Joint profiling of multiplex chromatin interactions, gene expression, and RNA-chromatin associations in single cells of the human brain

Authors

Zhifei Luo [1#], Xingzhao Wen [2#], Wenxin Zhao [1], Tri C. Nguyen [1], Xueyi Wan [1], John Lalith Charles Richard [3], Sheng Zhong [1,3,*]

Affiliation

1. Shu Chien-Gene Lay Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA
2. Program in Bioinformatics and Systems Biology, University of California San Diego, La Jolla, CA 92093, USA
3. Institute of Engineering in Medicine, University of California San Diego, La Jolla, CA 92093, USA

# Equal contribution

* Email: szhong@ucsd.edu

Abstract

The dynamically organized chromatin complexes often involve multiplex chromatin interactions and sometimes chromatin-associated RNA. Here, we introduce the Multi-Nucleic Acid Interaction Mapping in Single Cell (MUSIC) technique to enable simultaneous profiling of multiplex chromatin interactions, gene expression, and RNA-chromatin associations within a single nucleus. We applied MUSIC to profile >9,000 single nuclei in the human frontal cortex. MUSIC-derived single-nucleus transcriptomes provide a comprehensive categorization of cortical cell types, subtypes, and cellular states. The genomic sequences of highly expressed genes frequently co-complex with their flanking genomic regions, forming Gene-Expression-Associated Stripes (GEAS), which exemplify the intricate coordination between transcription and chromatin architecture at the single-cell level. Additionally, we observed significant heterogeneity among female cortical cells in the association between the XIST long non-coding RNA (lncRNA) and the X chromosome (XIST-chrX association, quantified as XAL). Cells with high XAL demonstrated a greater difference in spatial organization between the XIST-associated (Xi) and non-associated (Xa) X chromosomes compared to XAL-low cells. Notably, excitatory neurons displayed enrichment in XAL-high cells and exhibited a more pronounced disparity in spatial organization between Xi and Xa compared to other cell types. The MUSIC technique offers a powerful tool for future investigations into chromatin architecture and transcription at a cellular resolution within complex tissues.

Introduction

The three-dimensional (3D) folding of the genome is known to exhibit dynamic changes during cellular differentiation processes and demonstrates heterogeneity among terminally differentiated single cells [1–4]. While its regulatory role in the expression of specific genes has been well-established [5–7], the extent to which the 3D genome structure impacts the expression of most genes remains a topic of debate [8]. Given the pronounced heterogeneity observed in chromatin structure and gene expression across individual cells [9–11], a comprehensive understanding of the relationship between 3D genome structure and gene expression at the single-cell resolution is necessary. Therefore, developing single-cell multimodal technologies capable of simultaneously profiling chromatin conformation and gene expression is instrumental for elucidating these intricate relationships.

Single-cell multi-omic technologies have revolutionized our ability to collectively analyze epigenomic or chromatin conformation features and gene expression [12–16]. Various methodologies have been introduced to enable the joint profiling of specific aspects of the genome, transcriptome, or proteome. For instance, Snare-seq [13] and SHARE-seq [15] allow for simultaneously profiling chromatin-accessible regions (CARs) and gene expression (Table S1). Additionally, scNMT-seq [17] facilitates the combined analysis of CARs and DNA methylation, while scNOMeRe-seq [18] enables the joint examination of CARs, DNA methylation, and gene expression. Techniques such as Pair-tag provide insights into histone modification profiles and transcription levels [19], and sn-m3C-seq captures chromatin interactions and DNA methylation patterns [20]. CITE-seq, on the other hand, allows for the simultaneous measurement of surface protein and transcriptome [21]. Recently, HiRES technology has been introduced to the concurrent application of Hi-C and RNA-seq within individual cells [16] (Table S1). Furthermore, imaging tools like DNA seqFISH+ [10], ORCA [22], and MINA [23] have facilitated the joint analysis of chromatin traces and gene expression (Table S1). Despite these technical advancements, the simultaneous profiling of multiplex chromatin interactions (co-complexed DNA sequences), gene expression, and RNA-chromatin associations from a single cell remain challenging. To fill this gap, we developed the Multi-Nucleic Acid Interaction Mapping in Single Cell (MUSIC) technique, which enables the simultaneous profiling of gene expression and co-complexed DNA sequences with or without co-complexed RNA at the single-cell level.

The architecture of chromatin can encompass both pairwise and multiplex chromatin interactions, highlighting the intricate nature of chromatin complexes [9,24–27]. ChIA-Drop technology has facilitated the mapping of multiplex chromatin interactions at single-complex resolution from bulk cells, revealing that multiplex chromatin interactions are prevalent in *Drosophila* [24]. The MUSIC technique expands the capability to evaluate the composition of pairwise and multiplex chromatin interactions in individual human cells at single-cell resolution.

In addition to DNA, chromatin complexes can also encompass RNA molecules, introducing another layer of complexity to chromatin architecture [28–30]. Chromatin-associated RNA (caRNA) has been shown to contribute to gene expression regulation [29,31,32]. For instance, the accumulation of the XIST long noncoding RNA (lncRNA) on the X chromosome (XIST-chrX association) is crucial for the silencing of one of the two X chromosomes in female cells, a process known as X-

chromosome inactivation (XCI). Various human tissues exhibit both shared and tissue-specific incomplete XCI genes, which are expressed from the silenced X chromosome [33]. Genes with incomplete XCI can display higher expression levels in females, potentially contributing to sex differences in disease susceptibility [34]. Recent advancements have enabled genome-wide mapping of RNA-chromatin associations in bulk cells [29,35–39]. With the application of MUSIC, we can now obtain RNA-chromatin association maps at the single-cell level. Utilizing MUSIC, we uncover cellular heterogeneity in XIST-chrX association levels (XAL) within the female cortex and explore the co-variation between XAL and chromatin interactions among the female cortical cells.

Results

The MUSIC technology

MUSIC technology is a method developed to simultaneously profile gene expression, genomic interactions, and RNA-chromatin associations from the same nucleus. It achieves this through three design goals (Figure 1). Firstly, RNA and fragmented DNA from the same nucleus are sequenced together, identified by a unique Cell Barcode, enabling the matching of RNA-sequencing and DNA-sequencing data originating from the same nucleus. Secondly, a system is designed to distinguish between source molecules, whether RNA or DNA, via the different nucleotide sequences of RNA and DNA Linkers. Lastly, the data should identify any DNA-DNA and RNA-DNA associations, accomplished by tagging each molecular complex with a unique Complex Barcode.

The implementation of these goals is achieved in two steps. The first step involves the ligation of RNA and DNA Linkers to respective molecules and the addition of Cell Barcodes. This labels the RNA and DNA of the same nucleus and distinguishes the RNA and DNA inserts in the sequencing library. An innovation in this design is that the same procedure adds Cell Barcodes to both RNA and DNA, maximizing efficiency.

In the second step, molecular complexes are traced back to each nucleus by their Cell Barcodes, and these are labeled with a unique Complex Barcode. This is done using the 10X Genomics' Gel Bead-in-Emulsion (GEM) system and Index Adaptors, adding both 10X Barcodes and I7 Barcodes to RNA and fragmented DNA. This step prepares the molecular complexes to be constructed into the final sequencing library using PCR.

The final sequencing library is then sequenced, with the combination of the Cell and Complex Barcodes specifying every molecular complex. In the current release, MUSIC technology offers approximately 25 trillion possible barcode combinations, aiding the profiling of gene expression and genomic interactions at an unprecedented level.

Mapping multiplex chromatin interactions from single nuclei of human and mouse embryonic stem cells

We applied MUSIC to analyze a mixed population of H1 human and E14 mouse embryonic stem cells. The resulting mixed-species MUSIC library was sequenced on a NovaSeq platform, generating 3,067,956,666 read pairs. These read pairs resolved 533,233,356 uniquely mapped, non-duplicate, and barcode-complete (containing Cell Barcode, 10X Barcode, I7 Barcode, and either DNA Linker or RNA Linker) (UMNDBC) read pairs. Because as per the experimental design, each UMNDBC read pair contains only one DNA or RNA insert, we will refer to a UMNDBC read pair as a DNA read or an RNA read. This mixed-species dataset resolved 372,878,969 DNA reads and 22,714,916 RNA reads for humans (hg38), 128,687,584 DNA reads and 8,951,899 RNA reads for mice (mm10). All the subsequent analyses are based on these UMNDBC reads.

We used this mix-species dataset to evaluate MUSIC's resolution at single-cell and single-complex levels. First, MUSIC identifies the DNA reads and RNA reads that share the same Cell Barcodes (CB) as coming from the same cell. Based on an established approach and threshold [9], MUSIC resolved 2,546 human cells, 1,381 mouse cells, and 36 cells with mixed species content. This resulted in a species-mixing rate of 0.91% at the cell level (Figure S1a). Comparatively, this species-mixing rate is within the same order of magnitude as the species-mixing rate of 3% reported for scSPRITE under the same thresholds [9]. Next, MUSIC identifies the reads sharing the Cell Barcodes and the Complex Barcodes (CB+10X+I7) as a cluster. In the mixed-species dataset, 154,499,469 clusters were resolved, of which 154,291,286 clusters contained reads from a single species. This corresponds to a complex-level species-mixing rate of 0.13% (Figure S1b). Taken together, the low species-mixing rates at the cell level and the complex level support the ability of MUSIC to generate data at single-cell and single-complex resolutions.

We analyzed the human H1 cells regarding the numbers of reads, clusters, and pairwise contacts. First, the data solved 2,546 human H1 cells. Each H1 cell contained an average of 144,049 UMNDBC DNA reads (blue dots) and 11,384 UMNDBC RNA reads (orange dots) (Figure 2a). Second, clusters were defined as collections of reads sharing the same Cell Barcode and Complex Barcode (CB+10X+I7). In the perfect scenario where cluster-level species-mixing rate = 0, a cluster reflects a molecular complex. A cluster is called a DNA-only, an RNA-only, or an RNA-DNA cluster if it contains only DNA reads, only RNA reads, or both DNA and RNA reads. On average, each H1 cell had 7,036 DNA-only clusters (DD clusters), 232 RNA-only clusters (RR clusters), and 1,170 RNA-DNA clusters (RD clusters) (Figure 2b). Finally, co-complexed multi-way associations were projected to co-complexed read pairs based on a previously described procedure [9], resulting in an average of 2,639,302,084 co-complexed DNA-DNA pairs, 7,089,720 co-complexed RNA-RNA pairs, and 250,525,581 co-complexed RNA-DNA pairs per cell (Figure 2c).

In total, there were 18,144,410 non-singleton DD clusters, accounting for 55,670,578 DNA reads; 324,121 RR clusters, accounting for 835,184 RNA reads, and 2,401,392 RD clusters, accounting for 13,151,716 RNA reads and 216,515,595 DNA reads (Figure 2d). Among the non-singleton DD clusters, 13,111,228 (72.26%) contained two DNA reads that correspond to pairwise interactions,

and 5,033,182 (27.74%) contained three or more DNA reads that respond to multiplex interactions (Figure 2e, Figure S2). In the RD clusters, 1,009,706 (42.05%) contained two reads, i.e., one DNA read and one RNA read, 783,709 (32.64%) contained 3-10 reads, and 607,977 (25.32%) contained more than 10 reads (Figure 2e, Figure S2).

The TAD chromatin structure predominantly consists of multiplex chromatin interactions

We compared MUSIC's ensemble DNA reads with Micro-C data. At the chromosome level (1 Mb resolution), the contact map of DD clusters (lower triangle) reproduced the structures observed in the Micro-C derived contact map (4DN data portal: 4DNFI2TK7L2F, [41]) (upper triangle, Figure 3a). At a finer resolution (50 Kb), DD clusters reflected the Topologically Associating Domain (TAD) structure, however, the different-sized DD clusters exhibited some differences. Small DD clusters (2-10 DNA reads per cluster) primarily contained contacts within TADs (Figure 3b). Middle-sized (11-50 reads per cluster) and large clusters (51-100 reads per cluster) recapitulate the TADs and the nested TAD structure (Figures 3c, d). Large clusters revealed more contacts between the nested TADs within a larger TAD (Arrows, Figure 3d).

To visualize the clusters, we plotted each cluster in a row, with every DNA read of this cluster aligned to their respective genomic coordinates. We ordered the clusters by the genomic coordinates of their leftmost DNA reads. This way, we created a stacked map of the clusters (Figure 3g, h). By comparing the 2-D contact map based on ensemble MUSIC data to the stacked maps, we observed that pairwise interactions (clusters with two DNA reads) alone poorly reflected the TAD structure (Figure 3g), while multiplex interactions (clusters with three or more DNA reads) recapitulated the TAD structure (Figure 3h). This analysis corroborates the difference in the contact maps of different cluster sizes (Figure 3b-d) to suggest that the TAD chromatin structure predominantly consists of multiplex chromatin interactions.

