Cell-type specialization in the brain is encoded by specific long-range chromatin topologies

Warren Winick-Ng[1,#], Alexander Kukalev[1,#], Izabela Harabula[1,2,#], Luna Zea Redondo[1,2], Mandy Meijer[3], Leonid Serebreni[1,a], Simona Bianco[4], Dominik Szabo[1,2], Andrea M. Chiariello[4], Ibai Irastorza Azcarate[1], Luca Fiorillo[4], Francesco Musella[4], Christoph Thieme[1], Ehsan Irani[1,5], Elena Torlai Triglia[1,b], Aleksandra A. Kolodziejczyk[6,7,c], Andreas Abentung[8,d], Galina Apostolova[8], Eleanor J. Paul[9,e], Vedran Franke[10], Rieke Kempfer[1,2], Altuna Akalin[10], Sarah Teichmann[5,6], Georg Dechant[8], Mark A. Ungless[9], Mario Nicodemi[4,5], Gonçalo Castelo-Branco[3,11], Ana Pombo[1,2,5,*]

[1]Epigenetic Regulation and Chromatin Architecture Group, Berlin Institute for Medical Systems Biology, Max-Delbrück Centre for Molecular Medicine, 10115 Berlin, Germany
[2]Humboldt-Universität zu Berlin, 10117 Berlin, Germany
[3]Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 17177 Stockholm, Sweden
[4]Dipartimentio di Fisica, Università di Napoli Federico II, and INFN Napoli, Complesso Universitario di Monte Sant'Angelo, 80126 Naples, Italy
[5]Berlin Institute of Health, 10178 Berlin, Germany
[6]Cavendish Laboratory, University of Cambridge, JJ Thomson Ave, Cambridge CB3 0EH, UK
[7]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK
[8]Institute for Neuroscience, Medical University of Innsbruck, 6020 Innsbruck, Austria
[9]Institute of Clinical Sciences, Imperial College London, London SW7 2AZ, UK
[10]Bioinformatics Platform Group, Berlin Institute for Medical Systems Biology, Max-Delbrück Centre for Molecular Medicine, 10115 Berlin, Germany
[11]Ming Wai Lau Centre for Reparative Medicine, Stockholm node, Karolinska Institutet, 17177 Stockholm, Sweden

[a]current address: Research Institute of Molecular Pathology (IMP), 1030 Vienna Biocenter (VBC), Vienna, Austria
[b]current address: Broad Institute of MIT and Harvard, 02142 Cambridge, MA, USA
[c]current address: Immunology Department, Weizmann Institute of Science, 7610001 Rehovot, Israel
[d]current address: Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, 7491 Trondheim, Norway
[e]current address: Developmental Neurobiology, King's College London, London WC2R 2LS, UK

[#]These authors contributed equally to this work

*Corresponding author: ana.pombo@mdc-berlin.de

Abstract

Neurons and oligodendrocytes are terminally differentiated cells that perform highly specialized functions, which depend on cascades of gene activation and repression to retain homeostatic control over a lifespan. Gene expression is regulated by three-dimensional (3D) genome organisation, from local levels of chromatin compaction to the organisation of topological domains and chromosome compartments. Whereas our understanding of 3D genome architecture has vastly increased in the past decade, it remains difficult to study specialized cells in their native environment without disturbing their activity. To develop the application of Genome Architecture Mapping (GAM) in small numbers of specialized cells in complex tissues, we combined GAM with immunoselection. We applied immunoGAM to map the genome architecture of specific cell populations in the juvenile/adult mouse brain: dopaminergic neurons (DNs) from the midbrain, pyramidal glutamatergic neurons (PGNs) from the hippocampus, and oligodendrocyte lineage cells (OLGs) from the cortex. We integrate 3D genome organisation with single-cell transcriptomics data, and find specific chromatin structures that relate with cell-type specific patterns of gene expression. We discover abundant changes in compartment organisation, especially a strengthening of heterochromatin compartments which establish strong contacts spanning tens of megabases, especially in brain cells. These compartments contain olfactory and taste receptor genes, which are de-repressed in a subpopulation of PGNs with molecular signatures of long-term potentiation (LTP). We also show extensive reorganisation of topological domains where activation of neuronal or oligodendrocyte genes coincides with formation of new TAD borders. Finally, we discover loss of TAD organisation, or 'TAD melting', at long (>1Mb) neuronal genes when they are most highly expressed. Our work shows that the 3D

organisation of the genome is highly cell-type specific in terminally differentiated cells of the brain, and essential to better understand brain-specific mechanisms of gene regulation.

Introduction

The genome-wide study of three-dimensional (3D) organization has revealed its intrinsically association with gene regulation and cell function[1-3]. Genome-wide sequencing approaches such as Hi-C[4], GAM[5] and SPRITE[6] have shown that the genome is organized hierarchically, from specific promoter-enhancer contacts, to topological associating domains (TADs), to broader compartments of open and closed chromatin (compartments A and B, respectively), up to whole chromosome territories[4-7]**.** Dynamic changes of chromatin organization during neuronal development have been previously reported using *in-vitro* differentiated neurons or glutamatergic neurons which are abundant in the cortex, or from whole hippocampi, after tissue dissociation and FACS isolation from pools of mice[8-10]. However, it has been particularly difficult to assess chromatin structure of specific cell types from the brain, especially without disrupting tissue organization, and in single animals. Understanding 3D chromatin structure of rare cell populations is of exceptional importance in the brain, where connecting disease-associated genetic variants in non-coding genomic regions with their target genes remains particularly challenging[11-13], and where cell state and cell localization within the tissue can influence both transcriptional and physiological outcomes[14-17].

Here, we have developed immunoGAM, an extension of the GAM technology[5] that allows the genome-wide mapping of chromatin topology in specific cell

populations, leveraging its advantages to work directly in intact tissue without prior dissociation, and with small cell numbers (~400-1500 cells). To demonstrate the versatility of immunoGAM, we applied it in three different tissues of the brain to study specific cell types in the brain with diverse functions. We selected dopaminergic neurons (DNs) from the ventral tegmental area of the midbrain (VTA), pyramidal glutamatergic neurons (PGNs) from the *cornu ammonis 1* (CA1) region of the hippocampus, and oligodendrocytes (OLGs) from the somatosensory cortex, which we compared with mouse embryonic stem cells (mESCs). We find that chromatin organization reflects cell-type specialization at all genomic scales. GAM data from the three brain cell types shows that clusters of sensory receptor genes (e.g. olfactory receptor genes, *Olfrs*) form strong contacts that can span very long distances (~5 – 40 Mb), in contrast with dividing mESCs where *Olfrs* are occasionally expressed. Interestingly, we find that *Olfr* expression escape is more abundant in the CA1 PGNs where it coincides with neuronal activation as assessed by the expression of *Fos*, *Egr1* and *Nrxn2* and other activation related genes. We also detect a large reorganisation of TADs, with 67% of all TAD borders being found in at least one of the three brain cell types and not in mESCs, while only 9% of TAD borders are common to all cell types. Cell-type specific TAD borders contain genes expressed in the relevant cell types, and are related with cell identity and specialization, showing that 3D genome reorganisation is highly connected with cell-type specific gene expression. Finally, we investigated the 3D topology of ~1-Mb long neuronal genes, which span either large or more than one TAD, and detected significant loss of TAD structure, which we call 'TAD melting', in the cell type where long neuronal genes are most active.

Results

In-situ mapping of chromatin contacts in specific cell types of the brain

To study chromatin topology in specific cell populations in their native environment in the brain, we mapped genome-wide 3D chromatin contacts by adapting Genome Architecture Mapping (GAM)[5] to three brain areas. GAM is a ligation-free technology that maps 3D genome topology by extracting and sequencing the genomic content of ultrathin cryosections from nuclei, followed by spatial statistics to detected interactions between genomic loci from their increased probability of co-segregation in a large population of nuclear slices. GAM was previously applied to mESCs and shown to capture TADs, A/B compartments and pair-wise contacts across long genomic distances. Here, we developed immunoGAM to allow the selection of nuclear slices from specific cell types within a complex tissue. Nuclear sections from specific brain cell types were collected from mice perfused with fixative, by laser microdissection after immunofluorescence detection, followed by genomic DNA extraction and sequencing (**Fig. 1**).

After cryopreservation, tissues were sectioned into ~230 nm thin cryosections, which were incubated with specific antibodies to mark cells of interest, before laser microdissection of individual nuclear sections (nuclear profiles; NPs), collected in groups of three into individual PCR lids. A detailed flowchart of the data collection is presented in **Supplemental Fig. S1a**. After extraction, amplification and sequencing of the DNA content of each GAM sample, we produced GAM matrices that represent

the co-segregation of pairs of loci in the collection of GAM samples, and which inform about their 3D physical distance inside the nucleus[5]. GAM matrices were also used to map TAD borders, compartments A and B, and to investigate differential contacts between the cell types analysed (**Fig. 1b**). To explore how dynamic changes in chromatin topology relate with cell-type specific patterns of gene expression, we also collected published single-cell gene expression (scRNA-seq) from the relevant brain cell types, and produced scRNA-seq data from mESCs (**Fig. 1b**).

To select the cell types of interest from the complex brain samples, we used immunofluorescence to detect specific cell markers (**Fig. 1c**). We selected OLGs from 3-weeks old mouse somatosensory cortex based on immunofluorescence detection of GFP expression from a Sox10-cre-LoxP-GFP mice[18], and produced a GAM dataset consisting of 290 libraries each from three independent NPs, corresponding to a total of 870 cells. PGNs were selected from the CA1 region of the dorsal hippocampus based on their position within the CA1 pyramidal cell layer and immunofluorescence detection of pan-histone, as a nuclear marker. We collected PGN GAM samples from two separate mice in a total of 209 x 3NPs (627 cells) and 275 x 3NPs (825 cells). DNs were selected from the midbrain VTA based on immunofluorescence detection of tyrosine hydroxylase (TH), in most samples in combination with pan-histone, from two separate mice from two different genotypes (wild-type and with a TH-GFP reporter, used initially to guide the tissue dissection). We collected 585 x3NP (1755) and 291 x3NP (873) DNs from wildtype and TH-GFP midbrains, respectively. As a reference for comparison, we used a dataset produced from mESCs consisting of 249 x3NP (747) mESCs. To detect lower quality samples where DNA extraction or amplification was less efficient or where laser microdissection failed, we performed quality control analyses which assess different

parameters of each library, including the percentage of positive genomic windows surrounded by non-detected windows or the number of uniquely mapped reads; 86-96% of GAM libraries passed the quality control analyses (see Methods for more details; **Supplemental Table 1**). To assess the quality of sampling in each GAM dataset, we measured the frequency with which all possible intra-chromosomal pairs of genomic windows are found in the same GAM sample, and found that 98.8 – 99.9% of all mappable pairs of windows were sampled at least once at resolution 50 kb at all genomic distances (and 92.7-99.7% of all pairs of windows were sampled at least 5 times within the shorter distances of 5 Mb, except for OLGs with 83.3%).

To begin exploring the differences in chromatin contacts between mESCs, OLGs, PGNs and DNs, we plotted normalized chromatin contact matrices (using normalized pointwise mutual information, NPMI; see **Methods**) for each cell type and biological replicate, which represent the frequency of locus co-segregation normalized to the frequency of locus detection[5] (**Fig. 1d**). Visual inspection of contact matrices across a 60-Mb genomic region on chromosome 17 shows cell-type specific local domain structures. Further, we also detect strong patches of long-range contacts for all three brain cell types that are absent in mESCs, which are separated by ~35-Mb (**Fig. 1d**) and are reproduced in the biological replicates (**Supplemental Fig. S1d**). Closer inspection of local contacts, for example across a 5-Mb region on chromosome 13 centered on the *Ht1ra* gene, confirms clear cell-type specific 3D genome topologies (**Fig. 1e**). The *Ht1ra* gene encodes the 5-hydroxytryptamine (serotonin) receptor 1A, which is an important modulator of synaptic transmission in the hippocampus and has single-nucleotide polymorphisms (SNPs) associated with schizophrenia[19,20]. *Htr1a* is not expressed in mESCs nor OLGs, and is embedded in domains that change conformation in DNs and in PGNs. Our visual inspections of

contact matrices showed clear differences in pairwise contacts between all four cell types, with the PGNs and DNs having the most similarity between cell types, while contacts were conserved between biological replicates collected from different animals.

A/B compartments are altered and present different contact frequencies in brain cells, in comparison to mESCs

To further investigate the cell-type specific contacts detected across megabase distances in GAM datasets, we started by computing compartments A/B after performing PCA from GAM matrices and choosing the compartment eigenvectors most correlated with GC content, as previously described[5] (**Fig. 2a**). Visual comparisons of compartment eigenvectors show extensive variation between mESCs and the three brain cell types considered here. As expected from the developmental stage and origin of each cell type, compartment eigenvectors and compartment A/B assignment per genomic window were best correlated between biological replicates, next between the two neuronal cell types DNs and PGNs, followed by between neurons and oligodendrocyte lineage cells, and least between the different post-mitotic cells and mESCs (**Supplemental Fig. S2a-c**).

Comparisons between different chromosomes showed that the compartment B eigenvectors were much stronger in some chromosomes in the brain cell types than mESCs, especially in chromosomes 7 and 17 (**Fig. 2a**), though the total compartment A/B lengths were similar between all cell types (**Supplemental Fig. S2d**). Visual inspection of matrices and compartments revealed that the strong long-range patches of contacts observed in the brain cell types are often established

between B compartments and can be separated by linear genomic distances above 10 Mb and up to 50 Mb. Contacts between B compartments in chromosome 7 can be strongest in one brain cell type, but clearly visible in the other cell types (**Fig. 2b,c**).

To quantify the homotypic contact frequencies within A or B compartments, and heterotypic contacts between A and B compartments in the different cell types, we calculated the mean contacts score from normalized matrices per chromosome (**Fig. 2d**). For most chromosomes and in all cell types, the average normalized contact score is similar between A-A or B-B contacts, and A-B contacts are less intense, as expected. Interestingly, we found that average contacts between B compartment windows had higher contact scores especially on chromosomes 7 and 17 in the three brain cell types than in mESCs (**Fig. 2d**); these average values were confirmed across the whole range of compartment score intensities (**Fig. S2e**). B-B contacts in chromosome 7 in DNs have the highest average contact score of any chromosome or cell type (0.33 ± 0.16), followed by PGNs (0.25 ± 0.14) and OLGs (0.25 ± 0.14), while mESCs had the lowest B-B average NPMI contact score (0.16 ± 0.11).

To further explore the more extensive frequency of long-range contacts in different cell types and specifically in chromosome 7, we calculated the decay of average contact frequency across genomic distances for all chromosomes and chromosome 7 (**Supplemental Figs. S2f** and **S2g**, respectively). We observed that terminally differentiated brain cells have higher average contact frequency at long distances (> 10-Mb) than mESCs, while chromosome 7 displays more frequent long-range chromatin contacts especially in PGNs and in DNs than the average behaviour of all chromosomes in the corresponding cell types.

Strong B compartments in brain cells establish long-range interactions between clusters of sensory receptor genes

To explore possible biological roles of the strong B compartments of chromosome 7 observed in all three brain cell types, we searched for enriched Gene Ontologies (GOs) in the different cell types (**Supplemental Fig. 2h**). For all cell types, we found a significant enrichment for GO terms containing sensory receptor clusters. These included *Olfrs* and vomeronasal receptor (*Vmns*) genes, which are most often found in gene clusters[20,21]. In particular, *Olfrs* have been reported to be stochastically and minimally expressed in PGNs and substantia nigra DNs[22,23], have non-sensory organ functions, and are mis-expressed in several neurodegenerative diseases[23,24].

A sub-population of activated PGNs express a single Olfr gene, and co-express surrounding genes

To investigate the cell-type specific 3D genome topologies and how they relate with gene expression, we analysed single-cell transcriptomes of mature OLGs (8 subtypes) and OLG progenitors[15], CA1 PGN (3 subtypes)[25] and VTA DNs (4 subtypes)[14] from the brain regions of interest, and collected single cell mESC transcriptomes (**Fig. 3a**, **Supplemental Fig. S3a**). Separation of the different transcriptomes into cell-type specific clusters was confirmed by Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction, coloured according to the cell-type definitions provided in the published datasets (**Fig. 3b**) and validated by overlaying the expression of known marker genes for each cell type, which

confirmed that the different transcriptomes were grouped into cell-type specific clusters (**Supplemental Fig. S3b**).

To investigate further the strong B compartment features of *Olfr* genes, we started by exploring their expression state in single cells of each cell type (**Fig. 3c**). We found that 52% of mESCs express at least one *Olfr* gene (51/98 single cells), in comparison with 5/170 and 2/113 cells in OLGs and DNs, suggesting that the tighter B compartmentalization of *Olfr* genes might be important for their repression in the brain cell types. However, a large proportion of PGNs (19%; 165/875 cells) expressed one or more *Olfr* genes (up to 11). Most often the *Olfr* genes expressed in mESC or PGNs were found in chromosome 7 (20%, 10/51; and 29%, 47/165, respectively), which is consistent with the genome-wide distribution of *Olfrs* (21% *Olfr* genes in chromosome 7; **Supplemental Fig. S3c**). To better understand the significance of *Olfr* expression in mESCs and PGNs, we compared *Olfr* gene expression with the expression of all genes in each cell type. While *Olfr* expression is low in mESCs (mean TPM for all expressed *Olfr* genes: 23.0 ± 33.7), expressed *Olfr* genes in PGNs had higher expression (156.8 ± 121.5; **Fig. 3d**), suggesting that *Olfr* silencing mechanisms might be strongly disrupted in a subpopulation of PGNs, in contrast to being weakly expressed in many PGNs.

To determine whether *Olfr* escapee expression in PGNs was stochastic or related with functional differences between cell states within the PGN population, we measured differential gene expression between cells expressing at least one *Olfr* gene (*Olfr*-expressing PGNs) or none (non-expressing *Olfr* PGNs). Remarkably, we found 83 genes significantly upregulated in the PGNs expressing at least one *Olfr* gene, and 295 genes which were significantly downregulated (**Fig. 3e**). We asked

whether the differentially expressed genes were related to cell function by searching for enriched Gene Ontologies (GO) in each group of PGNs (**Fig. 3f**). We found that genes upregulated in *Olfr*-expressing PGNs are related to neuronal activation and long-term potentiation (LTP). For example, the GO term 'learning or memory' (GO:0007611) contained genes such as *Fos*, *Jun*, *Nrxn2*, *Egr1* and *Grin1*, which are all genes involved in the early response to neuronal activation and LTP[27-29]. Additional GO terms in upregulated genes were related to 'calcium ion transport', containing *Camk2a* and *Camk2b*, which are involved in calcium-dependent signaling during synaptic plasticity[30], and the calcium channel *Ryr3*, which is an epilepsy candidate gene[31]. Significantly downregulated genes mostly included mitochondrial related processes, which are coupled to and finely tuned during neuronal activation[32]. These data suggest that *Olfr* de-repression in PGNs coincides with the cascade of transcriptional responses to recent strong activation, which naturally occurs in CA1 PGNs during spatial tasks and is critical for long-term memory formation[33-35].

