




Simultaneous profiling of 3D genome structure and DNA methylation in single human cells

Dong-Sung Lee^{1,5}, Chongyuan Luo^{2,3,5}, Jingtian Zhou^{2,5}, Sahaana Chandran¹, Angeline Rivkin², Anna Bartlett², Joseph R. Nery¹ , Conor Fitzpatrick⁴, Carolyn O'Connor⁴, Jesse R. Dixon^{1*}  and Joseph R. Ecker^{1,2,3*} 

Dynamic three-dimensional chromatin conformation is a critical mechanism for gene regulation during development and disease. Despite this, profiling of three-dimensional genome structure from complex tissues with cell-type specific resolution remains challenging. Recent efforts have demonstrated that cell-type specific epigenomic features can be resolved in complex tissues using single-cell assays. However, it remains unclear whether single-cell chromatin conformation capture (3C) or Hi-C profiles can effectively identify cell types and reconstruct cell-type specific chromatin conformation maps. To address these challenges, we have developed single-nucleus methyl-3C sequencing to capture chromatin organization and DNA methylation information and robustly separate heterogeneous cell types. Applying this method to >4,200 single human brain prefrontal cortex cells, we reconstruct cell-type specific chromatin conformation maps from 14 cortical cell types. These datasets reveal the genome-wide association between cell-type specific chromatin conformation and differential DNA methylation, suggesting pervasive interactions between epigenetic processes regulating gene expression.

Three-dimensional (3D) genome architecture is a critical feature of gene regulation in metazoan organisms^{1–3}. Chromatin conformation profiling has revealed the existence of features such as topologically associated domains (TADs) and enhancer–promoter interactions^{4–9}. Despite the increasing use of these datasets, most existing chromatin conformation maps are generated from cell lines or from bulk tissues^{4,8–11}. While these data have helped to elucidate general principles of chromatin organization, it cannot fully represent the diversity of cell types that arise in vivo. Single-cell 3C or Hi-C represent attractive strategies to resolve cell-type heterogeneity^{12–14}. However, current single-cell Hi-C profiles from cultured cells primarily capture cell-cycle patterns^{12,15}. It remains unclear whether single-cell Hi-C profiles are suitable for partitioning constituent cell types.

In contrast to single-cell Hi-C data, single-cell DNA methylome datasets enable high-resolution cell-type classification from cell types in primary human tissues^{16,17}. Cytosine DNA methylation (mC) is unaltered in the basic protocol of 3C or Hi-C, therefore it may be feasible to detect both long-range ligation junctions and mC by combining 3C or Hi-C with bisulfite sequencing.

Here, we describe a method, single-nucleus methyl-3C sequencing (sn-m3C-seq), to jointly profile chromatin conformation and mC from the same cell. Bulk and single-cell m3C-seq profiles accurately recapitulate chromatin architectures of mouse embryonic stem cells (mESCs). Furthermore, we show that sc-m3C-seq can distinguish cultured mouse cell types as well as highly heterogeneous human brain cell populations. Using 4,238 sn-m3C-seq profiles, we identify 14 cell types from human frontal cortex by clustering of mC profiles and from these clusters identify cell-type specific 3D chromatin structures. We observe a strong, cell-type specific relationship between mC and 3D genome structure, suggesting extensive crosstalk between these epigenomic features.

Results

Joint profiling of chromatin conformation and DNA methylation from the same DNA molecule. In sn-m3C-seq, we first performed restriction enzyme digestion and ligation on fixed nuclei, as is typically performed in an in situ 3C experiments^{8,18,19}. The ligated 3C nuclei are dispensed into 384-well PCR plates using fluorescence-activated nuclei sorting (FANS) and subject to proteinase digestion and bisulfite conversion, and libraries are constructed similar to our previous snmC-seq2 method (Fig. 1)^{16,20}. When performed as a bulk assay (m3C-seq) ligated nuclei are not sorted into wells but treated in bulk.

To evaluate the quality of chromatin contact maps generated by m3C-seq, we compared bulk m3C-seq data to conventional bulk in situ 3C-seq and Hi-C profiles in mESCs. Both Hi-C/3C and bisulfite conversion present challenges for read alignment due to the presence of chimeric reads and the conversion of unmethylated cytosines to uracils, respectively. We developed TAURUS-MH (two-step alignment with unmapped reads using read splitting for methyl-Hi-C), a mapping pipeline for m3C-seq data using a hybrid of ungapped and read splitting alignments (Supplementary Fig. 1a). Sequencing reads were first mapped to an in silico bisulfite converted genome using Bismark calling an ungapped aligner (bowtie1)²¹, and unmapped reads are split into three segments followed by ungapped alignment. We compared the performance of TAURUS-MH to BWA-METH²², which is designed for bisulfite sequencing data alignment using BWA-MEM. This comparison was performed using typical Hi-C data with in silico simulated bisulfite conversion. We use the alignment of conventional Hi-C data²³ as our gold standard. When compared with BWA-METH, our pipeline showed 19.43% higher in mappability (86.12 versus 66.69%, Fig. 2a), 3.64% higher in accuracy (97.86 versus 94.22%, Fig. 2b)

¹Peptide Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA. ²Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA. ³Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA, USA. ⁴Flow Cytometry Core Facility, The Salk Institute for Biological Studies, La Jolla, CA, USA. ⁵These authors contributed equally: Dong-Sung Lee, Chongyuan Luo, Jingtian Zhou.

*e-mail: jedixon@salk.edu; ecker@salk.edu

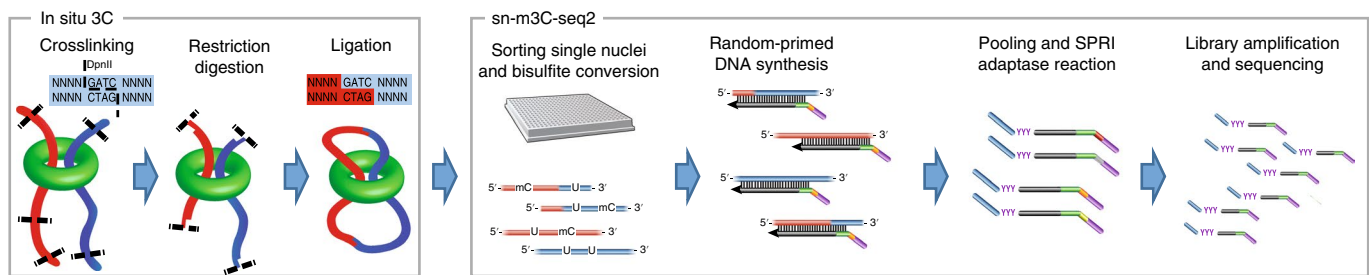


Fig. 1 | Outline of the single-nucleus methyl-3C sequencing (sn-m3C-seq) method. Samples are processed with a typical in situ 3C/Hi-C procedure followed by single-cell DNA methylome library preparation using sn-m3C-seq2.

and 13.41% higher long-range *cis*-contacts (42.79 versus 29.38% from total fragments and 49.68 versus 44.06% from mapped fragments, Fig. 2c).

We then analyzed chromatin contact data quality comparing bulk m3C-seq with a matched 3C-seq library. Bulk m3C-seq libraries showed a comparable fraction of long-range (>1 kb) intrachromosomal ligation events compared to the control 3C-seq library (26.6% in 3C-seq and 19.0% in m3C-seq) (Fig. 2d). We observed more frequent interchromosomal ‘contacts’ in m3C-seq (30.0% in m3C-seq and 12.24% in 3C-seq) (Fig. 2d). Since snmC-seq2 involves random-primed DNA synthesis²⁰, we speculate that the interchromosomal ‘contacts’ are an artifact caused by spurious hybridization and polymerase extension during random-primed DNA synthesis. We further hypothesized that spurious interchromosomal ligation is dependent on DNA concentration and the frequency of intermolecular interactions. Consistent with this hypothesis, in sn-m3C-seq, where the random-primed DNA synthesis reaction contains a much lower DNA concentration, we found a similar background level (15.11% in sn-m3C-seq and 12.24% in 3C-seq) of interchromosomal contacts.

To assess the methylome quality of bulk m3C-seq, we compared bulk m3C-seq with published whole-genome bisulfate sequencing (WGBS) profiles generated from mESCs²⁴. With comparable sequencing depth, the m3C-seq library showed more uniform genomic coverage compared to the WGBS library, covering more cytosines and more CpG sites and showed a narrower distribution of coverage at CpG sites (Supplementary Fig. 1b,c).