Data from bulk assays revealed that the probability of chromatin interactions (Pc) decreases as the genomic distance (s) between the two DNA fragments increases [40,42]. Furthermore, Pc(s) almost linearly correlates with s at the log-log scale [40,42] (Figure 3e). In MUSIC data, the Pc(s) curve derived from size-2 DD clusters exhibited a linear relationship at the log-log scale (size=2 curve, Figure 3e), consistent with proximity-ligation-based assays. The DD clusters of size 3 or greater also exhibit decreasing Pc(s) curves, which are no longer linear at the log-log scale. The larger the cluster size, the greater Pc(s) deviates from a straight line (Figure 3e). The linear relationship in the log-log scale is expected from a polymer model of chromosome [40]. The deviation suggests that a pure polymer model may not fully explain all the multiplex chromatin interactions. Moreover, compared to the pairwise contacts, the multiplex interactions displayed higher contact frequencies at submegabase to several megabase genomic distances, indicating enrichment of long-range chromatin interactions in the multiplex complexes.

MUSIC's RNA reads and RNA-chromatin associations correlate with bulk assay

We compared MUSIC's ensemble RNA reads (RNA ensemble) from 2,546 H1 cells to RNA measurements obtained from two bulk assays in H1. Using all 60,719 genes defined in GENCODE v36, we quantified the RNA level of each gene in terms of reads per kilobase (RPK).

The RPKs of MUSIC's RNA ensemble correlated with those of bulk RNA-seq (ENCSR000COU [43]) (Figure 3i, rho = 0.8, p-value = 2.2e-16; Figure 3k). Furthermore, iMARGI is a bulk assay of RNA-chromatin interactions, where the collection of the RNA reads from iMARGI (iMARGI's RNA reads) measures the transcriptome in the nuclei [30]. MUSIC's RNA ensemble also correlated with those of iMARGI's RNA reads (Figure 3j, rho = 0.9, p-value = 2.2e-16). This indicates that the gene expression levels quantified by MUSIC's ensemble RNA reads are consistent with those obtained from bulk RNA assays. Additionally, the RNA strands detected by MUSIC were consistent with the RNA strands observed in the bulk assay (Figure 3k).

Previous bulk assays studies have identified two key features of genome-wide RNA-chromatin associations. First, genomic regions with high levels of locally transcribed premature mRNA (pre-mRNA) (pre-mRNA-rich regions) tend to correspond to the A compartments defined by chromatin conformation assays such as Hi-C and Micro-C [29,44]. Second, these pre-mRNA-rich regions also serve as hotspots for long-range RNA-chromatin interactions involving non-coding RNAs MALAT1, 7SK, and snRNAs (collectively called nsaRNAs) [44]. In the MUSIC-identified RNA-DNA contacts (ensemble RD), the pre-mRNA associated genomic regions correlated with Micro-C derived A compartment, and more specifically with the "Speckle" nuclear compartmentalization jointly inferred from TSA-seq, DamID, and Hi-C data [45] (Figure 3l). To categorize an RNA read as pre-mRNA, we required it to span an intron-exon junction and contain at least 15 intronic nucleotides. Furthermore, the genomic regions targeted by long-range nsaRNA-chromatin associations (blue curve) overlapped with regions associated with high levels of locally transcribed pre-mRNA (p-value < 2.2e-16, red curve, Figure 3l, m). Thus, MUSIC recapitulated both prominent features of genome-wide RNA-chromatin associations.

A single-cell map of transcriptome and DD/RD complexes in the human frontal cortex

We generated a MUSIC dataset on 14 postmortem samples of the human frontal cortex (FC) from tissue donors aged 59 and above (Table S4). This dataset resolved 9,087 single nuclei, 755,123,054 UMNDBC DNA reads and 29,319,780 UMNDBC RNA reads (hg38). This dataset is referred to as the MUSIC FC dataset.

To evaluate the performance of MUSIC compared to other technologies, we compared the number of resolved nuclei and the number of identified DNA-DNA read pairs per nucleus. We compared MUSIC to published single-cell multi-omic datasets of scHiC_1 [46], sci-Hi-C [3], scHi-C_2 [1], snHi-C [2], Dip-C [47], scSPRITE [9] and multi-omics methods Methyl-Hi-C [48], sn-m3C-seq [20]. Most of these datasets contained up to 2,000 nuclei, while the sn-m3C-seq dataset had 6,500 nuclei (Figure S3b). We reused the numbers of median DNA-DNA contacts per cell of each technology from a recent review [49]. MUSIC FC and MUSIC H1 had 9,087 and 2,546 nuclei, respectively,

surpassing the number of nuclei in most other datasets (Figure S3b). Regarding DNA-DNA contacts per nucleus, MUSIC FC and MUSIC H1 ranked second and third in the median number of resolved pairwise DNA-DNA contacts per nucleus. MUSIC FC had a median of 2,651,577 DNA-DNA contacts, while MUSIC H1 had a median of 455,373,385 DNA-DNA contacts (Figure S3b). These numbers were comparable to the other technologies, whose datasets yielded a median of up to 1x10e6 DNA-DNA contacts per nucleus, except for scSPRITE, which had a median of 9.3x10e8 DNA-DNA contacts per nucleus (Figure S3b). Of note, previous work projected scSRITE's multiplex interactions to pairwise contacts and reported the projected pairwise contacts as the number of DNA-DNA contacts of scSPRITE [9]. We followed this approach to project MUSIC's multiplex data to pairwise contacts. The greater numbers reported here for scSPRITE and MUSIC are attributable to their ability to capture multi-way contacts, which does not mean that they are more sensitive in revealing pairwise chromatin interactions than the other methods.

We also compared the number of RNA reads per cell between MUSIC and other technologies. We compared the RNA reads in MUSIC H1 and MUSIC FC to single-nucleus RNA-seq data from human prefrontal cortex (snRNA-seq PFC) [50] and multi-omic datasets including CITE-seq data generated from human PBMC [21], SNARE-seq data generated from mouse brain [13], and PairTag data generated from mouse brain [12] (Table S2). snRNA-seq PFC resolved a median of 1,973 RNA reads per nucleus, corresponding to 1,348 genes. CITE-seq, SNARE-seq, and PairTag resolved a median of 1416, 1332 and 845 RNA reads per nucleus, corresponding to 770, 1203, and 626 genes (Figure S3a). MUSIC FC exhibited a median of 1,136 RNA reads per nucleus, corresponding to 853 genes, which was within the range of the other methods (Figure S3a). MUSIC H1 had a higher median of 6,559 RNA reads per nucleus, corresponding to 3,684 genes, likely attributed to the higher data quality from a cell line rather than the superiority of the technology itself (Figure S3a). Overall, MUSIC demonstrated a comparable read depth to snRNA-seq and other multi-omic methods in mapping single-nucleus transcriptomes.

Cell type characterization in human frontal cortex by MUSIC's RNA reads

In the MUSIC FC dataset, we performed clustering analysis on the 9,087 cortical nuclei based on their RNA expression levels (Figure 4a). The clusters were then assigned to specific cell types using known marker genes from previous studies (Table S3, [50,51]). A cluster of 1,457 nuclei was categorized as excitatory neurons (ExN) based on the expression of marker genes SLC17A7, CAMK2A, and NRGN [50] (Figure S4a). Within this cluster, two subclusters were identified: intratelencephalic (IT) neurons (on the left) and corticothalamic (CT) and near-projecting (NP) neurons (on the right) (Figures S5g-i). The IT subcluster further consisted of three subgroups: Layer 2/3 intratelencephalic (L2/3 IT) (597 nuclei), Layer 5 IT (403 nuclei), and Layer 6 IT (263 nuclei) neurons (Figures S5g-i). The CT and NP subcluster contained Layer 6 corticothalamic (L6 CT) (142 nuclei) and Layer 5/6 near-projecting (L5/6 NP) (52 nuclei) neurons, respectively [52] (Figures S5g-i).

A cluster of 771 nuclei was categorized as inhibitory neurons (InN) based on the expression of marker genes GAD1 and GAD2 [50] (Figure S4a). This cluster formed four subclusters (Figures S5j-l) corresponding to Lamp5, Vip, Pvalb, and Sst neurons. The two sub-clusters on the top

corresponded to Lamp5 (128 nuclei) and Vip (206 nuclei) neurons that develop from caudal ganglionic eminence [53]. The two sub-clusters on the bottom corresponded to Pvalb (232 nuclei) and Sst (205 nuclei) neurons that develop from medial ganglionic eminence (Figures S5j-l).

A cluster of 661 nuclei was categorized as microglia based on the expression of marker genes CD74, CSF1R, and C3 [50] (Figure S4a). This microglial cluster consisted of a major subcluster (573 nuclei) and a minor subcluster (88 nuclei) marked by different levels of Membrane-spanning 4A (MS4A) genes [54] (Figures S5m-o). In mice, the activation of MS4A genes marks the transition of microglia from a chemokine state to an interferon state [54]. Thus, the major and the minor sub-clusters may reflect the corresponding cellular states of microglia in humans.

A cluster of 4,539 nuclei was categorized as oligodendrocytes based on the expression of marker genes MBP, PLP1, and MOBP [50]. This cluster further formed a major subcluster (4,097 nuclei) and three minor subclusters (224, 132, and 86 nuclei) (Figure S4p-r). Another cluster of 760 nuclei was categorized as vascular cells based on the expression of marker genes EMCN, VWF, and FLT1 [51] (Figure 4a). Lastly, two distinct clusters with 571 and 328 nuclei were categorized as astrocytes (Ast) based on marker genes AQP4 and GFAP, and oligodendrocyte precursors (Opc) based on marker genes VCAN, PDGFRA, and CSPG4 [50].

The stratification analysis by sex (Figure S4a, S5e) and individual cortex sample (Figure S4d) did not significantly impact the proportions of cells in the clusters or subclusters, except for a higher number of oligodendrocytes in males compared to females (Figure S4c, e). Our data indicate a sex difference in the number of cortical oligodendrocytes in the elderly people (>=59 years of age), which aligns with previous studies showing “the lifespan of oligodendrocytes is shorter in females than in males” in mice [55]. In summary, MUSIC FC data formed clear clusters and these clusters correspond well with cortical cell types and cellular states.

Cellular heterogeneity in the relationship between chromatin interaction frequency and genomic distance

A well-established principle of chromatin conformation is that chromatin interaction frequency (Pc) decreases as the genomic distance (s) between chromatin regions increases in the bulk [40]. This reverse correlation is approximately linear on the log-log scale [40]. This trend is observed in MUSIC FC data as well, where the aggregate chromatin interaction frequency (Pc) in the ensemble of single cells decreases with increasing genomic distance (s) (Figure 4b). Hereafter we will refer to this trend as the "aggregate Pc-s relationship”.

At the single-cell level, most single cells exhibited a reverse correlation between Pc and s as well, whereas a minority of single cells exhibited the largest Pc not necessarily at the smallest s, a deviation from the aggregate Pc-s relationship (Figure 4c, d). To test if this observed cellular heterogeneity is compatible with the aggregate Pc-s relationship, we binned the genomic distances and counted the proportion of single cells that exhibit the largest Pc in each genomic distance bin (Figure 4d). The proportion of single cells is smaller in the bins of longer genomic distances, conforming to a reverse correlation that is approximately linear in the log-log scale

(Figure 4d). Thus, despite the high degree of cellular heterogeneity, the population summary of the single cells reproduces the previously reported aggregate relationship. Interestingly, the minority of single cells exhibiting high Pc at large genomic distances is required for reproducing the log-log linearity in the aggregate Pc-s relationship, suggesting this minority of single cells cannot be ignored in a complete characterization of chromatin conformation at the single-cell level.

While the different cell types exhibited similar aggregate Pc(s) curves, these Pc(s) curves are not identical (Figure 4b). These differences indicate cell-type variations in chromatin conformation. In particular, excitatory neurons exhibited more frequent chromatin interactions within the sub-Mb range of genomic distances than other cell types (Figure 4c). Consistent with this aggregate behavior, excitatory neurons had a larger proportion of single cells exhibiting the most frequent chromatin interactions at the sub-Mb range compared to the other cell types (Figure 4c, Figure S4f). Together, these observations highlight the influence of cellular composition in each cell type on the cell-type variation in chromatin conformation.

Gene-expression-associated stripes (GEAS), a joint feature of transcription and chromatin interaction

We observed cell-type variations in their chromatin contact maps (Figure 4e). For example, at the chromosomal scale (1 Mb resolution), we observed a stripe chromatin structure, where a specific sequence associates with a long stretch of chromosomal regions [56], on Chromosome 9 in excitatory neurons, which is obscure or absent in the other cell types (arrow, Figure 4e). For another example, a stripe is prominent at the Myelin Basic Protein (MBP) gene locus in oligodendrocytes but obscure or absent in the other cell types (50 Kb resolution) (Figure 4f). This cell-type-specific stripe is consistently observed in small (2-10 DNA reads), medium (11-20 DNA reads), and large (21-50 DNA reads) DD clusters. MBP is a maker gene of oligodendrocytes, and as expected, MBP is expressed at a high level in oligodendrocytes and low levels in the other cell types (bottom track, Figure 4f). In this example, we observed a chromatin stripe that co-localizes with a highly expressed gene.