As *Olfr* genes are often found in clusters, we asked whether their de-repression in PGNs was also related to the expression of the surrounding genes (Fig. 3f-i). We selected *Olfr* genes that were active in at least one PGN or mESC and are located in one of two *Olfr* gene clusters on chromosome 7 (**Fig. 3f-i**). In mESCs, we found that other genes in the region (e.g. genes from the tripartite motif-containing (*Trim*) family) were active in some single cells, in the absence of *Olfr* expression. Of the 23 single PGNs that expressed at least one *Olfr* gene, seven PGNs (30%) co-expressed several genes in their neighborhood of the two *Olfr* clusters (**Fig. 3h**). The co-expressed genes included several *Trim* genes, which are E3-ubiquitin ligases involved in immune responses[36], and are involved in neuron development and spatial memory formation[37]. In contrast, PGNs that did not express *Olfr* genes had almost no

detectable expression for the genes in the represented region (23 randomly selected PGNs are shown in **Fig. 3i**; full heatmap is presented in **Supplemental Fig. S3d**). Inspection of GAM contact matrices in the same region show that expression of *Trim* genes in mESCs coincides with a smaller local hub of contacts, whereas in PGNs the locus is more globally open than in OLGs and DNs (**Fig. 3j**). Taken together, these data suggest that although *Olfr* cluster repression in brain cell types can coincide with B compartmentalization, the repression is not strict across the cell population. We find that a small number of *Olfr* genes are de-repressed, especially in PGNs which are undergoing neuronal activation, which is known to induce chromatin structure reorganisation[10]. In the same activated PGNs, we find reactivation of neighbouring *Trim* genes that are critical immune and memory response genes in the brain. These results raise the possibility that activation of neighboring *Trim* genes may lead to de-repression of neighboring *Olfrs*, irrespectively of their strong long-range contacts and membership to heterochromatic compartments B.

Differentially expressed genes are most often found in compartment A

To investigate further how changes in compartments relate with cell-type specific patterns of gene expression, we extended our analysis to determine genes significantly differentially expressed between each brain cell type and mESCs (**Fig. 3j**, **Supplementary Fig. S3e,f**). We found 11,861 differentially expressed genes (p adjusted value < 0.05), which could be clustered into 7 categories, including genes specifically active in each cell type or expressed in one or more cell types. We classified each gene as belonging to either a strong ‘A’, strong ‘B’ or an intermediate weak ‘A/B’ compartment, based on determining terciles from the eigenvectors distributions for each chromosome. Most differentially expressed genes were found in the strong A compartment, followed by the weak A/B, regardless of the cluster and

cell type, suggesting that the vast majority of most significant differences in gene expression do not coincide with strong B compartmentalization (**Fig. 3k**). To further test this observation, we took advantage of published maps of lamina-associated domains (LADs) in mESCs and in neuronal progenitor cells (NPCs)[38], and confirmed that most differentially expressed genes do not fall in LADs in mESCs or in *in-vitro* differentiated NPCs (**Fig. 3k**). Although a small number of genes changed their compartment definition between cell types, the relative proportion of genes within strong A, weak A/B and strong B compartments remained the same for all cell types, and consistently infrequent in B compartments.

Differentially expressed genes are targets of regulation by Polycomb repressor complexes

To explore the mechanisms that regulate differential expression between cell types in compartments A or intermediate A/B, we investigated the involvement of Polycomb regulation. In mESCs, about one fifth of genes are marked by H3K27me3, a repressive histone mark catalysed by Polycomb Repressor Complex 2 (PRC2), which partially resolves through neuronal differentiation[26]. As most of H3K27me3 positive peaks fall in compartment A in mESCs (72%; **Supplemental Fig. S3g**), we asked whether the differentially expressed genes that are upregulated in the brain cell types (gene clusters 1, 4, 6 and 7) are under Polycomb repression in mESCs. We took advantage of published promoter-state classifications previously defined in mESCs[26] according to the occupancy of active form of RNA polymerase II (Active promoters), contained the initiating form of RNA polymerase II (only RNAPII promoters), associated with H3K27me3 (Polycomb promoters) or with none (Inactive) (**Supplemental Fig. S3h**). We find that the genes with upregulated expression in brain cell types are often Polycomb repressed in mESCs. For example, Polycomb

marks the promoters of up to 36% (688 of 1936) genes in cluster 1. In cluster 7, which contains genes highly expressed in DNs, 27% (342 of 1245) genes are Polycomb targets in mESCs.

To explore the biological functions of Polycomb-target genes in clusters 1, 4, 6 and 7 we performed Gene Ontology (GO) analysis (**Supplemental Fig. S3i**). For cluster 1, Polycomb-classified genes which become expressed in all three brain cell types are enriched for GO terms related to shared functions such as 'gated channel activity' (GO:0022836), 'glutamate receptor signaling pathway' (GO:0007215) and 'synapse organization' (GO:0050808). Importantly, Polycomb-classified genes in cell-type specific clusters are critical for cell specialization. For example, GO terms for Polycomb-classified genes in the OLGs cluster (cluster 4) include 'ensheathment of neurons' (GO:0007272), and 'regulation of cell adhesion' (GO:0030155). Similarly, Polycomb-classified genes expressed in PGNs (cluster 6) are enriched for GO terms such as 'regulation of synaptic plasticity' (GO:0048167) and 'learning or memory' (GO:0007611), both critical functions of PGNs[29,30]. Polycomb-classified DN-specific genes in cluster 7 have GO terms including 'response to cocaine' (GO:0042220) and 'response to alcohol' (GO:0097305), highlighting their role in dopamine-receptor responses related to reward-based learning[39,40]. Together, these data suggest that many differentially expressed genes critical for cell specialization can be found within compartment A irrespectively of cell type, and are regulated through specialized mechanisms such as Polycomb repression.

Genes important for cell specialization are found within topological domain boundaries in brain cells

We found striking changes in compartment organization in cell types from the brain, including the strengthening of compartments B, but also that many differentially expressed genes between mESCs, OLGs, PGNs and DNs are present in the open compartments A. To investigate local changes in chromatin topology that may relate with cell-type specific gene expression, we next mapped the organization of topologically associating domains (TADs) in the four cell types. We applied the insulation score method[5,41] and compared the overlaps between TAD borders allowing overlap by at least one nucleotide. We found extensive genome-wide reorganisation of TADs, with 67% of all TAD borders detected in all datasets being found in one or more of the three brain cell types, but not in mESCs, and only 9% of all detected TAD borders being common to all cell types (**Supplemental Fig. S4a**). Comparisons of TAD features between cell types show similar average TAD length (~ 1 Mb) and TAD border length (~ 350 kb) and good conservation between replicates (82-93% overlap and correlated insulation scores**; Supplemental Fig. S4b-d**).

To visualise the extensive re-organisation of topological domains between mESCs and brain cell types, we chose as exemplars a genomic region on chromosome 5 containing a group of GABA receptor (*Gabr*) genes which are not expressed in mESCs and most active in PGNs (**Fig. 4a**) and the *Hist1* locus on chromosome 13 which contains clusters of replication-dependent histone 1 genes interrupted by a cluster of *Vmn* genes (**Supplemental Fig. S4e**). In both cases, inspection of local contacts revealed topologies that are specific to each cell type.

Pairwise comparisons of TAD boundaries between brain cells and mESCs showed the most divergent differences, as expected from the developmental distance between the terminally differential and pluripotent cell types, with ~500 borders specific of each brain cell type in comparison to ~300 – 400 specific of

mESCs (**Fig. 4b**). Comparisons between the two types of neurons, PGNs and DNs showed the lowest number of cell-type specific borders (**Fig. 4c**), and other comparisons showed intermediate levels of overlap, as expected (**Supplemental Fig. S4f**).

To assess the strength of TAD boundaries in each case, we computed the insulation score centered at the lowest point within each TAD border, and plotted average insulation score profiles in ±1-Mb windows centered at the lowest insulation score for common or cell-type specific TAD boundaries (**Fig. 4d**). We find stronger insulation scores surrounding the TAD boundaries in brain cell types, with the common boundaries being consistently stronger than cell-type specific boundaries. Boundaries common to DNs and PGNs had similar insulation profiles for both common and unique boundaries, with common being strongest (**Fig. 4e**).

To investigate whether the cell-type specific boundaries contained genes with functions relevant to the corresponding cell type, we investigated the ontology of genes at unique TAD boundaries, and observed that they are enriched for GO terms relevant for the respective cell function (**Fig. 4f**). Borders that are specific to mESCs are enriched GOs for functions such as 'retinoid binding', important for the maintenance of pluripotency[42], tyrosine phosphate activity and insulin-like growth factor binding, that mediate and maintain self-renewal in stem cells[43,44]. Boundaries unique to OLGs are enriched GO terms related to 'synaptic vesicle transport' and 'NMDA receptors', containing genes that encode synaptic proteins and neurotransmitter receptors expressed in OLGs and with functions in the response to neuronal input and modulation of myelin formation[45]. We found similar specialization-related genes in PGNs, with GO terms including genes related to 'phospholipase A2

activity', involved in the modulation of glutamatergic synapses[46], and 'ephrin receptor binding', which can regulate glutamate receptor signaling in the hippocampus[47]. DN-specific borders also contained cell specialization related genes, such as genes related to circadian regulation. Midbrain DNs have an important role in the homeostatic circadian and sleep-related modulation of neuronal activity[48]. Remarkably, comparisons between the smaller numbers of TAD borders specific to PGNs and DNs, also resulted in enrichment for genes involved in neuronal functions (e.g. 'GPCR binding' for DNs, and 'neurotransmitter' or 'glutamate receptor activity' for PGNs) or related to neurodevelopment and maintenance of cell identity (e.g. transcription factors such as *Foxj1*[14,49] and *Neurod6*[14] in DNs). Together, these results show that the genome folds into unique TAD boundaries that are cell-type specific and which contain genes critical for cell specialization and cell identity, suggesting that chromatin architecture holds signatures of the functional state of the cell.

Long and highly expressed neuronal genes undergo extensive rearrangement and can lose their TAD structure

Next, we were interested in exploring the genomic organisation of long neuronal genes. Many neuronal genes which are involved in specialized cell processes, such as synaptic plasticity, are longer than 200 kb in genomic length, often have many alternative promoters, and produce a large number of isoforms as a result of complex splicing reactions[50,51]. Transcription of long neuronal genes is also sensitive to topoisomerase inhibition suggesting their expression is highly dependent on topological constraints[52]. For example, Neurexin-3 (*Nrxn3*) is a 1.54-Mb long gene that encodes a membrane protein involved in synaptic connections, plasticity, and the modulation of neurotransmitter release[53]. In mESCs, the *Nrxn3* gene is folded into

two strong TADs that span past the length of the gene on either side, with an intervening TAD border near the middle of the gene (**Fig. 5a**). In mESCs, the region was mostly found in B compartment, and no mRNA expression was detected from *Nrxn3* (TPM = 0). To estimate the contact density for the region, we visualized the heatmap of a range of insulation scores (see **Methods**), and found that both TADs had strong contact density in mESCs. With increasing *Nrxn3* gene expression in OLGs (TPM = 16.6) and PGNs (TPM = 56.8), the contact density became weaker in the first TAD, but was maintained in the second (**Fig. 5b,c**).In contrast, in VTA DNs, where the *Nrxn3* is most highly expressed and sequencing reads are abundantly detected (TPM = 138.3) until the end of the gene, the contact density of the first *Nrxn3* TAD was reduced though it was still observed, whereas the contact density of the second *Nrxn3* TAD was drastically reduced, with nearly no TAD structure observable (**Fig. 5d**).

These observations were confirmed in the other neuronal cell type, PGNs, and for other long neuronal genes, for example in *Rbfox1* and *Nlgn1* in PGNs (**Supplemental Fig. S5a** and **S5b,** respectively). *Rbfox1* and *Nlgn1* are not expressed in mESCs and are localized within a 1.3-Mb TAD or within a strongly insulated and large domain of ~2.5-Mb, respectively. In PGNs, the contact densities of both *Rbfox1* and *Nlgn1* are strongly reduced, coinciding with very high expression levels for both genes (*Rbfox1* TPM = 204.3; *Nlgn1* TPM = 55.0). The loss of chromatin contacts within localized genomic regions that coincide with highest gene expression suggest that these genes undergo extensive decondensation, a phenomenon that we report here as 'TAD melting', and is reminiscent of chromatin puffs that accompany high transcription levels in polytene chromosomes[54].

We further quantified the significant changes in contact density around the *Nrxn3* gene between the different cell types by determining the top 5% significant differential contacts between each cell type (**Fig. 5e-g**, see **Methods**). Few significant differential contacts are found between mESCs and OLGs (**Fig. 5e**), though all significant differential contacts localized within the first *Nrxn3* TAD were enriched in mESCs. To assess changes in the number of significant contacts more broadly across the entire chromosome for each genomic window, we measured how many differential contacts each genomic region has across the entire 5-Mb region, and observed a depletion of significant contacts in genomic windows across the entire gene body in OLGs. The same trend could be seen between mESCs and PGNs, though with a larger number of significantly differential contacts, especially concentrated within the first *Nrxn3* TAD (**Fig. 5f**). Comparisons between PGNs and DNs showed a few differences, with slightly increased contacts in DNs the first *Nrxn3* TAD (**Supplemental Fig. S5c**). The 'melting' of the second *Nrxn3* TAD was associated with significantly stronger contacts only for PGNs. Comparison between DNs and mESCs showed significant depletions of DN contacts across the entire *Nrxn3* gene body (**Fig. 5g**), with loss of contacts localized within the melted TAD, but also between the two *Nrxn3* TADs in DNs. As with the other comparisons between mESCs and brain cells, we observed a depletion of significant contacts for DNs across the entire gene body. Similar significant depletion of contacts was found for *Rbfox1* and *Nlgn1* in comparisons between mESCs and PGNs (**Supplemental Fig. S5d,e**). For both genes, contacts were depleted at the gene promoter, and for *Rbfox1* between the first half of the gene body and the downstream region. When considering genomic windows with significant contacts, we also observed that PGNs were depleted for significant contacts across the entire gene body for *Rbfox1* and *Nlgn1*.

Modelling of loss of Nrxn3 TAD structure shows decompaction of chromatin

To better understand how the loss of TAD structure could influence 3D chromatin folding, we performed a polymer-physics-based approach (PRISMR)[55] to model the 3D structure of the *Nrxn3* region in mESCs and DNs (**Fig. 5h-l**). Following the generation of 3D models, we used an *in-silico* GAM approach (see **Methods**) to randomly section a single nuclear profile for each model, and then reconstructed an "*in-silico* GAM" chromatin contact matrix for mESCs (**Fig. 5h**) and DNs (**Fig. 5i**). *In-silico* GAM matrices closely resembled the experimental data for both mESCs (Pearson r = 0.72) and DNs (r = 0.79). Importantly, the loss of contacts in the second *Nrnx3* TAD was captured by the *in-silico* GAM approach.

Next, we next inspected single polymer models, coloured according to the structure of domains found in DNs. In the mESC models, the structure of the region was globular and highly intermingled (**Fig. 5j**). The second *Nrxn3* TAD (in green) frequently interacted with the surrounding domains, though the highest interactions were found with the regions downstream of the gene (blue and yellow domains). Remarkably, the second *Nrxn3* TAD in DNs was highly extended in DNs (**Fig. 5k**). We observed that the melted TAD tended to envelope the downstream (blue) domain, which was reflected in the contacts between the melted domain and the downstream region, as seen in both the experimental and in-silico matrices (**Fig. 5i**). In the polymer models from DNs, the second *Nrxn3* TAD shows a higher range of gyration radii compared to the same genomic regions in mESCs, indicating that the region tends to be highly decondensed or 'melted' (**Fig. 5l**). Together, these data highlight the complicated restructuring of local and long-range chromatin organization for different activation states of long neuronal genes, which can result in the loss of

topological structure, decompaction, and melting of an entire TAD. These observations open new avenues to better understand how gene expression is regulated in terminally differentiated cells of the brain and especially the genome topologies that may underly the production of complex transcript isoforms from long neuronal genes.

Discussion

Here, we adapted the GAM method to capture genome-wide chromatin conformation states of selected cell populations in the brain that perform specialized functions. Using immunoGAM, we selected nuclear slices from specific brain cells by immunofluorescence prior to laser microdissection, and characterized the relationship between 3D chromatin organization and cell specialization in intact tissues. We discovered dramatic reorganization of B compartment contacts in brain cell types, and specific contacts which spanned mega-base distances. In neurons, the tight repression of *Olfr* genes within these B compartment regions was especially noteworthy, as *Olfr* genes have critical non-sensory organ functions in neurons related to specificity of neuronal activation[23-24]. These functions were emphasized in our study, as activation of PGNs coincided with immediate early gene and synaptic plasticity related gene expression, activation of *Trim* genes and de-repression of neighbouring *Olfr* genes. *Olfr* genes are highly repressed in mature olfactory sensory neurons, where they form a large inter-chromosomal hub to regulate specificity of single *Olfr* gene activation[56,57]. Our results expand these previous findings, highlighting that alternative long-range mechanisms of *Olfr* repression may reflect their specialized roles in different cell types of the brain.

Of particular importance was the observation that the highest expression levels of very long neuronal genes coincide with a massive reorganization of local and long-range structure. We show that the highest expression of *Nrxn3*, *Rbfox1* and *Nlgn1* coincides with the loss of chromatin contacts and unfolding of the entire domain, which we call 'TAD melting', reminiscent of the formation of chromatin puffs previously reported by microscopical assays[54]. Many long neuronal genes are regulated in a specialized manner, for example by the activity of topoisomerases[52], by the presence of long stretches of broad H3K27ac and H3K4me1 which act as enhancer-like domains[58], or through repressive mechanisms such as DNA methylation[59]. The regulation of these long neuronal genes is further complicated by intricate splicing dynamics[50,51], which require highly dynamic and adaptive responses based on neuronal activation state. It is therefore not surprising that for many of these long genes, including for *Nrxn3*, *Rbfox1* and *Nlgn1*, genetic variants are often associated with or directly result in neuronal diseases[60-62]. Thus, understanding how genome folding relates with the response to environmental challenges is increasingly important to further our understanding of the mechanisms of neurological disease.