Finally, we compared contact maps and mC profiles generated by bulk m3C-seq with conventional Hi-C and MethylC-seq data generated from mESCs (Fig. 2e,f)²⁵. We observed strong agreement between bulk m3C-seq and Hi-C (Fig. 2e, stratum adjusted correlation coefficient, SCC=0.91²⁶). Similarly, we observed strong concordance of methylation profiles from bulk m3C-seq with existing MethylC-seq datasets for mESCs (Fig. 2g, Pearson’s correlation = 0.82). We further compared bulk m3C-seq to multiple published bulk Hi-C and MethylC-seq datasets of mESCs and found strong correlations for both types of profile (Supplementary Fig. 2).

Fluorescence-activated nuclei sorting excludes nuclei multipliers.

To generate sn-m3C-seq profiles, in situ 3C treated nuclei were sorted using FACS into 384 well PCR plates followed by snmC-seq2 single-cell methylome library preparation. In control species mixture experiments (Supplementary Table 1), we found frequent (23.2%) interspecies nuclei multipliers (Supplementary Fig. 3a) due to formaldehyde crosslinking, whereas multiples were eliminated when crosslinking was performed separately for each species (Supplementary Fig. 3b). We found that performing crosslinking with a ten-fold diluted nuclei preparation or stringently selecting nuclei with a 2n genomic DNA content could largely eliminate interspecies nuclei multipliers (7.4% for dilution, 1% for 2n gating) (Supplementary Fig. 3c–e).

sn-m3C-seq generates high-quality single-cell DNA methylation profiles.

We systematically compared the technical characteristics of sn-m3C-seq with published single-cell methylome datasets generated using Zymo Pico Methyl-seq^{27,28} and scNMT-seq²⁹. sn-m3C-seq showed a superior read mapping rate ($72.4 \pm 3.6\%$) to the Pico Methyl-seq ($33 \pm 12.3\%$) or snNMT-seq ($32.2 \pm 9.4\%$, Supplementary Fig. 4a). The library complexity of sn-m3C-seq (maximumly $27.5 \pm 9.9\%$ of the mouse genome) is similar to that of scNMT-seq ($22.8 \pm 11\%$), and is greater than Pico Methyl-seq ($10 \pm 4.7\%$, Supplementary Fig. 4b). scBS-seq and its derivative scNMT-seq show coverage bias toward CpG islands (2.48 ± 0.82 -fold enriched compared to random permutation of the location of scNMT-seq reads)³⁰, whereas sn-m3C-seq (1.21 ± 0.15), snmC-seq2 (1.14 ± 0.09) and Pico Methyl-seq (1.57 ± 0.1) showed modest enrichment of CpG islands (Supplementary Fig. 4c). Last, sn-m3C-seq and Pico Methyl-seq show comparable coverage uniformity (Supplementary Fig. 4d). At a coverage of 1 million non-clonal reads, sn-m3C-seq covers 28.9% of 1 kb genomic bins and 89% of 10 kb bins, while Pico Methyl-seq covers 29.1% of 1 kb bins and 90.5% of 10 kb bins. Both assays are more uniform than scNMT-seq, which covers 23.5% of 1 kb bins and 78% of 10 kb bins with 1 million non-clonal reads. We observed a high correlation of the mC profiles between pooled sn-m3C-seq and previously generated bulk WGBS experiments with mESC specific hypomethylation at the promoter regions of pluripotent genes such as *Dppa4* and *Dppa2* (Pearson’s $r = 0.89$, sn-m3C-seq versus mESCs, ref.²⁵) (Fig. 3d and Supplementary Fig. 2)^{25,31–33}.

sn-m3C-seq profiles recapitulate chromatin conformation contact maps.

We have compared our data with previous single-cell Hi-C studies. We profiled comparable numbers of cells as the published single-cell Hi-C datasets with the largest numbers of profiled cells so far (Ramani et al. and Nagano et al., refs.^{12,34}; Ramani, 10,696; Nagano, 3,413; this study, 6,200 total cells between mESC, NMuMg and human brain data), yet we obtained 1.7-fold more contacts than Nagano et al. and 49.27-fold more contacts than Ramani et al. Other studies (Flyamer et al. and Tan et al., refs.^{35,36}) profiled lower numbers of cells (Tan = 35, Flyamer = 246) but with higher numbers of contacts (Tan, 1,165,296; Flyamer, 2,416,802; see Supplementary Fig. 5a,b and Online Methods for details of analysis). These data indicate that our method generates single-cell chromatin conformation data of comparable quality to existing unimodal methodologies.

Using hierarchical clustering on a matrix of SCC, we observed that contact maps from our mESC sn-m3C-seq clustered with Hi-C data from mESCs³⁷, while the cortical neurons and neural progenitor cell (NPC) datasets clustered separately (Fig. 3a–c)⁹. In both Hi-C and pooled sn-m3C-seq data, we observed mESC-specific contacts, such as enhancer-promoter contacts at the *Sox2* locus (Fig. 3c). We have observed additional cell-type specific hypomethylated regions in association with chromatin interaction differences in *Tbx5* and *Tfap2d* (Supplementary Figs. 6 and 7).

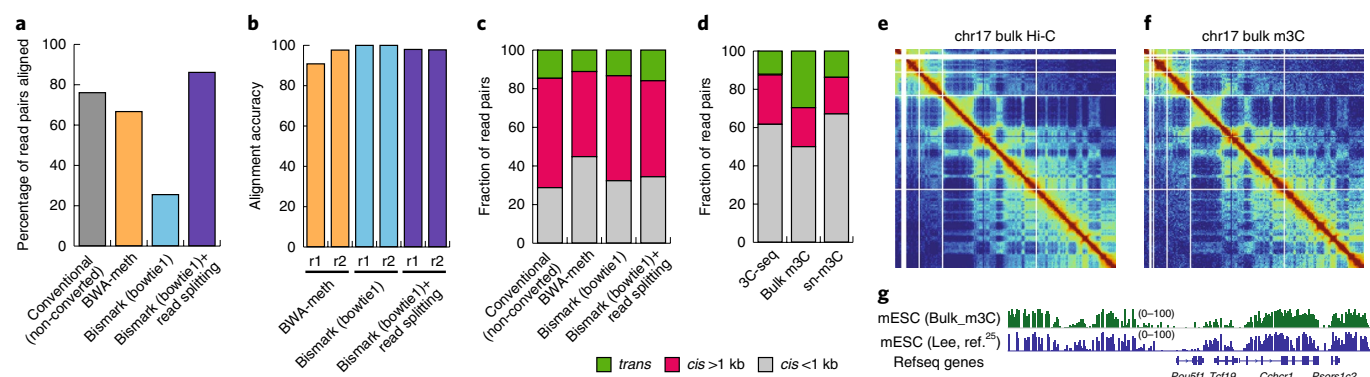


Fig. 2 | Data processing and analysis of m3C-seq sequencing reads. Reads derived from non-bisulfite treated regular Hi-C sequencing are converted C to T (r1, read1) and G to A (r2, read2) in silico and aligned using BWA-meth, Bismark (bowtie1) and Bismark (bowtie1) followed by split-read alignment. Alignment of non-converted reads using conventional alignment pipeline is used as a standard (conventional, non-converted). For **a–d**, the mapping algorithms were applied to a common test dataset ($n=1$) to make a fair comparison of their performance. **a**, Percentage of aligned reads as a pair. **b**, Alignment accuracy of different alignment strategies compared with conventional Hi-C alignment using in silico converted reads. **c**, Fraction of read pairs with *cis*- short-range reads (*cis* < 1 kb), *cis*-long-range interactions (*cis* > 1 kb) and *trans*-interactions (*trans*) using different alignment strategies. **d**, Similar to **c**, but for 3C-seq (without conversion), bulk m3C-seq (with conversion, from the same sample as bulk 3C-seq) and combined 192 single-nucleus m3C-seq results. **e**, Contact maps from chromosome 17 for conventional bulk Hi-C and bulk m3C data. **f**, mC profiles near the *Pou5f1* gene for conventional bulk MethylC-seq as well bulk m3C-seq. The experiment was repeated twice independently with similar results.

sn-m3C-seq profiles separate mouse cell types. To test the robustness of cell-type identification using sn-m3C-seq profiles, we performed t-distributed stochastic neighbor (t-SNE) embedding with CpG methylation levels in non-overlapping 100 kb bins from single cells, which shows a clear separation of mESCs and NMuMG cells (Fig. 4a). Using pooled sc-m3C-seq data, we identified distinct A/B compartment signatures between the two cell types (Fig. 4b) as well as local differences in Hi-C contacts (arrows in Supplementary Fig. 8a,b).