To check if the observed colocalization is a singular case, we identified the chromatin stripes in the genome for each cell type. The genes co-localized with the identified stripes exhibited greater expression levels than the other genes outside the stripes (Figure S5b). This genome-wide co-localization of chromatin stripes and highly expressed genes suggests a recurring feature in the joint space of the chromatin conformation and the transcriptome. We will refer to this feature as a gene-expression-associated stripe (GEAS).

We developed a bioinformatic pipeline called GEAS-scan to identify GEASs throughout the genome. GEAS-scan begins by calling stripes from the small, medium, and large DD clusters in each cell type (Figure S5a) and merging any co-localized stripes within each cell type. GEAS-scan then outputs the stripes that co-localize with highly expressed genes (top 20% in expression levels) in each cell type as the GEASs that cell type. Finally, GEAS-scan merges non-overlapping GEASs across the cell types. Using GEAS-scan, we identified a total of 2,227 GEASs in human

364 cortical cells, which include 1,474, 762, 397, 276, 352, 1,006, and 107 GEASs in excitatory
365 neurons (ExN), inhibitory neurons (InN), astrocytes (Ast), microglia (Mic), oligodendrocyte
366 precursors (Opc), oligodendrocytes (Oli), and vascular cells (Vas).

367 Cellular heterogeneity in XIST lncRNA-X chromosome association in the female
368 cortex

369 The XIST lncRNA is detected in female cortical cells but not in any male cells (figure 5a), which
370 is consistent with its expected presence in female somatic tissues and absence in male tissues
371 [57]. However, MUSIC did not detect XIST lncRNA in every female cell. There are two probable
372 causes for this variation. Technically, MUSIC may not be sensitive enough to detect XIST lncRNA
373 in every nucleus. Biologically, the expression level of XIST may be heterogeneous among the
374 female cells. Supporting the coexistence of both causes, XIST lncRNA is exclusively detected in
375 female cells, and as the threshold on the number of total RNA reads in a cell is increased, the
376 proportion of XIST-detected cells among the remaining female cells remains relatively stable with
377 a modestly increasing trend (Figure S6a). This analysis suggests that the RNA read count in a
378 cell is influenced by both the expression level of the molecule and the sensitivity of the technique
379 itself. To minimize the impact of false negatives on our analysis, we will apply filters on RNA read
380 count to reduce false negatives. More importantly, we will not make any conclusion about any
381 single cell. Instead, all the subsequent analyses are focused on learning the characteristics of cell
382 groups.
383
384 In the ensemble of female cells, the XIST lncRNA exhibited a strong association with the entire X
385 chromosome (Figures 5b, c), consistent with its known ability to spread across one of the X
386 chromosomes (the Xi chromosome) [58,59]. At the single-cell level, the female cortical cells exhibited
387 heterogeneous XIST association level (XAL), as measured by the number of RD clusters involving
388 XIST lncRNA and X chromosomal DNA in a nucleus (Figure S6c). Filtering the female cells based
389 on a threshold of total RNA reads per cell (> 5000) did not eliminate cellular heterogeneity,
390 indicating that the observed heterogeneity cannot be solely attributed to the limited sensitivity of
391 the technique. As expected, the XIST RNA read count in a cell correlated with XAL among the
392 female cells (XIST_ct column vs. Heatmap, Figure S6e), whereas the number of chromosome X
393 DNA reads remained relatively invariant, confirming that the total DNA read count of the X
394 chromosome is independent of XAL (log2 chrX DNA column, Figure S6e).
395
396 We compared the X-chromosomal clusters associated with XIST lncRNA (XIST+) and those not
397 associated with XIST lncRNA (XIST-). XIST+ clusters included RD clusters with at least one XIST
398 RNA read, while XIST- clusters were RD and DD clusters that did not contain any XIST RNA read.
399 At the chromosomal scale, most DNA-DNA contacts in XIST- clusters were concentrated near the
400 diagonal line in the chromatin contact map (Figure 5d), similar to the contact maps of autosomes.
401 However, XIST+ clusters exhibited not only near-diagonal contacts but also a significant number
402 of contacts spanning distances of 10 Mb or more (Figure 5d).
403
404 Consistent with the chromatin contact maps, the frequency of chromatin contacts (Pc) of XIST-
405 clusters is greater than that of the XIST+ clusters when the genomic distance (s) is smaller than

406 ~10Mb (Figure 5e). To determine if this separation in Pc(s) curves could be attributed to limited
407 sensitivity in detecting XAL at the single-cell level, a stratification analysis was performed. Female
408 cells were stratified into four groups based on zero, low, medium, and high XALs. Of note, the
409 zero XAL group (Group 1) only contains XIST- clusters, and Groups 2, 3, and 4 contain the same
410 number of cells. Pc(s) of the XIST- clusters (XIST- Pc(s)) is above XIST+ Pc(s) when s is smaller
411 than ~10Mb in Groups 2, 3, and 4 (Figure 5f). Importantly, the difference between XIST- and
412 XIST+ Pc(s) curves increased from Group 2 to Group 4, indicating that higher XALs in a cell group
413 led to a more pronounced chromatin conformation difference between XIST- and XIST+ clusters.
414 As a control, the Pc(s) curves of X chromosomal clusters associated with or without any RNA
415 (Any_RNA+ Pc(s) or Any_RNA- Pc(s)) in Group 1 (zero XAL group) were nearly indistinguishable
416 (Figure 5f). These data suggest that the active X chromosome (Xa) has a higher contact
417 frequency than the inactive X chromosome (Xi) in the sub-10 Mb range of genomic distances in
418 the female human cortex. In line with this finding, chromatin tracing analysis in the female mouse
419 cortex revealed that Xa has smaller 3D spatial distances than Xi when the genomic sequence is
420 below ~10Mb, and vice versa when the genomic distance is above ~10Mb (see Fig. S24g in [60]).
421
422 Different cell types exhibited different proportions of XAL-positive (XAL+) cells (Figure S6b), with
423 excitatory neurons having the highest proportion of XAL+ cells (Chi-square, p-value < 10e-16)
424 (Figure 5g). Considering the larger separation between XIST+ and XIST- Pc(s) curves (ΔPc(s))
425 in XAL-high cells, we anticipated seeing a cell-type difference in ΔPc(s), particularly with
426 excitatory neurons to exhibit a larger ΔPc(s) compared to other cell types. To test this idea, we
427 compared the three cell types with at least 45% of cells exhibiting non-zero XAL, namely excitatory
428 neurons (ExN), inhibitory neurons (InN), and astrocytes (Ast) (Figure S6b). As expected, ExN
429 displayed a larger ΔPc(s) than InN and Ast (Figure 5h). Specifically, ExN's XIST+ Ps(c) curve is
430 beneath the XIST- Ps(c) curve when the genomic distance is less than ~10Mb and traverses
431 above the XIST- Ps(c) curve at ~10Mb. Although this transversion was consistently observed in
432 InN and Ast, the gaps between the Ps(c) curves were narrower in InN and Ast. Consistently,
433 XIST+ chromatin exhibited more off-diagonal contacts in the chromatin contact maps than the
434 XIST- chromatin in all three cell types. However, the difference in off-diagonal contacts was more
435 pronounced in ExN compared to InN and Ast (Figure 5i). These data suggest a cell-type variation
436 in the conformations of the two X chromosomes in the female human cortex.
437
438 Aligned with our data, chromatin tracing analysis of the female mouse cortex demonstrated a
439 clearer difference between Xa and Xi in ExN compared to InN and Ast (see Fig. S24 in [60]).
440 Specifically, most single ExN cells exhibited a smaller intra-chromosomal spatial distance in Xi
441 than Xa at genomic distances greater than ~10Mb. In contrast, a notable fraction of InN and Ast
442 cells showed non-distinguishable differences in intra-chromosomal spatial distance between Xi
443 and Xa at any genomic distance (see Fig. S24g in [60]). Taken together, these data suggest a
444 conserved cell-type variation in the spatial organization of the two X chromosomes in the female
445 cortex in mice and humans.

Discussion

ChIA-drop analysis in *Drosophila* S2 cells identified approximately half of the chromatin interactions as multiplex interactions [24]. Consistent with this observation, the MUSIC H1 and MUSIC FC data suggest that approximately 30%-40% of chromatin complexes involve multiplex interactions in humans (Figure S2a). Moreover, multiplex interactions better recapitulate the TAD structure than pairwise interactions in the human MUSIC data. In addition to TADs, the stripe chromatin structure is more pronounced in the multiplex chromatin complexes. These data argue multiplex interactions are an important constituent of the chromatin structure, from *Drosophila* to humans.

Gene-Expression-Associated Stripes (GEAS) emerged as a frequent pattern in the joint space of single-cell transcriptome and chromatin conformation. GEAS exemplifies a correlation between the heterogeneous single-cell gene expression and single-cell chromatin conformation. MUSIC may facilitate the research community to discover other connections between gene expression and chromatin conformation.

After initial debates [61], chromatin-associated RNA (caRNA) has gradually been recognized as a structural component of chromatin [28]. The analysis of *Drosophila* and chicken cells suggested 2%–5% of total chromatin-associated nucleic acids are RNA [62]. Similarly, in the MUSIC H1 data, approximately 4.6% of chromatin-complex reads (in DD and RD clusters with at least 1 DNA read) are RNA reads, indicating relatively invariant proportions of RNA in chromatin-associated nucleic acids across species. Additionally, RNA is involved in approximately 11.7% of chromatin complexes (either RD clusters with at least 1 RNA read and 1 DNA read or DD clusters with at least 2 DNA reads). Notably, a larger fraction of RNA-containing chromatin complexes exhibit multiplex DNA-DNA contacts compared to RNA-free complexes (Figure S2c, d). This observation supports that RNA can contribute to spatial genome compartmentalization [29]. Moreover, the chromatin complexes with and without the XIST long non-coding RNA (lncRNA) display differences in their distributions of contact probabilities.

Our analyses revealed a correlation between the loss of XIST-chrX association and the diminishing conformation difference between Xa and Xi in a subset of single cells in the female cortex. Without a phased genome, MUSIC data cannot infer allelic expression and thus cannot directly pinpoint any gene expressed from the Xi chromosome (incomplete XCI gene). However, a survey conducted by the Genotype-Tissue Expression (GTEx) consortium, involving 5,500 transcriptomes from 449 individuals across 29 tissues, reported that incomplete XCI genes generally exhibit higher expression in females compared to males in the corresponding tissue, suggesting sex bias can be used as “a proxy for XCI status" [33]. The MUSIC data can infer sex-difference in gene expression at the single-cell level. To illustrate this point, for every female cell and every previously annotated incomplete XCI gene [33], we calculated the fold change between the RNA read count of this gene in this female cell and the average RNA read counts of this gene in all the male cells of the matching cell type (sex-fold-change) (Figure S6d, f). XAL- female cells exhibited greater sex-fold-changes in most of the incomplete XCI genes compared to XAL+ female cells, suggesting the loss of XIST-chrX association in single female cells correlates with larger sex-difference in the gene expression in the human cortex (Figure S6d, f). Taken together,

MUSIC provides a unique tool to jointly analyze gene expression, multiplex chromatin interactions, and RNA-chromatin associations with single-cell resolution from complex tissue.

Acknowledgments

We thank Riccardo Calandrelli for proofreading the manuscript. UCSD IGM Genomics Center for support on sequencing.

Funding

This work is funded by NIH grants DP1DK126138, R01GM138852, UH3CA256960, U01CA200147, R01HD107206, and a Kruger Research Grant.

Author contributions

N.C.T. and S.Z. conceived of the MUSIC experimental strategy. Z.L. developed the MUSIC workflow and performed the experiments. X.W. performed data analysis. W.Z. X.W. and J.R. contributed to generating data. X.W., Z.L., W.Z., and S.Z. wrote the manuscript. S.Z. managed and supervised the project.