Our results also indicate that chromatin architecture might allow the prediction of functional states of a different cell type. We found that genes involved in very specific cell functions as retinoid signaling in ES cells, synaptic proteins involved in myelin formation in OLGs, glutamate signaling in PGNs and circadian regulation in DNs were in unique TAD boundaries in the corresponding cell types.

Collectively, our work revealed that cell specialization in the brain and chromatin structure are intimately linked at multiple genomic scales. ImmunoGAM is uniquely positioned for future studies to probe questions related to cell state and

specialized function within the brain or in other complex tissues, as local tissue structure is maintained. ImmunoGAM also has the potential to be applied in multiple cell types within the same tissue while retaining the geographic positions of each cell type, a prospect that can further deepen our understanding of coordinated interactions between cells in complex diseases. As immunoGAM requires only very small cell numbers (~400 – 1500 cells), it may additionally have prognostic value in highly precious clinical patient samples, and to provide critical insights into the aetiology and progression of neurological disease.

Methods

Animal maintenance

Collection of GAM data from dopaminergic neurons was performed using C57Bl/6NCrl (RRID: IMSR_CR:027; WT) mice which were purchased from Charles River, and from TH-GFP (B6.Cg-Tg(TH-GFP)21-31/C57B6) animals were obtained as previously described[63,64]. All procedures involving WT and TH-GFP animals were approved by the Imperial College London's Animal Welfare and Ethical Review Body. Adult male mice of age 2–3 months were used. All mice had access to food and water *ad libitum* and were kept on a 12 h:12 h day/night cycle. C57Bl/6NCrl and TH-GFP mice received an intraperitoneal (IP) injection of saline 14 days or 24 h prior to the tissue collection, respectively, and they were part of a larger experiment for a different study.

Collection of GAM data from somatosensory oligodendrocyte cells was performed using Sox10::Cre-RCE::loxP-EGFP animals which were obtained by crossing Sox10::Cre animals[65] on a C57BL/6j genetic background with RCE::loxP-EGFP animals[66] on a C57BL/6xCD1 mixed genetic background, both available at

The Jackson Laboratories. The Cre allele was maintained in hemizygosity while the reporter allele was maintained in hemi- or homozygosity. Experimental procedures for Sox10::Cre-RCE::loxP-EGFP animals were performed following the European directive 2010/63/EU, local Swedish directive L150/SJVFS/2019:9, Saknr L150 and Karolinska Institutet complementary guidelines for procurement and use of laboratory animals, Dnr 1937/03-640. The procedures described were approved by the local committee for ethical experiments on laboratory animals in Sweden (Stockholms Norra Djurförsöksetiska nämnd), lic.nr. 130/15. Animals were sacrificed at P21. Mice were housed to a maximum number of 5 per cage in individually ventilated cages with the following light/dark cycle: dawn 6:00-7:00, daylight 7:00-18:00, dusk 18:00-19:00, night 19:00-6:00.

Collection of GAM data from hippocampal CA1 pyramidal glutamatergic neurons was performed using 19 weeks old male *Satb2*$^{flox/flox}$ mice. C57Bl/6NCrl (RRID: IMSR_CR:027; WT) mice were purchased from Charles River, Satb2$^{flox/flox}$ mice that carry the floxed exon 4 have been previously described[67]. The experimental procedures were done according to the Austrian Animal Experimentation Ethics Board (Bundesministerium für Wissenschaft und Verkehr, Kommission für Tierversuchsangelegenheiten). All mice had access to food and water ad libitum and were kept on a 12 h:12 h day/night cycle.

Tissue fixation and preparation

WT, TH-GFP, and *Satb2*$^{flox/flox}$ mice were anaesthetised under isoflurane (4 %), given a lethal IP injection of pentobarbital (0.08 µl; 100 mg/ml; Euthatal), and transcardially perfused with 50 ml of ice-cold phosphate buffered saline (PBS) followed by 50-100 ml of 4% depolymerised paraformaldehyde (PFA; Electron microscopy grade, methanol free) in 250 mM HEPES-NaOH (pH 7.4-7.6).

Sox10::Cre-RCE::loxP-EGFP animals were sacrificed with a ketaminol/xylazine intraperitoneal injection followed by transcardial perfusion with 20 ml PBS and 20 ml 4% PFA in 250 mM HEPES (pH 7.4-7.6). From C57Bl/6NCrl or TH-GFP mice, brains were removed and the tissue containing the VTA were dissected from each hemisphere at room temperature, and quickly transferred to fixative. For SATB2-Cre mice, the CA1 field hippocampus was dissected from each hemisphere at room temperature. For Sox10Cre/RCE mice, brain tissue containing the somatosensory cortex was dissected at room temperature. Following dissection, tissue blocks were placed in 4% paraformaldehyde (PFA) in 250 mM HEPES-NaOH (pH 7.4-7.6) for post-fixation at 4°C for 1 h. Brains were then placed in 8% PFA in 250mM HEPES and incubated at 4°C for 2-3 h. Tissue blocks were then placed in 1% PFA in 250 mM HEPES and kept at 4°C until tissue was prepared for cryopreservation (up to 5 days).

Cryoblock preparation and cryosectioning

Fixed tissue samples from different brain regions were further dissected to produce ~1.5x3 mm tissue samples suitable for Tokuyasu cryosectioning[5] (Fig. S1a), at room temperature in 1% PFA in 250 mM HEPES. For the hippocampus, the dorsal CA1 region was further isolated. Approximately 1-3 mm x 1-3 mm blocks were dissected from all brain regions and were further incubated in 4% PFA in 250 mM HEPES at 4°C for 1 h. The fixed tissue was transferred to 2.1 M sucrose in PBS and embedded 16-24 h, at 4°C, positioned at the top of copper stub holders suitable for ultracryomicrotomy and frozen in liquid nitrogen. Cryopreserved tissue samples are kept indefinitely immersed under liquid nitrogen.

Frozen tissue blocks were cryosectioned with a Ultracryomicrotome (Leica Biosystems, EM UC7), with an approximate thickness of 220-230nm[5]. Cryosections were captured in drops of 2.1 M sucrose in PBS solution suspended in a copper wire

loop and transferred to 10 mm glass coverslips for confocal imaging, or onto a 4.0 µm polyethylene naphthalate (PEN; Leica Microsystems, 11600289) membrane for laser microdissection.

Immunofluorescence detection for confocal microscopy

For confocal imaging, cryosections were incubated in sheep anti-tyrosine hydroxylase (TH, 1:500; Pel Freez #P60101-0), mouse anti-pan-histone H11-4 (1:500; MAB3422 EMD Milipore #2842169) or chicken anti-GFP (1:500; ab13970 Abcam) followed by donkey anti-sheep or goat anti-chicken IgG conjugated with AlexaFluor-488 (for TH and GFP; Invitrogen) or donkey anti-mouse IgG conjugated with AlexaFluor-555 (for pan-histone; Invitrogen).

For PGNs, cryosections were washed (3x, 30 min total) in PBS, permeabilized (5 min) in 0.3% Triton X-100 in PBS (v/v) and incubated (2 h, room temperature) in blocking solution (1% BSA (w/v), 5% FBS (w/v) (GibcoTM Cat#10270), 0.3% Triton X-100 (v/v) in PBS). After incubation (overnight, 4°C) with primary antibody in blocking solution, the cryosections were washed (3-5x; 30 min) in 0.025% Triton X-100 in PBS (v/v) and immunolabeled (1 h, room temperature) with secondary antibodies in blocking solution, followed by 3 (15 min) washes in in 0.025% Triton X-100 in PBS (v/v). Cryosections were then counterstained with 0.5 µg/ml 4',6'-diamino-2-phenylindole (DAPI; Sigma-Aldrich® Cat#D9542), and then rinsed in PBS. Coverslips were mounted in Mowiol® 4-88 solutuon in 5% glycerol, 0.1M Tris-Hcl (pH 8.5).

The number of SATB2 positive cells present in the hippocampal CA1 area of the *Satb2*$^{flox/flox}$ control mice was determined by counting nuclei positive for the SATB2 immunostaining (AB_10563678, Abcam). To avoid counting the same nuclei, only every 30th ultrathin section cut through the tissue was collected, and the

remaining sections discarded. Twenty-five nuclei were identified in the pyramidal neurons layer per image in the DAPI channel only and SATB2-positive cells counted. We confirmed that within the CA1 layer most cells (96%) are pyramidal glutamatergic neurons (data not shown).

For DNs and OLGs, cryosections were washed (3x, 30min total) in PBS, quenched (20 min) in PBS containing 0.0375g Glycine**,** then permeabilized (15 min) in 0.1% Triton X-100 in PBS (v/v). Cryosections were then incubated (1h, room temperature) in blocking solution (1% BSA (w/v), 0.2% fish-skin gelatin (w/v), 0.05% casein (w/v) and 0.05% Tween-20 (v/v) in PBS). After incubation (overnight, 4°C) with the antibody in blocking solution, the cryosections were washed (3-5x; 1 h) in blocking solution and immunolabeled (1 h, room temperature) with secondary antibodies in blocking solution, followed by 3 (15 min) washes in in 0.5% Tween-20 in PBS (v/v). Cryosections were then counterstained with 0.5 µg/ml, then rinsed in PBS. Coverslips were mounted in Mowiol® 4-88.

Digital images were acquired with a Leica TCS SP8-STED confocal microscope (Leica Microsystems) using a 63x oil-immersion objective (NA = 1.4) or a 20x oil-immersion objective using pinhole equivalent to 1 Airy disk. Images were acquired using 405 nm excitation and 420-480 nm emission for DAPI; 488 nm excitation and 505-530 nm emission for TH or GFP; and 555 nm excitation and 560 nm emission using a long-pass filter at 1024x1024 pixel resolution. Images were processed using Fiji (version 2.0.0-rc-69/1.52p), where adjustments included the optimization of the dynamic signal range with contrast stretching.

Immunofluorescence detection for laser microdissection

For laser microdissection, cryosections on PEN membranes were washed, permeabilized and blocked as for confocal microscopy, and incubated with primary

and secondary antibodies as indicated above except for the use of higher concentrations of primary antibodies, as follows: anti-TH (1:50), anti-pan-histone (1:50) or anti-GFP (1:50). Secondary antibodies were used at the same concentration. Cell staining was visualized using a Leica laser microdissection microscope (Leica Microsystems, LMD7000) using a 63x dry objective. Following detection of cellular sections of the cell types of choice containing nuclear slices (nuclear profiles; NPs), individual NPs were laser microdissected from the PEN membrane, and collected into PCR adhesive caps (AdhesiveStrip 8C opaque; Carl Zeiss Microscopy #415190-9161-000). Three NPs were collected into each adhesive cap, and the presence of NPs in each lid was confirmed with a 5x objective using a 420-480 nm emission filter. Control lids not containing nuclear profiles (water controls) were included for each dataset collection, and can be found in **Supplemental Table 1**.

Whole genome amplification of nuclear profiles

Whole genome amplification (WGA) was performed using an in-house protocol[68]. Briefly, NPs were lysed directly in the PCR adhesive caps for 4 or 24h at 60°C in 1.2x lysis buffer (30 mM Tris-HCl pH 8.0, 2 mM EDTA pH 8.0, 800 mM Guanidinium-HCl, 5 % (v/v) Tween 20, 0.5 % (v/v) Triton X-100), containing 2.116 units/ml QIAGEN protease (Qiagen, 19155). After protease inactivation at 75°C for 30 min, the extracted DNA was amplified using random hexamer primers with an adaptor sequence. The pre-amplification step was done using 2x DeepVent mix (2x Thermo polymerase buffer (10x), 400µm dNTPs, 4 mM $MgSO_4$ in ultrapure water), 0.5 µM GAT-7N primers and 2 units/µl DeepVent® (exo-) DNA polymerase (New England Biolabs, M0259L). Primers that anneal to the general adaptor sequence were then used in a second amplification reaction to increase the amount of product.

This step was done as previously described (ref Protocols) using 2x DeepVent mix, 10 mM dNTPs, 100 µM GAM-COM primers and 2 units/µl DeepVent (exo-) DNA polymerase. For some NPs from DNs (see **Supplementary Table 1**), WGA was performed using a WGA4 kit (Sigma-Aldrich) using the manufacturer's instructions

GAM library preparation and high-throughput sequencing

Following WGA, the samples were purified using SPRI beads (0.725 or 1.7 ratio of beads per sample volume). The DNA concentration of each purified sample was measured using the Quant-iT® Pico Green dsDNA assay kit (Invitrogen #P7589) according to manufacturer's instructions. GAM libraries were prepared using the Illumina Nextera XT library preparation kit (Illumina #FC-131-1096) according to manufacturer's instructions or from an in-house library preparation protocol[68]. Following library preparation, the DNA was purified using SPRI beads (1.7 ratio of beads per sample volume) and the concentration for each sample was measured using the Quant-iT® PicoGreen dsDNA assay. An equal amount of DNA from each sample was pooled together (up to 196 samples), and the final pool was additionally purified three times using the SPRI beads (1.7 ratio of beads per sample volume). The final pool of libraries was analyzed using DNA High Sensitivity on-chip electrophoresis on an Agilent 2100 Bioanalyzer to confirm removal of primer dimers and estimate the average size and DNA fragment size distribution in the pool. NGS libraries were sequenced on Illumina NextSeq 500 machine, according to manufacturer's instructions, using single-end 75 bp reads. The number of sequenced reads for each sample can be found in **Supplemental Table 1**.

GAM data sequence alignment

Sequenced reads from each GAM library were mapped to the mouse genome assembly GRCm38 (Dec. 2011, mm10) with Bowtie2 using default settings[69]. All non-uniquely mapped reads, reads with mapping quality < 20 and PCR duplicates were removed and excluded from further analyses.

GAM data window calling and sample quality control

Positive genomic windows that are present within ultrathin nuclear slices were identified for each GAM library. Briefly, the genome was split into equal-sized windows (50 kb), and number of nucleotides sequenced in each bin was calculated for each GAM sample with bedtools[70]. Next, we determined the percentage of orphan windows (i.e. positive windows that were flanked by two adjacent negative windows) for every percentile of the nucleotide coverage distribution and identified the percentile with the lowest percent of orphan windows for each GAM sample in the dataset. The number of nucleotides that corresponds to the percentile with lowest percent of orphan windows in each sample was used as optimal coverage threshold for window identification in each sample. Windows were called positive if the number of nucleotides sequenced in each bin was greater than the determined optimal threshold.

Each dataset was assessed for quality control by determining the percentage of orphan windows in each sample, number of uniquely mapped reads to the mouse genome, and correlations from cross-well contamination for every sample (**Supplemental Table 1**). Each sample was considered to be of good quality if they had < 70% orphan windows, > 50,000 uniquely mapped reads, and a cross-well contamination score (Jaccard index) of < 0.4. The number of samples in each cell type passing quality control is summarized in **Table 1**.

Publicly available GAM datasets from mouse embryonic stem cells (mESC)

For mESCs, GAM datasets were downloaded from 4D Nucleome portal. We used 249 x 3NP GAM datasets from mESCs (clone 46C) which were grown at 37°C in a 5% CO_2 incubator in Glasgow Modified Eagle's Medium, supplemented with 10% fetal bovine serum, 2 ng/ml LIF and 1 mM 2-mercaptoethanol, on 0.1% gelatin-coated dishes. Cells were passaged every other day. After the last passage 24 h before harvesting, mESCs were re-plated in serum-free ESGRO Complete Clonal Grade medium (Millipore). The list of 4DN sample IDs is provided in **Supplementary Table S2**.

Visualization of pairwise chromatin contact matrices

To visualize GAM data, contact matrices were calculated where the frequency of window detection is normalized using pointwise mutual information (PMI) for all pairs of windows genome-wide. PMI describes the probability of a pair of genomic windows being found in the same NP given both their co-segregation frequency and their individual frequency across all NPs. PMI was calculated by the following formula, where p(x) and p(y) are the individual distributions of genomic windows x and y, respectively; and p(x,y) are their joint distribution:

(1) $PMI = \log(p(x,y) / p(x)p(y))$

PMI can be bounded between -1 and 1, to produce a normalized PMI (NPMI) value given by the following formula:

(2) $NPMI = PMI / (-\log p(x,y))$

Identification of compartments A and B

For compartment analysis, matrices of co-segregation frequency were determined by the ratio of independent occurrence of a single positive window in

each sample over the pairwise co-occurrence of pairs of positive windows in a given pair of genomic windows. GAM co-segregation matrices at 250kb resolution were assigned to either A- or B-compartments, as previously described[5]. Briefly, each chromosome was represented as a matrix of observed interactions O(i,j) between locus i and locus j (co-segregation), and separately for E(i,j) where each pair of genomic windows are the mean number of contacts with the same distance between i and j. A matrix of observed over expected values O/E(i,j) was produced by dividing O by E. A correlation matrix C(i,j) is produced between column i and column j of the O/E matrix. Principal component analysis was performed for the first 3 components on matrix C, before extracting the component with the best correlation to GC content. Finally, loci with PCA eigenvectors with the same sign that correlate best with GC content were called A compartments, while regions with the opposite sign were B compartments. Compartments were considered common if they had the same compartment definition within the same genomic bin. Compartment changes between cell types were computed after considering compartments that were common between biological replicates, unless otherwise indicated.

Insulation score and topological domain boundary calling

TAD calling was performed by calculating insulation scores in NPMI normalized pairwise GAM contact matrices at 50-kb resolution, as previously described[41]. The insulation score was computed individually for each cell type and biological replicate, with insulation square sizes ranging from 100 - 1000 kb. TAD boundaries were called using a 500kb-insulation square size and based on local minima of the insulation score. Boundaries that were found overlapping by at least 1 genomic bin were merged. Boundaries separated by at least one genomic bin were considered different between datasets. In Figs. 4b, S4c, we considered the genomic

window within a boundary with the lowest insulation value as the boundary coordinate.

Gene Ontology

Gene ontology (GO) term enrichment analysis was performed using GOElite (version 1.2.4)[71]. In Figs. 2e, 3l, and 4c, all unique genes annotated to mm10 were used as the background dataset, and in Fig. 3f all expressed PGN genes were used. Default parameters were used for the GO enrichment: GO terms that were enriched above the background (significant permuted p-values < 0.05, 2000 permutations) were pruned to select the terms with the largest Z-score (> 1.96) relative to all corresponding child or parent paths in a network of related terms (genes changed > 2). GO terms which had permuted p-value ≥ 0.01, contained fewer than 6 genes per GO term, or from the 'cellular_component' ontology, were not reported. A full list of unfiltered GO terms can be found in **Supplemental Table 3**.

