates were merged and peaks called using MACS2. Peaks were merged and the number of reads in each peak was calculated using the BEDTools genomecov and multiroc. For visualization, bedgraph files were generated using the BEDTools genomecov with a scaling factor of 1/g (total number of reads in peaks). All analysis scripts are available at https://github.com/rbeagrie/alpha-tiledc.
DIG FITC [Roche 11207741910] 1:50 in blocking solution for 30 m at RT, washed 2× with TST, incubated with 100 µl rabbit anti-sheep FITC [Vector Fi-6000] 1:100 in blocking solution for 30 m at RT, washed 1× with PBS and post-fixed in 4% formaldehyde for 5 m at RT. Slides were then washed and dehydrated with successively PBS, 70% ethanol, 90% ethanol, and 100% ethanol before air drying. Finally, coverslips were mounted with 10 µl mounting media [1 mg/mL DAPI in Vectashield antifade mounting medium (Vector)]. Two negative controls were also included: brain tissue from a male adult CD1 mouse and Ter119+ (i.e., mature) fetal liver erythroid cells that were probed with secondary antibody but no primary antibody. Magnetically purified but not FACS-purified Ter119+ fetal liver erythroid cells were used as a positive control.

RNA-FISH—imaging equipment and settings. Widefield fluorescence imaging was performed at 20°C on a DeltaVision Elite system (Applied Precision) equipped with a ×100/1.40 NA UPLSAPO oil immersion objective (Olympus), a CoolSnap HQ2 CCD camera (Photometrics), DAPI (excitation 390/18; emission 435/40) and FITC (excitation 475/28; emission 525/45) filters. 12-bit image stacks were acquired with a z-step of 150 nm giving a voxel size of 64.5 × 64.5 × 150 nm.

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

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 author upon reasonable request. Tiled-C, ATAC-seq and single-cell RNA-seq data generated in this study have been deposited in the NCBI Gene Expression Omnibus under accession code GSE137477. All RNA-FISH image files are archived in the Imaging Data Resource under accession code idr0084. A UCSC hub for visualizing ATAC-seq and single-cell RNA-seq mean expression data is available at http://saro.molbio.iox.ac.uk/public/hugheslab/alpha-tiledc/hub.txt. The source data underlying Fig. 3e are provided as a Source Data file.

Code availability. Tiled-C data were analyzed with the CCseqBasic pipeline and custom scripts available at https://github.com/Hughes-Group/CCseqBasic/releases and https://github.com/oudelaar/TiledC. ATAC-seq data were analyzed with the NGseqBasic pipeline and scripts available at https://github.com/rbeagrie/alpha-tiledc. Single-cell RNA-seq data were analyzed with scripts available at https://github.com/Hooman/CITE-seq-Count and https://github.com/rbeagrie/alpha-tiledc. RNA-FISH analysis was performed using scripts available at https://github.com/rbeagrie/alpha-tiledc.