Reference

1. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
2. Flyamer, I. M. *et al.* Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
3. Ramani, V. *et al.* Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).
4. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
5. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
6. Dekker, J. *et al.* The 4D nucleome project. *Nature* **549**, 219–226 (2017).
7. Sun, J. H. *et al.* Disease-Associated Short Tandem Repeats Co-localize with Chromatin Domain Boundaries. *Cell* **175**, 224–238.e15 (2018).
8. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305–320.e24

517 (2017).
518 9. Arrastia, M. V. *et al.* Single-cell measurement of higher-order 3D genome organization
519 with scSPRITE. *Nat. Biotechnol.* **40**, 64–73 (2022).
520 10. Takei, Y. *et al.* Integrated spatial genomics reveals global architecture of single
521 nuclei. *Nature* **590**, 344–350 (2021).
522 11. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially
523 resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
524 12. Zhu, C., Preissl, S. & Ren, B. Single-cell multimodal omics: the power of many.
525 *Nat. Methods* **17**, 11–14 (2020).
526 13. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the
527 transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* **37**, 1452–1457
528 (2019).
529 14. Plongthongkum, N., Diep, D., Chen, S., Lake, B. B. & Zhang, K. Scalable dual-
530 omics profiling with single-nucleus chromatin accessibility and mRNA expression
531 sequencing 2 (SNARE-seq2). *Nat. Protoc.* **16**, 4992–5029 (2021).
532 15. Ma, S. *et al.* Chromatin Potential Identified by Shared Single-Cell Profiling of
533 RNA and Chromatin. *Cell* **183**, 1103–1116.e20 (2020).
534 16. Liu, Z. *et al.* Linking genome structures to functions by simultaneous single-cell
535 Hi-C and RNA-seq. *Science* **380**, 1070–1076 (2023).
536 17. Clark, S. J. *et al.* scNMT-seq enables joint profiling of chromatin accessibility
537 DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781 (2018).
538 18. Wang, Y. *et al.* Single-cell multiomics sequencing reveals the functional
539 regulatory landscape of early embryos. *Nat. Commun.* **12**, 1247 (2021).
540 19. Zhu, C. *et al.* Joint profiling of histone modifications and transcriptome in single
541 cells from mouse brain. *Nat. Methods* **18**, 283–292 (2021).
542 20. Lee, D.-S. *et al.* Simultaneous profiling of 3D genome structure and DNA

543 methylation in single human cells. *Nat. Methods* **16**, 999–1006 (2019).
544 21. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in
545 single cells. *Nat. Methods* **14**, 865–868 (2017).
546 22. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell
547 resolution. *Nature* **568**, 49–54 (2019).
548 23. Liu, M. *et al.* Multiplexed imaging of nucleome architectures in single cells of
549 mammalian tissue. *Nat. Commun.* **11**, 2907 (2020).
550 24. Zheng, M. *et al.* Multiplex chromatin interactions with single-molecule precision.
551 *Nature* **566**, 558–562 (2019).
552 25. Deshpande, A. S. *et al.* Identifying synergistic high-order 3D chromatin
553 conformations from genome-scale nanopore concatemer sequencing. *Nat. Biotechnol.* **40**,
554 1488–1499 (2022).
555 26. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome
556 Organization in the Nucleus. *Cell* **174**, 744–757.e24 (2018).
557 27. Dotson, G. A. *et al.* Deciphering multi-way interactions in the human genome.
558 *Nat. Commun.* **13**, 5498 (2022).
559 28. Thakur, J. & Henikoff, S. Architectural RNA in chromatin organization. *Biochem.*
560 *Soc. Trans.* **48**, 1967–1978 (2020).
561 29. Quinodoz, S. A. *et al.* RNA promotes the formation of spatial compartments in
562 the nucleus. *Cell* **184**, 5775–5790.e30 (2021).
563 30. Calandrelli, R., Wen, X., Charles Richard, J. L. & Luo, Z. Three-dimensional
564 organization of chromatin associated RNAs and their role in chromatin architecture in
565 human cells. *BioRxiv* (2021).
566 31. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu.*
567 *Rev. Biochem.* **81**, 145–166 (2012).
568 32. Lee, J. T. Gracefully ageing at 50, X-chromosome inactivation becomes a

569 paradigm for RNA and chromatin control. *Nat. Rev. Mol. Cell Biol.* **12**, 815–826 (2011).
570 33. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human
571 tissues. *Nature* **550**, 244–248 (2017).
572 34. Yan, Y. *et al.* X-linked ubiquitin-specific peptidase 11 increases tauopathy
573 vulnerability in women. *Cell* **185**, 3913–3930.e19 (2022).
574 35. Yan, Z. *et al.* Genome-wide colocalization of RNA–DNA interactions and fusion
575 RNA pairs. *Proceedings of the National Academy of Sciences* **116**, 3328–3337 (2019).
576 36. Wu, W. *et al.* Mapping RNA–chromatin interactions by sequencing with iMARGI.
577 *Nat. Protoc.* **14**, 3243–3272 (2019).
578 37. Li, X. *et al.* GRID-seq reveals the global RNA–chromatin interactome. *Nat.*
579 *Biotechnol.* **35**, 940–950 (2017).
580 38. Bell, J. C. *et al.* Chromatin-associated RNA sequencing (ChAR-seq) maps
581 genome-wide RNA-to-DNA contacts. *Elife* **7**, e27024 (2018).
582 39. Bonetti, A. *et al.* RADICL-seq identifies general and cell type-specific principles of
583 genome-wide RNA-chromatin interactions. *Nat. Commun.* **11**, 1018 (2020).
584 40. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions
585 reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
586 41. Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome
587 Architecture. *Mol. Cell* **78**, 554–565.e7 (2020).
588 42. Mirny, L. A. The fractal globule as a model of chromatin architecture in the cell.
589 *Chromosome Res.* **19**, 37–51 (2011).
590 43. ENCODE Project. *The ENCODE Project Encyclopedia of DNA Elements*. (2012).
591 44. Chen, W. *et al.* RNAs as Proximity-Labeling Media for Identifying Nuclear
592 Speckle Positions Relative to the Genome. *iScience* **4**, 204–215 (2018).
593 45. Wang, Y. *et al.* SPIN reveals genome-wide landscape of nuclear
594 compartmentalization. *Genome Biol.* **22**, 36 (2021).

595 46. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome
596 structure. *Nature* **502**, 59–64 (2013).
597 47. Tan, L., Xing, D., Chang, C.-H., Li, H. & Xie, X. S. Three-dimensional genome
598 structures of single diploid human cells. *Science* **361**, 924–928 (2018).
599 48. Li, G. *et al.* Joint profiling of DNA methylation and chromatin architecture in single
600 cells. *Nat. Methods* **16**, 991–993 (2019).
601 49. Zhou, T., Zhang, R. & Ma, J. The 3D Genome Structure of Single Cells. *Annu*
602 *Rev Biomed Data Sci* **4**, 21–41 (2021).
603 50. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease.
604 *Nature* **570**, 332–337 (2019).
605 51. Franzén, O., Gan, L.-M. & Björkegren, J. L. M. PanglaoDB: a web server for
606 exploration of mouse and human single-cell RNA sequencing data. *Database* **2019**,
607 (2019).
608 52. Bakken, T. E. *et al.* Comparative cellular analysis of motor cortex in human,
609 marmoset and mouse. *Nature* **598**, 111–119 (2021).
610 53. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical
611 areas. *Nature* **563**, 72–78 (2018).
612 54. Deming, Y. *et al.* The MS4A gene cluster is a key modulator of soluble TREM2
613 and Alzheimer's disease risk. *Sci. Transl. Med.* **11**, (2019).
614 55. Cerghet, M. *et al.* Proliferation and death of oligodendrocytes and myelin proteins
615 are differentially regulated in male and female rodents. *J. Neurosci.* **26**, 1439–1447 (2006).
616 56. Vian, L. *et al.* The Energetics and Physiological Impact of Cohesin Extrusion. *Cell*
617 **173**, 1165–1178.e20 (2018).
618 57. Weakley, S. M., Wang, H., Yao, Q. & Chen, C. Expression and function of a large
619 non-coding RNA gene XIST in human cancer. *World J. Surg.* **35**, 1751–1756 (2011).
620 58. Disteche, C. M. Dosage compensation of the sex chromosomes. *Annu. Rev.*

621 *Genet.* **46**, 537–560 (2012).

622 59. Lee, J. T. Epigenetic regulation by long noncoding RNAs. *Science* **338**, 1435–

623 1439 (2012).

624 60. Takei, Y. *et al.* Single-cell nuclear architecture across cell types in the mouse

625 brain. *Science* **374**, 586–594 (2021).

626 61. Holmes, D. S., Mayfield, J. E., Sander, G. & Bonner, J. Chromosomal RNA: its

627 properties. *Science* **177**, 72–74 (1972).

628 62. Rodríguez-Campos, A. & Azorín, F. RNA is an integral component of chromatin

629 that contributes to its structural organization. *PLoS One* **2**, e1182 (2007).

Figure 1. Overview of the MUSIC method. The DNA Linker (square) and RNA Linker (pentagon) are ligated to DNA and RNA, respectively (first column). Three sets of Cell Barcodes (green beads) are subsequently added (second column). Two sets of Complex Barcodes including the 10X Barcodes (10X) and I7 Barcodes (I7) are subsequently added to chromatin complexes (3rd and 4th columns).

Figure 2. Summary of MUSIC data in H1 cells. (a-c) The numbers of uniquely mapped non-duplicate reads (a), clusters (b), and pairwise contacts (c) in every H1 cell (column). DNA-DNA (DD, blue), RNA-DNA (RD, yellow), and RNA-RNA (RR, gray) clusters are separately counted. Multiplex interactions are projected to pairwise interactions and the numbers of pairwise contacts are reported in (c). (d) The distribution of DNA reads, and RNA reads (in gray) in DD, RD, and RR clusters. (e) Cluster size distributions of DD and RD clusters.

Figure 3. Chromatin contacts, gene expression, and RNA-chromatin association in H1 cells. (a) Comparison of Micro-C (upper triangle) and ensemble MUSIC (lower triangle) derived chromatin contact maps on Chromosome 1 at 1Mb resolution. (b-d) Chromatin contact maps based on the ensemble of small (b), middle-sized (c), and large (d) MUSIC DD clusters (lower triangles), and Micro-C (upper triangles) in human H1 cells. Resolution: 50Kb. Arrows: Contacts between nested TADs. (e) Pc(s) curves showing the frequency of chromatin contacts (Pc, y axis) vs. genomic distance (s, x axis) for MUSIC DD clusters with different sizes (shades of blue) and Micro-C (black). DNA cluster size: the number of DNA reads in a DD cluster. (f-h) Comparison of 2-D contact map from ensemble MUSIC data (f) with a cumulative view of every DD cluster (row) (g-h) Stacked maps of the clusters. Each row is a DD cluster. The yellow dots in the same row correspond to the genomic locations of the DNA reads of a DD cluster. The DD clusters are ordered by the smallest genomic coordinate of any DNA read in each cluster. The clusters composed of 2 DNA reads (pairwise interactions) (g) are separately plotted with the clusters with 3 or more DNA reads (multiplex interactions) (h). (i-j) Scatterplots of RNA levels measured by Reads Per Kilobase (RPK) for every gene (dot) in ensemble MUSIC (x axis) vs. in iMARGI (y axis) (i) and in RNA-seq (y axis) (j). R: Spearman correlation. (k) The RNA reads from RNA-seq, iMARGI, and MUSIC mapped to both strands ((+) and (-)). All data were generated from human H1 cells. Ensemble MUSIC and the MUSIC data of three single cells are shown. (l) Distribution of chromatin-associated pre-mRNA and nsaRNA on Chromosome 1 as measured by ensemble MUSIC in H1 cells. Micro-C derived A/B compartments are colored in red/blue. The “Speckle compartmentalization” derived from the SPIN model using Hi-C, TSA-seq, and DamID is denoted in the SPIN-Speckle track. The pre-mRNA enriched genomic regions tend to co-localize with the nsaRNA enriched genomic regions, A compartment, and the Speckle compartmentalization. (m) Scatter plot of normalized counts of pre-mRNA reads (y axis) and nsaRNA reads (x axis) in every 1 Mb genomic bin (dot) across the entire genome, based on Ensemble MUSIC data in H1 cells.

Figure 4. A single-cell map of transcriptome and DD/RD complexes in human frontal cortex. (a) UMAP representation of individual cortical cells based on MUSIC RNA reads. n: number of cells. (b) Chromatin contact frequency (Pc, y axis) versus genomic distance (s, x axis) for each cell type (color). (c) Histogram of the proportions of single excitatory neurons (ExN) (y axis) with their most frequent chromatin interactions in each genomic bin (x axis) is superimposed on the contact

frequency (color intensity) vs. genomic distance (x axis) plot for every ExN (row). (d) Chromatin contact frequency (color) vs. genomic distance (x axis) in individual cortical cells (rows). (e) 2-D contact maps for each cortical cell type for chromosome 9 with 1Mb resolution. A stripe structure is prominent in ExN (arrow). (f) 2-D contact maps for small, middle-size, and large clusters (rows) in each cell type (column) in a 5Mb region centered at the MBP gene. Resolution: 50Kb. Bottom track: Gene expression levels at the corresponding genomic locations.