Determining differential contacts between GAM datasets

To determine significant differences in pairwise contacts between a pair of GAM datasets, genomic windows with low detection, defined as less than 2% of the distribution of all detected genomic windows for each chromosome, were removed from both datasets to be compared. Next, NPMI contact frequencies at each genomic distance of each chromosome were normalized by computing the Z-score transformation. A differential matrix D was derived by subtracting the two Z-score normalized matrices. A 10-Mb distance threshold was applied after computing the normalized matrices. The top significant differential contacts were determined for each dataset by fitting a normal distribution to matrix D and defining the upper and lower 5% from the fitted curve. Finally, the number of top significant contacts was

quantified in each genomic window for each dataset. The difference between the number of significant contacts found at each genomic region was computed to define a differential window score for each genomic bin.

mESC cell culture for scRNA-seq

Mouse embryonic stem cells from the 46C clone, derived from E14tg2a and expressing GFP under *Sox1* promoter[72], were a kind gift of D. Henrique (Instituto de Medicina Molecular, Faculdade Medicina Lisboa, Lisbon, Portugal). mESCs were cultured as previously described[26], i.e. cells were routinely grown at 37°C, 5% (v/v) CO_2, on gelatine-coated (0.1% v/v) Nunc T25 flasks in GMEM medium (Invitrogen, Cat# 21710025), supplemented with 10% (v/v) Foetal Calf Serum (FCS; BioScience LifeSciences, Cat# 7.01, batch number 110006), 2000 U/ml Leukaemia inhibitory factor (LIF, Millipore, Cat# ESG1107), 0.1 mM beta-mercaptoethanol (Invitrogen, Cat# 31350-010), 2 mM L-glutamine (Invitrogen, Cat# 25030-024), 1 mM sodium pyruvate (Invitrogen, Cat# 11360039), 1% penicillin-streptomycin (Invitrogen, Cat# 15140122), 1% MEM Non- Essential Amino Acids (Invitrogen, Cat# 11140035). Medium was changed every day and cells were split every other day. mESC batches were tested for mycoplasma infection. Before collecting material for scRNA-seq, cells were grown for 48h in serum-free ESGRO Complete Clonal Grade Medium (Millipore, Cat# SF001- B) supplemented with 1000 U/ml LIF, on gelatine (Sigma, Cat# G1393-100ml)-coated (0.1% v/v) Nunc 10 cm dishes, with a medium change after 24h.

Single-cell mRNA library preparation

Single-cell mRNA-seq libraries were prepared according to Fluidigm manual "Using the C1 Single-Cell Auto Prep System to Generate mRNA from Single Cells

and Libraries for Sequencing". Cell suspension was loaded on 10-17 microns C1 Single-Cell Auto Prep IFC (Fluidigm, Cat# 100-5760, kit Cat# 100-6201). After loading, the chip was observed under the microscope to score cells as singlets, doublets, multiplets, debris or other. Chip was then loaded again on Fluidigm C1 and cDNA was synthesised and pre-amplified in the chip using Clontech SMARTer kit (Takara Clontech, Cat# 634833). Spike-In Mix 1 (Life Technologies, Cat# 4456740), diluted 1:1000, was added to batch B ESCs as from Fluidigm manual. Illumina sequencing libraries were prepared with Nextera XT kit (Illumina, Cat# FC- 131-1096) and Nextera Index Kit (Illumina, Cat# FC-131-1002), as previously described[73]. Libraries from each microfluidic chip (96 cells) were pooled and sequenced on 4 lanes on Illumina HiSeq 2000, 2x100bp paired-end (batch A) or 1 lane on Illumina HiSeq 2000, 2x125bp paired-end (batch B) in the Wellcome Trust Sanger Institute Sequencing Facility.

ScRNA-seq data processing: mapping and expression estimates

To calculate expression estimates, mRNA-seq reads were mapped with STAR (Spliced Transcripts Alignment to a Reference, v2.4.2a)[74] and processed with RSEM, using the 'single-cell-prior' option (RNA-Seq by Expectation- Maximization, v1.2.25)[75]. The references provided to STAR and RSEM were the gtf annotation from UCSC Known Genes (mm10, version 6) and the associated isoform-gene relationship information from the Known Isoforms table (UCSC), adding information for ERCCs sequences when present. Tables were downloaded from the UCSC Table browser (http://genome.ucsc.edu/cgi- bin/hgTables) and, for ERCCs, from ThermoFisher website (www.thermofisher.com/order/catalog/product/4456739). Gene-level expression estimates in "Expected Counts" from RSEM were used for the analysis.

ScRNA-seq data processing: quality control

Cells scored as doublets/ multiplets, debris or other during visual inspection of the C1 chip were excluded from the analysis. Datasets were also excluded if any of the following conditions were met: <500,000 reads (calculated using sam-stats from ea-utils.1.1.2-537)[76]; <60% of reads mapped (calculated with sam-stats); <50% reads mapped to mRNA (picard-tools-2.5.0, broadinstitute.github.io/picard/); >15% of reads mapped to chrM (sam-stats); if present, >20% of reads mapped to ERCCs (sam-stats).

Correlations between previously published mRNA-seq bulk[26] and single-cell RNAseq data collected from mESCs (clone 46C) were performed to assess quality of the single cell data. Correlations were performed as previously described[77]. Average single cell expression was highly correlated with bulk RNA-seq (**Supplemental Fig. S3a**).

ScRNA-seq integration and analysis

To integrate single cell transcriptomes from brain cell types of interest, we selected p21-22 Oligodendrocytes[15], p22-32 CA1 PGNs[25], and p21-26 VTA DNs[14], based on the cell type and subtype definitions provided in their respective publications. A summary of all published data used in this study can be found in **Supplemental Table 4**. The matrices of counts provided in each publication, along with the single-cell mESC transcriptomes produced here that passed QC, were normalized applying the *LogNormalize* method and scaled using Seurat[78]. The scaled data was used for a PCA analysis, followed by processing through dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP)[79]. Subsequent clustering was performed using the Seurat R package[78], with

default parameters. To calculate Transcripts Per Million (TPM) for each cell type, counts per gene for all cells belonging to the same cell type were pooled and normalized by sequencing depth and gene length. Effective gene length was calculated by combining the length of all exons for each gene.

Mean transcription values of olfactory receptor genes (*Olfr*, n=1120) and non-*Olfr* genes (n= 22167) were computed for cells with respective non-zero TPMs. Single cell TPM values were log2 transformed (pseudocount=1) and cells hierarchically clustered (mean agglomerative clustering with euclidean distance metrics) for generation of heatmaps and sample distances. The order of genes along the y-axis reflects their genomic position. A randomly chosen subset of cells (n=23) not expressing *Olfr* genes was drawn to illustrate transcription differences of *Olfr* clusters 1 + 2 and interspersed genes.

To identify co-expressed genes for both groups of PGNs (expressing and not-expressing *Olfr* genes) all PGN cells were pooled per group and partitioned into two pseudo-replicates. To determine differentially expressed (DE) genes, all 6 possible pairwise comparisons between samples were performed using DEseq2 with default parameters[80]. All genes classified as DE (adjusted p-value < 0.05; Benjamini-Hochberg test) in at least one comparison were considered for further analysis. A summary table for the DE expression analysis of *Olfr*-expressing PGNs can be found in **Supplemental Table 5.**

For differential expression analysis for all cell types, pseudo-bulk replicate samples were obtained by randomly partitioning the total number of single cells per dataset into three groups. DE expression analysis was performed as in the *Olfr* analysis and with the same parameters, except with 3 pseudo-replicates. To identify and visualize gene expression differences among DE expressed genes, *K-means clustering* was performed using the pheatmap R package with *k-means* = 7.

Expression values from DESeq per gene were centered and scaled for visualization. A summary table for the DE expression analysis of cell types can be found in **Supplemental Table 6.**

To generate bigwig tracks for visualization, raw fastq files from each cell type were pool into one fastq file. Reads were mapped to the mouse genome (mm10) using STAR[81] with default parameters but *--outFilterMultimapNmax 10*. Bam files were sorted and indexed using Samtools[82] and normalized (RPKM) bigwigs were generated using Deeptools[83] *bamCoverage*. To account for differences in the number of runs sequences for OLGs, cells were divided into groups by number of runs (1,2 and 6). The median of the reads for the group with the lowest sequencing depth was used as a threshold to normalize the other groups (i.e. the rest of the fastq files were randomly downsampled to that number of reads). The three groups of raw reads were pooled together and processed by applying the same method as for the other cell types.

Modeling and In-silico GAM

To reconstruct 3D conformations of the *Nrxn3* locus we employed the Strings & Binders Switch (SBS) polymer model of chromatin[84,85]. In the SBS, a chromatin region is modelled as a self-avoiding chain of beads, including different binding sites for diffusing, cognate, molecular binders. Binding sites of the same type can be bridged by their cognate binders, thus driving the polymer folding. The optimal SBS polymers for the *Nrxn3* locus in mESCs and DNs were inferred by using PRISMR, a machine learning based procedure which finds the minimal arrangement of the polymer binding sites best describing input pairwise contact data, such as Hi-C[55] or GAM[86]. Here, PRISMR was applied to the GAM experimental data by additionally considering the NPMI normalization. We focused on a 4.8Mb region around the *Nrxn3* gene (Chr12:

87,600,000-92,400,000; mm10) at 50kb resolution in mESCs and DNs. The procedure returned optimal SBS polymer chains made of 1440 beads, including 7 different types of binding sites, in both cell types.

Next, to generate thermodynamic ensembles of 3D conformations of the locus, we ran massive molecular dynamics simulations of the optimal polymers, performed with the freely available LAMMPS software[87]. In those simulations, the system evolves according to Langevin equation, with dynamics parameters derived in classical polymer physics studies[88]. Polymers are first initialized in self-avoiding conformations and then evolve up to reach their equilibrium globular phase[84]. Beads and binders have the same diameter $\sigma = 1$, expressed in dimensionless units, and experience a hard-core repulsion by use of a truncated Lennard-Jones (LJ) potential. Analogously, attractive interactions are modelled with short-ranged LJ potentials[84]. We sampled a range of affinities between beads and cognate binders in the weak biochemical range, from $3.0K_BT$ to $8.0K_BT$ (where K_B is the Boltzmann constant and T the system temperature). In addition, binders interact non-specifically with the polymer with a lower affinity, sampled from $0K_BT$ to $2.7K_BT$. For sake of simplicity, we used the same affinity strengths for all different binding site types. Total binder concentration was taken above the polymer coil-globule transition threshold[84]. For each of the considered cases, we derived ensembles of up to ~500 distinct equilibrium configurations. Full details about the model and simulations are discussed in[84,89].

In-silico GAM NPMI matrices were obtained from the ensemble of 3D structures by applying the *in-silico* GAM algorithm[90], here generalized to simulate the GAM protocol with 3NPs per GAM sample and to perform NPMI normalization. Specifically, we used the same number of slices of the GAM experiments, 249x3NPs for mESCs and 585x3NPs for DNs, respectively. Pearson correlation coefficients were used to compare the *in-silico* and experimental NPMI GAM matrices.

Example of single 3D conformations were rendered by a third-order spline of the polymer beads positions, with regions of interest highlighted in different colors. To quantify the size and variability of the 3D structures in mESCs and DNs, we measured the average gyration radius, R_g of the TAD encompassing the *Nrxn3* gene (chr12: 89,500,000-90,300,000, green region in **Figs. 5j**,**k**), expressed in dimensionless units σ in **Fig. 5l**.

Acknowledgments

The authors thank Sheila Q. Xie for help processing midbrain samples, and the Pombo lab members for helpful discussions. A.P. and A.A. acknowledge support from the Helmholtz Association (Germany). A.P. and M.N. acknowledge support from the National Institutes of Health Common Fund 4D Nucleome Program grant U54DK107977, and the Berlin Institute of Health (BIH). A.P. acknowledges support from the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) under Germany's Excellence Strategy – EXC-2049 – 390688087. G.C.B. acknowledges European Union Horizon 2020/European Research Council Consolidator Grant (EPIScOPE no. 681893), Swedish Research Council (no. 2015-03558; 2019-01360), Swedish Brain Foundation (no. FO2017-0075), Knut and Alice Wallenberg Foundation (grant 2019-0107), The Swedish Society for Medical Research (SSMF, grant JUB2019), Ming Wai Lau Centre for Reparative Medicine and Karolinska Institutet. G.D. and G.A. acknowledge support from the Austrian FWF through DK W1206 "Signal Processing in Neurons"and SFB F44 "Cell Signaling in Chronic CNS Disorders", P25014-B24. M.A.U. acknowledges funding by the Medical Research Council (UK) (U120085816) and a Royal Society University Research Fellowship. M.N. thanks support from CINECA ISCRA Grant HP10CYFPS5 and

HP10CRTY8P, by computer resources at INFN and Scope at the University of Naples (MN). I.H. was supported by a Boehringer Ingelheim Fonds PhD fellowship, L.Z.R. by Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) IRTG2403, and E.T.T. by an EMBO short-term fellowship (ASTF 336-2015).

Author Contributions

AP designed the concept for this work; WW, AK, IH, MM, LS and RK produced GAM datasets; WW, AK, IH, MM and LS optimised the experimental protocol; WW, AK, IIA, CT and EI developed computational pipelines for bioinformatics and QC analyses of GAM data; AK and IIA performed QC analyses of GAM data; WW and AK performed bioinformatics analyses of GAM data; WW, MM, AA, EJP and AP collected animal tissues; ETT performed mESC culture experiments; ETT and AAK produced single-cell RNAseq data; LZR and DS performed RNA-seq analysis; LZR performed differential gene expression analyses; SB optimized polymer modeling and performed PRISMR analysis; SB and AMC produced models for polymer modeling; AMC performed statistical analyses of polymer models; LF and FM performed in-silico GAM experiments; ST supervised scRNA-seq experiments; MM, GA, GD, MU and GCB supervised animal tissue collection and provided animal samples; AP supervised GAM experiments and bioinformatics analyses; VF, AA, and AP supervised the scRNA-seq analysis; MN supervised the polymer modeling and in-silico GAM; WW, AK, IH, LZR, MM, SB, AMC, MN, GCB and AP contributed to the interpretation of the results; WW wrote the first draft of the manuscript; WW and IH designed the figures; WW, AK, IH, LZR, SB, and AP wrote the manuscript. All authors provided critical feedback and helped revise the manuscript. The authors consider WW, AK and IH to have contributed equally to this work.

References

1. Lupiáñez, D.G., *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).

2. Medrano-Fernández, A. & Barco, A. Nuclear organization and 3D chromatin architecture in cognition and neuropsychiatric disorders. *Mol. Brain* **9**, 83 (2016).

3. Jung, I., *et al.* A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat. Genet.* **51**, 1442-1449 (2019).

4. Rao, S.S., *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).

5. Beagrie, R.A., *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519-524 (2017).

6. Quinodoz, S.A., *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757 (2018).

7. Dixon, J.R., *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-336 (2015).

8. Fraser, J., *et al.* Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol. Syst. Biol.* **11**, 852 (2015).

9. Bonev, B., *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572 (2017).

10. Fernandez-Albert, J., *et al.* Immediate and deferred epigenomic signatures of in vivo neuronal activation in mouse hippocampus. *Nat. Neurosci.* **22**, 1718-1730 (2019).

11. Parikshak, N.N., *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008-1021 (2013).

12. McCarthy, S.E., *et al.* De novo mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol. Psychiatry* **19**, 652-658 (2014).

13. Won, H., *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527 (2016).

14. La Manno, G., *et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167**, 566-580 (2016).

15. Marques, S., *et al.* Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326-1329 (2016).

16. Tiklová, K., *et al.* Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. *Nat. Commun.* **10**, 581 (2019).

17. Poulin, J.F., *et al.* Mapping projections of molecularly defined dopamine neuron subtypes using intersectional genetic approaches. *Nat. Neurosci.* **21**, 1260-1271 (2018).

18. Falcão, A.M., *et al.* Disease-specific oligodendrocyte lineage cells arise in multiple sclerosis. *Nat Med* **24**, 1837-1844 (2018)

19. Costa, L., Trovato, C., Musumeci, S.A., Catania, M.V. & Ciranna, L. 5-HT(1A) and 5-HT(7) receptors differently modulate AMPA receptor-mediated hippocampal synaptic transmission. *Hippocampus* **22**, 790-801 (2012).

20. Magklara, A., *et al.* An epigenetic signature for monoallelic olfactory receptor expression. *Cell* **145**, 555-570 (2011).

21. Kambere, M.B. & Lane, R.P. Co-regulation of a large and rapidly evolving repertoire of odorant receptor genes. *BMC Neurosci.* **8 Suppl 3**, S2 (2007).

22. Grison, A., *et al.* Mesencephalic dopaminergic neurons express a repertoire of olfactory receptors and respond to odorant-like molecules. *BMC Genomics* **15**, 729 (2014).

23. Garcia-Esparcia, P., *et al.* Functional genomics reveals dysregulation of cortical olfactory receptors in Parkinson disease: novel putative chemoreceptors in the human brain. *J. Neuropathol. Exp. Neurol.* **72**, 524-539 (2013).

24. Ansoleaga, B., *et al.* Dysregulation of brain olfactory and taste receptors in AD, PSP and CJD, and AD-related model. *Neuroscience* **248**, 369-382 (2013).

25. Zeisel, A., *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-142 (2015).

26. Ferrai, C., *et al.* RNA polymerase II primes Polycomb-repressed developmental genes throughout terminal neuronal differentiation. *Mol. Syst. Biol.* **13**, 946 (2017).

27. Minatohara, K., Akiyoshi, M. & Okuno, H. Role of Immediate-Early Genes in Synaptic Plasticity and Neuronal Ensembles Underlying the Memory Trace. *Front. Mol. Neurosci.* **8**, 78 (2016).

28. Duclot, F. & Kabbaj, M. The Role of Early Growth Response 1 (EGR1) in Brain Plasticity and Neuropsychiatric Disorders. *Front. Behav. Neurosci.* **11**, 35 (2017).

29. Born, G., *et al.* Genetic targeting of NRXN2 in mice unveils role in excitatory cortical synapse function and social behaviors. *Front. Synaptic Neurosci.* **7**, 3 (2015).

30. Murakoshi, H., *et al.* Kinetics of Endogenous CaMKII Required for Synaptic Plasticity Revealed by Optogenetic Kinase Inhibitor. *Neuron* **94**, 37-47.e5 (2017)

31. EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, & Epi4K Consortium. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *Am. J. Hum. Genet.* **95**, 360-370 (2014).

32. Kann, O. & Kovács, R. Mitochondria and neuronal activity. *Am. J. Physiol. Cell Physiol.* **292**, C641-657 (2007).

33. Zhang, S., Schönfeld, F., Wiskott, L. & Manahan-Vaughan, D. Spatial representations of place cells in darkness are supported by path integration and border information. *Front. Behav. Neurosci.* **8**, 222 (2014).