We also compared the ability of mC or Hi-C contacts to partition sc-m3C-seq into the relevant cell types. DNA methylation profiles could easily distinguish between ES and NMuMG cells using the first principal component (PC) alone, which explains 33.7% of total variance (Supplementary Fig. 8c,d). In contrast, PC analysis (PCA) using whole-genome Hi-C contacts at 100 kb could not distinguish between ES and NMuMG cells using the first two PCs (Fig. 4c), but the third PC did clearly separate these two cell types (Fig. 4d). PCA using local contacts (<2 Mb) was able to distinguish the two cell types using the second PC (Fig. 4e). We observed that the first PC was highly correlated with per cell sequencing depth (Fig. 4f), suggesting the power for cell-type identification using Hi-C contacts is highly dependent on sequencing coverage. In contrast, the ability to distinguish the two cell types using mC profiles is not sensitive to sequencing coverage (Supplementary Fig. 8e,f), indicating the robustness of cell-type classification from mC profiles. These results underscore the importance of jointly profiling mC along with chromatin conformation to reliably distinguish cell types in single-cell experiments.

sn-m3C-seq identifies cell-type specific chromatin interaction in human prefrontal cortex (PFC). To test whether sn-m3C-seq can be applied to complex human primary tissues, we generated 4,238 sn-m3C-seq profiles from the PFC region of two postmortem adult human brains (Supplementary Table 2). We first identified non-neuronal cell types using CG methylation (mCG) signature followed by fine clustering of neuronal subtypes using non-CG methylation (mCH), resulting in the identification of 14 major cell types in human PFC (Fig. 5a,b). We annotated the clusters based on the depletion of mCG and mCH at the gene body of known cell-type markers (Supplementary Fig. 9). The methylation profile is highly correlated in each cell type between the two individuals (Supplementary Fig. 10). Brain neuron subtypes (excitatory neuron

subtypes—L2/3, L4, L5 and L6; inhibitory neuron subtypes—Pvalb, Sst, Ndnf and Vip) can be identified with much greater resolution using mCH or mCG signatures, compared to only using chromatin interactions (Fig. 5a–c). However, clustering analysis using chromatin interactions alone or jointly with mCH can robustly resolve non-neuronal brain cell types (Fig. 5c and Supplementary Fig. 11).

Guided by the cell-type identification using mC signatures, we reconstructed brain cell-type chromatin interaction maps using sn-m3C-seq reads. We further identified 36,559 cell-type specific chromatin interactions using a negative binomial test-based method (edgeR³⁸, false discovery rate (FDR)=0.1%, see Supplementary Table 3) and 6,161 differential domain boundaries using the recently described HiCluster method (Supplementary Table 4)³⁹. We found drastic chromatin interaction dynamics at cell-type signature genes (Fig. 5d,e and Supplementary Figs. 12 and 13). For example, *SATB2* is a marker gene for excitatory neurons and shows reduced mCH and mCG in excitatory neuron clusters (Fig. 5f,g). A distinct chromatin loop between *SATB2* promoter region and the adjacent *LINC01923* locus located 1.15 Mb away is only found in excitatory, but not inhibitory neuron types (Fig. 5d). A specific pattern of increased domain boundary probability at the *SATB2* locus was also observed only in excitatory cells. Similarly, *PROX1* is a marker gene for inhibitory neuron subtypes (Vip and Ndnf) derived from caudal ganglionic eminence⁴⁰. *PROX1* locus shows reduced mCH and mCG in Vip and Ndnf clusters (Fig. 5h,i). Chromatin loops specific to Vip and Ndnf neurons were found between promoters of *PROX1* and *RPS6KC1*. Higher domain boundary probabilities were also observed at the promoter of *PROX1* in CGE-derived neurons.

Cell-type specific chromatin interactions are associated with differential DNA methylation signatures. The PFC sn-m3C-seq dataset allowed us to explore the relationship between chromatin architecture and mC across brain cell types. We found a significant overlap between cell-type specific chromatin interactions and the 115,137 differentially methylated regions (DMRs) identified across brain cell types ($P<0.0001$, two-sided permutation test, Fig. 6a and Supplementary Table 5). Examining the mC profiles over the anchor regions (k -means clustered, $k=15$) revealed a striking hypomethylation pattern at the sites of differential interacting regions in the cell types showing enriched interaction frequencies (Fig. 6b,c). Therefore, on a global scale, cell-type specific chromatin interactions

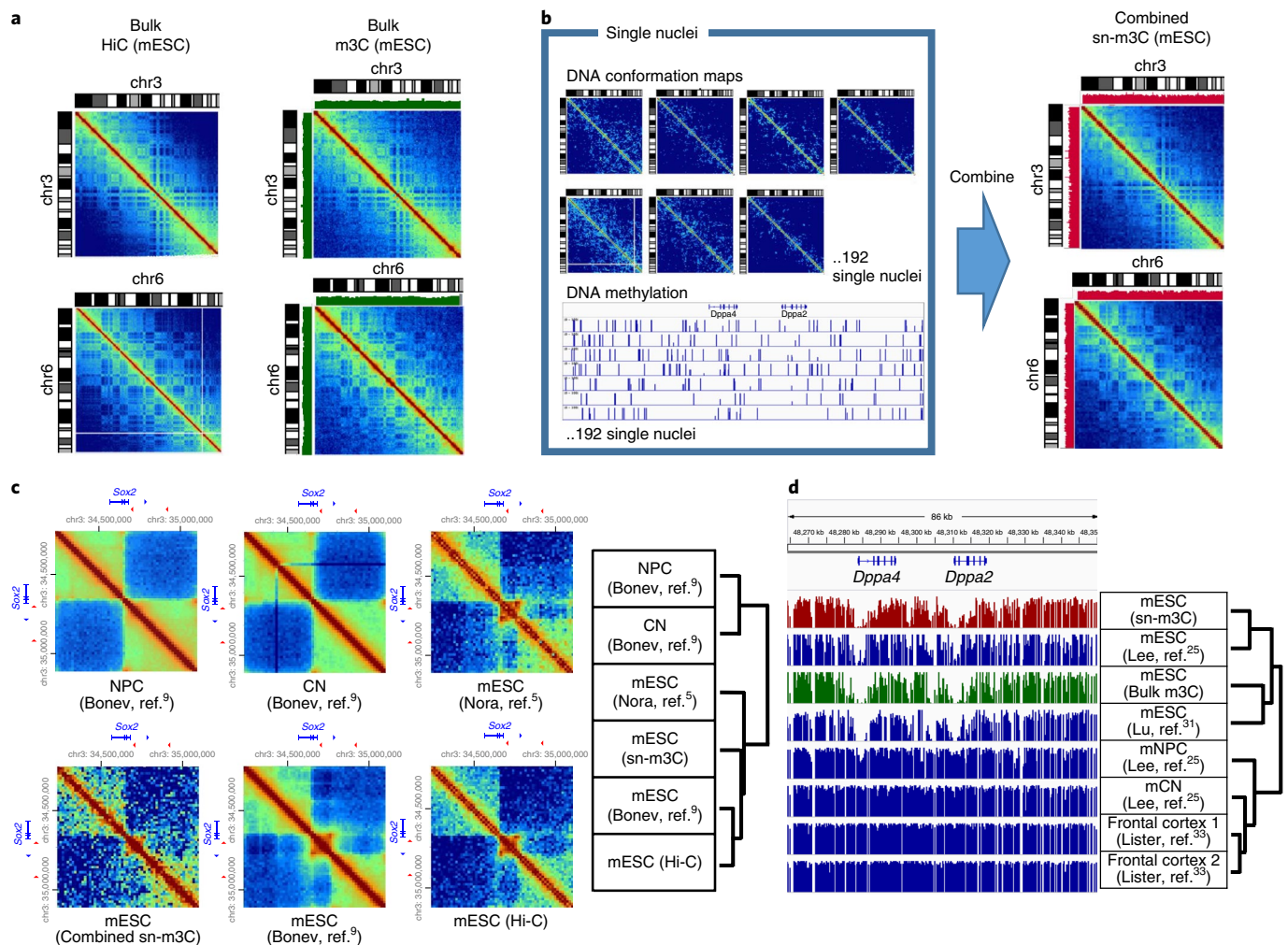


Fig. 3 | Bulk and single-nucleus m3C-seq of mouse embryonic stem cells. **a**, Comparison of Hi-C and bulk m3C-seq chromatin contact profiles. Green bar plot shows CpG methylation level from m3C-seq. **b**, Reconstructed mESC chromatin conformation map from sn-m3C-seq profiles compared to Hi-C or bulk m3C-seq. Red bar plot shows CpG methylation level from sn-m3C-seq. **c**, Bulk and single-nucleus m3C-seq chromatin contact profiles of the Sox2 locus in mESCs compared to published Hi-C data generated from mESC, cortical neurons (CN) and NPC. **d**, Bulk and single-nucleus m3C-seq DNA methylation profiles at Dppa2/4 locus compared to published methylome data generated from mESCs, mouse cortical neurons and frontal cortex. The experiment was repeated twice independently with similar results.

are associated with differential methylation patterns with matched cell-type specificity.