Figure 5. Cellular heterogeneity of XIST-chromatin interactions in female frontal cortex. (a) XIST expression level (y axis) of every single cell (dot) in each cell type (column) female (pink) and male (blue). (b) RNA-DNA contact map for Chromosome X based on the ensemble female cells. Each pixel represents the amount of RNA that is transcribed from the genomic bin (row) and is associated with the genomic bin (column). Resolution: 1Mb. (c) Distribution of chromatin-associated XIST RNA on every chromosome in the ensemble of female cells. Bin size: 1Mb. Y axis: the RNA attachment level (RAL) in a genomic bin among the female cells. (d) X chromosome contact maps for XIST-associated (XIST+), not-associated (XIST-) chromatin in female cells, and their contrast. Resolution: 1Mb. (e) Pc(s) curves showing the frequency of chromatin contacts (Pc, y axis) vs. genomic distance (s, x axis) for XIST+ (orange) and XIST- (blue) female X chromosomal chromatin complexes. (f) X chromosome Pc(s) curves in four female cell groups, with zero, low, medium, and high XAL (left to right panels). The Pc(s) curves for chromatin with (Any_RNA+, gray) and without any associated RNA (Any_RNA-, blue) do not exhibit a notable difference. The difference between XIST+ (darker color) and XIST- Pc(s) (lighter color) curves increases as XAL increases. (g) Genome-wide distribution of XIST-chromatin association in individual excitatory neurons (ExNs). Each row represents a single ExN. The RNA attachment level (RAL) of the XIST lncRNA (intensity of red color) at any genomic region on any chromosome is plotted with the corresponding genomic coordinates (x axis). Resolution: 1 Mb. Track at the bottom: cumulative RAL of XIST. Tracks on the right indicate the XIST RNA read count (XIST RNA), X chromosomal DNA read counts (chrX DNA) in log scale. (h) X chromosome Pc(s) curves in female excitatory neurons (ExN), inhibitory neurons (InN), and astrocytes (Ast). The difference between XIST+ (darker color) and XIST- Pc(s) (lighter color) curves is most pronounced in ExN. (i) Contrast contact maps between XIST+ and XIST- chromatins on chromosome X in female ExN, InN, and Ast. ExN exhibits more XIST+ chromatin interactions (red) off-diagonally than InN and Ast.

Figure 1.

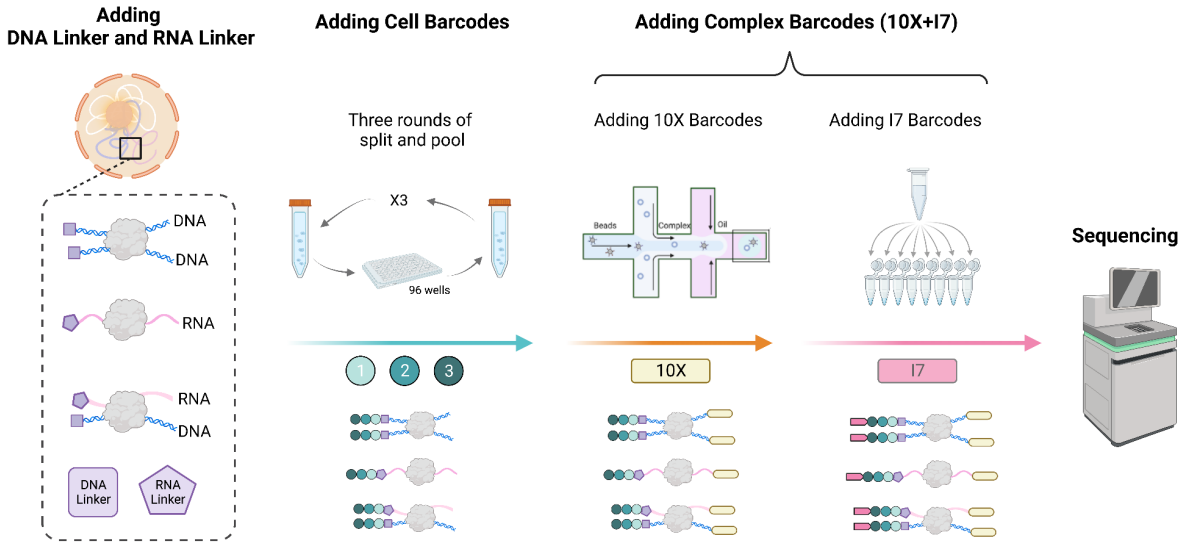

Adding
DNA Linker and RNA Linker
DNA
DNA
RNA
RNA
DNA
DNA Linker
RNA Linker
Adding Cell Barcodes
Three rounds of
split and pool
X3
96 wells
1
2
3
Adding Complex Barcodes (10X+I7)
Adding 10X Barcodes
Adding I7 Barcodes
Beads
Complex
Oil
10X
I7
Sequencing

Figure 2.

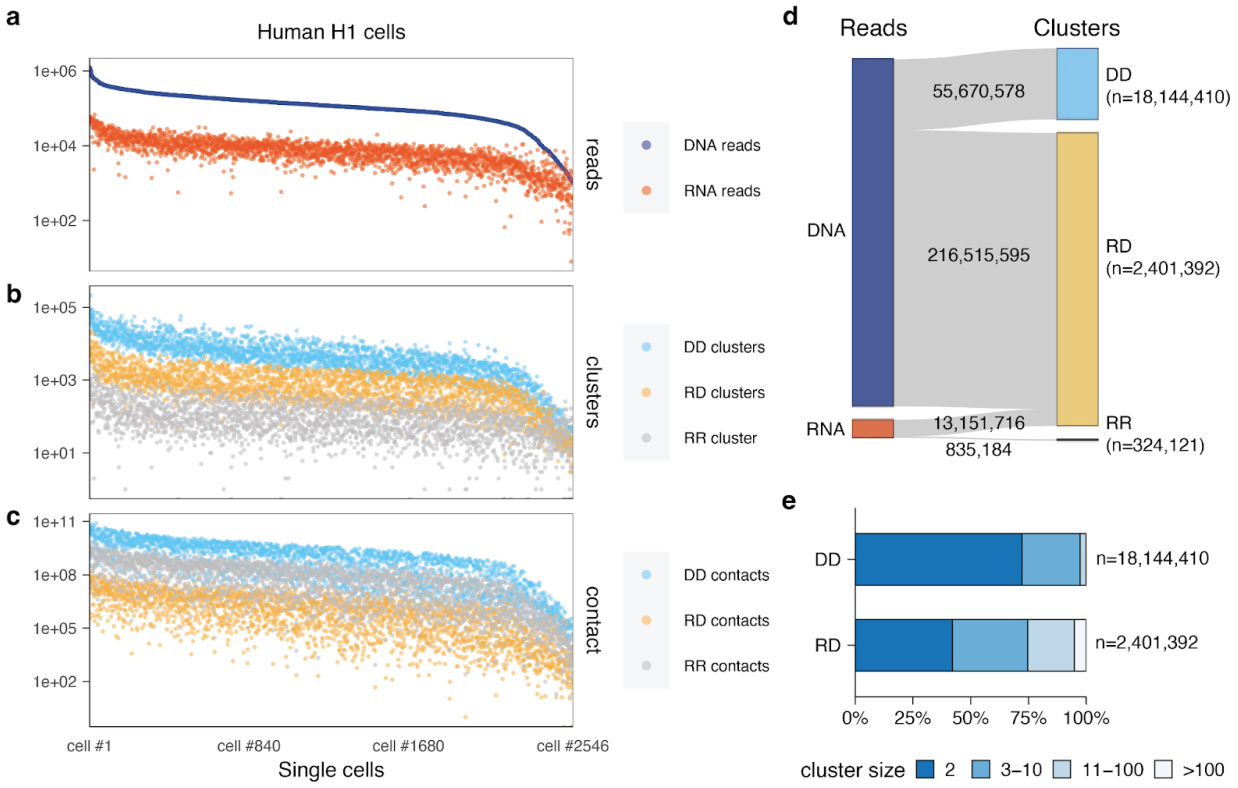

a
Human H1 cells
1e+06
1e+04
1e+02
reads
DNA reads
RNA reads
b
1e+05
1e+03
1e+01
clusters
DD clusters
RD clusters
RR cluster
c
1e+11
1e+08
1e+05
1e+02
contact
DD contacts
RD contacts
RR contacts
cell #1
cell #840
cell #1680
cell #2546
Single cells
d
Reads
Clusters
DNA
RNA
55,670,578
216,515,595
13,151,716
835,184
DD
(n=18,144,410)
RD
(n=2,401,392)
RR
(n=324,121)
e
DD
RD
n=18,144,410
n=2,401,392
0%
25%
50%
75%
100%
cluster size
2
3–10
11–100
>100

Figure 3.

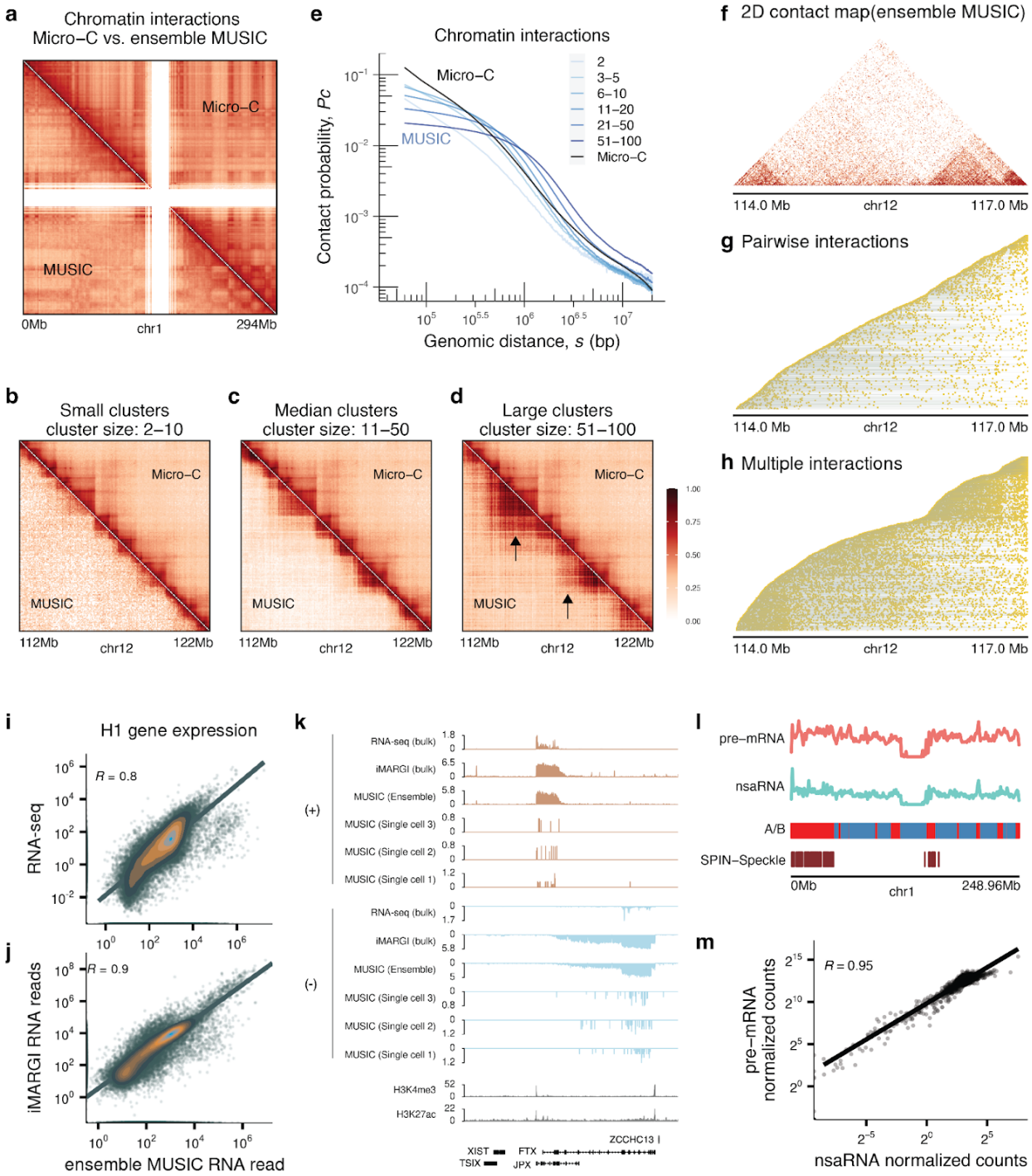

a
Chromatin interactions
Micro−C vs. ensemble MUSIC
Micro−C
MUSIC
0Mb
chr1
294Mb
e
Chromatin interactions
Contact probability, Pc
Micro−C
MUSIC
2
3−5
6−10
11−20
21−50
51−100
Micro−C
Genomic distance, s (bp)
f 2D contact map(ensemble MUSIC)
114.0 Mb
chr12
117.0 Mb
g Pairwise interactions
h Multiple interactions
b
Small clusters
cluster size: 2−10
c
Median clusters
cluster size: 11−50
d
Large clusters
cluster size: 51−100
112Mb
122Mb
i
H1 gene expression
R = 0.8
RNA-seq
j
R = 0.9
iMARGI RNA reads
ensemble MUSIC RNA read
k
RNA-seq (bulk)
iMARGI (bulk)
MUSIC (Ensemble)
MUSIC (Single cell 3)
MUSIC (Single cell 2)
MUSIC (Single cell 1)
(+)
(-)
H3K4me3
H3K27ac
XIST
TSIX
FTX
JPX
ZCCHC13
l
pre−mRNA
nsaRNA
A/B
SPIN−Speckle
0Mb
chr1
248.96Mb
m
R = 0.95
pre−mRNA normalized counts
nsaRNA normalized counts

Figure 4.