34. Epsztein, J., Brecht, M. & Lee, A.K. Intracellular determinants of hippocampal CA1 place and silent cell activity in a novel environment. *Neuron* **70**, 109-120 (2011).

35. Chan, R.W., *et al.* Low-frequency hippocampal-cortical activity drives brain-wide resting-state functional MRI connectivity. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6972-E6981 (2017).

36. Rajsbaum, R., Stoye, J.P. & O'Garra, A. Type I interferon-dependent and -independent expression of tripartite motif proteins in immune cells. *Eur. J. Immunol.* **38**, 619-630 (2008).

37. Boyer, N.P., Monkiewicz, C., Menon, S., Moy, S.S. & Gupton, S.L. Mammalian TRIM67 Functions in Brain Development and Behavior. *eNeuro* **5** (2018).

38. Peric-Hupkes, D., *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell*, **38**, 603-613 (2010).

39. Peciña, M., *et al.* DRD2 polymorphisms modulate reward and emotion processing, dopamine neurotransmission and openness to experience. *Cortex* **49**, 877-890 (2013).

40. Solecki, W., *et al.* Effects of brief inhibition of the ventral tegmental area dopamine neurons on the cocaine seeking during abstinence. *Addict. Biol.* **2**, e12826 (2019).

41. Crane, E., *et al.* Condensin-Driven Remodeling of X-Chromosome Topology during Dosage Compensation. *Nature* **523**, 240-244 (2015).

42. Khillan, J.S. Vitamin A/retinol and maintenance of pluripotency of stem cells. *Nutrients* **6**, 1209-1222 (2014).

43. Wang, L., *et al.* Self-renewal of human embryonic stem cells requires insulin-like growth factor-1 receptor and ERBB2 receptor signaling. *Blood* **110**, 4111-4119 (2007).

44. Kobayashi, M., *et al.* PRL2/PTP4A2 phosphatase is important for hematopoietic stem cell self-renewal. *Stem Cells* **32**, 1956-1967 (2014).

45. Huhges, A.N. & Appel, B. Oligodendrocytes express synaptic proteins that modulate myelin sheath formation. *Nat. Commun.* **10**, 4125 (2019).

46. Yue, H.Y., Fujita, T. & Kumamoto, E. Phospholipase A2 activation by melittin enhances spontaneous glutamatergic excitatory transmission in rat substantia gelatinosa neurons. *Neuroscience* **135**, 485-495 (2005).

47. Antion, M.D., Christie, L.A., Bond, A.M., Dalva, M.B. & Contractor, A. Ephrin-B3 regulates glutamate receptor signaling at hippocampal synapses. *Mol. Cell. Neurosci.* **45**, 378-388 (2010).

48. Fifel, K., Meijer, J.H. & Deboer, T. Circadian and Homeostatic Modulation of Multi-Unit Activity in Midbrain Dopaminergic Structures. *Sci. Rep.* **8**, 7765 (2018).

49. Bifsha, P., Balsalobre, A. & Drouin, J. Specificity of Pitx3-Dependent Gene Regulatory Networks in Subsets of Midbrain Dopamine Neurons. *Mol. Neurobiol.* **54**, 4921-4935 (2017).

50. Treutlein, B., Gokce, O., Quake, S.R. & Südhof, T.C. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E1291-1299 (2014).

51. Schreiner, D., *et al.* Targeted combinatorial alternative splicing generates brain region-specific repertoires of neurexins. *Neuron* **84**, 386-398 (2014).

52. King, I.F., *et al.* Topoisomerases facilitate transcription of long genes linked to autism. *Nature* **501**, 58-62 (2013).

53. Traunmüller, L., Bornmann, C. & Scheiffele, P. Alternative splicing coupled nonsense-mediated decay generates neuronal cell type-specific expression of SLM proteins. *J. Neurosci.* **34**, 16755-16761 (2014).

54. Adelman, K., *et al.* Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol Cell*, **17**, 103-112.

55. Biano, S., *et al.* Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat Genet.* **50**, 662-667 (2018).

56. Monahan, K., *et al.* Cooperative interactions enable singular olfactory receptor expression in mouse olfactory neurons. *Elife* **6**, e28620 (2017).

57. Monahan, K., Horta, A. & Lomvardas, S. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* **565**, 448-453 (2018).

58. Zhao, Y.T., *et al.* Long genes linked to autism spectrum disorders harbor broad enhancer-like chromatin domains. *Genome Res.* **28**, 933-942 (2018).

59. Gabel, H.W., *et al.* Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89-93 (2015).

60. Vaags, A.K., *et al.* Rare deletions at the neurexin 3 locus in autism spectrum disorder. *Am. J. Hum. Genet.* **90**, 133-141 (2012).

61. Hamada, N., *et al.* Essential role of the nuclear isoform of RBFOX1, a candidate gene for autism spectrum disorders, in the brain development. *Sci. Rep.* **6**, 30805 (2016).

62. Tristán-Clavijo, E., *et al.* A truncating mutation in Alzheimer's disease inactivates neuroligin-1 synaptic function. *Neurobiol. Aging* **36**, 3171-3175 (2015).

63. Sawamoto, K., *et al.* Generation of dopaminergic neurons in the adult brain from mesencephalic precursor cells labeled with a nestin-GFP transgene. *J. Neurosci.* **21**, 3895-3903 (2001).

64. Matsushita, N., *et al.* Dynamics of tyrosine hydroxylase promoter activity during midbrain dopaminergic neuron development. *J. Neurochem.* **82**, 295-304 (2002).

65. Matsuoka, T., *et al.* Neural crest origins of the neck and shoulder. *Nature* **436**, 347-355 (2005)

66. Sousa, V.H., *et al.* Characterization of Nkx6-2-derived neocortical interneuron lineages. *Cereb. Cortex* **19 Suppl 1**, i1-10 (2009).

67. Jaitner, C., *et al.* Satb2 determines miRNA expression and long-term memory in the adult central nervous system. *Elife* **5**, e17361 (2016).

68. Kukalev, A., *et al.* Genome Architecture Mapping: an improved pipeline for efficient ligation-free analyses of 3D genome topology. In preparation.

69. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359 (2012).

70. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

71. Zambon, A.C., *et al.* GO-Elite: a flexible solution for pathway and ontology over-representation. *Bioinformatics* **28**, 2209-2210 (2012).

72. Ying, Q.L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol.* **21**, 183-186 (2003)

73. Kolodziejczyk, A.A., *et al.* Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell* **17**, 471-485 (2015).

74. Dobin, A., *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

75. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

76. Aronesty, E. ea-utils: Command-line tools for processing biological sequencing data. *The open bioinformatics journal* http://code.google.com/p/ea-utils: **7**, 602 (2011)

77. Kar, G., *et al.* Flipping between Polycomb repressed and active transcriptional states introduces noise in gene expression. *Nat Commun.* **8**, 36 (2017).

78. Macosko, E.Z., *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).

79. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *The Journal of Open Source Software* doi: 10.21105/joss.00861

80. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

81. Dobin, A., *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013)

82. Li, H., *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009)

83. Ramirez, F., *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-165 (2016)

84. Barbieri, M. *et al.* Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci.* **109**, 16173–16178 (2012).

85. Chiariello, A. M., Annunziatella, C., Bianco, S., Esposito, A. & Nicodemi, M. Polymer physics of chromosome large-scale 3D organisation. *Sci. Rep.* **6**, 29775 (2016).

86. Fiorillo, L. *et al.* Inference of chromosome 3D structures from GAM data by a physics computational approach. *Methods* (2019). doi:10.1016/j.ymeth.2019.09.018

87. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).

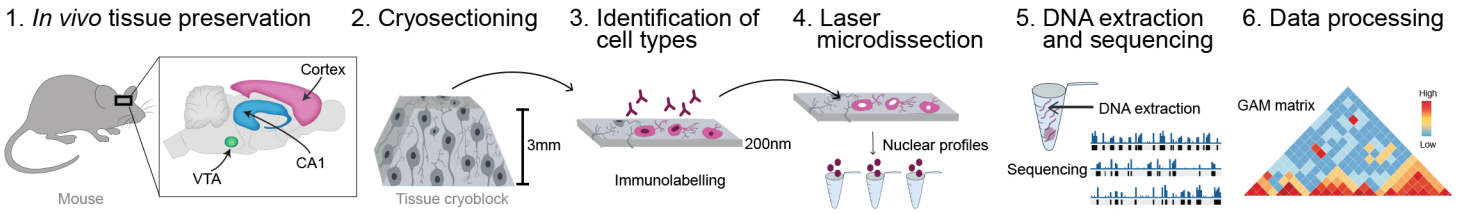
88. Kremer, K. & Grest, G. S. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *J. Chem. Phys.* **92**, 5057–5086 (1990)

89. Annunziatella, C. *et al. Molecular Dynamics simulations of the Strings and Binders Switch model of chromatin*. *Methods* **142**, (2018).

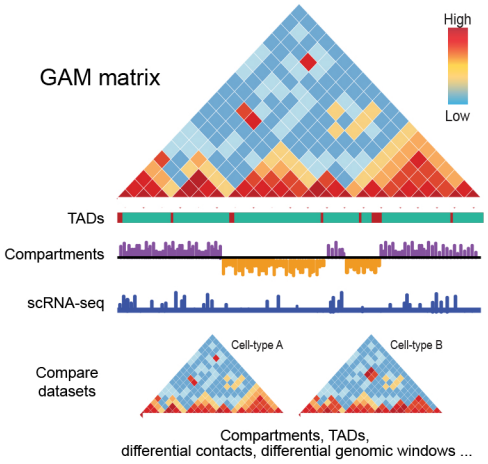
90. Fiorillo L., *et al.* Comparison of the Hi-C, GAM and SPRITE methods by use of polymer models of chromatin. In preparation.

Winick-Ng, Figure 1.

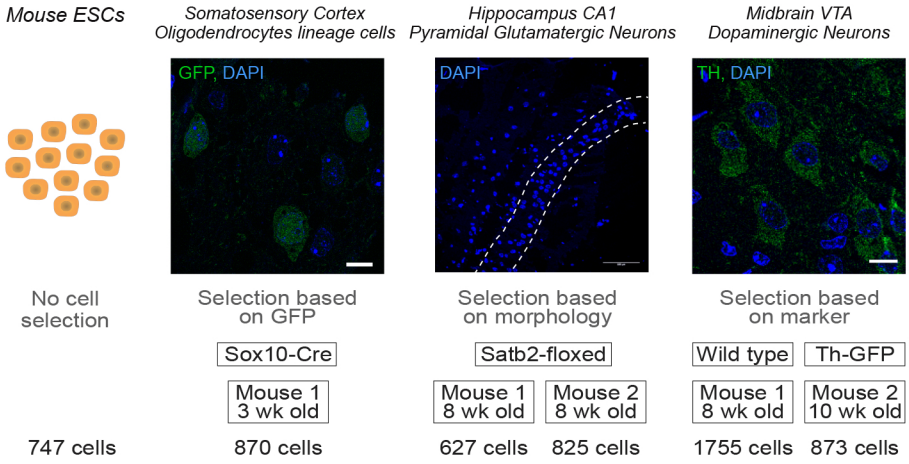
a *ImmunoGAM experimental workflow*



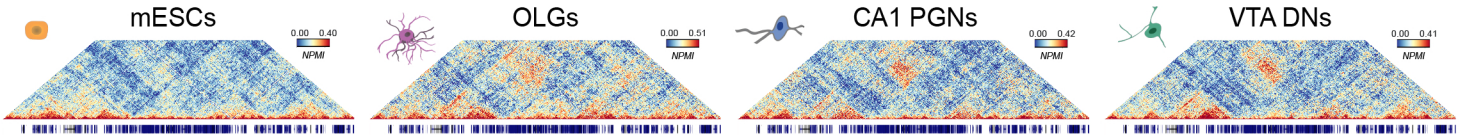
b *Data analysis summary*



c *GAM in brain summary*

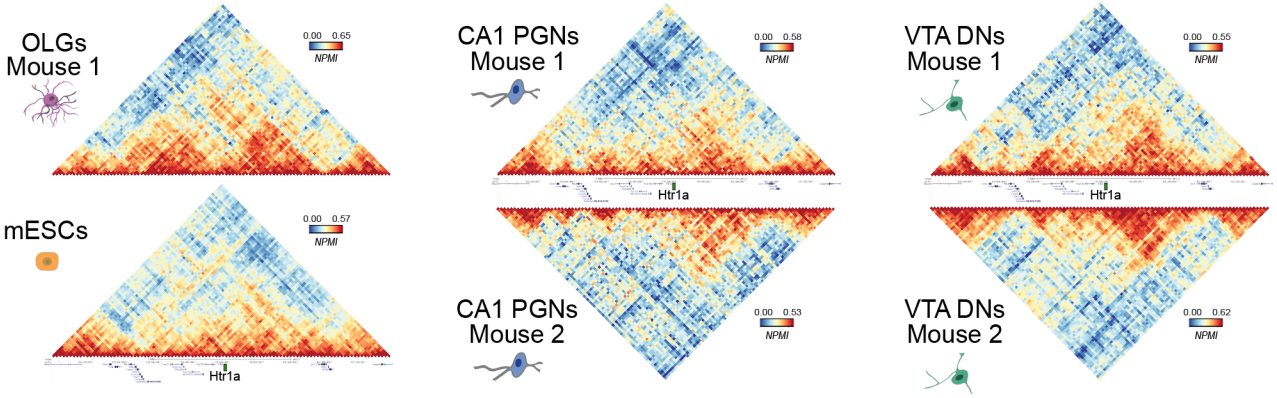


d *Examples of GAM matrices long-range contacts*



Chr17: 0-60Mb

e *Examples of GAM matrices short-range contacts*



Chr13: 102.9-107.9Mb

Figure 1. GAM captures cell-type specific chromatin contacts in the mouse brain.

a, Schematic representation of the immunoGAM experimental workflow in brain tissues. Upon transcardial perfusion with fixative (1), tissues are dissected, cryoprotected, and ultrathin cryosectioned (2). After immunodetection using cell-type specific markers (3), single nuclear profiles are dissected from independent cells and collected in groups of three (4). After extraction of genomic DNA and sequencing (5), the frequency of co-segregation of pairs of genomic loci depends on their true physical distance inside the nucleus, and is used to map chromatin contacts (6).

b, Schematic representation of immunoGAM data analysis of TADs, compartments A/B and contact matrices, which were developed in the present study, together with integration with cell-type specific gene expression extracted from scRNA-seq data from the same cell types.

c, Cell types used for the present study included cultured mouse ESCs (mESCs), oligodendrocytes (OLGs) from the somatosensory cortex (scale bar: 10 μm), pyramidal glutamatergic neurons (PGNs) from the CA1 region of the hippocampus (scale bar: 100 μm), dopaminergic neurons (DNs) from the midbrain ventral tegmental area (scale bar: 10 μm). mESCs were not selected, OLGs were selected based on GFP report expression, PGNs on tissue morphology (dotted line), and DNs on expression of TH.

d, Observed GAM contact matrices at 50-kb resolution (Chr17: 0-60,000,000) for each cell-type used in this study. A long-range patch of contacts involving two large genomic regions separated by ~35 Mb was observed for *in-situ* brain cell-types, but not for mESCs. GAM matrices represent co-segregation frequencies of pairs of genomic windows using Normalized Point Mutual Information (NPMI; see Methods).

e, GAMcontact matrices at 50-kb resolution show cell-type specific differences in local contacts (Chr13: 102,900,000-107,900,000), and reproducible between biological replicates taken from single animals.

Winick-Ng, Figure 2

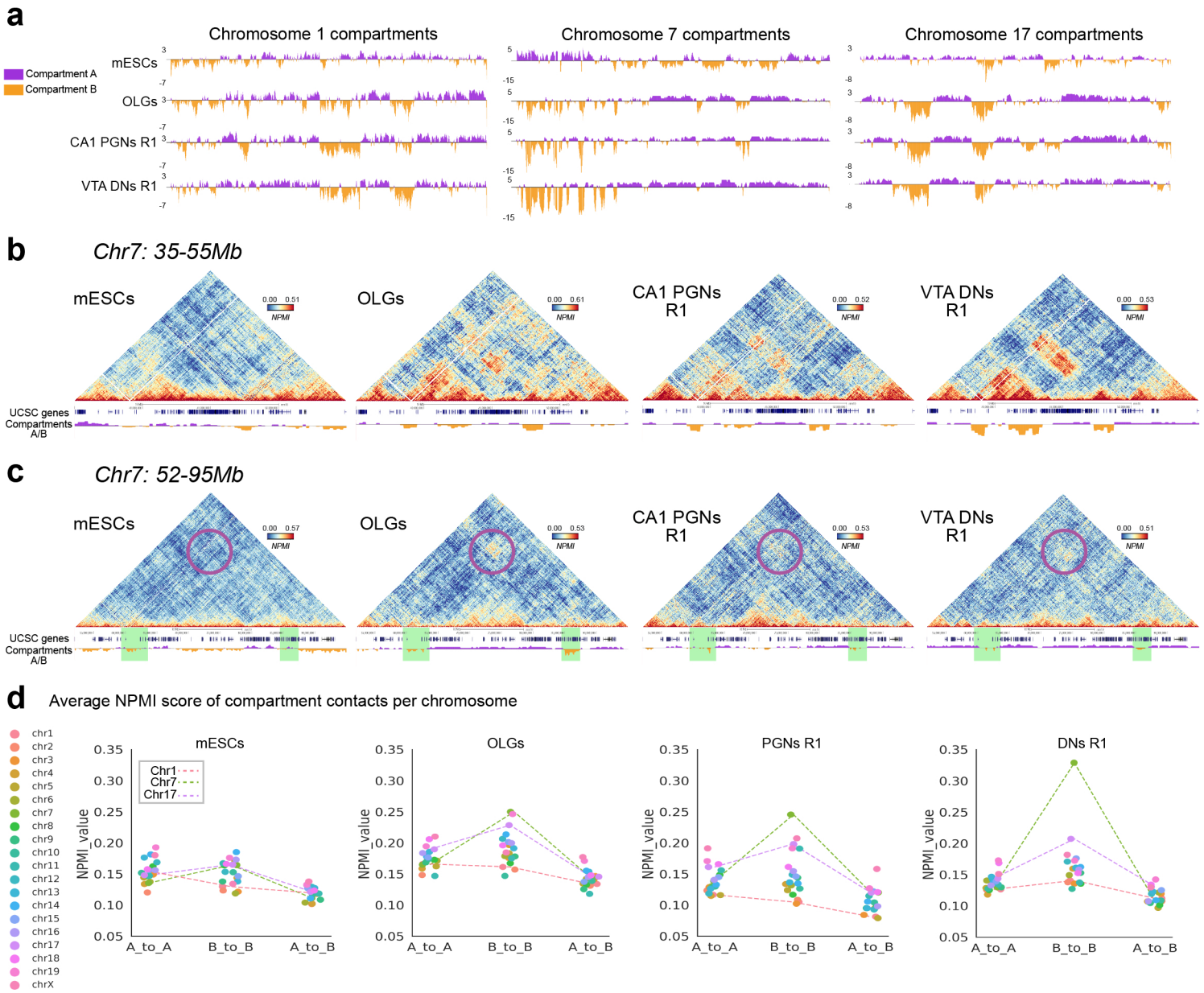


Figure 2. Mouse brain cells have strong B compartment interactions between clusters of sensory-receptor genes

a, Open and closed chromatin compartments (A and B, respectively) had different genomic distributions in mESCs, OLGs, PGNs and DNs. PCA eigenvectors were computed from GAM datasets from mESCs and for brain cell types, and used to classify compartments shown here for chromosomes 1, 7 and 17. R1; Mouse replicate 1. Purple, compartments A; orange, compartments B.

b, GAM contact matrices at 50-kb resolution for a 20Mb (Chr7: 35,000,000-55,000,000) region that shows strong interactions separated by 10 Mb that involve B compartments in OLGs, PGNs and DNs, but not mESCs.

c, GAM contact matrices at 50-kb resolution for a 43Mb region (Chr7: 52,000,000-95,000,000), showing strong interactions between B compartment regions in OLGs, PGNs and DNs (purple circle), but not mESCs, at a 30Mb distance.

d, Average NPMI score for compartment interactions in each chromosome. Whereas A-A and B-B interactions have similar average intensity per chromosome in mESCs, brain cells show increased B-B interactions especially for chromosomes 7 (green dashed line) and 17 (purple dashed line). A-B compartment interactions were detected, but proportionally less abundant, for all chromosomes in all cell types, as expected.