Higher-order chromatin structure is regulated by an interplay of genomic architectural proteins⁴¹, including the methylation sensitive DNA-binding protein CTCF^{42,43}. We examined whether differential interacting sites also showed variable methylation of the CTCF motif within CTCF binding sites defined by neuronal chromatin immunoprecipitation (ChIP)-seq⁴⁴. Within each cluster of differential interacting regions, CTCF binding sites were generally hypomethylated in the corresponding cell types showing increased chromatin interaction frequency (Fig. 6d).

We also investigated whether differential methylation of the methyl sensitive base at position 4 (refs. 42,43) in the CTCF motif is associated with differential chromatin interactions. We identified CTCF motifs where position 4 showed variable cytosine methylation across the 14 cell types (highest methylation >80%, lowest <20%). Only a small minority (1,141/57,740) of CTCF motifs showed variable methylation, indicating that a minority of CTCF binding sites may be subject to regulation by variable DNA methylation. One possible reason that such a small portion of CTCF motifs show variable methylation is that variably methylated CTCF motifs are more likely

to contain a CpG dinucleotide at position 4–5 relative to the genome-wide occurrence of CTCF motifs (Supplementary Fig. 14a), despite the fact that such CpG containing CTCF motifs represent the minority total CTCF motif occurrences in the genome (Supplementary Fig. 14b, 10.99%). We observed that motifs that have variable methylation of position 4 are more likely to be found in variable interacting regions of the genome (Supplementary Fig. 14c, $P = 1.7 \times 10^{-6}$, two-sided Fisher's exact test). These results indicate that a portion of variable interacting regions may be regulated by differential methylation of the CTCF motif, and underscore the importance of multi-omic profiling mC and chromatin conformation.

Finally, we examined the relationship between differential domain boundaries and mC. Here, 73% of the differential domain boundaries colocalize with differential interaction anchors ($P < 1 \times 10^{-300}$, two-sided hypergeometric test), and 46% of the differential domain boundaries overlap with DMRs ($P < 1 \times 10^{-50}$, two-sided hypergeometric test). Within a given cell type, we found the mC levels of CTCF motifs located at domain boundaries has significantly lower mC levels compared with non-boundary sites (Supplementary Fig. 15c). Genes whose transcription start sites (TSS) locate to within 2 kb of the boundaries also showed depletion

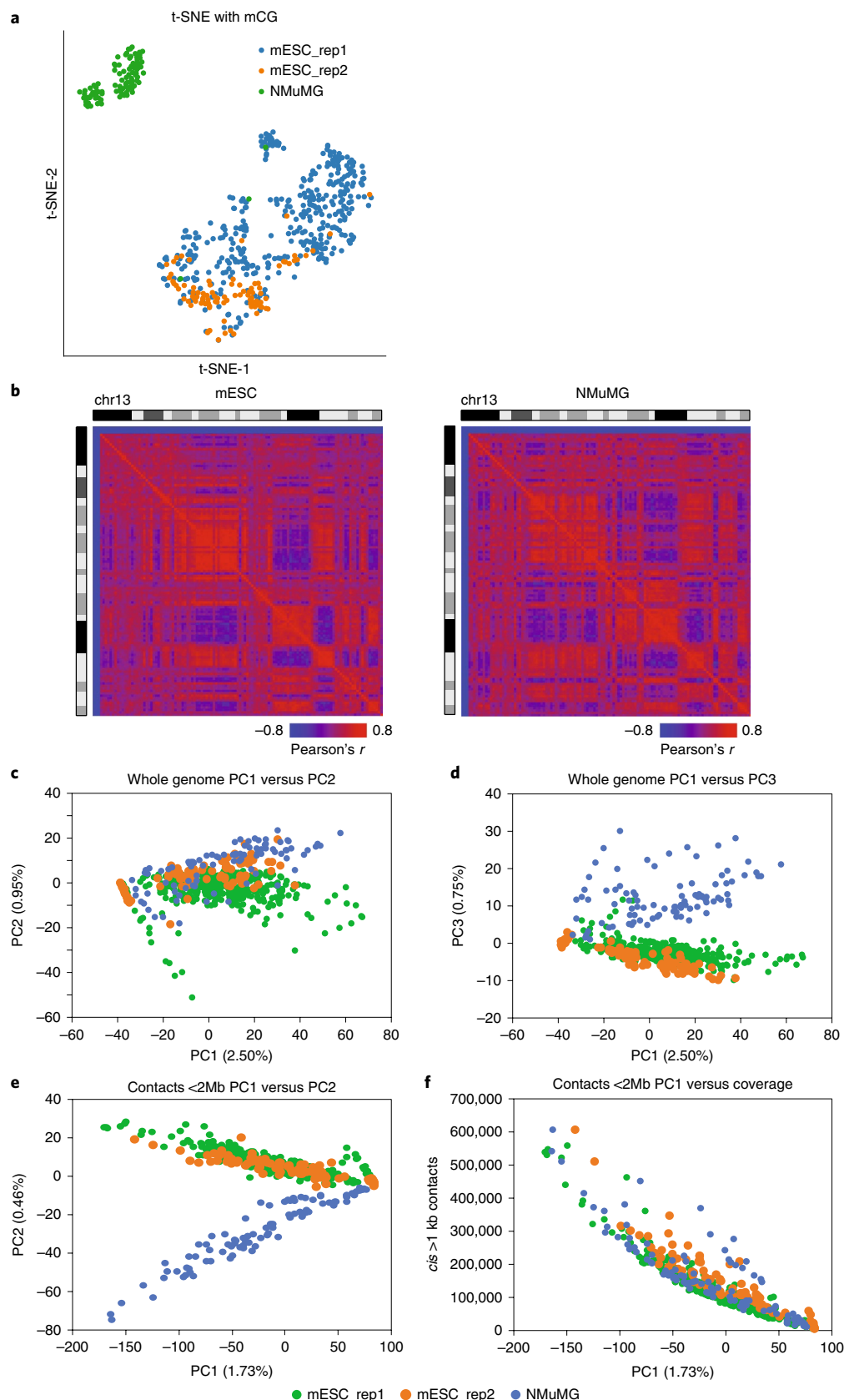


Fig. 4 | Single-nucleus m3C-seq reconstructs cell-type specific chromatin conformation maps. **a**, t-SNE of single-cell mC profiles of mESCs and NMuMG cells. **b**, Chromosome-wide Pearson's correlation matrix from pooled sc-m3C-seq maps for ES cells and NMuMG cells. **c,d**, PCA of whole-genome contact matrices from sc-m3C-seq from ES from NMuMG cells (percentages of variance are marked on the axis). PC1 and PC2 are shown in **c**; PC1 and PC3 are shown in **d**. **e**, PCA of local interactions (<2 Mb) from sc-m3C-seq data from NMuMG cells showing PC1 and PC2. **f**, Correlation of PC1 and per cell contacts. For **a** and **c–f**, $n=2$ independently prepared mESC cultures were analyzed. The two mESC replicates each contained 379 and 93 cells. One ($n=1$) replicate of NMuMG cells containing 96 cells was analyzed.