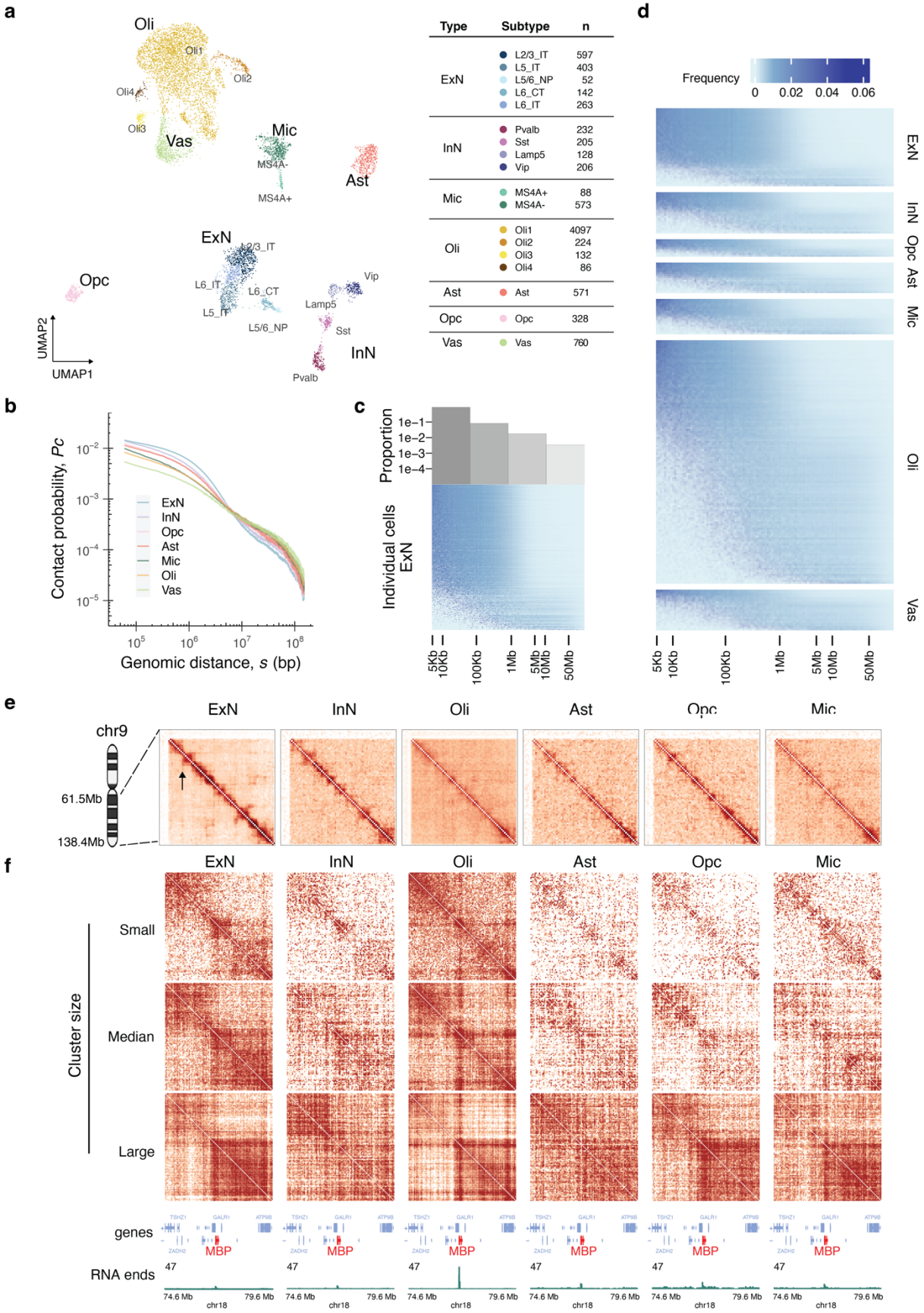

a
Oli
Oli1
Oli2
Oli4
Oli3
Vas
Mic
MS4A-
MS4A+
Ast
ExN
L2/3_IT
L6_IT
L6_CT
L5_IT
L5/6_NP
Opc
Vip
Lamp5
Sst
Pvalb
InN
UMAP2
UMAP1
Type Subtype n
ExN L2/3_IT 597
L5_IT 403
L5/6_NP 52
L6_CT 142
L6_IT 263
InN Pvalb 232
Sst 205
Lamp5 128
Vip 206
Mic MS4A+ 88
MS4A- 573
Oli Oli1 4097
Oli2 224
Oli3 132
Oli4 86
Ast Ast 571
Opc Opc 328
Vas Vas 760
b
Contact probability, Pc
Genomic distance, s (bp)
ExN
InN
Opc
Ast
Mic
Oli
Vas
c
Proportion
Individual cells
ExN
5Kb
10Kb
100Kb
1Mb
5Mb
10Mb
50Mb
d
Frequency
0 0.02 0.04 0.06
ExN
InN
Opc
Ast
Mic
Oli
Vas
e
chr9
61.5Mb
138.4Mb
ExN
InN
Oli
Ast
Opc
Mic
f
Cluster size
Small
Median
Large
genes
MBP
RNA ends
47
74.6 Mb
79.6 Mb
chr18

Figure 5.

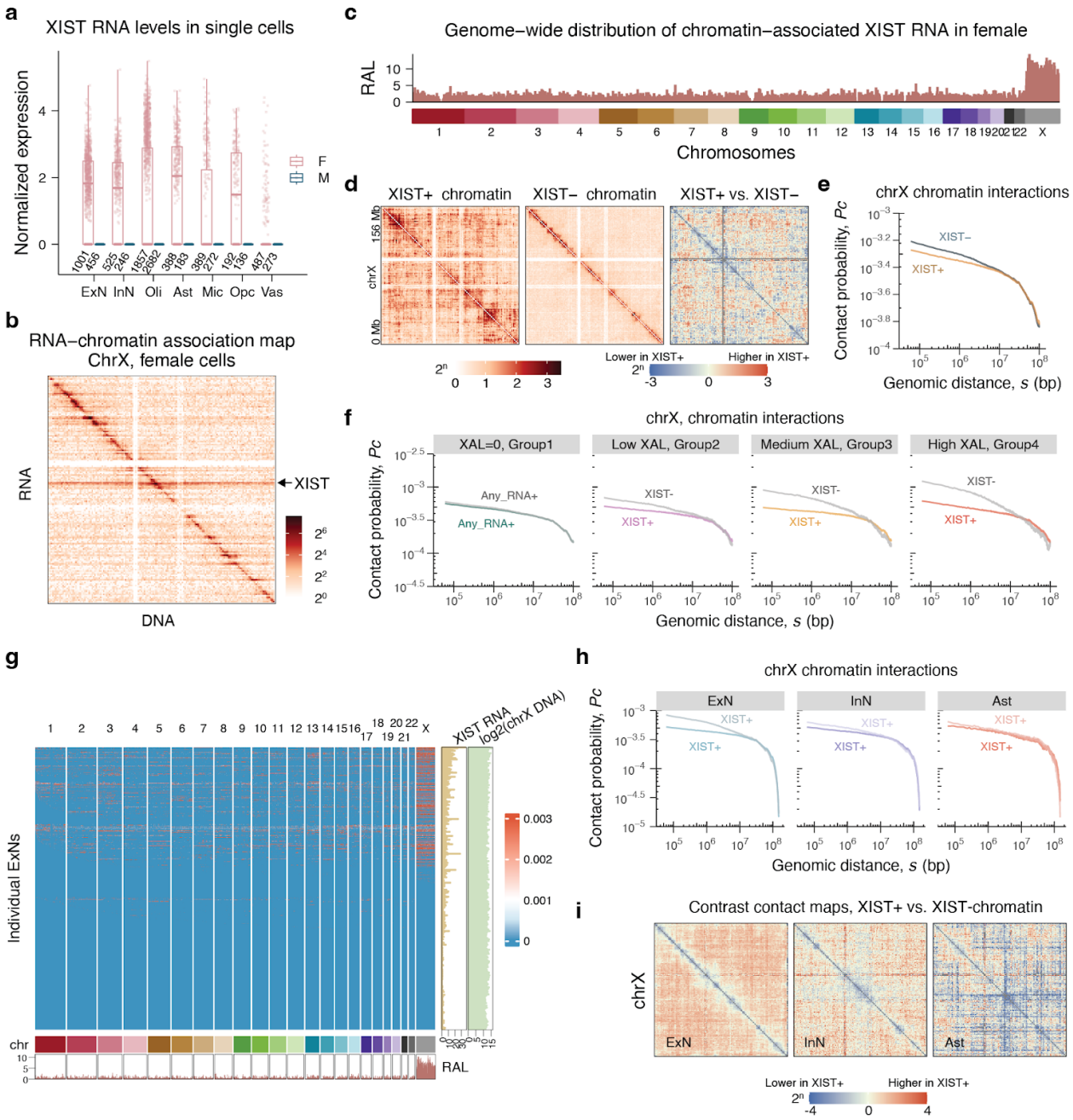

a
XIST RNA levels in single cells
Normalized expression
F
M
ExN InN Oli Ast Mic Opc Vas
b
RNA–chromatin association map
ChrX, female cells
RNA
DNA
XIST
c
Genome–wide distribution of chromatin–associated XIST RNA in female
RAL
Chromosomes
d
XIST+ chromatin
XIST− chromatin
XIST+ vs. XIST−
156 Mb
0 Mb
chrX
Lower in XIST+
Higher in XIST+
e
chrX chromatin interactions
Contact probability, Pc
XIST−
XIST+
Genomic distance, s (bp)
f
chrX, chromatin interactions
XAL=0, Group1
Low XAL, Group2
Medium XAL, Group3
High XAL, Group4
Any_RNA+
XIST-
XIST+
Contact probability, Pc
Genomic distance, s (bp)
g
Individual ExNs
XIST RNA
log2(chrX DNA)
chr
RAL
h
chrX chromatin interactions
ExN
InN
Ast
XIST+
Contact probability, Pc
Genomic distance, s (bp)
i
Contrast contact maps, XIST+ vs. XIST-chromatin
chrX
ExN
InN
Ast
Lower in XIST+
Higher in XIST+

Figure S1. | Mixed-species analysis. (a) The cells (dots) with more than 1000 reads (read count filter = 1000) are colored coded to red (human), blue (mouse) if 95% or more of the reads are mapped to a single species (purity filter = 95%), or yellow (mixed) if otherwise. The cells with less than 1000 reads (gray dots, ambient) are not used in the calculation of mixed-species rate. (b) The complexes (dots) are colored coded to red (human), blue (mouse) when 99% or more of the reads are mapped to a single species, or yellow (mixed) if otherwise.

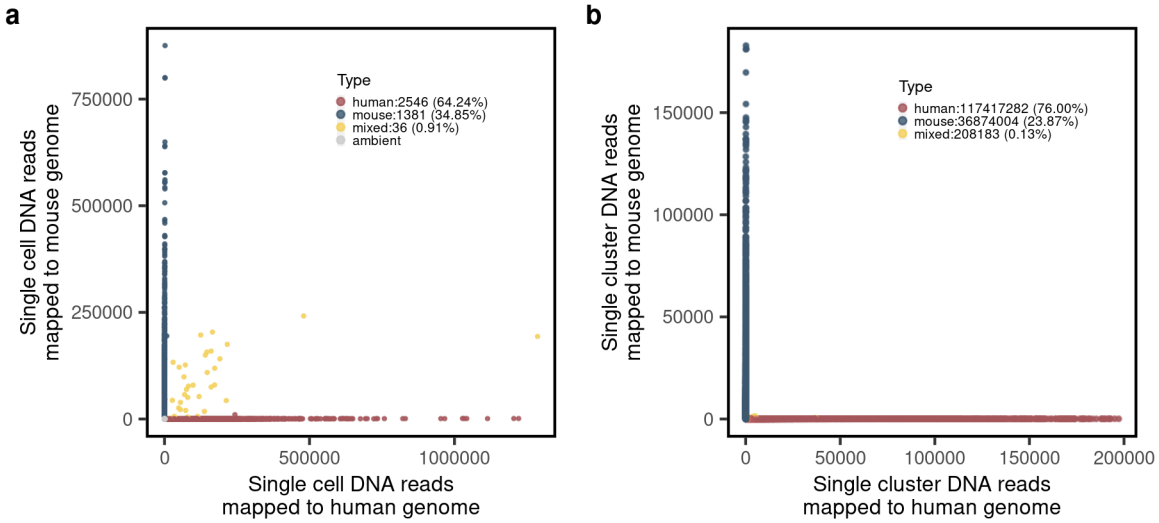


Figure S2. | Proportions of pairwise or multiplex interactions. The proportions of clusters (a) and DNA read counts (b) that consist of multiplex chromatin interactions (DNA reads >=3, light blue) in each cell type. Both RD and DD clusters are included in this analysis. (c) Proportions of clusters with pairwise (2 DNA reads, dark blue) or multiplex chromatin interactions (3 or more DNA reads, light blue) in DNA-only (DD) and RNA-DNA (RD) clusters in each cell type. (d) Proportions of DNA read counts with clusters corresponding to pairwise (dark blue) or multiplex (light blue) chromatin interactions.