Received: 6 February 2020; Accepted: 13 May 2020; Published online: 01 June 2020

References
1. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. Science 361, 1341–1345 (2018).
2. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. Nat. Rev. Genet. 20, 437–455 (2019).
3. Stadhouder, R., Filion, G. J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decisions. Nature 569, 345–354 (2019).
4. Andrey, G. & Mundlos, S. The three-dimensional genome: regulating gene expression during pluripotency and development. Development 144, 3646–3658 (2017).
5. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. Nature 502, 499–506 (2013).
6. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. Nat. Rev. Mol. Cell Biol. 34, 903–916 (2019).
7. Phillips-Cremins, J. E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 153, 1281–1295 (2013).
8. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. Nature 518, 331–336 (2015).
9. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. Cell 171, 557–572.e524 (2017).
10. Stadhouder, R. et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat. Genet. 50, 238–249 (2018).
11. Montavon, T. et al. A regulatory archipelago controls Hox genes transcription in digits. Cell 147, 1132–1145 (2011).
12. Williamson, L. Lettice, L. A., Hill, R. E. & Bickmore, W. A. Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. Development 143, 2994–3001 (2016).
13. Andrey, G. et al. Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. Genome Res. 27, 3259–3271 (2017).
14. Freire-Pritchett, P. et al. Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. eLife 6, e21926 (2017).
15. Rubin, A. J. et al. Lineage-specific dynamic and pre-established enhancer–promoter contacts cooperate in terminal differentiation. Nat. Genet. 49, 1522–1528 (2017).
16. Sun, F. et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503, 290–294 (2013).
17. Ghavi-Helm, Y. et al. Enhancer loops appear stable during development and are associated with paused polymerase. Nature 513, 89–100 (2014).
18. Paliou, C. et al. Preformed chromatin topology assists transcriptional robustness of Shh during limb development. Proc. Natl. Acad. Sci. USA 116, 12390–12399 (2019).
19. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. Science 295, 1306–1311 (2002).
20. Oudelaar, A. M. et al. Single-allele chromatin interactions identify regulatory landscape hubs in dynamic chromatin organization. Nature 538, 1348–1354 (2016).
21. Zhao, Z. et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat. Genet. 38, 1341–1347 (2006).
22. Kolovos, P. et al. Targeted Chromatin Capture (T2C): a novel high resolution high throughput method to detect genomic interactions and regulatory elements. Epigenetics Chromatin 7, 10 (2014).
23. Davies, J. O. J. et al. Multiplexed analysis of chromosome conformation at a globally improved sensitivity. Nat. Methods 8, 1202–1210 (2015).
24. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–293 (2009).
25. Simonis, M. et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. 38, 1348–1354 (2006).
26. Hay, D. et al. Genetic dissection of the α-globin super-enhancer in vivo. Nat. Genet. 48, 895–903 (2016).
27. Oudelaar, A. M. et al. Single-allele chromatin interactions identify regulatory hubs in dynamic chromatin compartmentalized domains. Nat. Genet. 50, 1744–1751 (2018).
28. 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).
29. Chiarlello, A. M. et al. A dynamic folded hairpin conformation is associated with α-globin activation in erythroid cells. Cell Rep. 30, 2125–2135.e2125 (2020).
30. Pop, R. et al. A key commitment step in erythropoiesis is synchronized with the cell cycle clock through mutual inhibition between PU.1 and S-phase progression. PLoS Biol. 8, e1000484 (2010).
31. Tusi, B. K. et al. Population snapshots predict early haematopoietic and erythroid hierarchies. Nature 555, 54–60 (2018).
32. Stockeius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods 14, 865–868 (2017).
33. Buenrostro, J. D., Giresi, P. G., Zaba, J. 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).
34. Oudelaar, A. M. et al. A revised model for promoter competition based on multi-way chromatin interactions at the α-globin locus. Nat. Commun. 10, 5412–5418 (2019).
35. Oudelaar, A. M. et al. Between form and function: the complexity of genome folding. Hum. Mol. Genet. 26, R208–R215 (2017).
36. Anguita, E. et al. Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. EMBO J. 23, 2841–2852 (2004).
37. Palstra, R. J. et al. The beta-globin nuclear compartment in development and erythroid differentiation. Nat. Genet. 35, 190–194 (2003).
41. Vernimmen, D., De Gobbi, M., Sloane-Stanley, J. A., Wood, W. G. & Higgs, D. R. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J.* 26, 2041–2053 (2007).
42. Brown, J. M. et al. A tissue-specific self-interacting chromatin domain forms independently of enhancer-promoter interactions. *Nat. Commun.* 9, 376 (2018).
43. Allahyar, A. et al. Enhancer hubs and loop collisions identified from single-cell topologies. *Nat. Genet.* 50, 1296–1303 (2018).
44. Chen, H. et al. Dynamic interplay between enhancer-promoter topology and gene activity. *Nat. Genet.* 50, 1296–1303 (2018).
45. Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. A. Emerging evidence of chromosome folding by loop extrusion. *Cold Spring Harb. Symp. Quant. Biol.* https://doi.org/10.1101/sqb.2017.82.034710 (2018).
46. Kagey, M. H. et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435 (2010).
47. Schmidt, D. et al. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* 20, 578–588 (2010).
48. Pietras, E. M. et al. Functionally distinct subsets of lineage-biased multipotent progenitors control blood production in normal and regenerative conditions. *Cell Stem Cell* 17, 35–46 (2015).
49. Dostie, J. et al. Chromosome Conformation Capture Carbon Copy (SC3): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16, 1299–1309 (2006).
50. Oudelaar, A. M., Downes, D., Davies, J. & Hughes, J. Low-input Capture-C: a chromosome conformation capture assay to analyze chromatin architecture in small numbers of cells. *Bio-protocol* https://doi.org/10.21769/BioProtoc.2645 (2017).
51. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16, 259 (2015).
52. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* 9, 999–1003 (2012).
53. Kruse, K., Hug, C. B., Hernández-Rodríguez, B. & Vaquerizas, J. M. TADtool: visual parameter identification for TAD-calling algorithms. *Bioinformatics* 32, 3190–3192 (2016).
54. Yang, T. et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* 27, 1939–1949 (2017).
55. Kerpedjiev, P. et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* 19, 125–112 (2018).
56. Telenius, J. & Hughes, J. R. Nseq2Basic—a single-command UNIX tool for ATAC-seq, DNase-seq, Cut-and-Run, and ChIP-seq data mapping, high-resolution visualisation, and quality control. *bioRxiv*, https://doi.org/10.1101/393413 (2018).
57. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137 (2008).
58. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. (Oxf., Engl.)* 26, 841–842 (2010).
59. Wrobel, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* 8, 281–291.e289 (2019).
60. Brown, J. M. & Buckle, V. J. Detection of nascent RNA transcripts by fluorescence in situ hybridization. *Methods Mol. Biol.* 659, 33–50 (2010).