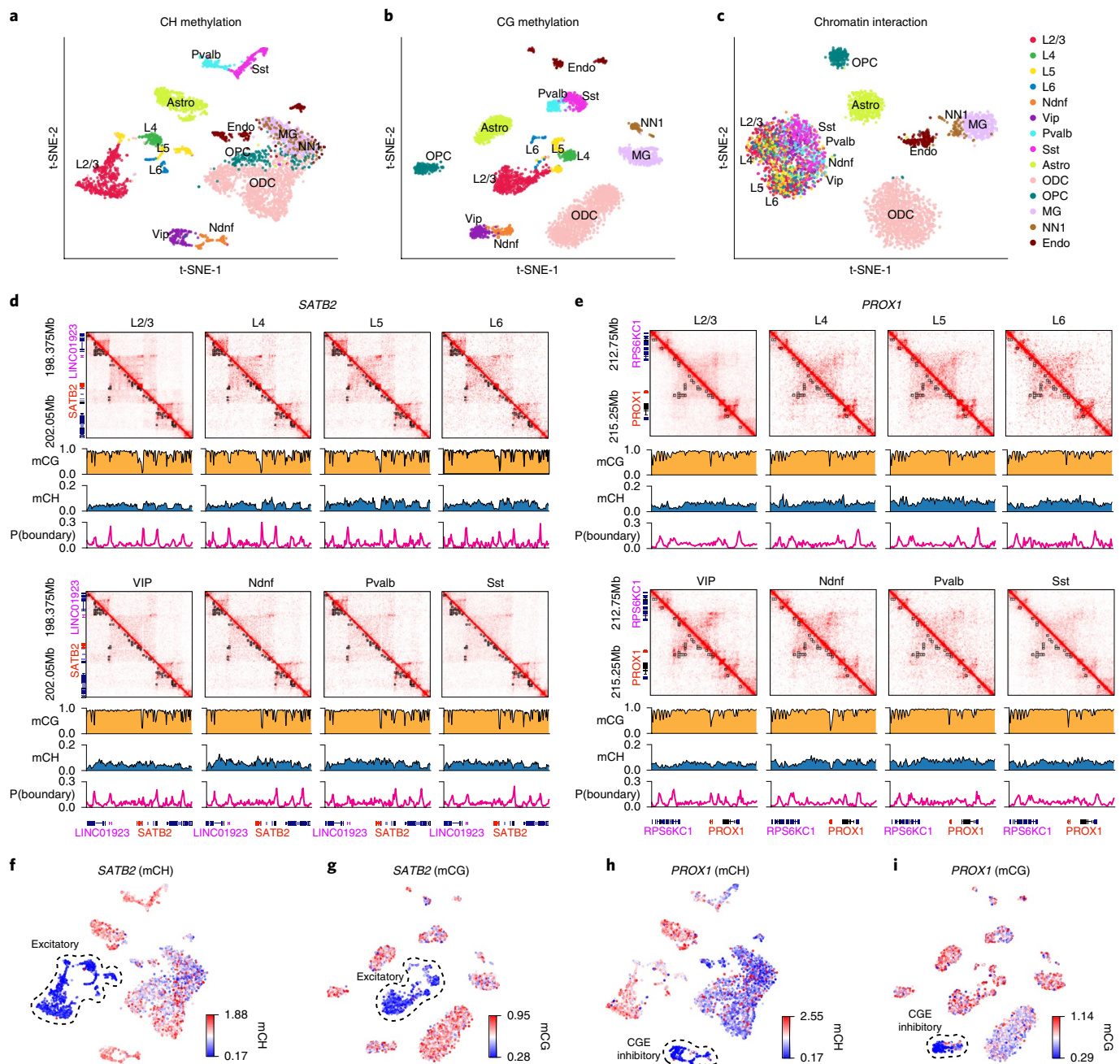


Fig. 5 | Single-nucleus m3C-seq in human brain PFC. **a–c**, Dimension reduction (t-SNE) visualization of single human PFC cells using mCH (**a**) and mCG (**b**) of non-overlapping 100 kb genomic bins, or chromatin interaction at 1Mb resolution (**c**). L2/3, L4, L5 and L6, excitatory neuron subtypes located in different cortical layers. Ndnf and Vip, CGE-derived inhibitory subtypes. Pvalb and Sst, medial ganglionic eminence-derived inhibitory subtypes. Astro, astrocyte. ODC, oligodendrocyte. OPC, oligodendrocyte progenitor cell. MG, microglia. NN1, non-neuronal cell type 1. Endo, endothelial cell. **d,e**, looping between the *SATB2* and *LINC01923* locus in excitatory neuron (L2/3, L4, L5 and L6) (**d**) and chromatin looping between *PROX1* and *RPS6KC1* region in CGE-derived inhibitory cell types—Vip and Ndnf (**e**). **f,g**, mCH (**f**) and mCG (**g**) levels at *SATB2* locus in excitatory neuron clusters. **h,i**, mCH (**h**) and mCG (**i**) levels at *PROX1* locus in CGE-derived inhibitory neuron clusters. All analyses were performed with 4,238 sn-m3C-seq profiles generated from $n=2$ independent human specimens.

of methylation at their gene bodies compared to the genes at non-boundary sites (Supplementary Fig. 15d), indicating that the gene is more likely to be active when a TAD boundary is identified at their promoters. Taken these together, we have observed strong correlations between 3D genome interactions and mC.

Discussion

Cell-type specific chromatin conformation maps can potentially provide a valuable addition to other single-cell modalities for the

creation of cell-type atlases⁴⁵. This information complements single-cell transcriptomes and the annotation of regulatory elements using single-cell epigenomic profiles, to provide a more comprehensive description of gene regulatory activities. However, it is currently unclear how well single-cell Hi-C/3C methods alone can distinguish unique cell types in a heterogeneous population. To enhance the cell-type signature in single-cell chromatin conformation data, we devised a method to allow jointly profiling of chromatin interaction and mC from a single nucleus. Consistent with previous single-cell

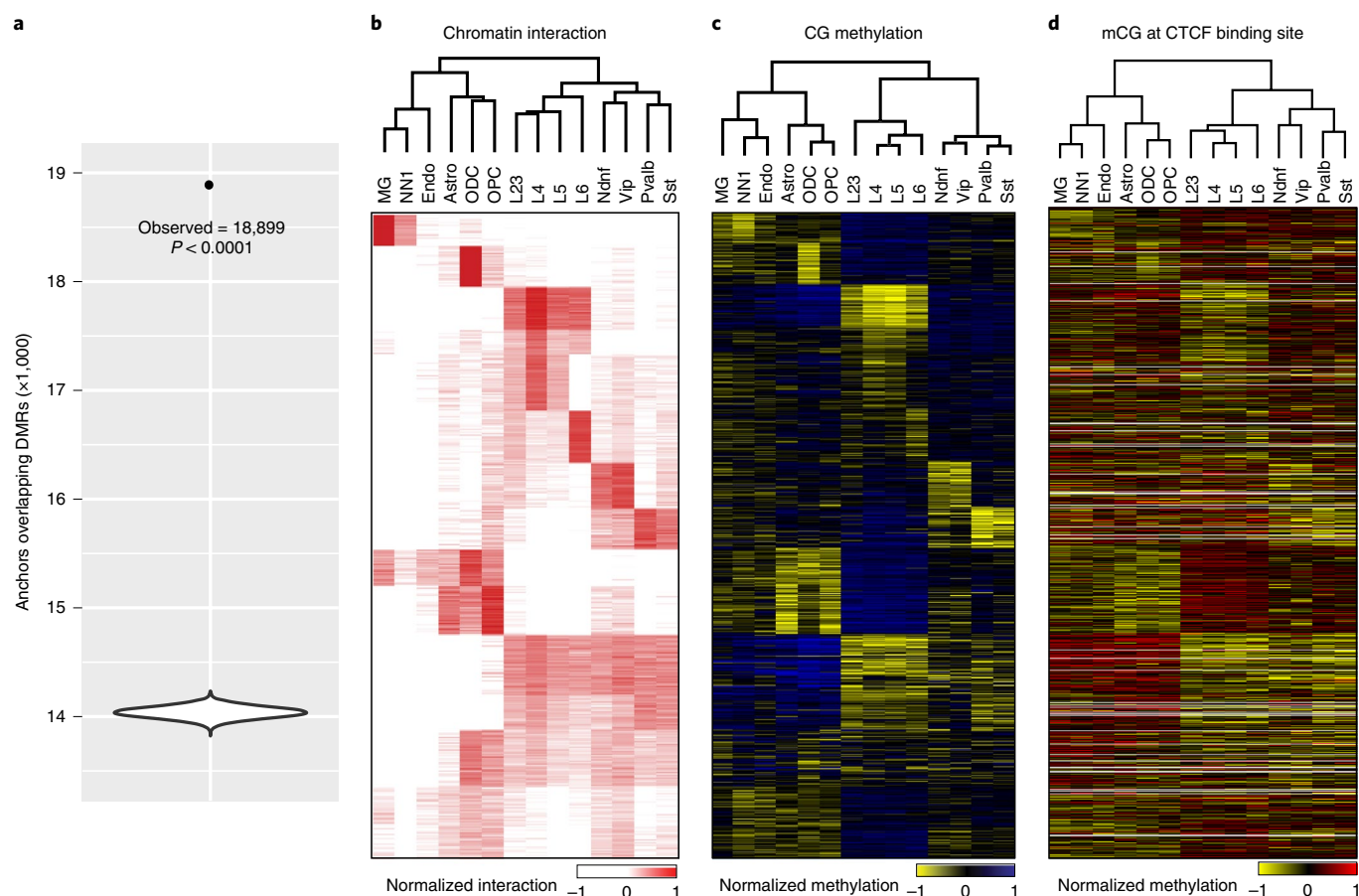


Fig. 6 | Differential mC signature associated with cell-type specific chromatin interactions. **a**, The violin plot of the distribution of the overlap between permuted differential interacting region anchor sites and DMRs; the labeled point indicates the observed overlap. Violin plot elements, maximum = 14,234; minimum = 13,822; mean = 14,038.7 (DMRs, $P < 0.0001$, two-sided permutation test, $n = 10,000$ permutations). **b–d**, Heatmap visualization of cell-type specific chromatin interaction (**b**), CG methylation at anchor regions (**c**) and CG methylation at CTCF binding sites overlapping with the anchor regions (**d**).

methylome studies, sn-m3C-seq allows unequivocal clustering of cell types, which can then guide the reconstruction of high-quality cell-type specific chromatin conformation maps.