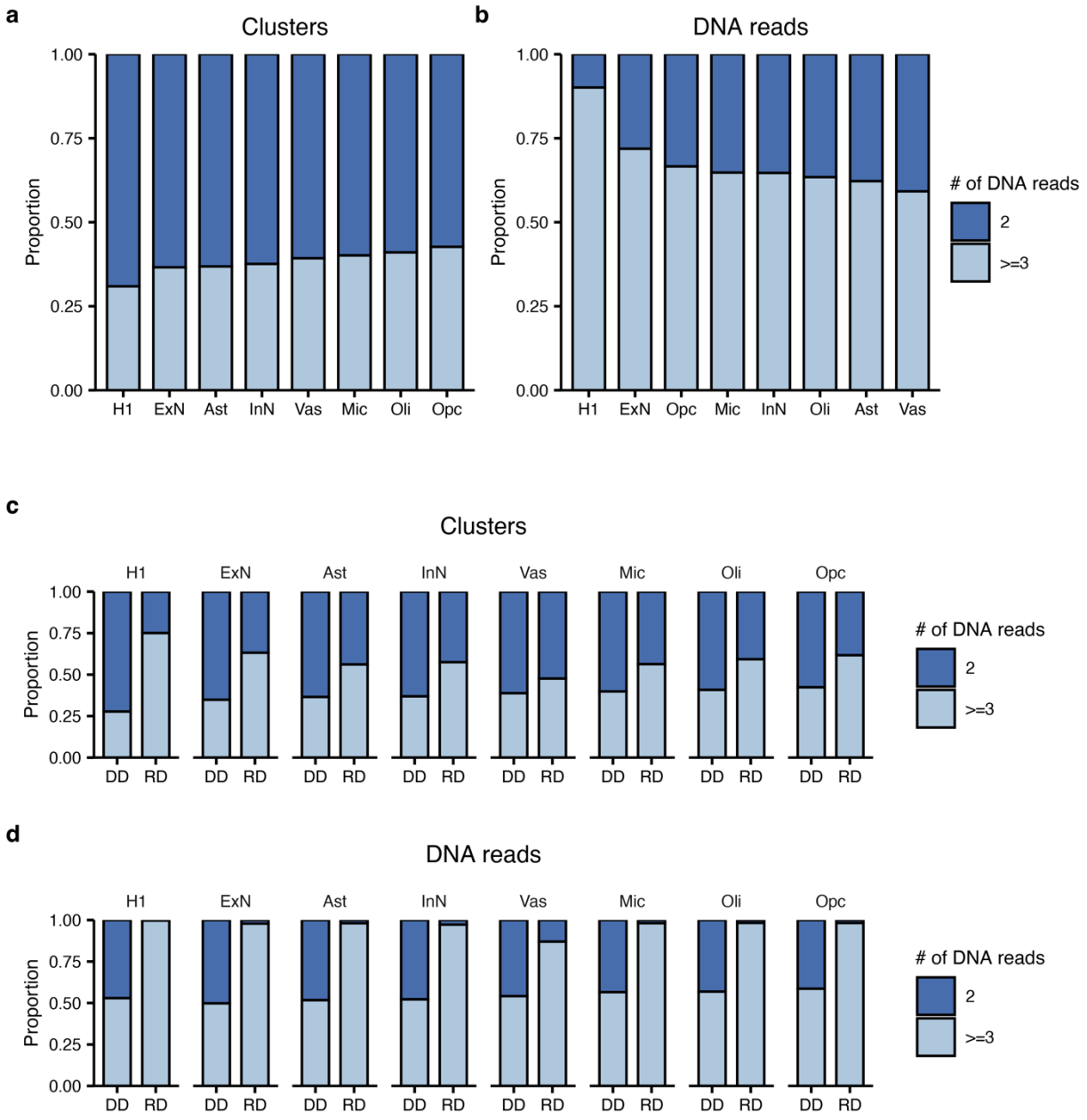


Figure S3. | Comparison of technologies. (a) The number of RNA reads per cell (x axis) vs. the number of detected genes per cell (y axis) of CITE-seq, SNARE-seq, PairTag, snRNA-seq, SNARE-seq, MUSIC H1 cells, and MUSIC frontal cortex. The central dot represents the average number, and the bars indicate the first and the third quartile of the distribution of all the cells (x axis) and per cell total detected genes (y axis). (b) The number of cells (x axis) vs. the median number of DNA-DNA contacts per cell (y axis) for each technique. For MUSIC and scSPRITE, pairwise interactions are projected from multiway complexes.

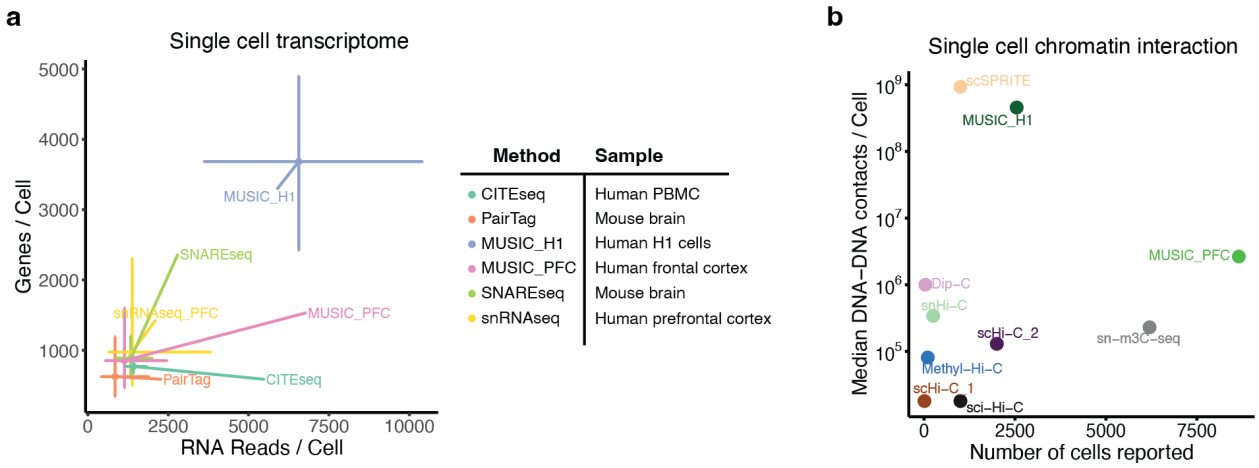


Figure S4. | MUSIC analysis of human frontal cortex. Marker gene expression in brains (a), number of cells (b), percent of female (pink) and male cells (blue) (c), and the relative proportion of each sample (d) in every single-cell cluster, and the assigned cell type and subtype for each cluster. Assigned subtypes and marker gene expression in excitatory neurons (g-i), inhibitory neurons (j-l), microglias (m-o), and oligodendrocytes (p-r). (e) Female (pink) and male (blue) cells in the UMAP embedding, showing a sex stratification. (f) Histogram of the proportions of single cells (y axis) with their most frequent chromatin interactions in each genomic bin (0-1Mb, 1Mb-150Mb) (x axis) in each cell type.

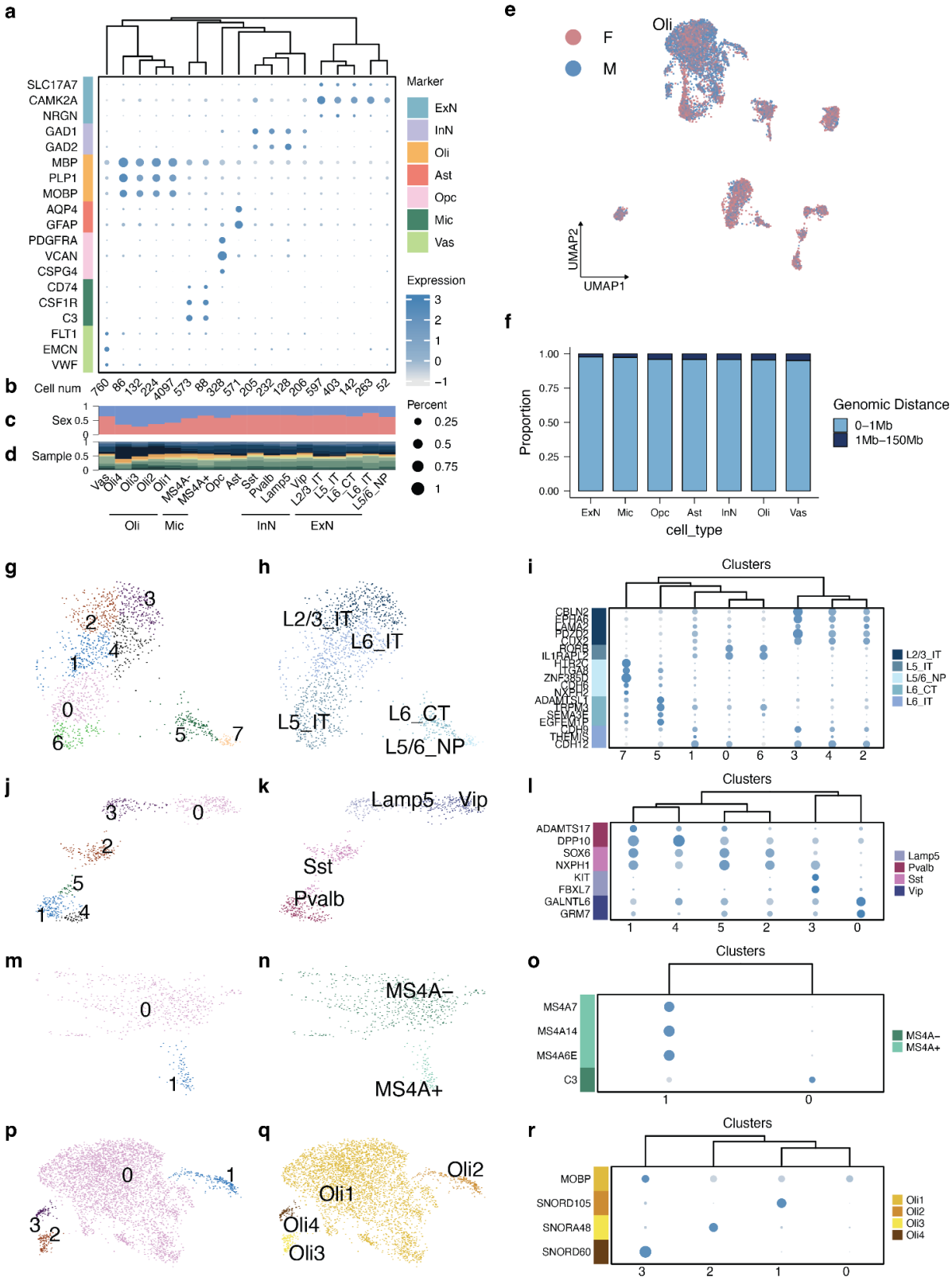

a
SLC17A7
CAMK2A
NRGN
GAD1
GAD2
MBP
PLP1
MOBP
AQP4
GFAP
PDGFRA
VCAN
CSPG4
CD74
CSF1R
C3
FLT1
EMCN
VWF
Marker
ExN
InN
Oli
Ast
Opc
Mic
Vas
Expression
b
Cell num
760 86 132 224 4097 573 88 328 571 205 232 128 206 597 403 142 263 52
c
Sex
d
Sample
Vas Oli4 Oli3 Oli2 Oli1 MS4A- MS4A+ Opc Ast Sst Pvalb Lamp5 Vip L2/3_IT L5_IT L6_CT L6_IT L5/6_NP
Oli Mic InN ExN
Percent
0.25
0.5
0.75
1
e
F
M
Oli
UMAP2
UMAP1
f
Proportion
cell_type
ExN Mic Opc Ast InN Oli Vas
Genomic Distance
0–1Mb
1Mb–150Mb
g
h
L2/3_IT
L6_IT
L5_IT
L6_CT
L5/6_NP
i
Clusters
L2/3_IT
L5_IT
L5/6_NP
L6_CT
L6_IT
j
k
Lamp5
Vip
Sst
Pvalb
l
Clusters
ADAMTS17
DPP10
SOX6
NXPH1
KIT
FBXL7
GALNTL6
GRM7
Lamp5
Pvalb
Sst
Vip
m
n
MS4A–
MS4A+
o
Clusters
MS4A7
MS4A14
MS4A6E
C3
MS4A–
MS4A+
p
q
Oli1
Oli2
Oli3
Oli4
r
Clusters
MOBP
SNORD105
SNORA48
SNORD60
Oli1
Oli2
Oli3
Oli4

Figure S5. | GEAS. (a) A schematic representation of GEAS-scan. A kernel representing the stripe pattern slides over the genome to produce a correlation score with the 2-D contact matrix in every sliding step. (b) Boxplots of the gene expression levels, measured in RPKM (Reads Per Kilobase Million), for the genes that colocalize (in_stripe) and do not colocalize (other) with chromatin stripes in each cell type. Wilcoxon test p-value of the two distributions of each cell type is provided on the top.

