Transformation of single-cell three-dimensional genome structure during postnatal development of the mammalian brain

Longzhi Tan[1,2*], Dong Xing[1,3,4], Nicholas Daley[1,5,6], X. Sunney Xie[3,4*]

[1]Department of Chemistry & Chemical Biology, Harvard University, Cambridge, MA, USA. [2]Present address: Department of Bioengineering, Stanford University, Stanford, CA, USA. [3]Present address: Beijing Innovation Center for Genomics, Peking University, Beijing, China. [4]Present address: Biomedical Pioneering Innovation Center, Peking University, Beijing, China. [5]Belmont Hill School, Belmont, MA, USA. [6]Present address: Harvard College, Cambridge, MA, USA. *e-mail: tttt@stanford.edu; sunneyxie@pku.edu.cn

Abstract

After birth, the mammalian brain undergoes numerous molecular changes that underlie cognitive plasticity and maturation. However, little is known about the dynamics of three-dimensional (3D) genome structure during this period. Here we generated a 3D genome atlas of 1,954 single cells from the developing mouse cortex and hippocampus, using our diploid chromatin conformation capture (Dip-C) method. In adult tissues, genome structure alone delineates major cell types such as cortical excitatory and inhibitory neurons, and hippocampal granule cells. During development, a major transformation was observed between the first and fourth week after birth—coincident with synaptogenesis and de novo DNA methylation, leading to 3D re-organization across multiple genomic scales. Using reciprocal crosses, we systematically examined allele-specific structure of imprinted genes, revealing chromosome-wide difference at the Prader-Willi/Angelman syndrome locus. These findings thus uncover a previously unknown dimension of postnatal brain development.

Main Text

The three-dimensional (3D) organization of the mammalian genome plays a critical role in the development of nervous systems. Dysregulation of chromatin organization in the brain leads to various neurodevelopmental disorders such as autism spectrum disorder, intellectual disability, and schizophrenia. It is therefore crucial to understand the dynamics of 3D genome folding during brain development.

Recently, bulk chromatin conformation capture (3C/Hi-C) experiments revealed extensive 3D genome re-folding during neuronal differentiation *in vitro* and in the embryonic (day 14.5) mouse brain (*1*). However, it remains unknown whether the 3D genome continues to transform after birth, as the brain undergoes dramatic changes in synaptic connectivity and neural plasticity. Furthermore, bulk Hi-C cannot capture the diversity of cell types—both neurons and glia—in the brain, necessitating a single-cell approach.

Using our high-resolution single-cell method, termed diploid chromatin conformation capture (Dip-C) (*2*), we have previously characterized 3D genome re-organization in sensory neurons during postnatal development of the mouse eye and nose (*3*). Here we applied Dip-C to the developing mouse forebrain and uncovered a major transformation of 3D genome structure in the first month after birth.

To create a 3D genome atlas of the developing mouse forebrain, we performed Dip-C on 1,954

single cells from the cortex and hippocampus at different ages throughout the first postnatal year: on postnatal days (P) 1, 7, 28, 56, 309, and 347 (Fig. 1a, Fig. 1b). All animals were filial 1 (F1) hybrids from a CAST/EiJ♀ × C57BL/6J♂ cross (denoted the initial cross) except the P309 animal, which was from a reciprocal cross (C57BL/6J♀ × CAST/EiJ♂) (Fig. 1b) for parent-of-origin analysis of allele-specific structures.

We obtained an average of 401 k chromatin contacts per cell (SD = 116 k, minimum = 101 k, maximum = 1.20 m) (Extended Data Fig. 1, Supplementary Table 1). Among the cells, 834 (43%) yielded high-quality 3D structures at 20-kb resolution, defined as root-mean-square deviation (r.m.s.d.) ≤ 1.5 particle radii (average = 0.90, SD = 0.28 among those cells; each 20-kb particle has a radius of ~100 nm) (Supplementary Table 1). These high-resolution contact maps and 3D structures provide a unique opportunity to study cell-type-specific genome structure and its developmental time course.

Single-cell A/B chromatin compartment values along the genome (*2, 3*) enabled us to separate major cell types of the brain based on 3D genome structure alone ("structure typing"). We have previously used principal component analysis (PCA) of A/B compartment values to distinguish cell types in the human blood (*2*) and in the mouse eye and nose (*3*). However, cell types are more complex in the brain; for example, a recent study could not tell apart single neurons of different sub-types from contact maps alone (*4*). To solve this problem, we performed t-distributed stochastic neighbor embedding (t-SNE) on the first 20 principal components (PCs) of single-cell A/B compartment values (Fig. 1c), in a manner similar to those widely used in transcriptomic studies. This approach clearly visualized multiple cell-type clusters, each with distinct age and tissue profiles (Fig. 1c left and middle). Our result is robust against down-sampling contacts, choosing different bin sizes or numbers of PCs, or substituting t-SNE with uniform manifold approximation and projection (UMAP) (Extended Data Fig. 2).

Hierarchical clustering of single-cell A/B compartment values (the first 20 PCs) identified 13 major clusters and 3 smaller clusters (Fig. 1d). Based on projection of published bulk Hi-C data (*1, 5-7*) onto the t-SNE plot (Fig. 1c right) and multi-modal integration of other omics data (see below), we classified 8 of the major clusters as neurons (denoted N1–8), 2 as astrocytes (A1–2), 2 as oligodendrocytes (O1–2), and the other microglia and endothelial cells (M) (Fig. 1d). Surprisingly, between P7 and P28, cluster composition of cells shifted drastically and concomitantly for neurons (from N1–2 to N3–8), astrocytes (from A1 to A2), and oligodendrocytes (from O1 to O1–2), indicating a major transformation of the underlying 3D genome "structure types" in all 3 lineages in the first postnatal month (Fig. 1e).

Integration with published transcriptomic (*8-11*) (Supplementary Table 2), methylomic (*12*), and chromatin accessibility (*13, 14*) data provided a detailed look into each cluster (Fig. 2a). We have previously shown that cell-type-specific genes (when averaged) tend to show higher A/B compartment values (namely, more A-like, more open, and more euchromatic) specifically in single cells of the corresponding cell type (*3*). Here we extend this correlation from gene expression to accessible chromatin regions and un-methylated DNA regions (Fig. 2a).

In oligodendrocytes, our A/B compartment analysis of enriched genes (*8*) and accessible regions (*13*) identified the O1 cluster as oligodendrocyte progenitors, and O2 as mature oligodendrocytes (Fig. 2a second section). The observed 3D genome change in oligodendrocytes therefore reflects the emergence of an additional cell type (O2: mature oligodendrocytes) between P7 and P28, rather than a change of existing cell types.

In astrocytes and neurons, however, the concomitant 3D genome changes led to a nearly complete transformation of their neonatal structure types (N1–2 and A1 at P1–7) into adult types (N3–8 and A2 from P28 onward) (Fig. 1e middle two rows). In astrocytes, both the neonatal (A1) and adult (A2) clusters exhibited similar A/B compartment values for astrocyte-specific genes (*8*); they differed, however, in average A/B compartment values for genes that are up- and down-regulated during postnatal astrocyte development, respectively (*8*) (Fig. 2a middle section). We also found adult astrocytes to intermingle their chromosomes less and to exhibit weaker A/B compartments (Fig. 2b). This suggests a switch in 3D genome structure for all astrocytes from P7 to P28.

We now focus on the neurons. Interestingly, only adult neurons (P28 onward) clearly separated into distinct sub-types on the t-SNE plot, while neonatal ones (P7 and earlier) clustered relatively close to each other (Fig. 1d). This suggests that, although already transcriptionally distinct at birth, 3D genome structure of neurons is relatively homogeneous in neonates, and only further differentiates into mature structure types in adults.

Among adult neurons, we found the 6 clusters (N3–8) to correspond to major neuronal sub-types of the two brain regions. We first inferred the hippocampus-specific N8 cluster (Fig. 1e) to be granule cells of the dentate gyrus (DG), because they exhibited the highest A/B compartment values for genes (*10*) and accessible regions (*14*) specific to those cells (Fig. 2a bottom section); this suggests that the other hippocampus-specific cluster, N3, is presumably pyramidal neurons of the *Cornu Ammonis* (CA) areas. A cluster shared by both cortex and hippocampus, N7, was similarly inferred to be inhibitory neurons by compartment analysis of enriched genes (*10*), accessible regions (*13*), and un-methylated regions (*12*) (Fig. 2a bottom section). The remaining clusters, N4–6, were presumably pyramidal neurons of various cortical layers. Representative 3D structures are shown in Fig. 2c. These 6 clusters therefore provide the structural basis for functional and anatomical sub-types of neurons in the brain.

We then ask how neonatal and adult neurons—completely separated on the t-SNE plot—differ in 3D genome structure. Genes that are up-regulated during early postnatal development (P2–11) (*11*) were more compartment A-like in adult neurons (N3–8) than in neonatal ones (N1–2) (Fig. 2a bottom row), similar to what we observed in olfactory sensory neurons (*3*). We also found N1—the cluster enriched for the youngest neurons (P1)—to intermingle their chromosomes more than the other 7 clusters (Fig. 2b).

The most prominent difference between neonatal and adult neurons, however, is radial positioning (preference for the nuclear periphery or interior) along the genome—a feature inaccessible to regular Hi-C experiments. Although the 6 adult clusters (N3–8) each exhibited cell-type-specific radial positioning, they differed consistently from the neonatal clusters (N1–2) (Fig. 3a). Many regions, most notably the entire Chr 7 and the centromeric half of Chr 17 (asterisks in Fig. 3a), were relocated significantly inward from neonates to adults. Fig. 3b shows the distribution of radial positions of Chr 7 among single cells. Such radial transformation was absent from astrocytes and oligodendrocytes (Fig. 3a bottom), and is therefore highly specific to neurons.

Regions that moved inward during postnatal neuronal development were highly enriched in previously identified “methylated CpH (mCH) deserts” (*15*) (Fig. 3c). These mCH deserts are low in CpG frequency and chromatin accessibility, sometimes harbor large gene clusters such as olfactory receptors (ORs) and vomeronasal receptors (VRs), and most importantly, escape

neuron-specific global DNA CpH methylation—a major epigenomic event at the same time (P7–28) (*15*) as our observed 3D genome transformation. This highlights a complex relationship between DNA methylation and 3D genome structure, because regions that moved radially are primarily those left untouched by de novo DNA methylation.

Such specific radial re-configuration of CpG-poor regions (Fig. 3c, Extended Data Fig. 3)—including OR and VR gene clusters—is reminiscent of our observation in olfactory sensory neurons, where ORs also move inward during development (*3, 16*). Here we also observed increased inter-chromosomal interactions between the relocated regions (Extended Data Fig. 4); however, these regions did not increase in their overall extent of inter-chromosomal interactions (Extended Data Fig. 5), unlike the olfactory case (*17*). This suggests both conserved and divergent mechanisms behind 3D genome transformation of central and peripheral nervous systems.

Genome is also re-folded on finer scales during this developmental transformation. We frequently observed local re-wiring of compartment B-like regions, such as gene deserts and gene clusters (for example, ORs and VRs), and strengthening of intra-chromosomal interactions between them (Fig. 3d top, Extended Data Fig. 6). We quantified such structural transformation by both changes in the fraction of cells where each pair of 20-kb particles are nearby in 3D (distance $\leq$ 3.0 radii) (*3*) for short-distance interactions (equivalent to contact maps in bulk Hi-C, but with an absolute scale and clear 3D interpretation), and changes in their median 3D distance for longer-distance interactions (which cannot be obtained with regular Hi-C) (Fig. 3d). This allowed us to also visualize complex 3D changes, which frequently involved increased isolation between nearby domains and changes in chromatin loops, near genes critical for synaptogenesis and/or linked to neurodevelopmental disorders (for example, *Kcnma1*, *Kalrn*, *Adcy5*, *Sema5b*, *Map2*, *Unc80*, *Dlg4*, *Ryr2*, *Rbfox1*, *Grin2a*, *Syne1*, *Syt1*, and *Nav3*) (Fig. 3d, Extended Data Fig. 6).

Finally, we look at allele-specific genome structure around all known imprinted genes in the brain. Genomic imprinting is crucial for development and linked to multiple disorders; however, its allele-specific 3D structure has not been examined systematically across the genome. We have previously visualized allele-specific chromatin loops at an imprinted locus (*H19*/*IGF2*) in the human blood based on a single donor (*2, 18*), which is insufficient for stringent discoveries. Here we used reciprocal crosses to eliminate confounding factors (primarily artifacts caused by strain-specific copy-number/structural variations, but also potentially strain-specific 3D structure) and revealed true parent-of-origin effects (Fig. 4a). Out of all 115 known imprinted genes in the mouse brain (*19*), which form ~29 clusters along the genome, we found strong evidence for parent-of-origin-specific 3D structure in 7 clusters and potential evidence in 1 (Fig. 4b). We quantified allele-specific difference with a non-parametric statistic—the fraction of cells where the paternal 3D distance is greater than the maternal one—for each pair of particles (Fig. 4c, Extended Data Fig. 7, Extended Data Fig. 8), which cannot be obtained with regular Hi-C.

All our observed parent-of-origin effects were relatively local, with one notable exception: the Prader-Willi/Angelman syndrome (PWS/AS) locus on Chr 7 (corresponding to human Chr 15). Two alleles of this gene cluster not only exhibited different domain organization within the ~3-Mb cluster (Fig. 4d bottom), but also interacted differently with multiple regions across the chromosome (Fig. 4d top). In particular, only the maternal allele interacted with multiple Mb-sized, compartment B-like regions across Chr 7, up to tens of Mb away. Such difference is also

apparent from raw contact maps, which did not involve any haplotype imputation or 3D modeling (Extended Data Fig. 9). This surprisingly long-range effect offers new insights into the mechanism behind genomic imprinting and its disorders.

Mammals are born with a highly plastic brain that continues to develop its myriad cognitive functions over time. Our discovery of cell-type-specific 3D genome transformation during this period provides a new molecular basis for neural plasticity, and opens up opportunities for 3D genome-based treatments for neurodevelopmental and genomic imprinting disorders.

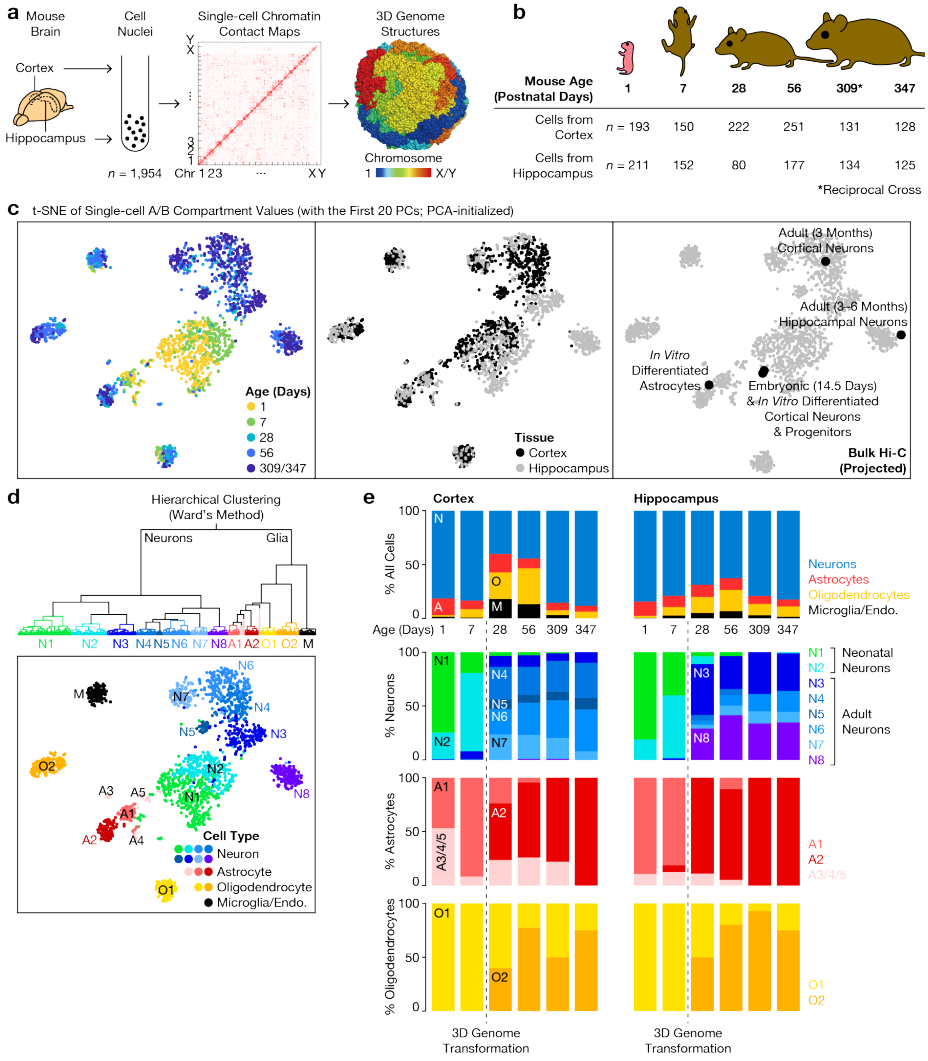


Fig. 1 | Diploid chromatin conformation capture (Dip-C) of the mouse forebrain during postnatal development. a, We performed Dip-C on 1,954 single cells from the developing mouse cortex and hippocampus, obtaining their chromatin contact maps (middle) and 834 high-quality 3D genome structures at 20-kb (~100 nm) resolution (right). **b**, Number of single cells from each tissue and age. Animals were filial 1 (F1) hybrids from a CAST/EiJ♀ × C57BL/6J♂ cross (the initial cross) except for the one at postnatal day (P) 309 (asterisk), which was from a reciprocal cross (C57BL/6J♀ × CAST/EiJ♂) for parent-of-origin analysis (Fig. 4). P1 animals were female; the rest were male. **c**, Major cell types—each with distinct age (left) and tissue (middle) profiles—can be separated by t-distributed stochastic neighbor embedding (t-SNE) of the first 20 principal components (PCs) of single-cell A/B compartment values along the genome (defined for each 1-Mb bin as the average CpG frequency of bins that it contacted, rank-normalized to 0–1 in each cell (*3*)). To avoid sex effects, only values along the autosomes were used. Right: bulk Hi-C data from (*1, 5-7*) are projected onto the t-SNE plot. **d**, Hierarchical clustering of A/B compartment values (the first 20 PCs) identified 13 major clusters (N1–8, A1–2, O1–2, and M) and 3 smaller ones (light red; A3–5), shown on both a dendrogram (top) and the t-SNE plot (bottom). **e**, Composition of clusters for each tissue and age, shown both for all cells (top; *n*: see b) and for each of the 3 main lineages (bottom; *n* = 157, 125, 89, 111, 112, 113, 178,

120, 55, 111, 106, 103 from left to right for neurons; n = 30, 12, 38, 23, 9, 7, 28, 16, 9, 19, 10, 8 for astrocytes; n = 3, 12, 55, 84, 6, 8, 4, 12, 12, 35, 14, 12 for oligodendrocytes). A major shift in structure-type clusters was observed between P7 and P28 (dashed line), indicating a 3D genome transformation.

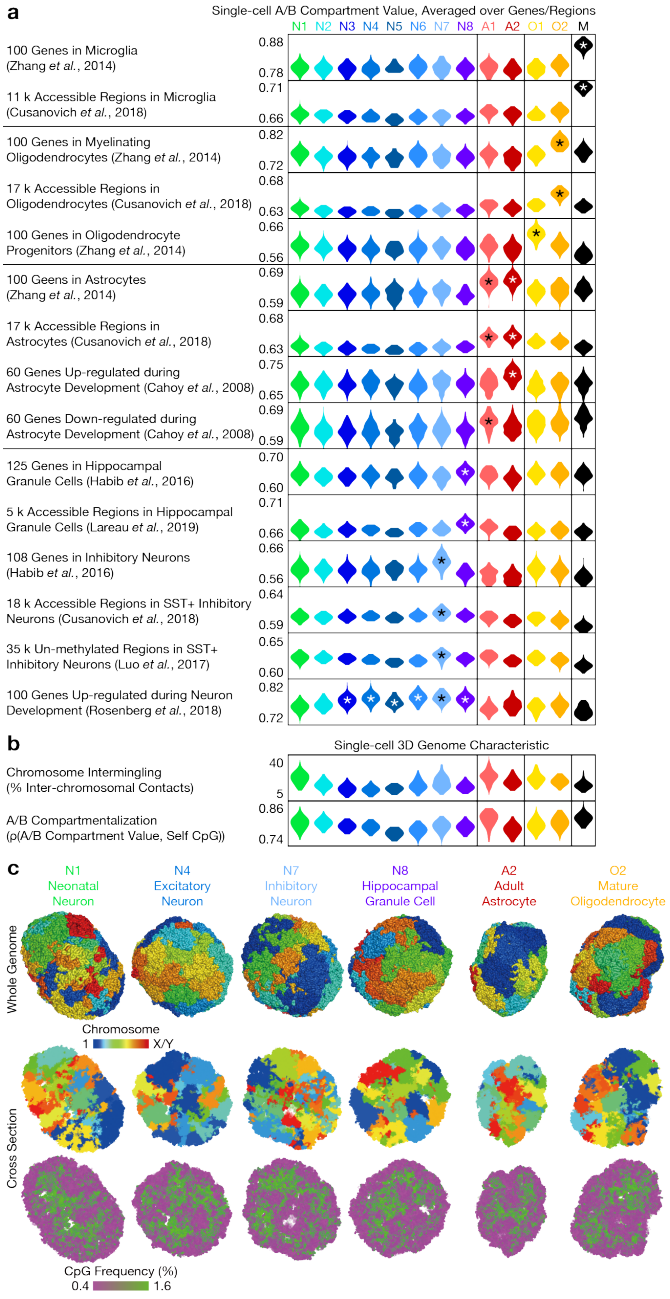


Fig. 2 | Cell-type-specific features of 3D genome structure. a, Cell-type-specific genes and regions from published transcriptomic (*8-11*), methylomic (*12*), and chromatin accessibility (*13, 14*) data showed different patterns of single-cell A/B compartment values across the 13 major cell-type clusters. Histograms of average A/B compartment values are shown as violin plots. Higher A/B compartment values (rank-normalized to 0–1 in each cell) indicate more compartment-A-like, more open, and more euchromatic properties. Cell numbers are n = 346, 239, 189, 135, 37, 187, 111, 136, 75, 93, 105, 152, 108 from left to right. Bandwidth (kernel window size) is 0.005 for gene lists and 0.01 for region lists. Asterisks denote cell-type clusters where values are the highest. **b**, Basic characteristics of 3D genome structure (*2*), calculated from single-cell contact maps, among the 13 major clusters. Bandwidth is 2% (top) and 0.01 (bottom). **c**, Representative 3D structures of 6 cell-type clusters. A/B compartmentalization (bottom) is

shown by CpG frequency as a proxy (high CpG for A, and low for B). Each cell is chosen to have a median extent of chromosome intermingling among its cell type.

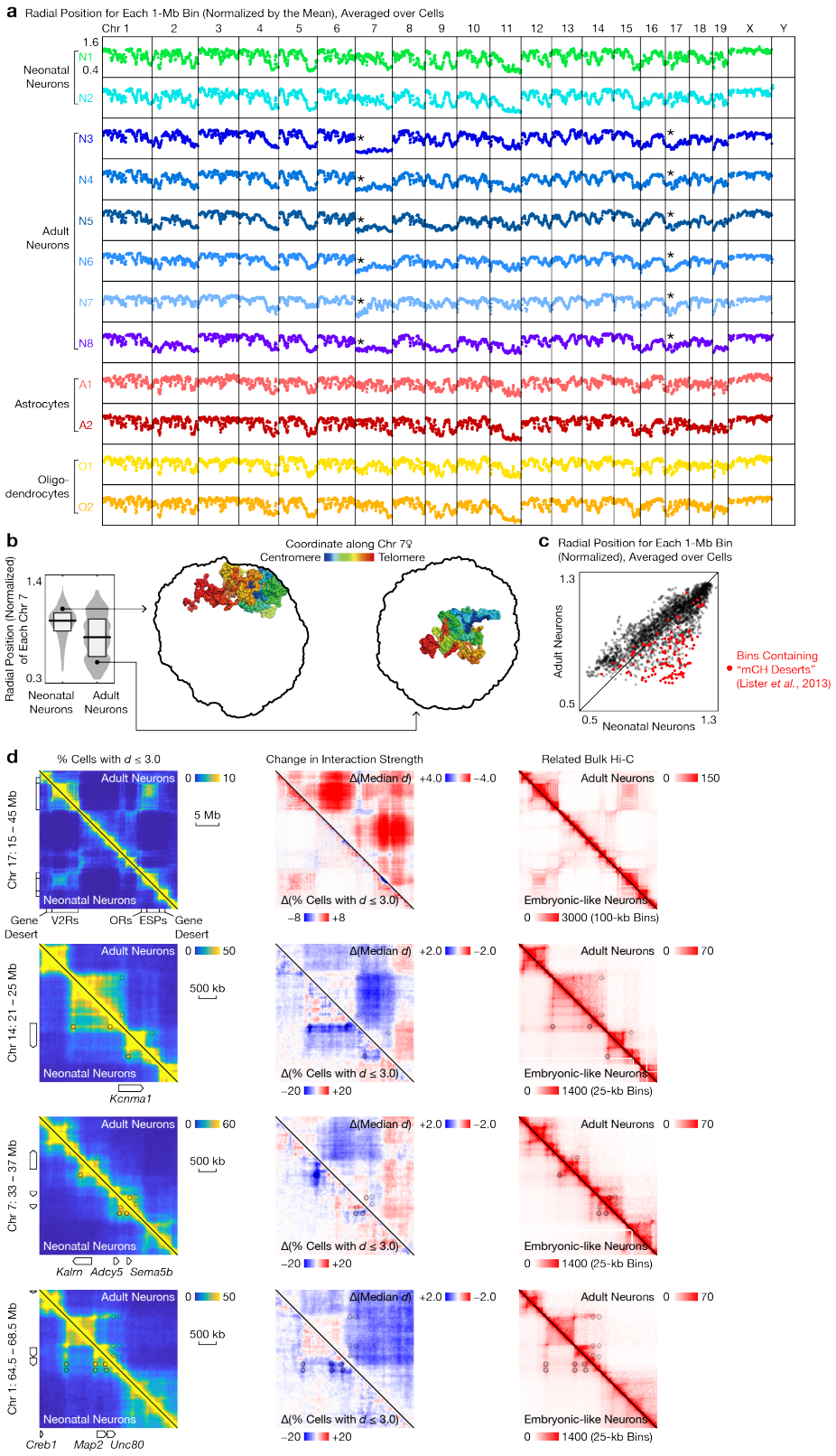


Fig. 3 | 3D genome is re-organized across multiple scales during postnatal neuronal development. a, Radial preference (distance to the nuclear center of mass, normalized by the

mean in each cell, and averaged across cells (*3*)) of each 1-Mb bin along the genome in each cell-type cluster. In neurons, many regions—most notably the entire Chr 7 and the centromeric half of Chr 17 (asterisks)—were relocated significantly inward (smaller radial values) from neonates to adults. Averages are calculated from n = 237, 110, 74, 37, 19, 90, 57, 86, 27, 19, 20, 41 cells × 2 alleles per cell, from top to bottom. If a bin is missing in some cells, an average is still calculated if $n \geq 5$ alleles. Note than values along sex chromosomes may not be directly comparable between clusters, because P1 animals were female and the rest were male. **b**, Distribution of the radial position of each Chr 7 (averaging all its 1-Mb bins) among single cells, shown as violin plots (left) and two example cells with peripherally and centrally positioned Chr 7, respectively (right). Horizontal lines and boxes denote the median and quartiles, respectively. n = 347 × 2 and 363 × 2 for neonatal and adult neurons, respectively. Bandwidth is 0.03. **c**, Comparison of radial preference between neonatal and adult neurons. Regions that moved inward during postnatal neuronal development were highly enriched in "methylated CpH (mCH) deserts" (*15*) (red). *n*: see b. If a bin is missing in some cells, an average is still calculated if $n \geq$ 20. **d**, Example developmental changes in local 3D genome structure (more in Extended Data Fig. 6). Left: comparison of the fraction of cells where 3D distance (d) is within 3.0 particle radii (~300 nm) for each pair of 20-kb particles, between neonatal (lower triangle) and adult (upper triangle) neurons. Circles denote chromatin loops that changed strengths (solid: stronger). *n*: see b. Two alleles in each cell are counted twice; only intra-homologous distance is analyzed. Some, but not all, genes and gene clusters are shown. V2R: type-2 vomeronasal receptor; OR: olfactory receptor; ESP: exocrine-gland-secreting peptide. Middle: changes in chromatin interactions are quantified by changes of the left heatmaps for short-range interactions (lower triangle) and changes in median 3D distances for longer-range interactions (upper triangle). Right: bulk Hi-C data from (*1*) for *in vitro* differentiated, embryonic-like neurons (no data are available for neonatal neurons) and (*5*) for adult neurons, visualized with Juicebox.js (*20*) (balanced normalization). Color scales are tuned (a 20:1 ratio) to best match the two samples.