**Acknowledgements**

We thank J. Sloane-Stanley and J. Sharpe of the WIMM Transgenics Facility for mouse breeding and fetal liver provision, K. Clark, S.-A. Clark and C. Waugh of the WIMM Flow Cytometry Facility for cell sorting services, N. Ashley of the WIMM Single Cell facility for antibody conjugation and 10X library preparation services, J. Brown for help with RNA-FISH, D. Downes for help with ATAC-seq and C. Garnett for advice on low-input ATAC-seq. We wish to acknowledge the Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, and University of Oxford for use of their services in this project. This work was supported by the Medical Research Council (MC_UU_00016 to D.R.H., J.R.H., and V.J.B. and MR/N0969X/1 to J.R.H., T.B., and V.J.B.), Wellcome Trust (209181Z/17/Z to R.A.B.), the National Institutes of Health (NIH NIDDK R01DK100915 to M.S.), Biotechnology and Biological Sciences Research Council (BB/K001694/1 and BB/R008655/1 to A.H.E.-S. and BB/M025624/1 to A.S.) and a Stevenson Junior Research Fellowship at University College, Oxford (to A.M.O.).

**Author contributions**

A.M.O. and R.A.B. performed experiments and bioinformatic analyses and wrote the manuscript. M.G., E.G., D.H., and J.C. performed Capture-C; A.M.O. and R.A.B. performed experiments. S.d.O., A.S., A.H.E.-S., and T.B. synthesized oligonucleotides. J.K. wrote the oligonucleotide design tool. J.M.T. performed bioinformatic analyses. V.J.B. and M.S. provided conceptual advice. D.R.H. and J.R.H. supervised the project and wrote the manuscript.

**Competing interests**

J.R.H. is founder and shareholder of Nucleome Therapeutics.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-16592-7.

Correspondence and requests for materials should be addressed to D.R.H. or J.R.H.

**Peer review information** *Nature Communications* thanks the 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.

© The Author(s) 2020