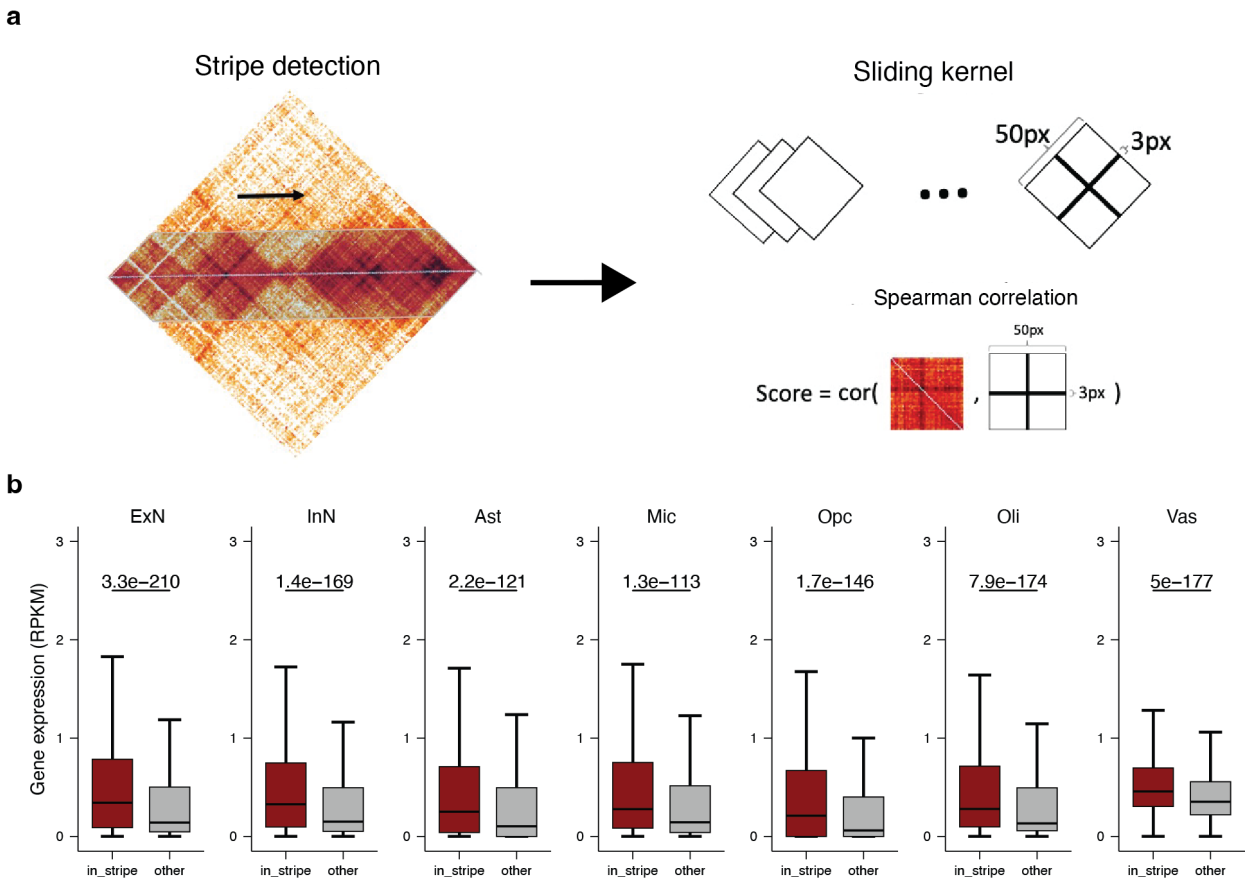
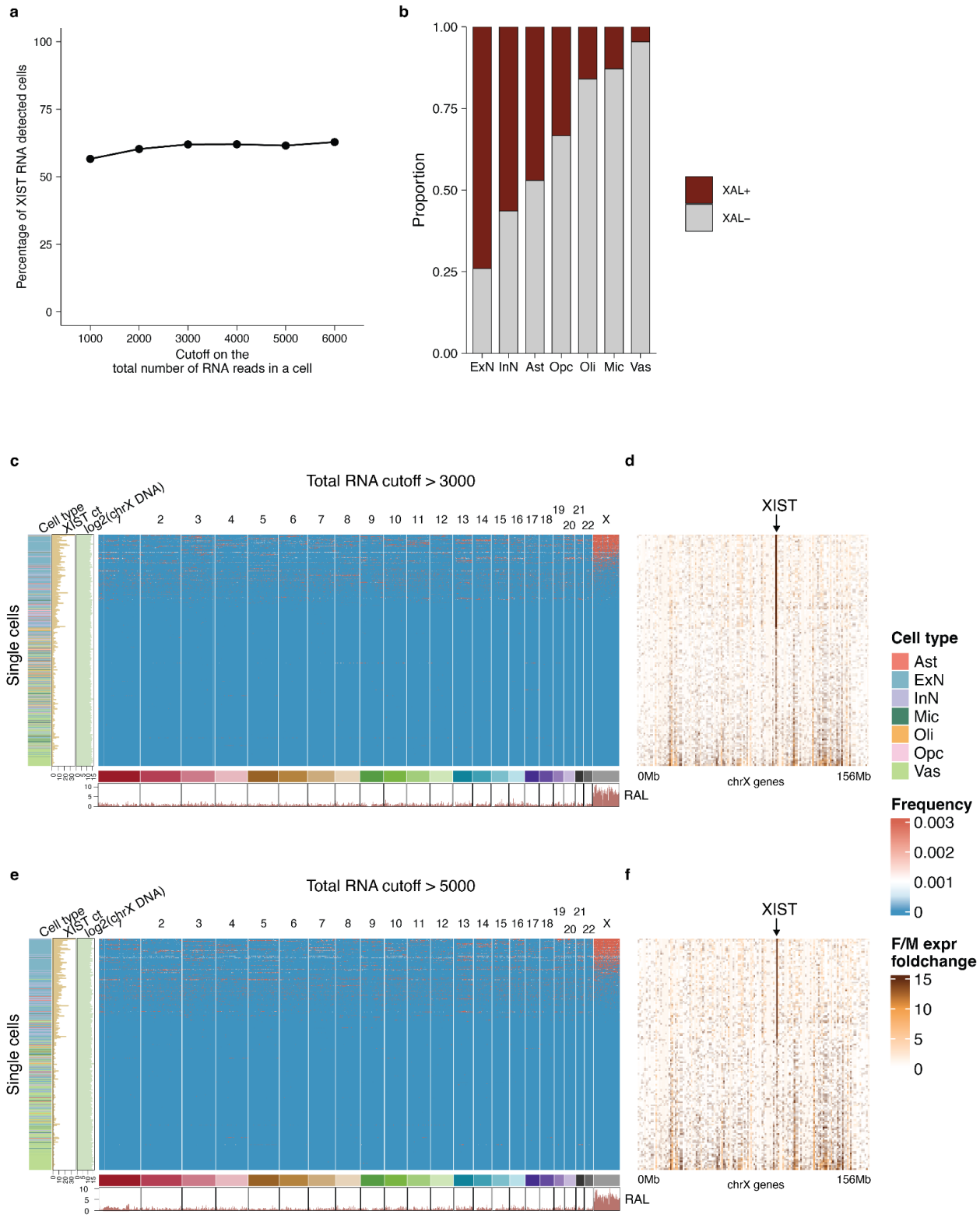


Figure S6. | XIST-chromatin association in single female cells. (a) The change of the percentage of XIST lncRNA detected cells (y axis) in the female cortical cells that satisfy the threshold on the total number of RNA reads in a cell (x axis). (b) The variation of the proportions of the observed XAL+ cells across cell types (columns) in the female cortex under total RNA larger than 5000 cutoff. (c-f) The female cells are filtered so that every cell has at least 3000 RNA reads (c-d) or at least 5000 RNA reads (e-f). (c, e) Genome-wide distribution of XIST-chromatin association in individual female cortical cells. Each row represents a single cell. The RNA attachment level (RAL) of the XIST lncRNA (intensity of red color) at any genomic region on any chromosome is plotted with the corresponding genomic coordinates (x axis). Resolution: 1 Mb. Track at the bottom: cumulative RAL of XIST. Tracks on the left indicate the cell type (color), XIST RNA read count (XIST RNA), X chromosomal DNA read counts (chrX DNA) in log scale. (d, e) sex-fold-change of a previously identified gene with incomplete XCI (column) between this female cell (row) and the average expression of the male cells of the matched cell type. Only the genes expressed in at least 20 cells are plotted.


a
Percentage of XIST RNA detected cells
100
75
50
25
0
1000
2000
3000
4000
5000
6000
Cutoff on the
total number of RNA reads in a cell
b
Proportion
1.00
0.75
0.50
0.25
0.00
ExN InN Ast Opc Oli Mic Vas
XAL+
XAL−
c
Total RNA cutoff > 3000
Cell type
XIST ct
log2(chrX DNA)
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
Single cells
RAL
d
XIST
0Mb
chrX genes
156Mb
Cell type
Ast
ExN
InN
Mic
Oli
Opc
Vas
Frequency
0.003
0.002
0.001
0
F/M expr
foldchange
15
10
5
0
e
Total RNA cutoff > 5000
Cell type
XIST ct
log2(chrX DNA)
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
Single cells
RAL
f
XIST
0Mb
chrX genes
156Mb

Supplementary Tables

Supplementary Table 1 | A partial list of single-cell multi-modal technologies. “CARs”: chromatin accessible regions. “Co-complex pairs”: variations of Hi-C technologies that sample a pair of DNA fragments from a chromatin complex involving either pairwise or multiplex interactions. “Multiplex interactions”: technologies that can measure multiplex interactions and thus differentiate pairwise and multiplex interactions.

Method	Imaging/Omics	Surface protein	CARs	Histone modification	DNA methylation	3D genome organization			Gene expression	RNA-chromatin association
						Chromatin traces	Co-complex pairs	Multiplex interactions		
DNA seqFISH+	Imaging			✓		✓			✓	
ORCA	Imaging					✓			✓	
MINA	Imaging					✓			✓	
Snare-seq	Omics		✓						✓	
SHARE-seq	Omics		✓						✓	
scNMT-seq	Omics		✓		✓					
scNOMeRe-seq	Omics		✓		✓				✓	
Pair-tag	Omics			✓					✓	
sn-m3C-seq	Omics				✓		✓			
CITE-seq	Omics	✓							✓	
HiRES	Omics						✓		✓	
MUSIC	Omics							✓	✓	✓

Supplementary Table 2 | Transcriptome profiling statistics from multi-modal single cell sequencing technologies. All statistics for each technique are computed based on the raw count matrix downloaded from each published dataset (CITE-seq: GSE100866_PBMC, SNARE-seq: GSE126074_AdBrainCortex, PairTag: GSE152020, snRNA-seq: Synapse: syn18485175).

Method	Median RNA reads / cell	Medium genes / cell	Number of cells in Raw count table	Number of genes in Raw count table	Sample
snRNA-seq	1,973	1,348	80,660	18,192	Human prefrontal cortex
CITE-seq	1,416	770	7,985	29,929	Human PBMC cells
SNARE-seq	1,332	887	10,309	33,160	Mouse adult brain cerebral cortex
Pair-Tag	845	626	64,849	52,781	Mouse frontal cortex and hippocampus
MUSIC_H1	6,559	3,684	2,546	39,007	Human embryonic H1 cell line
MUSIC_PFC	1,136	853	9,067	22,373	Frozen human frontal cortex

Supplementary Table 3 | Marker genes for each cell type and subtype.

Type	Marker genes	Subtype	Marker genes
ExN	SLC17A7, CAMK2A, NRGN	L2/3_IT	CBLN2, EPHA6, LAMA2, PDZD2, CUX2
		L5_IT	RORB, IL1RAPL2
		L6_IT	CDH9, THEMIS, CDH12
		L5/6_NP	NPSR1.AS1, HTR2C, ITGA8, ZNF385D, CDH6, NXPH2
		L6_CT	ADAMTSL1, TRPM3, SEMA3E, EGFEM1P
InN	GAD1, GAD2	Vip	GALNTL6, GRM7
		Lamp5	KIT, FBXL7
		Sst	SOX6, NXPH1
		Pvalb	ADAMTS17, DPP10
Mic	CD74, CSF1R, C3	MS4A-	N/A
		MS4A+	MS4A7, MS4A14, MS4A6E
Oli	MBP, PLP1, MOBP		
Opc	PDGFRA, VCAN, CSPG4		
Ast	AQP4, GFAP		
Vas	FLT1, EMCN, VWF		

Supplementary Table 4 | Brain sample metadata.

Sample ID	Sex	Age at death	ApoE	PlaqueTotal	TangleTotal	Braak score
B1	M	63	3/4	14.5	15	VI
B2	M	64	3/3	14.5	15	VI
B3	M	71	3/3	0	4	III
B4	M	73	3/3	0	2.25	II
B5	F	75	3/4	15	15	VI
B6	F	75	3/3	0	2.5	II
B7	F	82	3/3	15	14.5	VI
B8	F	82	2/3	1	1.5	I
B9	M	71	4/4	14	15	VI
B10	M	79	2/3	0	1	II
B11	F	59	3/3	1	0.5	I
B12	F	60	3/3	15	15	VI
B13	M	≥90	2/4	14.5	14	VI
B14	F	≥90	3/3	0	6.5	III