Winick-Ng, Figure 3

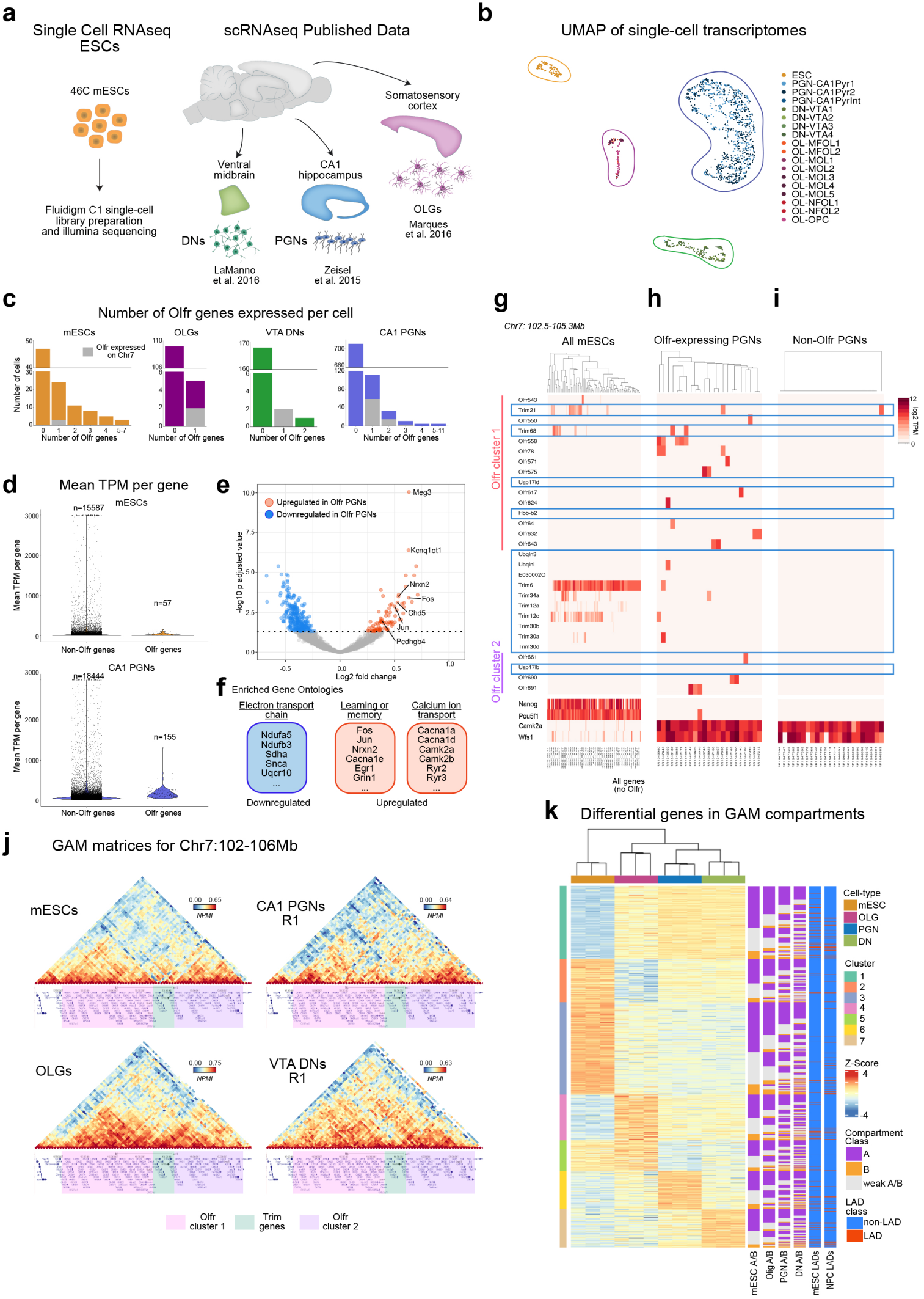


Figure 3. Single-cell gene expression of brain cell types is related to compartment definition.

a, Schematic representation of scRNA-seq datasets used in this study. We generated 98 single-cell transcriptomes from mESCs, and in parallel processed published scRNA-seq datasets from specific cell types in brain regions that were matched with GAM data.

b, Single-cell transcriptomes, visualized by UMAP clustering and colored by cell types that were defined by published data. Contours are drawn to contain each cell type.

c, Number of single cells expressing *Olfr* genes for each cell type. 52% of mESCs and 19% of PGNs expressed mostly one or two, but up to 7 or 11 *Olfr* genes, in single cells. Vast majority of OLGs and DNs did not express *Olfr* genes, with only five and two cells, respectively, that expressed one or two *Olfr* genes.

d, Mean TPM per gene for all gene (not *Olfr*) or *Olfr* genes in mESCs (top) and PGNs (bottom). *Olfr* genes expressed in PGNs had higher expression in single cells, than in mESCs.

e, Volcano plot of differentially expressed (DE) genes between Olfr-expressing and non-expressing PGNs. Genes with an adjusted p-value < 0.05 were considered DE.

f, Selected enriched gene ontologies (GOs) for DE genes in *Olfr*-expressing PGNs. GO terms of upregulated genes were enriched for neuronal activation, while downregulation with mitochondria metabolism.

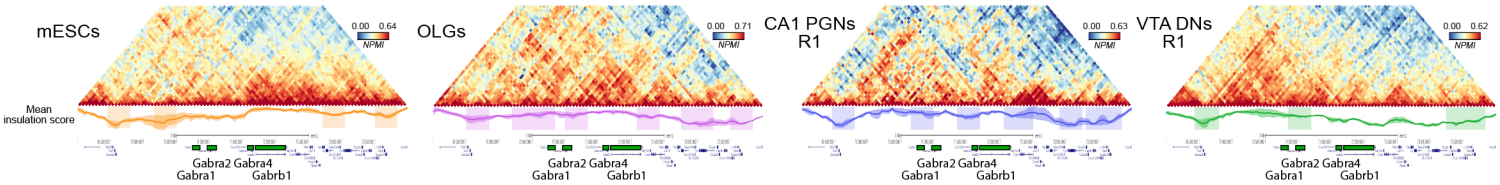
g-i, Heatmap of single-cell expression for genes within a *Olfr*- and *Trim*-rich region in chromosome 7 (Chr7: 102,476,700-105,337,714) that are expressed in at least one PGN or mESC. In mESCs, genes within or between the two *Olfr* clusters (green boxes) were expressed irrespective of *Olfr* expression (**g**). In some PGNs that express *Olfr* genes, genes within or between the two *Olfr* gene clusters were co-expressed (**h**). PGNs that do not express *Olfr* genes have no expression of genes within or between the two *Olfr* gene clusters (**i)**; exemplar cells were 23 randomly selected PGNs.

j, GAM contact matrices at 50-kb resolution for Chr7: 102,000,000-106,000,000 containing the two *Olfr* clusters and an intervening region with *Trim* genes discussed in (i). Expression of *Trim* genes in mESCs coincides with a smaller local hub of contacts, whereas in PGNs the locus is more globally open than in OLGs or DNs.

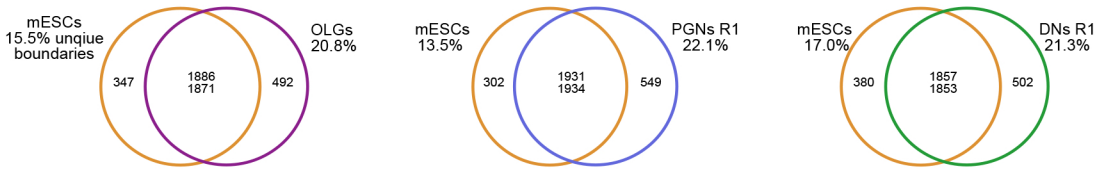
k, Heatmap of differentially expressed (DE) genes, clustered by cell type. Compartments are shown classified into terciles- strong A, strong B or an intermediate weak A/B. LADs are shown for mESCs and NPCs[38]. Most DE genes were found in A compartment regardless of cluster or cell type.

Winick-Ng, Figure 4

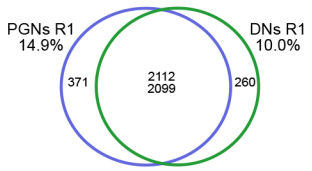
a Chr5: 69-74Mb



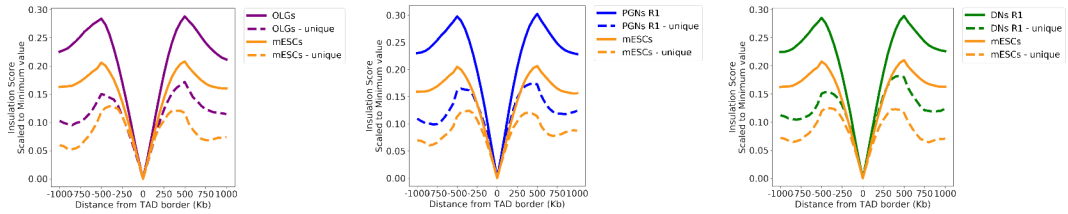
b Pairwise TAD border overlaps between cell-types



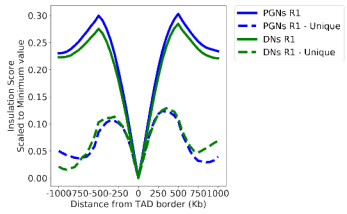
c



d Insulation score profile surrounding TAD borders



e



f

Selected Enriched Gene Ontologies	(p-val)	Genes (examples)
Retinoid binding	(0.002)	Aldh1a2, Lcn12, Rbp3, Rbp4
Growth factor activity	(0.004)	Bmp7, Fgf18, Gdf2, Nrg3
Regulation of synaptic vesicle transport	(0.001)	Grik5, Lrrk2, Nlgn1, Stxbp1
Regulation of NMDA receptor activity	(0.002)	Mef2c, Nlgn3, Oprm1, Ptk2b

Selected Enriched Gene Ontologies	(p-val)	Genes (examples)
Hormone activity	(0)	Edn3, Gip Prl*
Protein tyrosine phosphate activity	(0.003)	Dusp3, Ptpn11, Ptprb, Ptprg
Phospholipase A2 activity	(0.001)	Pla2g10, Pla2g2a, Pla2g7
Ephrin receptor binding	(0)	Efna1, Efna3, Grb2, Ntrik1

Selected Enriched Gene Ontologies	(p-val)	Genes (examples)
Insulin-like growth factor binding	(0.001)	Crim1, Esm1, Igfbp1, Igfbp3
Nucleosome organization	(0)	Arid1a, Hist1*
Circadian regulation of gene expression	(0.002)	Ciart, Cry1, Per2 Per3, Rora
Regulation of ERBB signaling	(0.001)	Cdh13, Gper1, Pde6h, Sos1

Selected Enriched Gene Ontologies	(p-val)	Genes (examples)
GPCR binding	(0.001)	Pde4b, Ppp1r1b, Wnt2, Wnt11
Sequence-specific DNA binding	(0.003)	Bcl6, Foxj1, Neurod6, Med1
Neurotransmitter transporter activity	(0)	Slc6a1, Slc6a18, Slc6a9
Glutamate receptor activity	(0)	Gria1, Grid1, Grin2b, Park2

Figure 4. Cell-type specific topological domains in the brain contain genes relevant for cell specialization.

a, GAM contact matrices at 50-kb resolution showing cell-type differences in local contacts and topological domains for a 5Mb region containing a GABA receptor cluster (Chr5: 69,000,000-74,000,000). Mean insulation square is plotted below matrices for each cell type and replicate, and shaded regions indicate called TAD borders.

b, Overlap between TAD boundaries for pairwise cell-type comparisons. OLGs, CA1 PGNs, and VTA DNs had ~500 unique TAD borders when compared to mESCs.

c, Overlap between TAD boundaries for DNs and PGNs. The neurons had higher TAD border conservation, with only 260 and 371 unique borders for DNs and PGNs, respectively.

d, Average insulation score ± 1Mb surrounding the lowest insulation point within a TAD border, for pairwise cell-type comparisons. For both shared and unique cell borders, all brain cell types had stronger TAD boundaries compared to mESCs.

e, Average insulation score surrounding TAD borders for PGNs and DNs. The two neuronal cell types had similar border strengths for both common and unique borders.

f, Selected gene ontologies (GOs) for genes found at unique TAD borders for pairwise cell-type comparisons. For all brain cell types, enriched GOs were relevant for cell specialization.

Winick-Ng, Figure 5

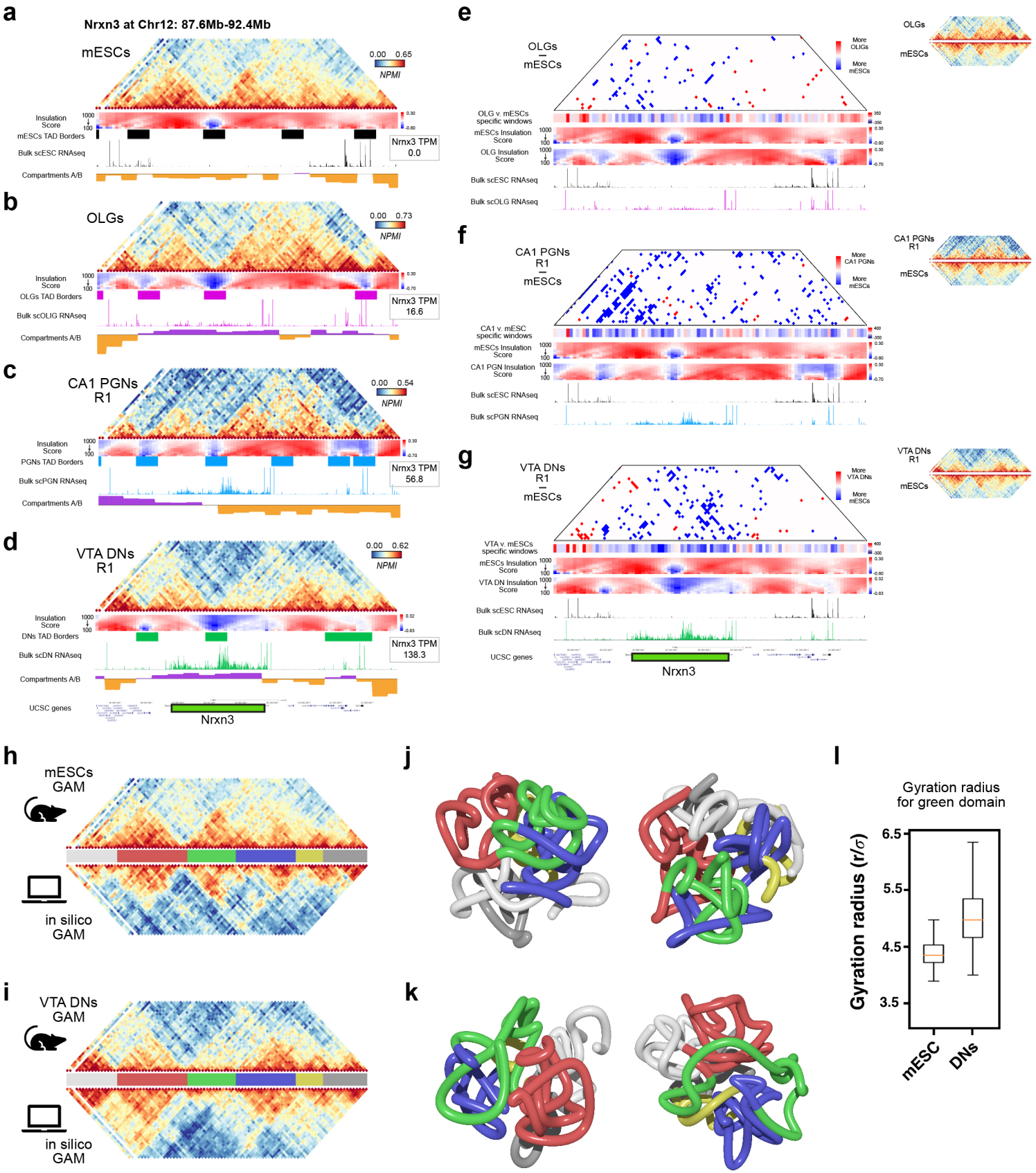


Figure 5. Extensive remodeling of chromatin structure at long and highly expressed neuronal genes.

a, GAM contact matrices at 50-kb resolution showing a 5Mb region (Chr12: 87,600,000-92,400,000) surrounding the *Nrxn3* gene. In mESCs, the gene is not expressed and is found within two TADs with high contact density as indicated by the insulation score heatmap. The *Nrxn3* gene is not expressed in mESCs

b, For OLGs, which has a low level of *Nrxn3* expression, the contact density is lower in the first than in the second TAD. Both TADs are visible, and the second TAD has stronger inter-TAD contacts with the downstream region.

c, In PGNs, *Nrxn3* expression is increased (16.6 TPM) and shows a similar contact density to OLGs, with lower contact density in the first TAD compared to the second. Both TADs are maintained.

d, In DNs, highest levels of *Nrxn3* expression (138.3) coincided with a loss of contact density and a loss of domain structure in the second TAD.

e-g, Differential contact matrices comparing different cell types for the *Nrxn3* region as shown in (**a**). OLGs and mESCs had only a few differential contacts, which were mostly found in the first Nrxn3 TAD (**e**). Genomic windows contained more mESC significant contacts as compared to OLGs across *the Nrxn3* gene body. For PGNs, contacts were significantly depleted in the first Nrxn3 TAD as compared to mESCs, and genomic windows were also depleted for PGN specific contacts within the first TAD (**f**). For DNs, contacts were significantly depleted in the second *Nrxn3* TAD, coinciding with the increase in Nrxn3 expression (**g**). Genomic windows were also significantly depleted for DN contacts across the entire *Nrxn3* gene body.

h, *In-silico* GAM performed from PRISMR modelling of the *Nrxn3* locus in mESCs based on experimental GAM data. The in-silico and experimental GAM matrices were well correlated (Pearson r = 0.72). Colour bars indicate domains as defined in the DNs.

i, *In-silico* GAM of the *Nrxn3* locus in DNs. The in-silico matrices were well correlated (Pearson r = 0.79) with the experimental GAM data. Importantly, TAD melting could be seen in the green domain of the in-silico matrix, corresponding to the second *Nrxn3* TAD.

j, Two examples of polymer models for the *Nrxn3* locus in mESCs shows the locus is organized as a globular structure. The white and red domains interact closely, while the green, blue and yellow domains interact with one another.

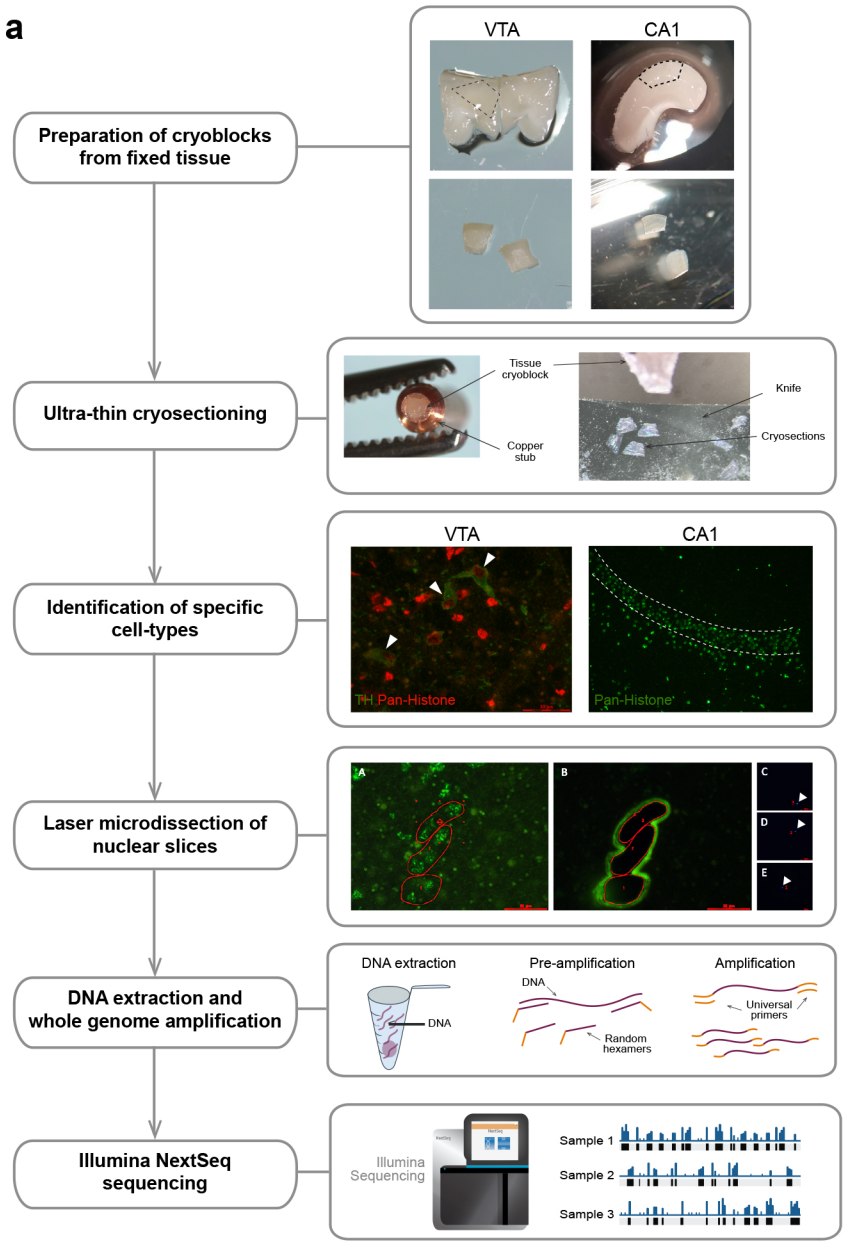
k, Two examples of polymer models for the *Nrxn3* locus in DNs. Note the extended green domain which loops away from the structure, corresponding to the *Nrxn3* melting TAD.

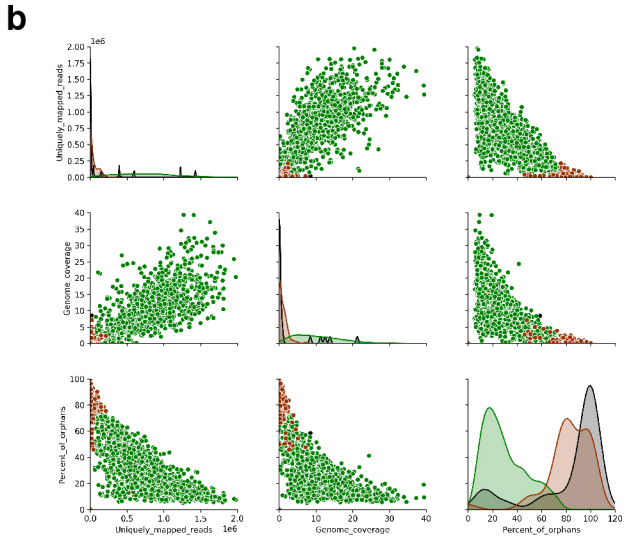
l, Distribution of gyration radii of the green domain in polymer models for mESCs and DNs, shows that the green domain, corresponding to the second *Nrxn3* TAD, has a range of higher gyration radii in DNs than mESCs, helping to visualise the decondensed, melted state of the region in DNs.

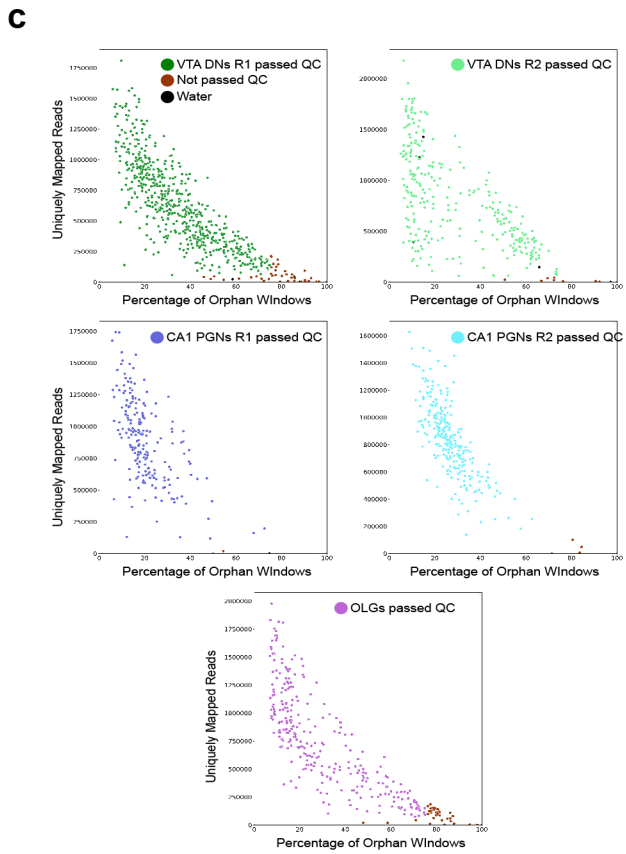
Table 1. Summary of GAM datasets generated in this study.

Dataset ID	Total samples	Samples passing QC	OW > 70%	UMRs < 50000	Cross-Contaminated	Water
VTA_DNs_R1	656	**585**	58	11	4	13
VTA_DNs_R2	316	**291**	19	6	3	6
CA1_PGNs_R1	218	**209**	7	2	0	2
CA1_PGNs_R2	288	**275**	7	1	0	6
OLIGs	336	**290**	46	4	3	0

Winick-Ng, SI Figure1







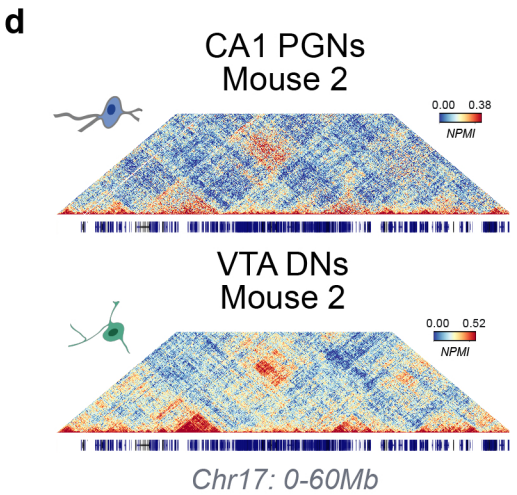


Figure S1. GAM method and quality controls.

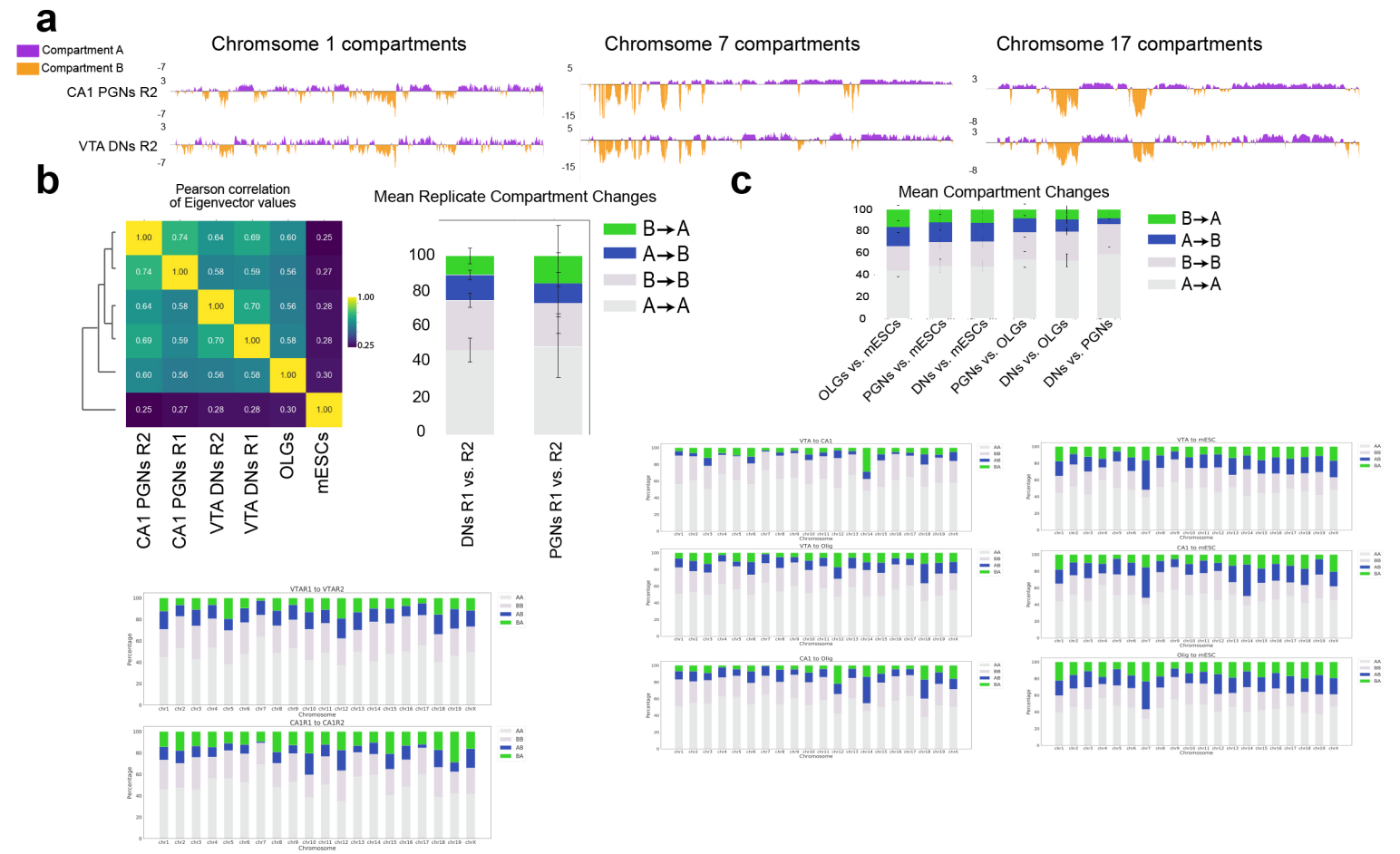
a, GAM experimental pipeline. VTA and CA1 dissections and cryoblock preparations are shown as examples. Tissue is dissected and cryopreserved before sectioning on an ultracryomicrotome. Pan-histone was used to identify cell slices that contained nuclei on the laser microdissection microscope, with additional TH labelling to identify DNs. PGNs were selected using the morphology of the pyramidal neuron layer. Three nuclear slices were selected and laser microdissected from the tissue to fall into the same PCR lid. DNA was then extracted from each sample and amplified using whole-genome amplification, followed by Illumina NextSeq sequencing.

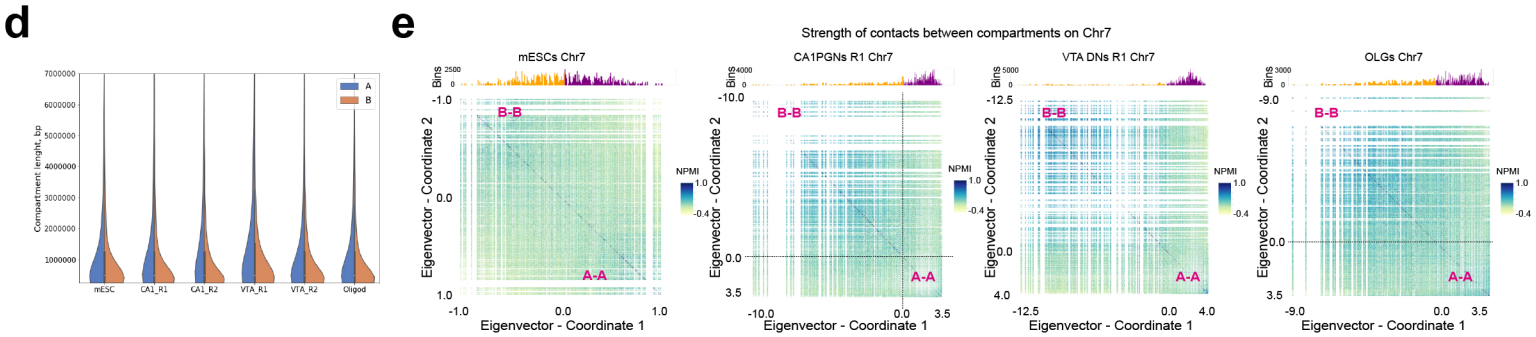
b, Quality control (QC) distributions for uniquely mapped reads, genome coverage, and percentage of orphan window for all samples in GAM brain cell types. Each data point represents a GAM sample. Samples passing QC are shown in green, samples not passing QC in red.

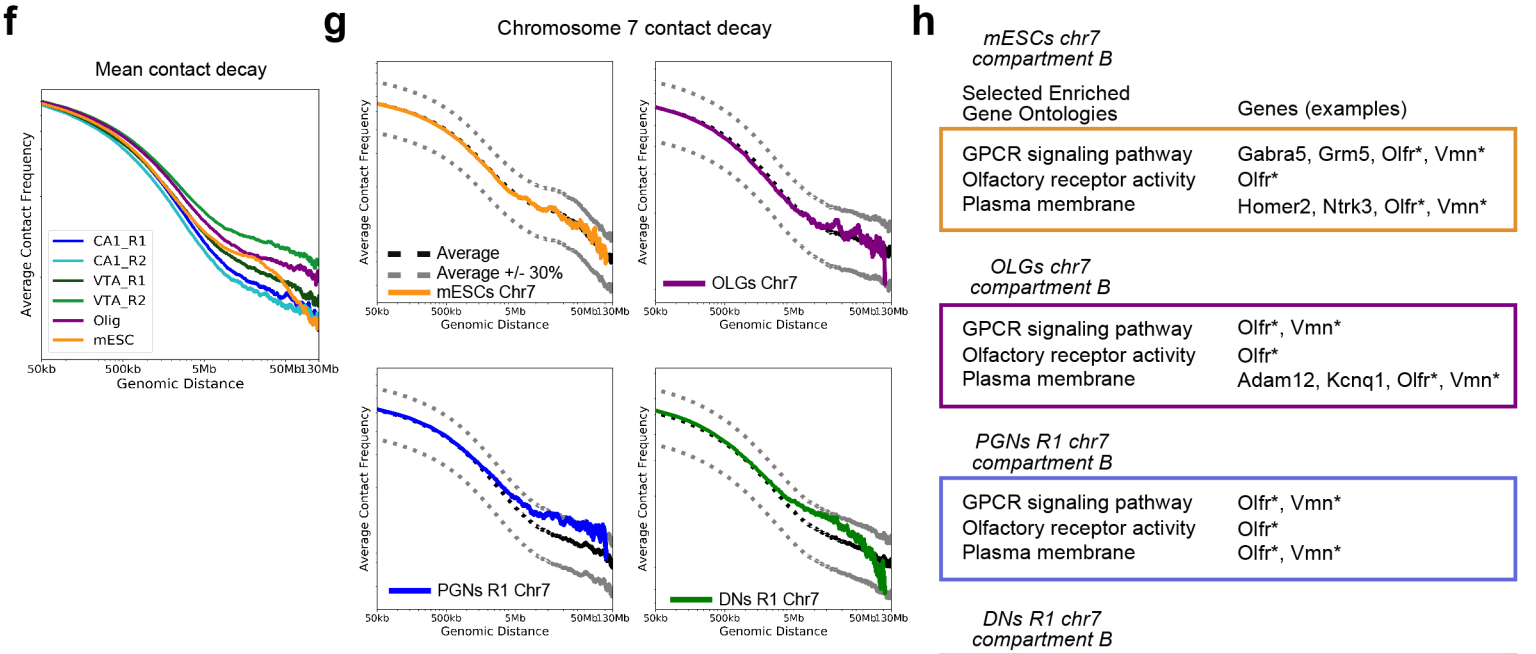
c, QC metrics of uniquely mapped reads versus orphan windows plotted for each GAM sample separately for each GAM brain cell type and replicate. Samples not passing QC are shown in red, water control samples (without a nuclear profile) are shown in black.

d, Example matrices for CA1 PGN and VTA DN replicate 2 for Chr17: 0-60,000,000 at 50-kb resolution.

Winick-Ng, SI figure2







h

mESCs chr7 compartment B

Selected Enriched Gene Ontologies	Genes (examples)
GPCR signaling pathway	Gabra5, Grm5, Olfr*, Vmn*
Olfactory receptor activity	Olfr*
Plasma membrane	Homer2, Ntrk3, Olfr*, Vmn*

OLGs chr7 compartment B

GPCR signaling pathway	Olfr*, Vmn*
Olfactory receptor activity	Olfr*
Plasma membrane	Adam12, Kcnq1, Olfr*, Vmn*

PGNs R1 chr7 compartment B

GPCR signaling pathway	Olfr*, Vmn*
Olfactory receptor activity	Olfr*
Plasma membrane	Olfr*, Vmn*

DNs R1 chr7 compartment B

GPCR signaling pathway	Olfr*, Vmn*
Olfactory receptor activity	Olfr*
Plasma membrane	Olfr*, Vmn*

Figure S2. Identification of compartments and differences between cell types.

a, Open and closed chromatin compartments (A and B, respectively) shown for biological replicates of PGNs and DNs biological replicates 2.

b, Correlation of compartments in GAM cell types and replicates. Pearson correlation of eigenvectors shows the largest differences between mESCs and brain cell types. Mean compartment changes, and comparisons by chromosome show good overlap between replicates.

c, Mean compartment changes for each cell type comparison. Only compartments common to both replicates were used in the comparison. Brain cell types had lower overlap with mESCs then to each other. PGNs and DNs had the most overlap.

d, Violin plots showing the distribution of compartment lengths for each cell type, show similar lengths between cell types.

e, Heatmaps of NPMI score for each genomic window pair on chromosome 7, sorted by eigenvectors. The histogram of eigenvectors is shown in the upper panel, coloured by compartment definition (orange, compartment B; purple, compartment A). Note the strong B compartment values in the brain cell types, and higher NPMI scores.

f, Average contact decay with genomic length for all chromosomes in logarithmic bins, for each cell-type. mESCs have a more rapid decay after 25 Mb, as compared to brain cell types.

g, Average contact decay for chromosome 7 across genomic distances in logarithmic bins, for each cell type. mESCs and OLGs have decay curves similar to the mean value for all chromosomes (black dashed line), while neurons have higher average contact frequencies at large genomic distances (> 5 Mb). Grey dashed lines indicate mean values ± 30%.

h, Selected enriched gene ontologies (GOs) from compartment B in chromosome 7. Enriched GOs were the same in all cell types, and contained multiple olfactory (*Olfr*) or vomeronasal (*Vmn*) receptor genes. *indicates multiple genes.

Winick-Ng, SI figure3

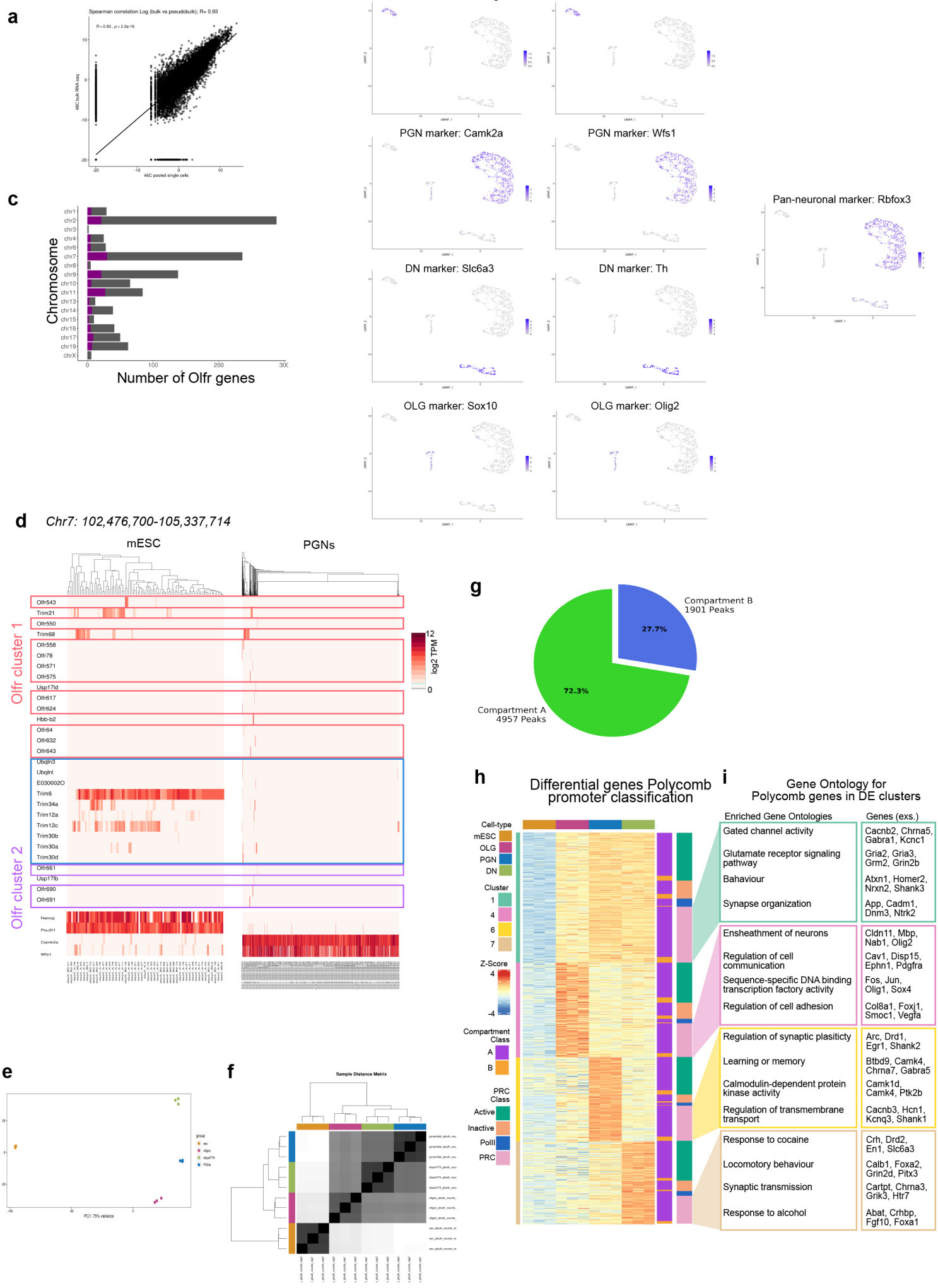


Figure S3. Integration of mESCs scRNA-seq with published brain cell-types.

a, Correlation plot of gene expression between published bulk[26] versus single-cell mESC. Average single cell expression was highly correlated with bulk RNA-seq.

b, Cell-type marker expression, represented by UMAP of single cell transcriptomes. *Nanog* and *Pou5f1* were used as markers for mESCs, *Camk2a* and *Wfs1* for PGNs, *SLc6a3* and *Th* for DNs, and *Sox10* and *Olig2* for OLGs. *Rbfox3* was used as a pan-neuronal marker. All markers were highly expressed in their respective cell types.

c, Histogram showing the distribution of *Olfr* genes across the genome (grey bars). *Olfrs* expressed in PGNs are shown in purple.

d, Heatmap of single-cell gene expression for genes that were expressed in at least one PGN or mESC, and were within or between two *Olfr* gene clusters (Chr7: 102,476,700-105,337,714). In mESCs, genes within or between the two *Olfr* clusters (green boxes) were expressed irrespective of *Olfr* expression. In some PGNs that expressed *Olfrs*, genes within or between the two *Olfr* clusters were co-expressed. PGNs that do not express *Olfrs* have no expression of genes within or between the two *Olfr* clusters.

e, PCA of cell types split into three replicates for differential expressionanalysis (see **Methods**). Replicates clustered separately by cell type.

f, Correlation plot of cell type and replicates for differential gene expression analysis. Replicates correlate most highly with one another, followed by brain cell types.

g, Proportion of H3K27me3 ChIP-seq peaks[26] found within compartments A/B in mESCs. 72.3% of H3K27me3 peaks were in compartment A.

h, DE genes were classified by their promoter state of activation for mESCs[25] in clusters 1, 4, 6 and 7. Most DE genes were found in compartment A, and a large subset of genes were classified as Polycomb regulated in mESCs.

i, Selected enriched gene ontologies (GOs) for Polycomb-regulated genes in DE clusters from (i). All selected GOs had a permuted p-value of 0. Enriched GOs for each brain cell type was specific for their specialized function.

Winick-Ng, SI figure4

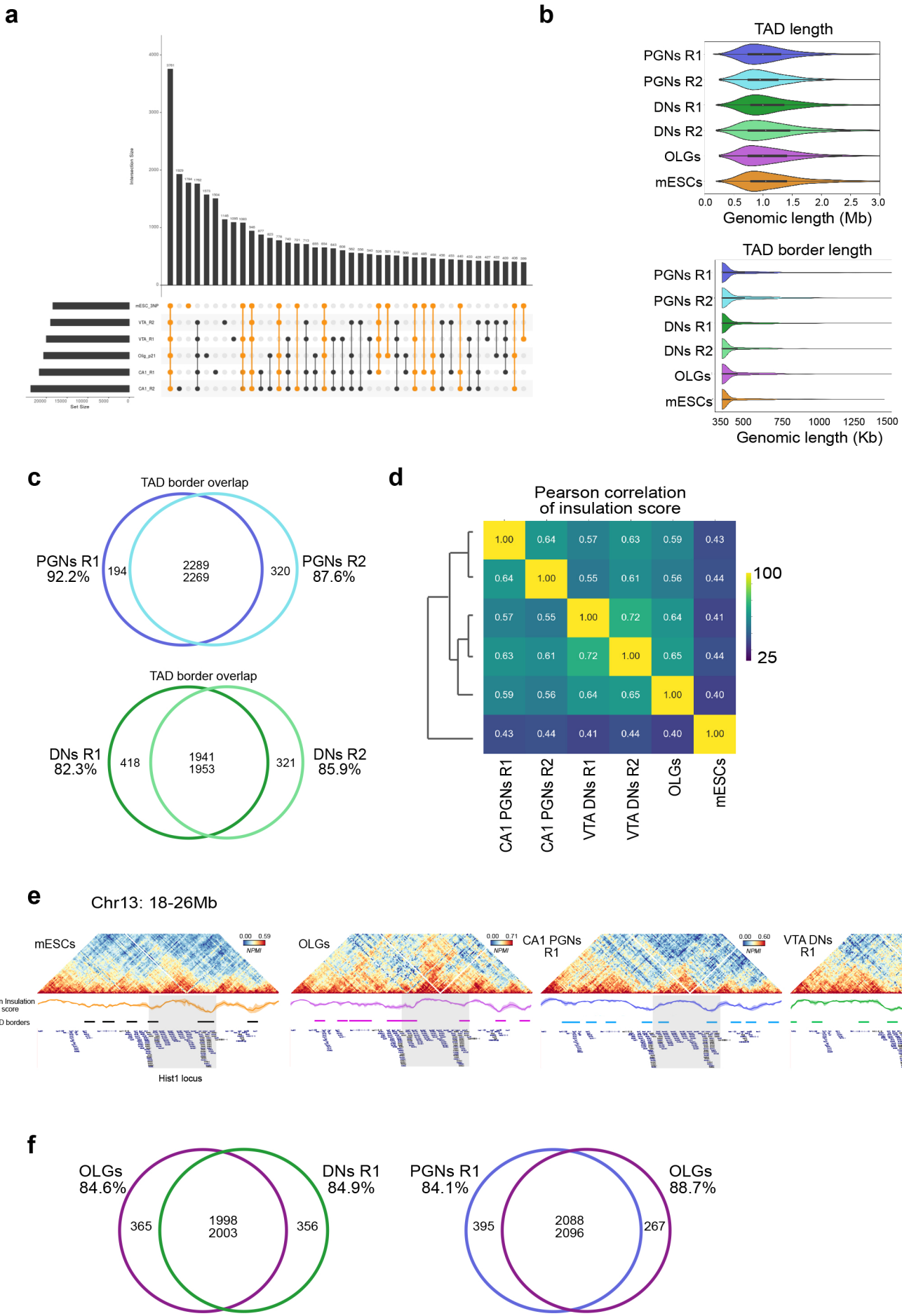


Figure S4. Identification of TADs and TAD boundaries, and differences between cell-types.

a, Upset plot showing TAD boundary overlap for 50-kb genomic bins with at least one cell type containing a TAD border. Sets that overlapped, or were unique for, mESCs are coloured in orange.

b, Violin plots showing the distribution of TAD and TAD boundary lengths. TAD length was calculated as the distance between two boundaries points (defined as lowest insulation score point within a boundary).

c, Overlap between TAD boundaries for pairwise PGN or DN replicate comparisons.

d, Pearson correlation of insulation scores in GAM cell types and replicates. The largest differences were found between mESCs and brain cell types. DN and PGN replicates had the highest correlations.

e, GAM contact matrices at 50-kb resolution showing an 8Mb region (Chr13: 18,000,000-26,000,000) surrounding the *Hist1* locus. Mean insulation scores are plotted for each cell type and replicate. The *Hist1* locus is indicated by the shaded grey area.

f, Overlap between TAD boundaries for pairwise cell-type comparisons between PGNs or DNs and OLGs.

Winick-Ng, SI figure5

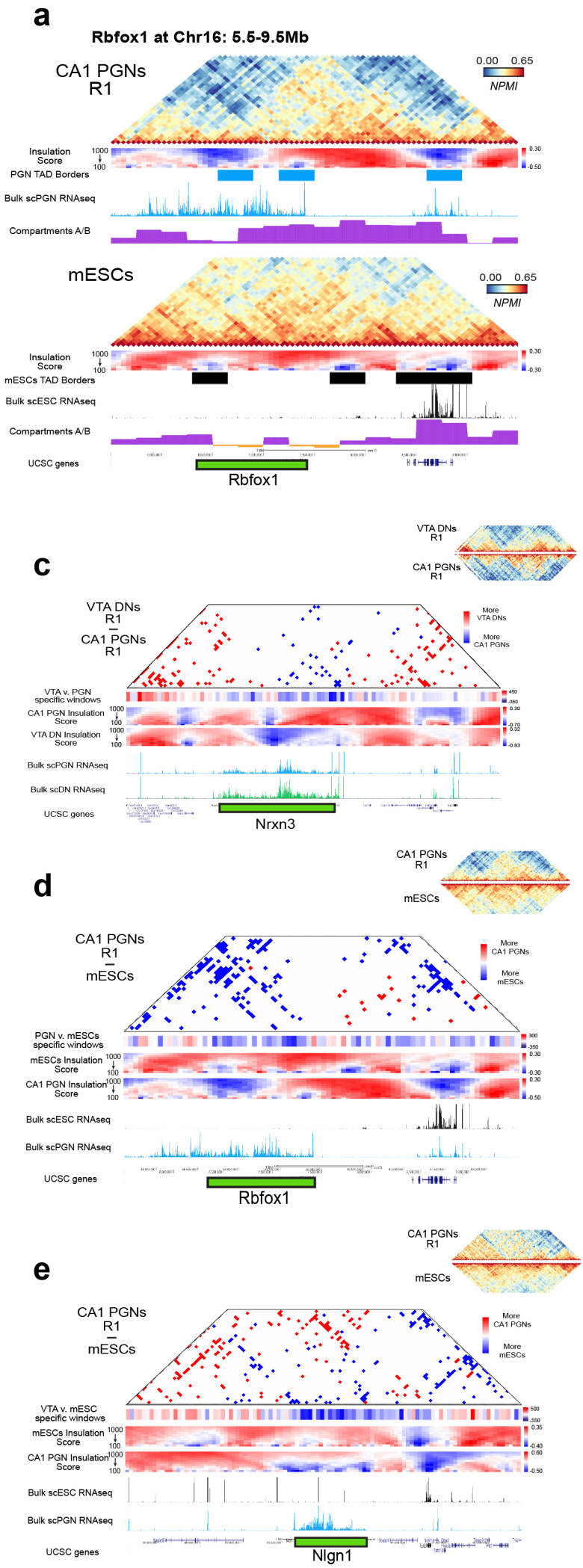


Figure S5. TAD melting at Rbfox1 and Nlgn1.

a, GAM contact matrices at 50-kb resolution showing a 4Mb region (Chr16: 5,500,000-9,500,000) surrounding the *Rbfox1* gene. In mESCs, the gene is not expressed and was found within a TAD with high contact density as indicated by the insulation score heatmap. In PGNs, high expression of *Rbfox1* coincided with a loss of contact density and a loss of domain structure in the TAD.

b, GAM contact matrices at 50-kb resolution showing a 5Mb region (Chr3: 23,300,000-28,300,000) surrounding the *Nlgn1* gene. In PGNs, high expression of *Ngln1* also coincided with a loss of contact density and a loss of domain structure.

c, Differential contact matrices for PGNs and DNs at the *Nrxn3* locus. Significant DN contacts were found outside the *Nrxn3* region, and there were few significant contacts in the first TAD. The second TAD had some PGN specific contacts. Genomic windows were depleted for DN significant contacts in the second TAD

d, Differential contact matrices for mESCs and PGNs at the *Rbfox1* locus, showing significantly different mESC contacts at the *Rbfox1* promoter, contacting the rest of the gene and a downstream region. Genomic windows contained more mESC significant contacts as compared to PGNs across the *Rbfox1* gene body.

e, Differential contact matrices for *Nlgn1*. Note the differential mESC contacts within the *Nlgn1* gene. Genomic windows contained more mESC significant contacts as compared to PGNs across the entire *Nlgn1* gene body, and downstream region.