Our results indicate that single-cell contact profiles alone can distinguish between drastically different cell types such as mESC and NMuMG. However, the confidence in cell-type separation is highly dependent on sample coverage and downstream processing. Indeed, in the human PFC sn-m3C-seq dataset, it is possible to use contact maps to distinguish between non-neuronal cells and neurons, but contact maps alone do not robustly distinguish neuronal subtypes beyond coarsely separating excitatory and inhibitory cells. Our strategy of using mC signatures to define 14 cell types from human PFC followed by the identification of cell-type specific chromatin interactions clearly demonstrated the advantage of our multi-omic approach.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at <https://doi.org/10.1038/s41592-019-0547-z>.

Received: 28 December 2018; Accepted: 2 August 2019;

Published online: 09 September 2019

References

- Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: the unit of chromosome organization. *Mol. Cell* **62**, 668–680 (2016).
- Rowley, M. J. & Corces, V. G. The three-dimensional genome: principles and roles of long-distance interactions. *Curr. Opin. Cell Biol.* **40**, 8–14 (2016).
- Dekker, J. & Heard, E. Structural and functional diversity of Topologically Associating Domains. *FEBS Lett.* **589**, 2877–2884 (2015).
- Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
- Sexton, T. et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
- Phillips-Cremins, J. E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
- Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
- Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e24 (2017).
- Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
- Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.* **17**, 2042–2059 (2016).
- Nagano, T. et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
- Nagano, T. et al. Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat. Protoc.* **10**, 1986–2003 (2015).
- Nagano, T. et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
- Liu, J., Lin, D., Yardimci, G. G. & Noble, W. S. Unsupervised embedding of single-cell Hi-C data. *Bioinformatics* **34**, i96–i104 (2018).
- Luo, C. et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).

17. Hui, T. et al. High-resolution single-cell DNA methylation measurements reveal epigenetically distinct hematopoietic stem cell subpopulations. *Stem Cell Rep.* **11**, 578–592 (2018).
18. Lee D. S., et al. Single-cell multi-omic profiling of chromatin conformation and DNA methylation. *Protoc. Exch.* <https://doi.org/10.21203/rs.2.11454/v1> (2019).
19. Sajjan, S. A. & Hawkins, R. D. Methods for identifying higher-order chromatin structure. *Annu. Rev. Genom. Hum. Genet.* **13**, 59–82 (2012).
20. Luo, C. et al. Robust single-cell DNA methylome profiling with snmC-seq2. *Nat. Commun.* **9**, 3824 (2018).
21. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
22. Pedersen, B. S., Eyring, K., De, S., Yang, I. V. & Schwartz, D. A. Fast and accurate alignment of long bisulfite-seq reads. Preprint at <https://arxiv.org/abs/1401.1129v2> (2014).
23. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997> (2013).
24. Habibi, E. et al. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).
25. Lee, D.-S. et al. An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator. *Nat. Commun.* **5**, 5619 (2014).
26. Yang, T. et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **27**, 1939–1949 (2017).
27. Gravina, S., Dong, X., Yu, B. & Vijg, J. Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol.* **17**, 150 (2016).
28. Yu, B. et al. Genome-wide, single-cell DNA methylomics reveals increased Non-CpG methylation during human oocyte maturation. *Stem Cell Rep.* **9**, 397–407 (2017).
29. Clark, S. J. et al. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781 (2018).
30. Smallwood, S. A. et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014).
31. Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. & Zhang, Y. Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev.* **28**, 2103–2119 (2014).
32. Lee, S.-M. et al. Intragenic CpG islands play important roles in bivalent chromatin assembly of developmental genes. *Proc. Natl Acad. Sci. USA* **114**, E1885–E1894 (2017).
33. Lister, R. et al. Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
34. Ramani, V. et al. Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).
35. Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
36. Tan, L., Xing, D., Chang, C.-H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924–928 (2018).
37. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017).
38. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
39. Zhou, J. et al. Robust single-cell Hi-C clustering by convolution- and random-walk-based imputation. *Proc. Natl Acad. Sci. USA* **116**, 14011–14018 (2019).
40. Miyoshi, G. et al. Prox1 regulates the subtype-specific development of caudal ganglionic eminence-derived GABAergic cortical interneurons. *J. Neurosci.* **35**, 12869–12889 (2015).
41. Merkenschlager, M. & Nora, E. P. CTCF and cohesin in genome folding and transcriptional gene regulation. *Annu. Rev. Genom. Hum. Genet.* **17**, 17–43 (2016).
42. Wang, H. et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–1688 (2012).
43. Zimmermann, B., Bilusic, I., Lorenz, C. & Schroeder, R. Genomic SELEX: a discovery tool for genomic aptamers. *Methods* **52**, 125–132 (2010).
44. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
45. Regev, A. et al. Science forum: the human cell atlas. *eLife* **6**, e27041 (2017).

Acknowledgements

This work was supported by the NIH (grant nos. 5U19MH114831 and 5R21HG009274 to J.R.E. and DP5OD023071 to J.R.D.). J.R.E. is a Howard Hughes Medical Institute investigator. J.R.D. is also supported by the Leona M. and Harry B. Helmsley Charitable Trust (grant no. 2017-PG-MED001) and a grant from the Salk Institute Innovation Research Fund. This work was also supported by the Flow Cytometry Core Facility of the Salk Institute with funding from NIH-NCI CCSG (grant no. P30 014195). We would like to thank the ENCODE consortium and the laboratory of M. Snyder from the Department of Genetics, Stanford University for the generation of CTCF ChIP-seq data used in this manuscript (GSE127577, ENCODE accession ENCSTR822CEA).

Author contributions

J.R.E., J.R.D. and C.L. conceived the study. J.R.E. and J.R.D. oversaw the study. J.R.D. and C.L. designed the strategy. S.C., A.R., A.B., J.R.N., C.F. and C.O. performed the experiments. D.S.L., J.Z. and C.L. analyzed the data. C.L. and J.R.D. drafted the manuscript. D.S.L., J.Z. and J.R.E. edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41592-019-0547-z>.

Reprints and permissions information is available at www.nature.com/reprints.

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

Peer review information: Nicole Rusk was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Methods

Cell culture. The mESCs (E14TG2a) were purchased from American Type Culture Collection (ATCC CRL-1821). ESCs were grown in DMEM media (Corning 10-013-CV) supplemented with 10% HyClone FBS (Fisher SH3007003E), 1× MEM non-essential amino acids (ThermoFisher 11140050), 1× glutamax supplement (ThermoFisher 35050061), 1× β-mercaptoethanol (Millipore ES-007-E), 100 U ml⁻¹ penicillin-streptomycin (ThermoFisher 15140122) and 1,000 U ml⁻¹ leukemia inhibitory factor (Millipore ESG1107). ESCs were cultured in feeder free conditions on 0.5% gelatin coated plates.

GM12878 cells were obtained from Coriell Institute for Medical Research. GM12878 cells were grown in RPMI-1640 medium (ThermoFisher 11875093) supplemented with 15% fetal bovine serum (Corning 35-010-CV) and 100 U ml⁻¹ penicillin-streptomycin (ThermoFisher 15140122).

NMuMg cells (RBRC-RCB2868) were obtained from the RIKEN BioResource Center. NMuMg cells were grown in DMEM (Corning 10-013-CV) with 10% fetal bovine serum (Corning 35-010-CV), 10 μg ml⁻¹ Insulin (ThermoFisher 12585014) and 100 U ml⁻¹ Penicillin-Streptomycin (ThermoFisher 15140122).

All cell lines were routinely tested for mycoplasma contamination and tested negative.

Human brain tissue. Postmortem human brain biospecimens were obtained from the National Institutes of Health (NIH) NeuroBioBank at the University of Maryland Brain and Tissue Bank. sn-m3C-seq was applied to BA10 cortical tissues of a 21-year-old Caucasian male (UMB5577) with a postmortem interval of 19h, as well as a 29-year-old Caucasian male (UMB5580) with a postmortem interval of 8h.