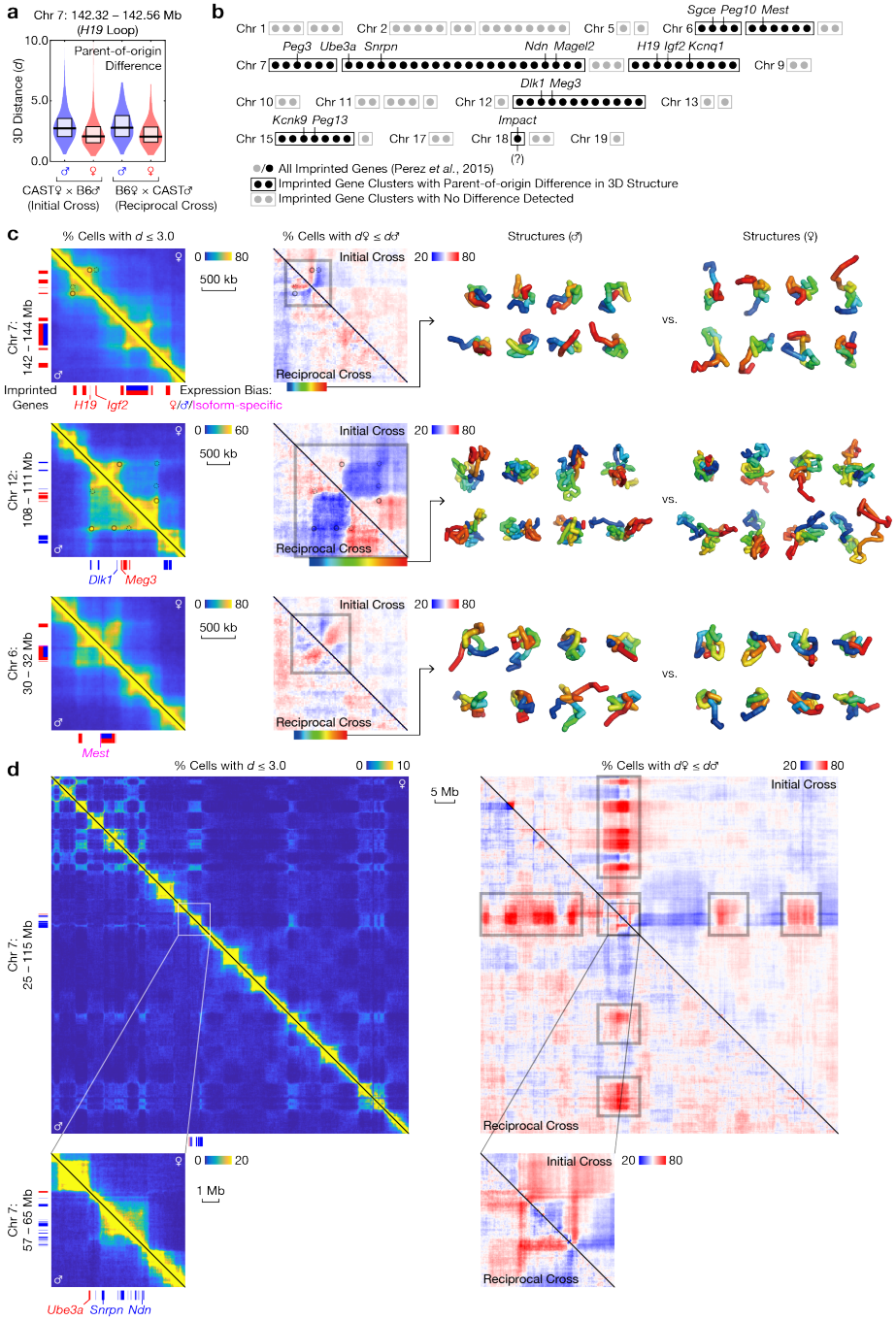


Fig. 4 | Parent-of-origin-specific 3D genome structure in multiple imprinted gene clusters. **a**, True parent-of-origin effects were separated from confounding factors (strain-specific effects) by reciprocal crosses: only allele-specific differences that are consistent between the two crosses are retained. Example histograms of the 3D distance (*d*) between two ends of a known maternal-specific chromatin loop at the *H19*/*Igf2* locus (*2, 18*) are shown as violin plots. Horizontal lines and boxes denote the median and quartiles, respectively. $n = 666$ and 168 for the initial and reciprocal crosses, respectively. **b**, Summary of detected (black) parent-of-origin differences among all imprinted gene cluster. Dots and boxes denote genes and gene clusters, respectively. Question mark denotes weak difference. Known imprinted genes in the brain are from (*19*). **c**,

Example parent-of-origin-specific 3D structures (more in Extended Data Fig. 7). Left: similar to Fig. 3d left, but comparing the paternal (lower triangle) and maternal (upper triangle) alleles. *n* = 834. Imprinted genes are shown with colors indicating parent-of-origin expression biases (*19*). Middle: allele-specific difference is quantified by the fraction of cells where the paternal 3D distance is greater than the maternal one for each pair of particles, in the initial (upper triangle) and reciprocal (lower triangle) crosses. True parent-of-origin effects show the same color in the two crosses, while confounding (strain-specific) effects show opposite colors. Gray boxes highlight regions with clear parent-of-origin-specific structures. Rainbows denote regions to be visualized on the right. *n*: see a. Right: example 3D structures. Cells are selected randomly. **d**, Similar to c, but showing chromosome-wide structural differences at the Prader-Willi/Angelman syndrome (PWS/AS) locus on Chr 7. Note that some regions (for example, between *Ube3a* and *Snrpn*) are unmappable, which may lead to poor estimation of 3D distances.

References

1. B. Bonev *et al*., Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e524 (2017).
2. L. Tan, D. Xing, C. H. Chang, H. Li, X. S. Xie, Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924-928 (2018).
3. L. Tan, D. Xing, N. Daley, X. S. Xie, Three-dimensional genome structures of single sensory neurons in mouse visual and olfactory systems. *Nat Struct Mol Biol* **26**, 297-307 (2019).
4. D. S. Lee *et al*., Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat Methods* **16**, 999-1006 (2019).
5. Y. Jiang *et al*., The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nat Genet* **49**, 1239-1250 (2017).
6. S. Sofueva *et al*., Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J* **32**, 3119-3129 (2013).
7. J. Fernandez-Albert *et al*., Immediate and deferred epigenomic signatures of in vivo neuronal activation in mouse hippocampus. *Nat Neurosci* **22**, 1718-1730 (2019).
8. Y. Zhang *et al*., An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci* **34**, 11929-11947 (2014).
9. J. D. Cahoy *et al*., A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* **28**, 264-278 (2008).
10. N. Habib *et al*., Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **353**, 925-928 (2016).
11. A. B. Rosenberg *et al*., Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176-182 (2018).
12. C. Luo *et al*., Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600-604 (2017).
13. D. A. Cusanovich *et al*., A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324.e1318 (2018).
14. C. A. Lareau *et al*., Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat Biotechnol* **37**, 916-924 (2019).
15. R. Lister *et al*., Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
16. E. J. Clowney *et al*., Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell* **151**, 724-737 (2012).
17. K. Monahan, A. Horta, S. Lomvardas, LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* **565**, 448-453 (2019).
18. S. S. Rao *et al*., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
19. J. D. Perez *et al*., Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *Elife* **4**, e07860 (2015).
20. J. T. Robinson *et al*., Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Syst* **6**, 256-258 e251 (2018).
21. B. Lacar *et al*., Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat Commun* **7**, 11022 (2016).
22. S. R. Krishnaswami *et al*., Using single nuclei for RNA-seq to capture the transcriptome

of postmortem neurons. *Nat Protoc* **11**, 499-524 (2016).

Methods

Animals

Animal protocols were approved by the Institutional Animal Care and Use Committee (IACUC) at Harvard University. F1 Hybrids of CAST/EiJ (JAX 000928) and C57BL/6J (JAX 000664) were generated by in-house breeding. The P1 sample (a pool of 2 pups) was female (inferred from sequencing data); all other samples (a single animal each) were male.

Isolation of cell nuclei

Cell nuclei were isolated based on protocols from (*21, 22*) with minor modifications. In particular, Tris buffer would react with formaldehyde, and was therefore substituted with equal molarities of HEPES buffer. See details below:

Cortex and hippocampus were dissected in ice-cold PBS, and placed in 2 mL nuclei isolation medium with Triton (0.25 M sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM HEPES pH 8.0, 1 uM DTT, 0.1% Triton X-100) in a 2 mL Dounce homogenizer (Sigma D8938). Note that DTT concentration was not specified in (*22*); here we used 1 uM from (*21*).

Tissues were homogenized with 5 strokes of the loose (A) pestle, and 15 strokes of the tight (B) pestle. The homogenate was centrifuged for 8 min at 94 g, 4 C, and the supernatant removed carefully. The pellet was re-suspended in 1 mL nuclei isolation medium without Triton (0.25 M sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM HEPES pH 8.0, 1 uM DTT).

The tube was centrifuged for 8 min at 94 g, 4 C, and the supernatant removed carefully. The pellet was re-suspended in 1 mL nuclei storage buffer (0.1665 M sucrose, 5 mM MgCl2, 10 mM HEPES pH 8.0), and filtered with 40 um. Note that sucrose concentration was not specified in (*22*); here we used 0.1665 M from (*21*). Also note that the recipe in (*21*) was inconsistent between ingredient volumes and final concentrations; here we followed the final concentrations.

Fixation of cell nuclei

Nuclei were fixed according to our Dip-C protocol (*3*) with minor modifications. In particular, nuclei were fixed directly in the 1 mL nuclei storage buffer (with HEPES instead of Tris buffer; see above) rather than PBS; and centrifugation was performed at 1,000 g rather than 600 g. See detailed below:

Nuclei were fixed directly in the nuclei storage buffer (see above; 1 mL) with the addition of 66.7 uL 32% PFA (EMS 15714) to a final concentration of 2%. The tube was rotated for 10 min at room temperature. Then 200 uL 1% BSA in PBS was added. The tube was centrifuged for 5 min at 1,000 g, 4 C, and the supernatant removed. The pellet was re-suspended in 1 mL ice-cold 1% BSA in PBS. Nuclei were counted and aliquoted to up to 0.5 million per tube. Each tube was centrifuged for 5 min at 1,000 g, 4 C, and the supernatant removed. The pellet was stored at -80 C.

Diploid chromatin conformation Capture (Dip-C)

A pellet of fixed cell nuclei was thawed on ice. We then proceeded directly to the step where 50 uL 0.5% SDS was added in the Dip-C protocol (*3*) and followed through with no modifications.

Cells were sorted on a Beckman Coulter MoFlo Astrios, and amplified with homemade Nextera. A total of 26.5 96-well plates were processed, leading to a total of 2,544 samples (Extended Data Fig. 1, Supplementary Table 1).

If assembling one's own Tn5 transposome is not an option, a roughly equal amount of "TTE Mix V50" (~0.015 uL per cell) from a TD501 kit (Vazyme) can be used in place of homemade Nextera transposome during the transposition step (unpublished data). In this way, the Dip-C procedure (the Nextera version) can be performed entirely with commercially available reagents, without the need for Tn5 transposase production.

Published Data

Reference genome. The mouse reference genome (GRCm38) and gene annotations (ALL) were downloaded from the GENCODE (https://www.gencodegenes.org/mouse/) M19 release.

Bulk Hi-C data. For projection onto the t-SNE plot (Fig. 1c right), the first 10 million reads of each data set were downloaded from the SRA with "fastq-dump" ("fastq-dump --split-files --gzip -X 10000000 [SRR accession]"). The accession numbers were: SRR5617733 for adult (3 months) cortical neurons (*NeuN*+) (*5*), SRR8441362 for adult (3–6 months) hippocampal excitatory neurons (*Camk2a*+) after saline injection (*7*), SRR8441366 for adult (3–6 months) hippocampal excitatory neurons (*Camk2a*+) 1 h after kainic acid injection (*7*), SRR8441370 for adult (3–6 months) hippocampal excitatory neurons (*Camk2a*+) 48 h after kainic acid injection (*7*), SRR5339908 for embryonic (E14.5) cortical neurons (*Hes5*-, *Dcx*+) (*1*), SRR5339876 for embryonic (E14.5) cortical neural progenitors (*Hes5*+, *Dcx*-) (*1*), SRR5339832 for *in vitro* differentiated cortical neurons (DIV12+9; *Tau*+, G0/G1 phase) (*1*), SRR5339786 for *in vitro* differentiated cortical neural progenitors (DIV12+2; *Sox1*+, G0/G1 phase) (*1*), and SRR941305 for *in vitro* differentiated astrocytes (*6*).

For visualization of bulk contact maps, data of *in vitro* differentiated, embryonic-like cortical neurons ("CN": DIV12+9; *Tau*+, G0/G1 phase) (*1*) was directly visualized through the Juicebox.js web browser (http://www.aidenlab.org/juicebox/) (*20*). Note that their *in vivo* embryonic cortical neurons ("ncx CN": E14.5; *Hes5*-, *Dcx*+) would be a better match, but were not available through the web browser. Data of adult (3 months) cortical neurons (*NeuN*+) was downloaded from the SRA of (*5*) (accession numbers: SRR5617731 and SRR5617733 for the two replicates, respectively; both male, and originally mapped to the mm9 reference genome), and processed in the same manner as our single-cell data. The resulting "contacts.hic" file was then visualized with Juicebox.js (*20*).

Lists of cell-type-enriched genes and regions. The following procedure is for Fig. 2a. See Supplementary Table 2 for all gene names.

The top 100 genes in microglia, in myelinating oligodendrocytes, in oligodendrocyte progenitors, and in astrocytes were obtained from the website of (*8*) (https://web.stanford.edu/group/barres_lab/brain_rnaseq.html; unfortunately no longer available by the time of this submission) by querying the top 100 genes (ranked by fold change) with FPKM $\geq$ 10 that are enriched in each cell type of interest (versus all other cell types), respectively.

The 10,984 accessible regions in microglia, 18,664 in oligodendrocytes, 17,418 in astrocytes, and 18,470 in SST+ inhibitory neurons were obtained by downloading the differential

accessibility file ("atac_matrix.binary.da_results.sig_open.txt") from the website of (*13*) (http://atlas.gs.washington.edu/mouse-atac/) and separating the peaks based on cell types ("16.3" for microglia, "21.1" for oligodendrocytes, "19.1" for astrocytes, and "15.3" for SST+ inhibitory neurons), respectively.

The top 60 genes up- and down-regulated during astrocyte development (comparing ages P1–8 to P17–30) were obtained from Figs. 6A and 6C, respectively, of (*9*).

The 125 genes in hippocampal granule cells, and 108 in (hippocampal) inhibitory neurons were obtained by first downloading Table S1 of (*10*). The TPM values were then back-calculated from the "log TPM values" (which we inferred to mean $\log_2$ (TPM + 1) based on descriptions in the paper). Finally, enriched genes were identified as TPM $\geq$ 10 and fold change $\geq$ 2 between the cell type of interest ("Granule cells DG" and "GABAergic interneurons", respectively) and each of the 4 other neuronal types.

The 4,896 accessible regions in hippocampal granule cells were obtained by downloading Table S5 of (*14*) and identifying peaks with their highest values in the cell type of interest ("EN16").

The 35,230 un-methylated regions in SST+ inhibitory neurons were obtained by downloading Table S5 of (*12*) and navigating to the tab for the cell type of interest ("mSst_1").

The top 100 genes up-regulated during cortical pyramidal cell development (comparing ages P2 to P11; note that this does not exactly match the time of our 3D genome transformation, but is the closest we can find) were obtained by first downloading the gene expression matrix ("GSM3017261_150000_CNS_nuclei.mat.gz", in MATLAB format) from the GEO accession (GSE110823) of (*11*). Cortical pyramidal cells were then extracted by their cell-type clusters ("cluster_assignment" containing "CTX Pyr" or "CLAU Pyr"; see Fig. 3C of (*11*)), and separated by age ("sample_type" being "p2_brain" or "p11_brain"). Expression values ("DGE") were normalized into TPM values (so that each cell sums to 1,000,000), and averaged among P2 and P11 cortical pyramidal cells, respectively. Finally, up-regulated genes were identified as the top 100 genes (ranked by fold change between P11 and P2) with TPM $\geq$ 100 in P11 cells. Note that similar to our previous observation in olfactory sensory neurons (*3*), the top developmentally down-regulated genes on average did not show changes in single-cell A/B compartment values.

Outdated gene names were converted to modern ones by looking up the Mouse Genome Informatics (MGI) website (http://www.informatics.jax.org/).

mCH desserts. The following procedure is for Fig. 3c and Extended Data Figs. Genomic coordinates of mCH desserts were obtained by downloading Table S2 of (*15*), navigating to the mouse section, and lifting over from mm9 coordinates to mm10 ones with the online LiftOver tool of the UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgLiftOver; parameters were: "Minimum ratio of bases that must remap" = 0.01 and "Allow multiple output regions", because of the repetitive nature of these regions).

Imprinted genes. The following procedure is for Fig. 4 and Extended Data Figs. Imprinted genes in the mouse brain (cerebellum), their genomic coordinates, and their expression biases were obtained from tab "(G)" of Supplementary File 1 of (*19*).

Data Analysis

Generation of contact maps. Single-cell contact maps were generated as we previously described (*3*), with "hickit" and "dip-c" packages.

Filtering out empty wells. Out of all 2,544 samples (26.5 96-well plates), we removed 590 (23%) that yielded < 100 k contacts per sample (Supplementary Table 1, Extended Data Fig. 1). Most of these "empty samples" were wells missed by the flow cytometer.

Single-cell A/B compartment values. Single-cell A/B compartment values were calculated as we previously described, from contact maps with "dip-c color2" (with parameters "-b1000000 -H -c color/mm10.cpg.1m.txt" for 1-Mb bins, or "-b100000 -H -c color/mm10.cpg.100k.txt" for 100-kb bins in Extended Data Fig. 2) (*3*).

Only autosomal bins that were present in all 1,954 cells were retained. A/B compartment values were rank-normalized to 0–1 in each cell with MATLAB "tiedrank" ("(tiedrank(compartment_values_in_cell)-1)./(num_bins_in_cell-1)").

After PCA with MATLAB "pca" ("pca(compartment_value_matrix)"), only the first 20 PCs ("pca_score(:, 1:20)") were retained for further visualization and clustering.

t-SNE was performed with MATLAB "tsne", initialized by the first 2 PCs ("tsne(pca_score(:, 1:20), 'InitialY', pca_score(:, 1:2))"). UMAP (Extended Data Fig. 2) was performed with MATLAB "run_umap" ("run_umap(pca_score(:, 1:20))").

Hierarchical clustering was performed with MATLAB "linkage", on Euclidean distances of the first 20 PCs with Ward's method ("linkage(pca_score(:, 1:20), 'ward', 'euclidean')"). The 13 major and 3 minor (unknown) cell-type clusters were then manually defined from the linkage tree with the t-SNE plot as a visual aid (Fig. 1d). We tentatively classified the 3 rare clusters (A3–5) as "astrocytes" based on the linkage tree; but this is uncertain because of their limited sample sizes.

Projection of bulk Hi-C data onto the t-SNE space. The first 10 million reads of each bulk Hi-C data were processed in the same manner as our single-cell data, and projected onto the PCA space as we previously described (*3*), with MATLAB. Subsequently, because additional data points cannot be directly projected onto an existing t-SNE space, the final plot (Fig. 1c right) was generated by performing t-SNE jointly on both bulk and single-cell data with MATLAB "tsne".

Integrative analysis of single-cell A/B compartment values, transcriptomes, methylomes, and chromatin accessibility. For each gene or region of interest, the 1-Mb bin closest to its mid-point (according to GENCODE for genes) was first identified. The same filtering as above (only autosomal bins that were present in all cells were retained) was then applied.

For each list of genes (Supplementary Table 2) or regions, normalized A/B compartment values of the above bins were averaged in each cell, producing a single-cell, list-averaged value. Note that for some lists, each bin might be counted more than once (especially for accessible or un-methylated regions, because there were only ~3,000 1-Mb bins across the genome); in this case, the average A/B compartment value was a weighted average of bins.

The distribution of these average A/B compartment values were visualized as a violin plot for each cell-type cluster (Fig. 2a), with MATLAB "ksdensity" ("ksdensity(data_to_plot, 'Bandwidth', bandwidth)"). Each violin plot was trimmed to the data range, and normalized by the highest value.

Basic characteristics of single-cell contact maps. For each cell, the extent of chromosome intermingling was quantified as the fraction of inter-chromosomal contacts (*3*) (Fig. 2b top, Supplementary Table 1). Here we calculated this fraction from raw contacts before removal of duplicates, which has the advantage of not depending on sequencing depths, and differs only slightly from that in duplicate-removed contacts.

The extent of A/B compartmentalization was quantified as the Spearman correlation between the single-cell A/B compartment value of each 1-Mb bin and its own CpG frequency (*2*) (Fig. 2b bottom), with MATLAB "corr" ("corr(compartment_values_in_cell, self_cpg_frequencies, 'type', 'spearman')"). Here we calculated A/B compartment value from contact maps rather than 3D structures.

Generation of 3D structures. Single-cell 3D genome structures were generated at 20-kb resolution as we previously described (*3*), with "hickit" and "dip-c" packages.

We generated 5 replicate structures per cell with different random seeds (1–5). After the same filtering as we previously described (the bottom 6% of 20-kb particles that harbored the least numbers of haplotype-resolved contact legs within ±0.5 Mb were removed; these regions were typically unmappable and thus difficult to model) (*2, 3*), a root-mean-square (across pairs of replicates) root-mean-square (across 20-kb particles) deviation (r.m.s.d.) value was calculated for each cell. Only cells with r.m.s.d. ≤ 1.5 particle radii (high-quality 3D structures) were retained for further analysis.

Radial positioning. The radial preference of each 1-Mb bin was calculated as we previously described (*3*), with "dip-c color -C". Note that this value (3D distance to the nuclear center of mass) is not on a uniform scale, because a spherical shell of the same thickness contains more chromatin (larger volume) at a larger radius.

Changes in 3D genome structure during postnatal neuronal development. We visualized pairwise 3D chromatin interactions between 20-kb particles by first calculating a matrix of pairwise 3D distances within a genomic region (or between two regions) in each cell with "dip-c pd" ("dip-c pd -1 region.leg" within one region; or "dip-c pd -1 region_1.leg -2 region_2.leg" between two regions, such as inter-chromosomal analysis in Extended Data Fig. 4a). The two alleles were calculated separately.

A matrix of fractions of cells where each pair of 20-kb particles are nearby in 3D (distance ≤ 3.0 radii) (for example, Fig. 3d left) (*3*) was then generated by binarizing each single-cell matrix of pairwise 3D distances and averaging them, with "scripts/threshold_np_float.py" of the "dip-c" package ("scripts/threshold_np_float.py 3").

Separately, a matrix of median 3D distances between each pair of 20-kb particles was generated by calculating element-wise medians, with "scripts/median_np_float.py" of the "dip-c" package.

Developmental changes of the two above matrices were calculated by subtracting the neonatal matrix from the adult one, with "scripts/subtract_np_float.py" of the "dip-c" package.

Chromosome intermingling in 3D structures. The extent of chromosome intermingling of each 1-Mb bin (Extended Data Fig. 5) was calculated as we previously described (*3*), with "dip-c color -i3" (note than this calculates the fraction of intra-chromosomal, or intra-homologous to be exact, 3D neighbors; the fraction of inter-chromosomal neighbors is 1 minus this value).

Parent-of-origin-specific 3D genome structure. Similar to our developmental analysis (see above), matrices of fractions of cells where each pair of 20-kb particles are nearby in 3D (distance ≤ 3.0 radii) and of median 3D distances between each pair of 20-kb particles were generated with "dip-c pd", "scripts/threshold_np_float.py", and "scripts/median_np_float.py", but for the maternal and paternal alleles (for example, Fig. 3c left), and for parent-of-origin analysis, further separated between the initial and reciprocal crosses.

Within the initial and reciprocal crosses, respectively, allelic differences (for example, Fig. 3c middle) were calculated with "scripts/subtract_np_float.py" of the "dip-c" package.

For representative 3D structures (Fig. 3c right), the region of interest was extracted from the full 3D genome file ("20k.1.clean.3dg") with "dip-c reg3" ("dip-c reg3 -i region.reg"), and converted to an mmCIF file with "dip-c color" and "dip-c vis" ("dip-c color -l color/mm10.chr.len" and "dip-c vis -c") for visualization in PyMol.

Haplotype-resolved raw contact maps (Extended Data Fig. 9) were generated by extracting contacts with haplotype-informative SNPs on both legs (both "0" phases for paternal, both "1" for maternal) from contact maps ("contacts.pairs.gz") before any haplotype imputation (before "hickit -u").

Data Availability

Raw sequencing data were deposited at the National Center for Biotechnology Information with accession number PRJNA607329 at https://www.ncbi.nlm.nih.gov/sra/PRJNA607329. Processed data, including files for interactive viewing in the Juicebox web browser, were deposited at with GEO Series accession number GSE146397 at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE146397.

Code Availability

Code is available at GitHub (https://github.com/tanlongzhi/dip-c and https://github.com/lh3/hickit).

Acknowledgements

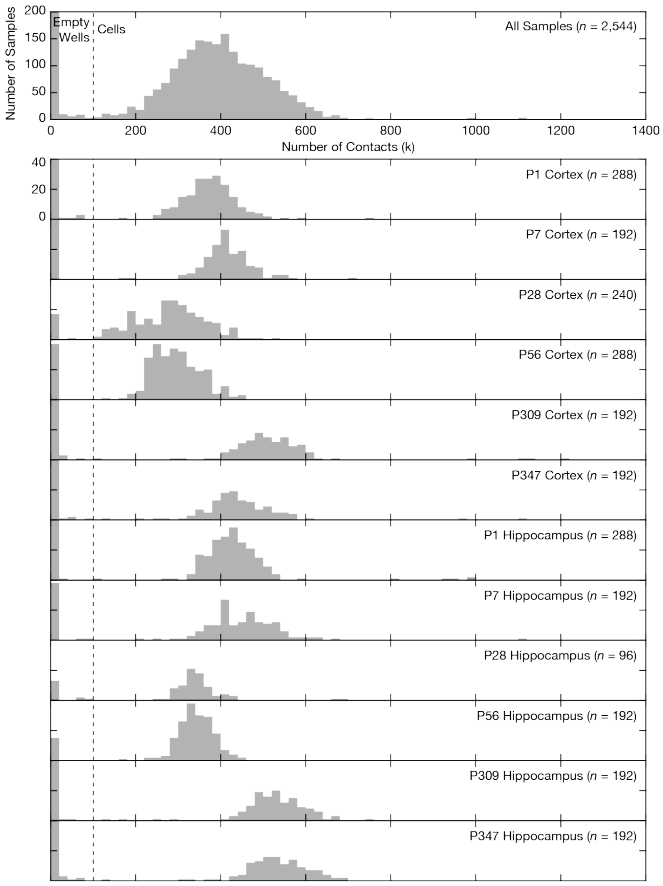
The authors thank the Bauer Core Facility at Harvard University (Z. Niziolek, J. Nelson, and C. Reardon) for flow sorting and PCR machines, F. Alt, P. Wei, F. Gage, and S. Parylak for advice on nuclei isolation, and H. Li and X. Jin for helpful discussions. This work was supported by Beijing Advanced Innovation Center for Genomics at Peking University, and a generous gift grant from Xianhong Wu to Harvard University.

Author Contributions

L.T., D.X., N.D., and X.S.X. designed the experiments. L.T., D.X., and N.D. performed the experiments. L.T. analyzed the data. L.T. and X.S.X. wrote the manuscript.

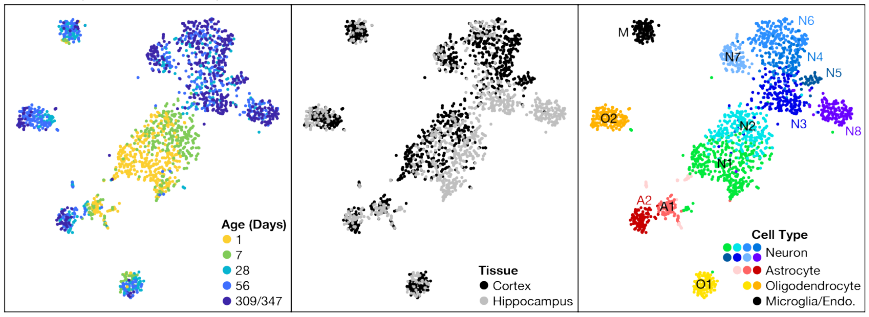
Competing Interests

L.T., D.X., and X.S.X. are inventors on a patent WO2018217912A1 filed by President and Fellows of Harvard College that covers Dip-C.

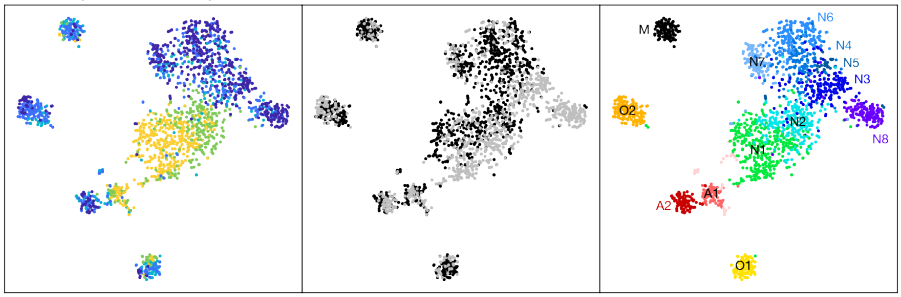


Extended Data Fig. 1 | Histograms of numbers of contacts per cell. Bin size is 20 k. Dash lines denote a threshold of 100 k, below which a sample is considered an empty well (most likely missed by the flow cytometer). The leftmost bins (0–20 k) are truncated in height for visual clarity.

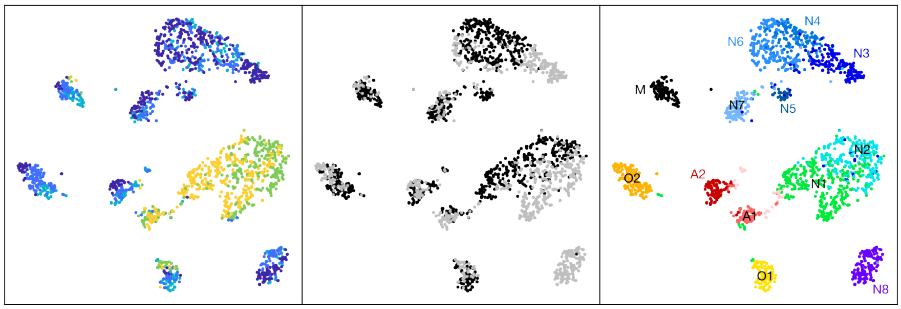
Down-sampled to 100 k Contacts per Cell



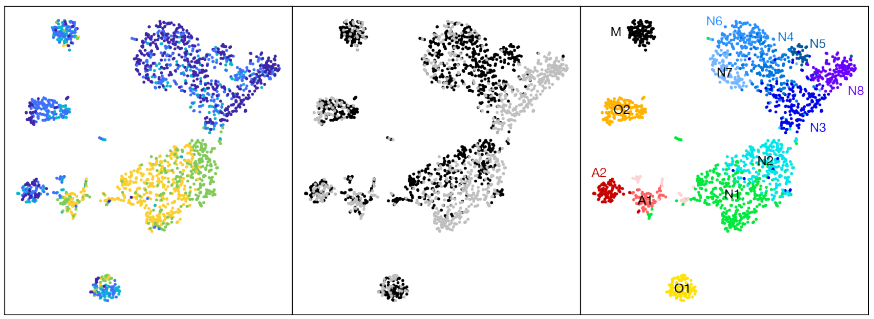
Down-sampled to 50 k Contacts per Cell



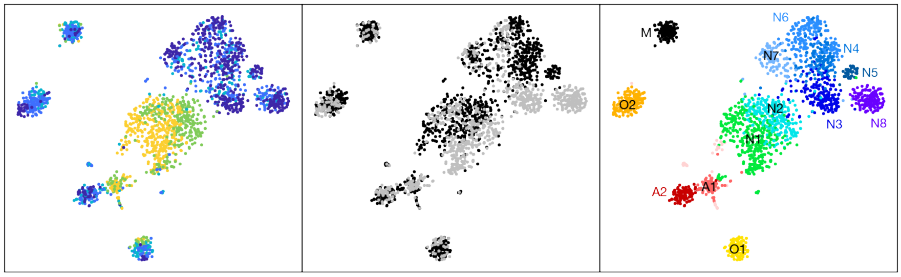
100-kb Bins Instead of 1-Mb



First 10 PCs Instead of 20

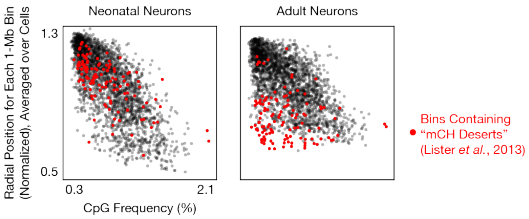


First 50 PCs Instead of 20

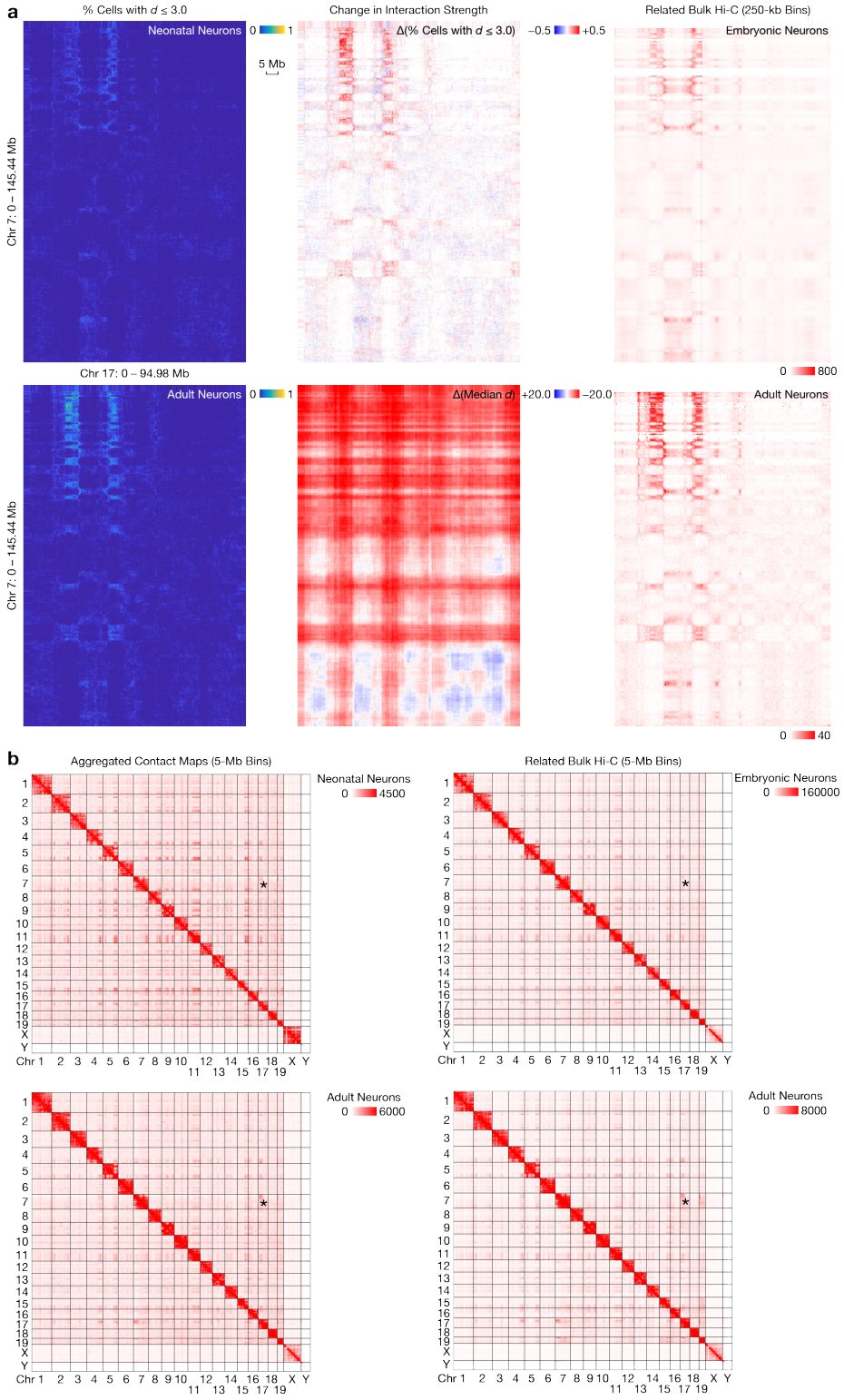


UMAP Instead of t-SNE

Extended Data Fig. 2 | Separation of cell-type clusters is robust against different visualization parameters or methods. Similar to Fig. 1c and Fig. 1d, but with alternative parameters or methods. Coloring of cell-type clusters (right) is still based on the method described in the main text (Fig. 1d), rather than alternative methods.

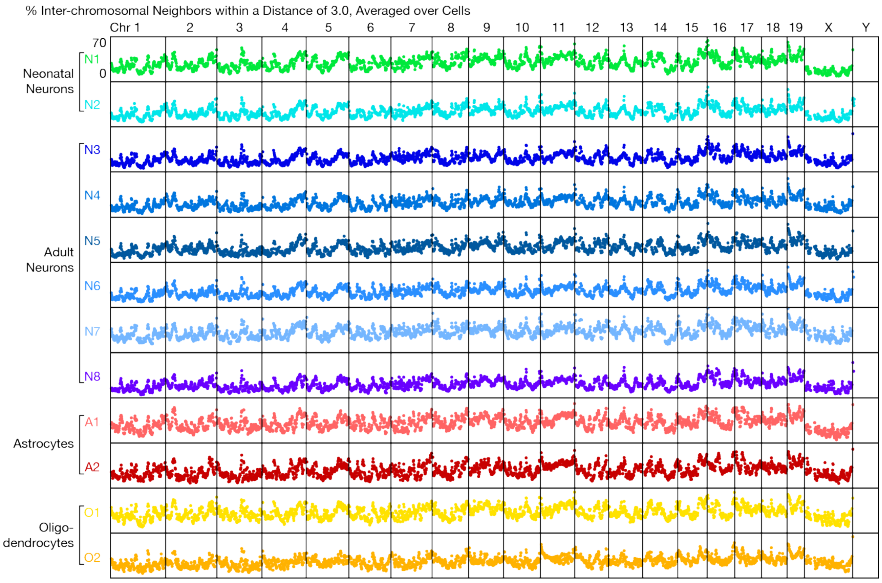


Extended Data Fig. 3 | CpG-poor regions were radially relocated during postnatal neuronal development. Similar to Fig. 3c, but plotting the neonatal and adult radial profiles separately, each against CpG frequencies along the genome. Neonatal neurons (left) followed a usual, negative correlation between CpG frequency and radial positioning. Adult neurons (right), however, positioned many CpG-poor regions into the nuclear interior, similar to olfactory sensory neurons (Fig. 3c of (*3*)).

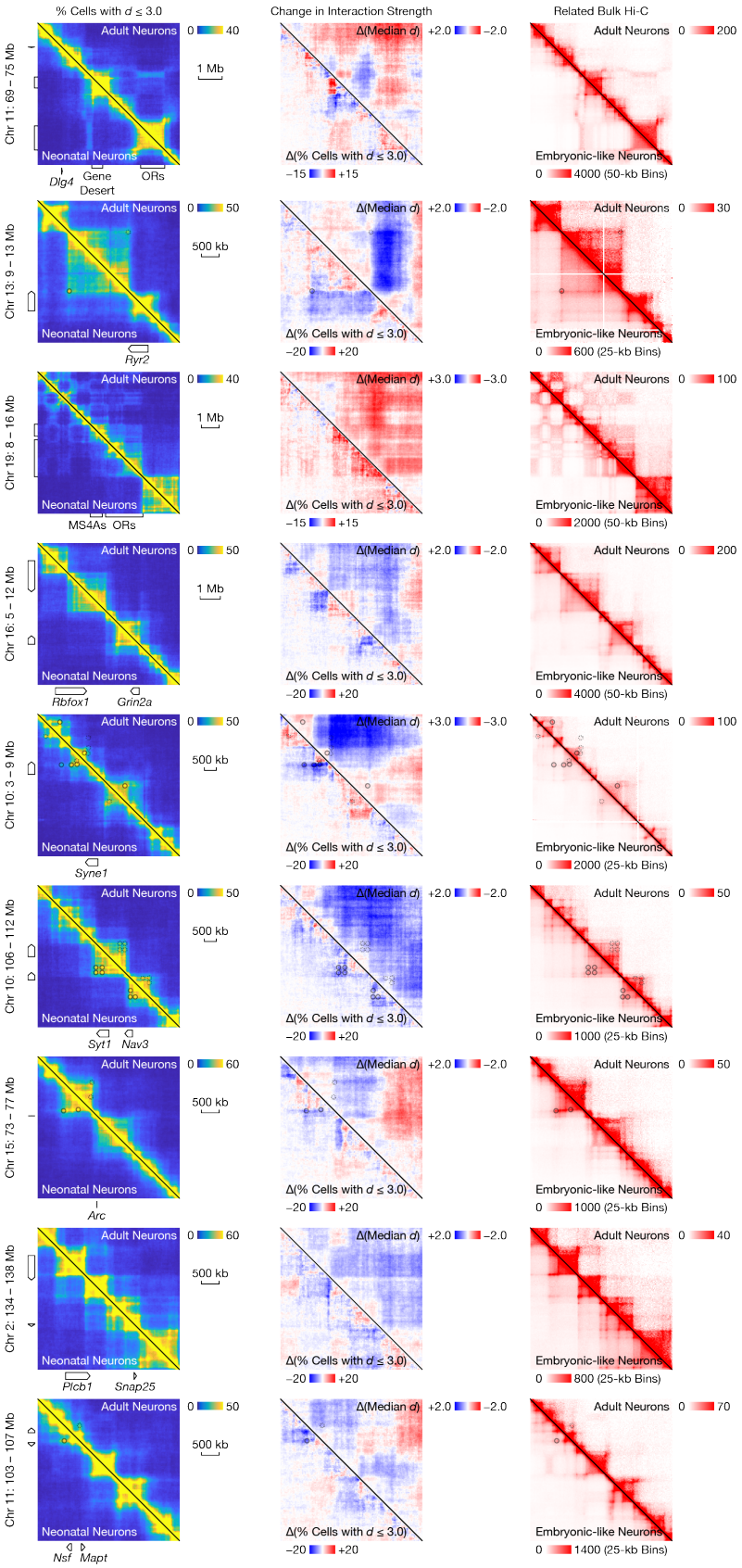


Extended Data Fig. 4 | Increased inter-chromosomal interaction between radially relocated regions during postnatal neuronal development. a, Similar to Fig. 3d, but for inter-chromosomal interaction between Chr 7 and Chr 17. Note that unlike intra-chromosomal interaction, here matrices are not symmetric, and thus cannot be represented by upper or lower

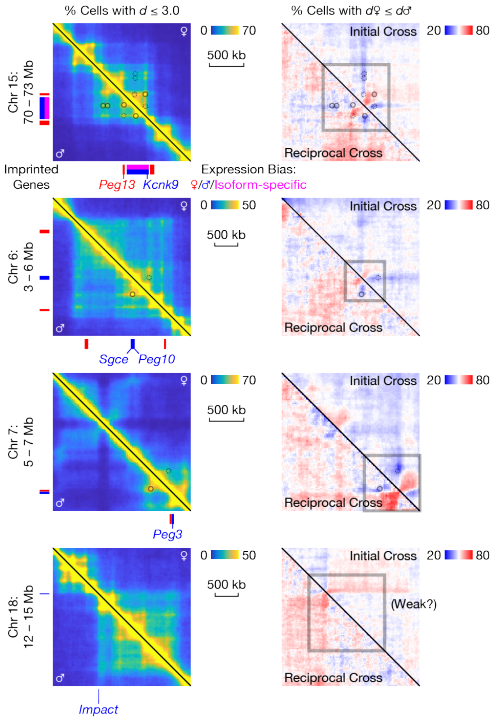
triangles. n = 347 cells × 4 allele pairs (♂♂, ♂♀, ♀♂, and ♀♀) and 363 × 4 for neonatal and adult neurons, respectively. To save computational costs, matrices of median 3D distances (bottom middle) is calculated from a random sub-sample of n = 100 each. **b**, Genome-wide view of inter-chromosomal contacts. In addition to the region in a (asterisks), increased inter-chromosomal interaction was also observed between other radially relocated regions (for example, the telomeric end of Chr 4), whereas the overall extent of inter-chromosomal interaction (namely, chromosome intermingling) decreased during postnatal neuronal development (Fig. 2b). Single-cell contact maps are aggregated from n = 585 and 795 cells for neonatal and adult neurons, respectively; their color scales are set proportionally (left). Note than values along sex chromosomes may not be directly comparable between clusters, because P1 animals were female and the rest were male.



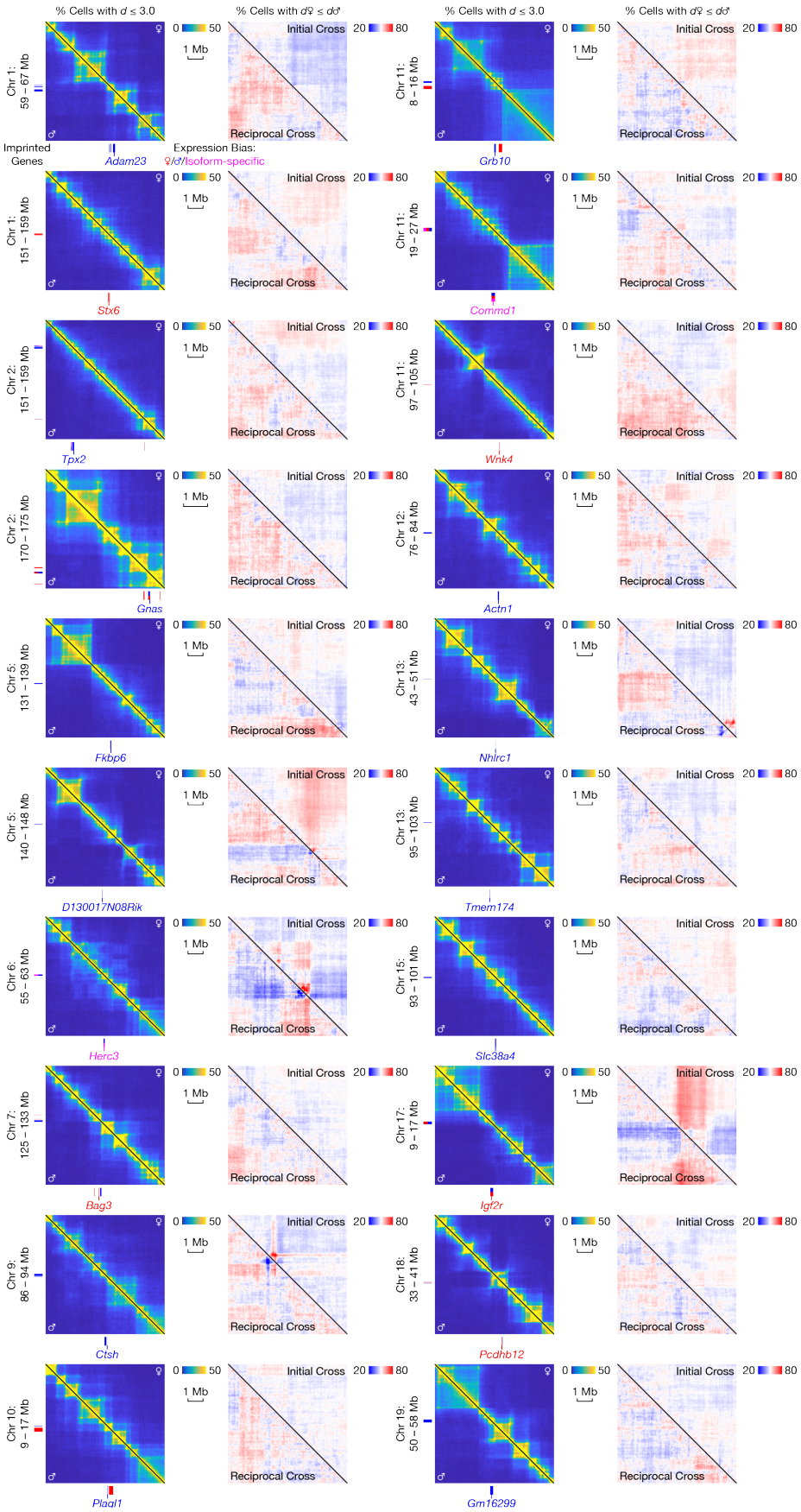
Extended Data Fig. 5 | Overall extent of chromosome intermingling did not preferentially increase for radially relocated regions during postnatal neuronal development. Similar to Fig. 3a, but for the extent of chromosome intermingling along the genome (defined as the fraction of inter-chromosomal neighbors near each particle, binned every 1 Mb; two homologs count as two different chromosomes). Unlike olfactory sensory neurons (*17*), radially relocated regions did not particularly intermingle more with other chromosomes. Relative extents of chromosome intermingling between different cell-type clusters are generally consistent with Fig. 2b, although among inhibitory neurons (N7), high-quality 3D structures were somewhat biased towards more intermingled cells.



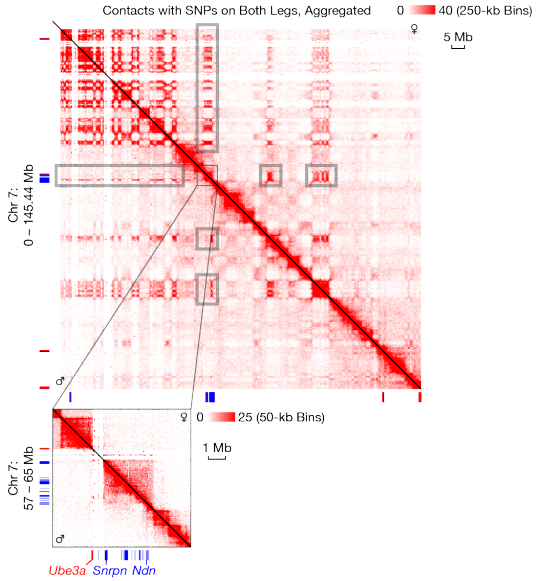
Extended Data Fig. 6 | Developmental changes in local 3D genome structure, continued. Similar to Fig. 3d. OR: olfactory receptor; MS4A: membrane-spanning 4-domains subfamily A.



Extended Data Fig. 7 | Parent-of-origin-specific 3D structure around imprinted genes, continued. Similar to Fig. 4c, but for 4 other imprinted genes or gene clusters. For the *Impact* locus (bottom), it is unclear with our current sample size whether the observed difference represents a true, yet weak, effect or noise.



Extended Data Fig. 8 | Parent-of-origin-specific 3D structure was not detected around the other imprinted genes. Similar to Fig. 4c and Extended Data Fig. 7, but for the other imprinting genes (showing two regions in a row), where parent-of-origin effect was not detected. Note that parent-of-origin effect may still exist, but was not detected because of our limited sample size. Each plot shows 8 Mb, except for the *Gnas* locus, where only 5 Mb is shown because of a nearby repetitive (unmappable) region.



Extended Data Fig. 9 | Chromosome-wide parent-of-origin-specific 3D structure was also apparent from raw contact maps. Similar to Fig. 4d, but based on raw contact maps (aggregated from all n = 1,954 cells) rather than 3D models (and for the entire Chr 7 in the top panel). Only contacts with haplotype-informative SNPs on both legs are used; therefore, no haplotype imputation or 3D modeling is involved.

Supplementary Table 1 | Information about each single cell.

Supplementary Table 2 | Lists of genes used in single-cell A/B compartment analysis.