In situ Hi-C and 3C. In situ Hi-C was performed as previously described using the MboI restriction enzyme⁸. In situ 3C experiments were performed based on the in situ Hi-C protocol with minor modifications. Briefly, before fixation, adherent cells were trypsinized, counted and collected by centrifugation; suspension cells were counted and collected by centrifugation. Cells were resuspended in culture media at a concentration of 1 × 10⁶ cells per ml of media and fixed in 1% formaldehyde for 10 min at room temperature with shaking. For standard species mixture experiments, equal numbers of mouse and human cells were combined into a single tube before fixation. For the 1:10 dilution species mixture experiment, cells were resuspended at a concentration of 1 × 10⁵ cells per ml of media before fixation. For the species mixture experiments where samples were mixed after fixation, each cell type was fixed independently as described above and combined at later stages in the protocol. In situ Hi-C samples were digested with the MboI restriction enzyme and processed as described previously. For in situ 3C experiments, samples were digested with DpnII or NlaIII enzyme overnight at 37 °C with gentle mixing. The following day, the sample was incubated at 62 °C for 10 min to inactivate the restriction enzyme. The typical biotin fill in step in the Hi-C protocol was omitted. The sample was then ligated for 4 h at room temperature with T4 DNA ligase in the same manner as in in situ Hi-C experiments. The sample was then stained with Hoechst (0.1 μg μl⁻¹) for the final 30 min of the ligation step. The sample was then passed through a 40 μm nylon cell strainer (Corning 431750) into a fluorescence-activated cell sorting tube before sorting. As a quality control step, 10% of the sample was taken for conventional library preparation and sequenced using shallow sequencing on a MiSeq.

FANS. FANS was performed at the Salk Institute Flow Cytometry Core Facility using a BD Influx cell sorter. A 100 μm nozzle tip was used, with 1× PBS as sheath fluid (sheath pressure was set to 18.5 psi) with sample and collection cooling set to 4°. The gating strategy for selecting intact, single, Hoechst labeled nuclei from debris was as follows: nuclei were first gated based on forward scatter and side scatter pulse height, then multiplet exclusion gating was applied (forward scatter and side scatter pulse width). Finally, nuclei of specific DNA content were selected (for example, 2N) by virtue of Hoechst fluorescence intensity. Individual nuclei were deposited into wells of 384-well plate using the Single Cell (1-drop single) mode. In preparation for 384-well plate deposition, 20–30 particles (for example, calibration beads) were sorted onto a transparent plastic plate cover for alignment calibration. Then, 20–30 particles are then directly sorted into the wells for final visual confirmation of alignment precision.

Bulk and single-cell methylome library preparation. Libraries for bulk and single-cell methylomes were generated using snmC-seq2. A detailed step-by-step bench protocol for snmC-seq2 is provided as supplementary methods in Luo et al.²⁰. Bulk methylome libraries were prepared manually using individual tubes. Single-cell methylome libraries were prepared using a Tecan Evo 100 robotic platform as described in Luo et al.²⁰. Libraries for mESC and NMuMG samples were sequenced using Illumina MiSeq and HiSeq 4000 instruments in PE150 mode. Libraries for human PFC sample were sequenced using Illumina HiSeq 4000 and Novaseq 6000 instruments in PE150 mode.

Data processing. The mESC and NMuMG datasets were mapped to mm10 reference genome. GM12878 data was mapped to hg38 reference genome.

Human PFC data were mapped to hg19 reference genome. C to T converted and G to A converted reference genomes were prepared for each reference genome using `bismark_genome_preparation`. The first (upstream) 25 base pairs and last (downstream) 3 bp were trimmed from both read1 and read2 to remove random primer sequence and the low complexity tail introduced by Adaptase. Read1 and read2 were mapped separately using Bismark with Bowtie1 with read1 as complementary (always G to A converted) and read2 (always C to T converted) as original strand^{21,46}. After the initial ungapped alignment, unmapped reads were split into three subreads by 40, 32 and 40 bp after removing 5 bp of both ends results in having six reads (the resulting subread IDs were converted to 1-1,1-2,1-3,2-1,2-2,2-3 for the later steps). Six subreads derived from unmapped reads were mapped separately using Bismark with Bowtie1. All aligned reads were merged into a BAM file using `picard` and were sorted by query name. The fragments with all the mapped reads aligned to the same positions were considered as duplicates and removed before `allelic` files were generated. For each fragment, the outermost aligned reads were chosen for the chromatin conformation map generation.

Contact matrix generation and visualization. From the contact files, Cooler was used for generating the contact matrix for different sized bins and Hiclass was used for visualization^{47,48}.

1. Generate fixed-width genomic bins:
`cooler makebins reference_chrNameLength_file BINSIZE > Genomic_bin_file`
2. Sort and index a contact list:
`cooler csort --nproc 2 -c1 2 -p1 3 -c2 4 -p2 5 -o output_file input_file reference_chrNameLength_file`
3. Cool file generation:
`cooler cload pairix Genomic_bin_file input_file output_file`

Comparison of m3C CpG methylation data with published WGBS data. We downloaded six publicly available WGBS datasets of different cell types including mESC, mCN, mNPC and fetal mouse brain tissues^{25,31,33,49}. The methylation level of CpG sites was computed after merging coverage from strands. Average methylation level between 2 kb upstream and 2 kb downstream of known gene TSS were computed ($n=63,759$). Complete linkage hierarchical clustering with Euclidean distance was performed based on Pearson's correlation coefficients (Fig. 3d and Supplementary Fig. 2b).

Comparison of m3C chromatin conformation data with published Hi-C data. We downloaded four publicly available Hi-C datasets of different cell types including mNPC, mCN and mESC^{3,37}. Genomic contact matrix at 1 Mb resolution was generated for each dataset. SCCs were calculated using HiCRep for intrachromosomal interactions across the whole genome²⁶ (Fig. 3c and Supplementary Fig. 2a).

Comparison of bulk m3C-seq and sn-m3C-seq data with published methylome datasets. The technical characteristics of bulk m3C-seq were compared to a mESC WGBS dataset (SRX202087)²⁴. Fastq files downloaded from SRX202087 were mapped to mm10 reference genome using Bismark with Bowtie2 aligner. The resulting BAM file was downsampled to match the coverage of bulk m3C-seq data. Published single-cell methylome datasets were downloaded from SRP069120, SRP062328 and SRP131024 (refs. 27–29). The fastq files were mapped to mm10 reference genome using Bismark with Bowtie2 aligner. `Preseq`⁵⁰ was used to estimate library complexity using forward reads with `Preseq gc_extrap` function with options `-e 5e+09 -s 1e+0712`. Library complexity values shown in this study were estimated for the sequencing depth of 50 million read pairs. To determine the enrichment of CpG islands in single-cell methylome data, the fraction of CpG islands on mouse chromosome 1 covered by a single-cell methylome was compared to shuffled regions with matching sizes. The shuffling was carried out using `bedtools shuffle` and was repeated five times and the average fraction of regions covered by reads was used. Bulk MethylC-seq data downsampled to 1 million non-clonal reads for this analysis. For computing the number of genomic regions covered by reads at different sequencing coverage, 1 and 10 kb bins were generated using `bedtools makewindows` across the mouse genome. The bins were intersected with bulk MethylC-seq and single-cell methylomes downsampled to 100,000 to 1.5 million reads.

Comparison of sn-m3C-seq chromatin conformation data with published single-cell Hi-C data. We downloaded multiple previous single-cell Hi-C datasets^{12,14,34–36,51} to compare with our sn-m3C-seq method. To allow for an unbiased comparison across methods, we processed each dataset uniformly using previously described alignment pipeline⁵². For each dataset, we quantified multiple metrics of data quality, including the number of reads sequenced per cell, the number of mapped pairs per cell, the PCR duplication rate per cell and the fraction of reads that align as short *cis*-fragments (<1 kb), long-range (>1 kb) *cis*-fragments and interchromosomal pairs. We use a threshold of 1 kb as a cutoff for defining *cis*-contacts to eliminate any possibility that two paired reads align to different restriction fragments as a result of either failed digestion or re-ligation.

For the Ramani et al.²⁴ dataset, we only used data deposited in Gene Expression Omnibus (GEO), specifically the ML3 dataset (GSM2254217). Also, since the Ramani dataset is a species mixture experiment, we aligned reads to both the human and mouse genomes and only considered reads if they aligned to either the human or mouse genome, but not both. We also noted that Flyamer et al.³⁵ perform multiple additional filtration steps after alignment due to the use of whole-genome amplification (WGA) to limit the possibility that a given ligation fragment is represented twice in a single cell due to WGA. We did not perform similar filtration steps, as we believed that using a single analytical pipeline as opposed to bespoke sample specific filtration was the least biased approach to compare across datasets and methods. Therefore, our pipeline reports more contacts per cell (2,416,802 long-range *cis*-contacts) than is reported by Flyamer et al.³⁵ (1,900,000) in their manuscript.

Cell-type identification using DNA methylation signature. CG methylation levels (mCG) are computed for every non-overlapping 100 kb bins across the genome in each single cell. The bins with more than 20 CG basecalls in more than 90% of cells were selected for further analysis. Bin-level mCG levels were normalized by global mCG of each cell. Similar to our previous analysis using snmC-seq, we imputed the mCG in each bin with fewer than 20 CG basecalls by using the mean mCG of that bin across all the cells having more than 20 CG basecalls in the bin¹⁶.

Cell-type identification using 3D genome structure. We generated a contact map using 100 kb bins in each cell. The interaction frequency of each bin is normalized by dividing the average interaction frequency of the bins at the same distance interactions. The bins that are covered with more than 100 cells ($n = 19,357$) were used for the PCA analysis shown in Fig. 4c,d. The bins with an interaction distance of less than 2 Mb ($n = 18,004$) were used for the PCA analysis shown in Fig. 4e,f.

Quality control for human cortical sn-m3C-seq profiles. We filtered the cells by total non-clonal reads >500,000, global mCCC <3%, total autosomal cytosines covered <100 M and total long-range (>10,000) *cis*-contacts >5,000. We also required total long-range *cis*-contacts of each chromosome >X, for a chromosome of length × Mb.

Visualization and clustering of human cortical cells based on methylation. Both CG and non-CG methylation level (mCG and mCH) were computed for every non-overlapping autosomal 100 kb bins. Bin-level mCG levels were normalized by global mCG of each cell. For each individual and each sequencing batch, the bins with more than 20 CG basecalls in more than 90% of cells were selected for further analysis. We imputed the mCG in each bin with fewer than 20 CG basecalls by using the mean mCG of that bin across all the cells having more than 20 CG basecalls in the bin. The mCG matrices of different individuals and batches were integrated together using Scanorama³³ using all the bins with default parameters. The first 50 dimensions of Scanorama embedding were used for t-SNE visualization and clustering. For clustering, we used the Euclidean distance of the embedding to generate the binary k -nearest neighbor graph A of all the cells with $k = 20$, where A_{ij} is 1 if cell j is one of the 20 nearest neighbors of cell i . Then A was weighted by the jaccard similarity matrix of A ⁵⁴. Specifically, the weight of the edge between cell i and j in the final graph was the jaccard similarity between A_i and A_j . Louvain clustering was performed on the weighted graph with resolution 1.6. We used mCG to cluster all the 4,238 cells, and then selected all the neurons (MEF2C+) to perform another round of clustering using mCH. The mCH matrices were processed in the same way except the basecall cutoff was set to be 100 in 99% of cells. We merged some of the clusters to have enough cells in each cluster, but still separating the known cell types. The clusters were annotated as known cell types based on the gene body depletion of mCG and mCH.

Visualization of human cortical cells based on chromosome interactions. The contact matrices were generated at 1 Mb resolution. We used scHiCluster for embedding the single-cell intrachromosomal contact matrices with default parameters³⁹. The first 20 dimensions of the embedding were used for t-SNE visualization.

Identification of DMRs. The single-cell methylation profiles at base resolution were merged for each cluster. CG sites from the two strands were merged to enhance the statistic power. DMRs were identified using methylpy DMRfind⁴⁹.

Identification of differential interaction regions. Single-cell contact maps were binned at a resolution of 50 kb. We retained bins if they contain non-zero values in at least 10% of single cells. Differential interactions were called using edgeR. Specifically, each single cell was treated as a replicate for each corresponding cluster. The data were normalized using calcNormFactors and dispersions were estimated using estimateDisp. We performed quasi-likelihood F -tests identifying differentially interacting regions across all samples using glmQLFit and glmQLFTest. Benjamini–Hochberg corrections were applied for multiple testing, and we retained bins with FDR <0.1%. Finally, we applied additional filters

such that we required the maximum cluster-wide average interaction frequency to be at least two-fold higher than the minimum cluster-wide average interaction frequency, and that the percentage of single cells with contacts had to be at least three-fold greater in the highest cluster than the lowest.

Identification of differential domain boundaries. For this analysis, we only used cells with more than 50,000 contacts. The TAD-like structures (TLSs) in single cells were identified using TopDom after scHiCluster imputation at 25 kb resolution³⁹. For the j -th bins, we counted the number of cells where the bin was identified as TLS boundaries in cluster i , denoted as c_{ij} . Then we computed the boundary probability by $p_{ij} = \frac{c_{ij}}{t_i}$, where t_i is the number of cells in cluster i . For each bin, we used the contingency table O to compute the P value by chi-square test, where $O_{i0} = c_{ij}$ and $O_{i1} = t_i - c_{ij}$, and performed multiple test correction using Benjamini–Hochberg procedure. For differential domain boundaries, we required FDR <0.01, $\min p_{ij} < 0.05$ and $\max p_{ij} - \min p_{ij} > 0.1$. To eliminate the effect of limited resolution for TAD identification (which could shift for 1 or 2 bins), we expanded 50 kb on both sides of the selected differential boundaries and repeated the test. We used c_{ij} to denote the number of cells in which one of the 25 kb bins in the 125 kb region surrounding the j -th selected boundary was called TLS boundaries in cluster i , and p_{ij} as the corresponding boundary probability. To filter for the significant differential boundaries, we required FDR <0.01, $\min_i p'_{ij} < 0.3$ and $\max_i p'_{ij} - \min_i p'_{ij} > 0.1$.

Comparison of methylation levels at differential domain boundaries. In each cell type i , we separated 25 kb bins into boundaries ($p_{ij} > 0.15$) and non-boundaries ($p_{ij} < 0.05$). We compared the mCG or mCH level between the boundaries and non-boundaries at (1) those bins, (2) CTCF motifs that overlapping with those bins and (3) the gene bodies whose TSS are within 2 kb of those bins.

CTCF methylation analysis. For analysis of CTCF ChIP-seq binding site methylation, we downloaded data from in vitro differentiated neurons generated by ENCODE (ENCSR822CEA)^{44,55}. Methylation levels were calculated within peak regions. For comparison with differential interacting regions, the CTCF binding site methylation levels were averaged across all CTCF sites within the pair of interacting bins.

For CTCF motif analysis, we used position weight matrices generated by SELEX⁵⁶. The reason we use SELEX defined motifs for this analysis is that we wanted to limit any biases that may result from using ChIP-seq defined motifs due to the possibility of CpG methylation changing the relative likelihood of observing specific motifs due to the known sensitivity of CTCF to bind to methylated CpGs at the fourth position in the motif. Motifs were identified in the genome using Homer scanMotifGenomeWide using a log-odds detection threshold of six⁵⁷. We identified variably methylated site motifs as those that showed at least one cell type with methylation of position 4 >80% and one cell type <20%, where at least ten reads cover the cytosine.

Statistics. The following statistical tests were used in this manuscript: Fisher's Exact test and hypergeometric tests were used to compute associations in contingency tables. Wilcoxon rank-sum tests were used to test for differences between groups and does not make any assumptions regarding the distribution of the underlying data. Pearson's correlations were used to evaluate the linear relationship between samples, in particular related to replicate experiments for reproducibility. The SCC was used to compare Hi-C datasets for reproducibility²⁶. EdgeR was used to analyze differential count data between groups³⁸, namely Hi-C contact frequencies. It assumes the data follows an underlying negative binomial distribution. All statistical tests were two-sided.

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

Data availability

Raw data and processed data for culture mouse cells mESC and NMuMG are available from NCBI GEO accession code GSE124391. Raw data and processed data for human PFC are available from GEO accession code GSE130711. Intermediate files for DNA methylation and chromatin contacts can be downloaded from <https://github.com/dixonlab/scm3C-seq>. An AnnoJ browser for DNA methylation data can be accessed at http://neomorph.salk.edu/snm3Cseq_human_FC.php. A public HiGlass genome browser session for the human PFC data can be accessed from <https://dixon.salk.edu/projects/snm3Cseq/>.

Code availability

The source code used is publicly available at <https://github.com/dixonlab/Taurus-MH> and <https://github.com/dixonlab/scm3C-seq>.

References

- Langmead, B. Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinforma.* **32**, 11–17 (2010).

47. Abdennur, N. & Mirny, L. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btz540> (2019).
48. Kerpedjiev, P. et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125 (2018).
49. Schultz, M. D. et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015).
50. Daley, T. & Smith, A. D. Predicting the molecular complexity of sequencing libraries. *Nat. Methods* **10**, 325–327 (2013).
51. Stevens, T. J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
52. Dixon, J. R. et al. Integrative detection and analysis of structural variation in cancer genomes. *Nat. Genet.* **50**, 1388–1398 (2018).
53. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* **37**, 685–691 (2019).
54. Levine, J. H. et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015).
55. Davis, C. A. et al. The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
56. Jolma, A. et al. DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).
57. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|---|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Raw data and processed data are available from NCBI GEO accession GSE124391.

Data analysis

The source code used is publicly available at <https://github.com/dixonlab/Taurus-MH>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw data and processed data are available from NCBI GEO accession GSE124391.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	One sample each was analyzed for bulk Hi-C, 3C and m3C-seq samples. At least 91 cells were analyzed for any sn-m3C-seq experiments.
Data exclusions	No Data exclusion was performed
Replication	Two replications were performed for each sn-m3C-seq experiment
Randomization	N/A
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging