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,&], Dominik Szabo[1,2,&], Mandy Meijer[3], Leonid Serebreni[1,a], Yingnan Zhang[4], Simona Bianco[5], Andrea M. Chiariello[5], Ibai Irastorza-Azcarate[1], Christoph J. Thieme[1], Thomas M. Sparks[1,2], Sílvia Carvalho[1,6,7,8], Luca Fiorillo[5], Francesco Musella[5], Ehsan Irani[1,9], Elena Torlai Triglia[1,b], Aleksandra A. Kolodziejczyk[10,11,c], Andreas Abentung[12,d], Galina Apostolova[12], Eleanor J. Paul[13,e], Vedran Franke[14], Rieke Kempfer[1,2], Altuna Akalin[14], Sarah A. Teichmann[10,11], Georg Dechant[12], Mark A. Ungless[13], Mario Nicodemi[5,9], Lonnie Welch[4], Gonçalo Castelo-Branco[3,15], Ana Pombo[1,2,9,*]

[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]School of Electrical Engineering and Computer Science, Ohio University, 45701 Athens, OH, USA
[5]Dipartimentio di Fisica, Università di Napoli Federico II, and INFN Napoli, Complesso Universitario di Monte Sant'Angelo, 80126 Naples, Italy
[6]UCIBIO, Department of Life Sciences, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2825-149 Caparica, Portugal
[7]Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, 4050-313 Porto, Portugal
[8]Graduate Program in Areas of Basic and Applied Biology, Universidade do Porto, 4050-313 Porto, Portugal
[9]Berlin Institute of Health, 10178 Berlin, Germany
[10]Cavendish Laboratory, University of Cambridge, JJ Thomson Ave, Cambridge CB3 0EH, UK
[11]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK
[12]Institute for Neuroscience, Medical University of Innsbruck, 6020 Innsbruck, Austria
[13]Institute of Clinical Sciences, Imperial College London, London SW7 2AZ, UK
[14]Bioinformatics and Omics Data Science Platform, Berlin Institute for Medical Systems Biology, Max-Delbrück Centre for Molecular Medicine, 10115 Berlin, Germany
[15]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: Center for Developmental Neurobiology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London WC2R 2LS, UK;

MRC Center for Neurodevelopmental Disorders, King's College London, London SE1 1UL, UK

[#]These authors contributed equally to this work
[&]These authors contributed equally to this work

*Joint corresponding authors: ana.pombo@mdc-berlin.de; warren.winick-ng@mdc-berlin.de

Abstract

Neurons and oligodendrocytes are terminally differentiated cells that sustain cascades of gene activation and repression to execute highly specialized functions, while retaining homeostatic control. To study long-range chromatin folding without disturbing the native tissue environment, we developed Genome Architecture Mapping in combination with immunoselection (immunoGAM), and applied it in the adult murine brain to midbrain dopaminergic neurons (DNs), hippocampal pyramidal glutamatergic neurons (PGNs), and cortical oligodendroglia (OLGs). We find specialized 3D chromatin structures that relate with cell-type specific patterns of gene expression at multiple genomic scales. We discover extensive decondensation, or 'domain melting', at long genes (> 400kb) when they are most highly expressed and/or with high chromatin accessibility. In neurons, melting of chromatin domains most often coincides with preferential intra-chromosomal contacts. By mining the most differential contacts between specific neuron subtypes, we find genes associated with specialized functions such as addiction, synaptic plasticity and memory, which are connected by the presence of predicted binding sites for multiple neuronal transcription factors within putative regulatory regions. Lastly, sensory receptor genes are preferentially found in heterochromatic compartments in brain cells, where they establish strong contacts across tens of megabases. Overall, we uncover highly specific chromatin conformations in terminally differentiated brain cells which are tightly related with mechanisms of gene regulation and specialized cell functions.

25 ***Introduction***

26 The study of three-dimensional (3D) chromatin organization has revealed its
27 intrinsic association with gene regulation and cell function[1]. Genome-wide
28 sequencing approaches such as Hi-C[2], GAM[3] and SPRITE[4] have shown that the
29 genome is organized hierarchically, from specific promoter-enhancer contacts, to
30 topologically associating domains (TADs), to broader compartments of open and
31 closed chromatin (A/B compartments, respectively), up to whole chromosome
32 territories[2-5]. Dynamic changes of chromatin organization during neuronal
33 development have been reported using Hi-C from in-vitro differentiated neurons, from
34 glutamatergic neurons dissociated from cortical tissue from pools of animals and
35 FACS isolated, or from dissociated mixed cell populations from whole hippocampi[6-
36 8].However, it has been difficult to assess the genome-wide chromatin structure of
37 specific cell types from the brain, especially without disrupting tissue organization,
38 and in single animals. Furthermore, large-scale changes in chromatin organisation
39 captured by imaging approaches, such as the relocation of *Bdnf* from the nuclear
40 periphery to the nucleoplasm interior[9], are invisible in 3C-based approaches[10].
41 Understanding 3D chromatin structure of specialized cell populations is of
42 exceptional importance in the brain, where connecting disease-associated genetic
43 variants in non-coding genomic regions with their target genes remains challenging,
44 and where cell state and localization within the tissue can influence both
45 transcriptional and physiological outcomes.

46

47 ***ImmunoGAM reveals features of 3D genome architecture in specific mouse***
48 ***brain cells***

49 Here we developed immunoGAM, an extension of the Genome Architecture
50 Mapping (GAM) technology[3] to perform genome-wide mapping of chromatin topology

in specific cell populations selected by immunostaining. GAM is a ligation-free technology that maps 3D genome topology by extracting and sequencing the genomic DNA content from nuclear cryosections. 3D chromatin interactions are identified from the increased probability of co-segregation of genomic loci across a collection of nuclear slices, each from a single nucleus. GAM was previously applied to mouse embryonic stem cells (mESCs) and shown to capture TADs, A/B compartments and pair-wise contacts across long genomic distances[3,11]. With immunoGAM, we now considerably expand the scope of GAM to directly work on selected cell types present within a complex tissue, without the need for prior dissociation or flow sorting to isolate cell types of interest, both of which can alter cell physiology[12], and overcoming limitations of Hi-C and SPRITE technologies. ImmunoGAM retains the power of using very small cell numbers (~1000 cells)[3,13], essential for studying precious biological samples from single individuals.

To explore how genome folding is related with cell specialization, we applied immunoGAM in three specific mouse brain cell types with diverse functions (**Fig. 1a**). We selected (1) oligodendroglia (oligodendrocytes and their precursors; OLGs) from the somatosensory cortex, which are involved in myelination of neurons and influence neuronal circuit formation[14]; (2) pyramidal glutamatergic neurons (PGNs) from the *cornu ammonis 1* (CA1) region of the dorsal hippocampus, which have specialized roles in temporal and spatial memory formation and consolidation, and are activated intrinsically and frequently[15,16]; and (3) dopaminergic neurons (DNs) from the ventral tegmental area of the midbrain (VTA), which are activated during cue-guided reward-based learning[17]. Finally, we compared the immunoGAM data from brain cell types to publicly available GAM data from mESCs[11] (see **Methods, Supplemental Table 1**).

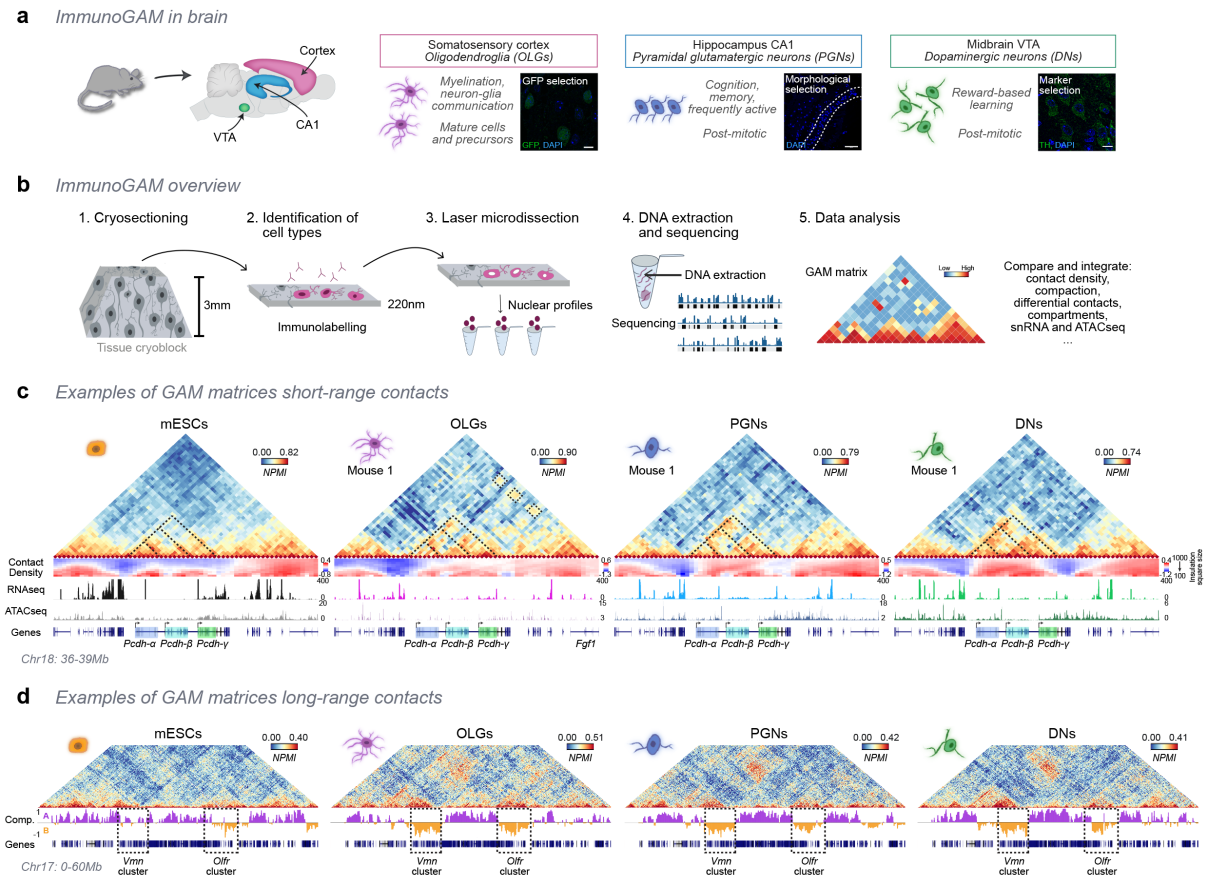

Figure 1. ImmunoGAM captures cell-type specific chromatin contacts in the mouse brain.

a, Immuno-GAM was applied to oligodendrocytes (OLGs), pyramidal glutamatergic neurons (PGNs), and dopaminergic neurons (DNs) (scale: 10, 100 and 10 µm, respectively). OLGs were selected based on GFP expression, PGNs using tissue morphology (dotted lines), and DNs by TH expression.

b, The immunoGAM experimental workflow. Upon transcardial perfusion with fixative, brain tissues are dissected, cryoprotected, and ultrathin cryosectioned (1). After immunodetection using cell-type specific markers (2), single nuclear profiles are laser microdissected (3), each from a single cell, and collected in groups of three, as described for multiplex-GAM[11]. After extraction of genomic DNA and sequencing (4), the frequency of co-segregation of pairs of genomic loci is used to calculate matrices of pairwise contacts, and to compare chromatin contact density, compaction, differential contacts and compartments (5). GAM data was also integrated with single-cell RNA- and ATAC-seq data.

c, Examples of contact matrices showing cell-type specific differences in local contacts for mESCs, OLGs, PGNs and DNs within a 3-Mb genomic region centered on the *Pcdh-α, -β,* and *-γ* regions. GAM matrices represent co-segregation frequencies of pairs of 50-kb genomic windows using normalized pointwise mutual information (NPMI). Dashed lines illustrate differences in contacts between *α*, *β* and *γ* regions for different cell types. NPMI scale bars range between 0 and the 99th percentile of NPMI values for each cell type. Contact density heatmaps represent insulation scores obtained with 100 – 1000kb square sizes. RNAseq tracks were obtained from scRNAseq the brain cell types of interest or mESCs, and converted into pseudobulk datasets. ATAC-seq tracks were obtained from scATAC-seq produced from brain tissues and converted to pseudobulk tracks for the cells of interest, or from bulk ATAC-seq from mESCs.

d, Examples of cell-type specific differences in long-range contacts within a 60-Mb genomic region in chromosome 17. Normalized PCA eigenvectors were computed from GAM

105 *(Fig. 1 legend, cont.)* datasets, and used to classify open and closed chromatin
106 compartments (A and B, respectively). Large groups of strong contacts between clusters of
107 vomeronasal (*Vmn*) and olfactory (*Olfr*) genes within B compartment regions, separated by
108 ~35 Mb, are observed in brain cells but not in mESCs.
109

110 In the immunoGAM experimental pipeline, tissues were collected from fixative-
111 perfused animals, cryoprotected by sucrose embedding, and frozen in liquid nitrogen
112 (**Fig. 1b**). After cryosectioning, the thin (~220 nm) tissue slices were immunostained
113 with specific antibodies to detect cell markers: TH for DNs, GFP following lineage
114 tracing for OLGs, and nuclear pan-histone to identify morphological layer of PGNs in
115 the CA1. We adopted the recently described multiplex-GAM collection approach[11],
116 where each GAM sample contains three nuclear slices each from a single cell in the
117 population and collected by laser microdissection. Next, the DNA content of each
118 sample was extracted, amplified and sequenced. Finally, after computing positive
119 genomic windows (50kb) in each dataset, matrices of pairwise contacts were
120 computed and normalized. A detailed flowchart of GAM data collection and quality
121 control measures are presented in **Extended Data Fig. 1a-c, and Supplemental**
122 **Table 2**. The final GAM datasets contained nuclear slices from ~850 cells (**Extended**
123 **Data Fig. 1d**). To generate GAM contact matrices, we corrected for statistical
124 fluctuations in window sampling typical of GAM data using normalized pointwise
125 mutual information (NPMI; **Extended Data Fig. 2a**). We found that NPMI more
126 robustly normalizes contacts involving slightly under- or over-detected window pairs
127 compared with the previous GAM normalization (D-prime)[3] (**Extended Data Fig. 2b**);
128 all GAM datasets showed low absolute biases in GC content, mappability and
129 window detection frequency (**Extended Data Fig. 2b,c**).
130

131 To explore differences in chromatin contacts between cell types, we calculated
132 contact densities and topological domains with the insulation square method[18],

133 previously shown to recapitulate Hi-C-like TAD borders in GAM data[3,11]
134 (**Supplemental Tables 3,4**), and compartment domains associated with open
135 (compartment A) and closed chromatin (compartment B) by performing principal
136 component analyses (PCA; **Supplemental Table 5**)[3]. We started by inspecting the
137 contacts of the protocadherin gene cluster known to have specific 3D topologies in
138 neuronal committed cells[19]. The *Pcdh* cluster is a 3-Mb region on chromosome 16,
139 containing three clusters of cell adhesion genes (α, β, and γ), with a unique
140 conserved promoter which can be bound by CTCF in mESCs[20] (see **Extended Data**
141 **Fig. 2d**). In mESCs, where only the γ locus is lowly expressed, the entire *Pcdh* locus
142 is contained in two TADs (**Fig. 1c**), as previously described[3,20]. Visualization of
143 contact densities across scales (100-1000kb insulation scores) also shows that the
144 *Pcdh* locus is generally open at scales above 500kb. We find that the increased
145 expression of *Pcdh-*α and -β in neurons[19] and OLGs[21] is accompanied by increased
146 long-range contacts between with the α, β and γ loci, separated by ~1Mb (**Fig. 1c,**
147 see **Extended Data Fig. 2d** for biological replicates). In OLGs, additional long-range
148 contacts are seen between all three *Pcdh* clusters and the downstream *Fgf1* gene,
149 which is more highly expressed in OLGs and has important roles in myelination[22], but
150 not with other intervening active genes. These observations suggest that cell-type
151 specific gene activity can coincide with the establishment of long-range chromatin
152 contacts between specific genes.
153
154 ImmunoGAM data also showed striking contacts spanning tens of megabases
155 between clusters of *Vmn* and *Olfr genes* exclusively in brain cell types, and not in
156 mESCs. An example is shown on chromosome 17 where a cluster of strong contacts
157 connect two large regions (~3 and ~5Mb wide) separated by 35 Mb is seen in the
158 three brain cells but not in mESCs (**Fig. 1d**), and reproduced in independent mice

159 (**Extended Data Fig. 2e**). The occurrence of Mb-scale clusters of contacts coincides
160 with a transition between compartment A-to-B of a region containing *Vmn* genes and
161 which associate with a neighbouring B compartment containing *Olfr* genes (**Fig. 1d)**.
162 Thus, application of immunoGAM directly to specific cell types in intact tissues
163 reveals large rearrangements of the 3D chromatin architecture, at both short- and
164 long-range genomic lengths. Next, we explored in more detail how genome-wide
165 changes in 3D genome topology at the level of domains, preferred contacts and
166 compartments, relate with cell-type specific patterns of gene expression.
167

168 ***Single-cell expression and chromatin accessibility maps for brain cells and***
169 ***mESCs***

170 To investigate how cell-type specific 3D genome topologies relate with gene
171 expression and chromatin accessibility, we produced and/or collected published
172 single-cell RNAseq and single-cell ATAC-seq from the cortex, hippocampus and
173 midbrain, and selected the cell populations equivalent to the cell types captured by
174 immuno-GAM (see **Methods, Supplemental Table 6**). We also generated scRNA-
175 seq and bulk ATAC-seq data from mESCs.
176

177 To obtain cell-type specific gene expression data, we analysed published
178 scRNA-seq datasets from the relevant tissues, from which we selected eight
179 subtypes of mature OLGs and OLG progenitors[23], three subtypes of PGNs from
180 CA1[24] and four DN subtypes from the VTA[25], together with mESC scRNA-seq data
181 produced here (**Extended Data Fig. 3a,b**). Separation of cell-type specific
182 transcriptomes into distinct clusters was visualized by Uniform Manifold
183 Approximation and Projection (UMAP) dimensionality reduction, and confirmed by
184 expression of cell-type marker genes. scRNA-seq data from mESCs was shown to

185 correlate with published bulk RNA-seq data[26] (**Extended Data Fig. 3c**). Cell cluster
186 identities were further verified by visualizing the expression of additional cell-type
187 marker genes on UMAPs (**Extended Data Fig. 3d-e**). Bulk gene expression values
188 were obtained for each cell type by pooling single-cell transcriptomes into pseudo-
189 bulk RNAseq datasets, where expression was defined by a threshold of regularized
190 log (R-log) ≥2.5 (see **Methods**, **Extended Data Fig**. **3f**). Bulk expression profiles of
191 marker genes in all cell types confirm cell-type specificity (**Extended Data Fig. 3g**).
192
193 To obtain cell-type specific chromatin accessibility data, we mined published
194 scATAC-seq by selecting mature OLGs from the prefrontal cortex[27], and one
195 population of PGNs from the CA1[28]. For DNs, we generated single-nucleus ATAC-
196 seq (snATAC-seq) from the VTA using single-nucleus isolation (see **Methods**). We
197 additionally produced bulk ATAC-seq for mESCs. All ATAC-seq data showed good
198 quality distribution of fragment sizes and the expected enrichment at gene start sites
199 (**Extended Data Fig. 3h-i**). The selection of DN-specific cell clusters from the newly-
200 generated snATAC-seq data in the VTA showed good quality, as assessed by the
201 large number of fragments per cell, the appropriate detection of reads at the
202 promoters of DN-expressed genes, and the enrichment for gene ontology (GO) terms
203 related to DN function, such as 'dopamine metabolic process' and 'locomotory
204 behaviour' (**Extended Data Fig. 3j-l**). Pseudo-bulk ATAC-seq datasets were derived
205 for each cell type by pooling single-cell accessibility maps.
206
207 ***Extensive reorganization of chromatin domains occurs at cell specialization***
208 ***genes***
209 We next investigated cell-type specific changes in chromatin domains,
210 including how preferred contacts differ between cell types at the level of TADs. For

211 example, we found complex and extensive changes in the contact density at a 4-Mb
212 region containing *Scn* genes that encode for sodium voltage-gated channel type II
213 subunits, expressed in brain cells at different levels during development, including
214 *Scn2a*, an autism candidate gene[29], and *Scn3a,* implicated in epilepsy[30] (**Fig**. **2a**, see
215 **Extended Data Fig. 4a** for biological PGN and DN replicates).
216
217 To explore the complex changes in domain organisation genome-wide, we
218 compared the position of TAD borders across cell types mapped with the insulation
219 score method[18] (**Fig. 2b**; 500kb insulation square). Similar numbers and lengths of
220 TADs were found in all cell types (n = 2209-2487, median length ~ 1Mb; **Extended**
221 **Data Fig. 4b**). TAD borders were defined as the lowest point of insulation expanded
222 by one genomic window on either side, to give borders 150kb wide; TAD borders
223 were considered to be conserved when separated by less than one genomic bin
224 (50kb). We found that TAD borders varied extensively between cell types (**Fig**. **2b**;
225 see **Extended Data Fig**. **4c** for less frequent combinations). One third of all borders
226 are unique to one cell type (7-10% for each cell type), and show statistically
227 significant lower insulation scores than in the other three cell types (**Fig. 2b**; Mann-
228 Whitney test for central TAD border window, $p < 1.8 \times 10^{-7}$ for all comparisons). In
229 contrast, only 14% of the total set of borders is common to all cell types (for example,
230 only 29% of mESC TAD borders are found in brain cells), and only 8% of TAD
231 borders are shared between the three brain cell types. Interestingly, TAD borders
232 common to all cell types show significantly stronger insulation in brain cell types than
233 in mESCs (Mann-Whitney test for central TAD border window, $p < 8.6 \times 10^{-10}$;
234 **Extended Data Fig. 4d**), possibly reflecting structural domain stabilization upon
235 terminal differentiation. TAD border conservation between PGN or DN biological
236 replicates was 59-65% (**Extended Data Fig. 4e**).

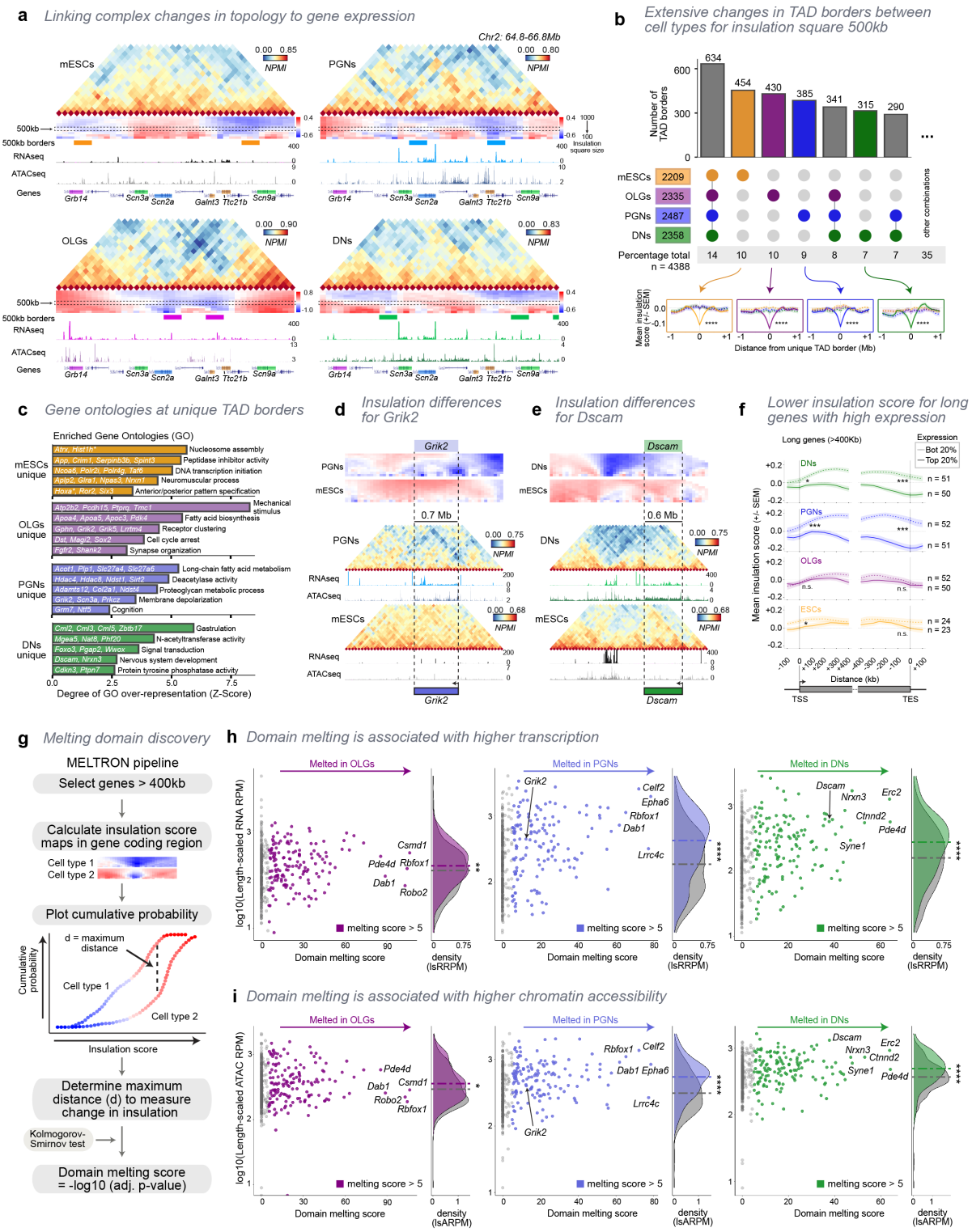

Figure 2. Broad remodeling of chromatin structure and domain melting of long, highly expressed, and/or highly accessible genes in the brain.

a, ImmunoGAM contact matrices show cell-type specific differences in local contacts and contact density within a 2-Mb region containing genes with cell-type specific expression, including sodium voltage-gated channel type II (*Scn*) subunit genes. Contact density maps (50-kb resolution), TAD borders, pseudo-bulk scRNAseq and scATACseq tracks are shown. Mouse replicate 1 is shown for brain cells. Scale bars range between 0 and the 99th percentile of NPMI values for each cell type.

(Fig. 2 legend, cont.) **b,** Upset plots show comparison of TAD boundaries between cell types. A total of 634 (14%) TAD borders are common to all cell types, followed by ~315-454 (7-10%) borders that are cell-type specific (for all combinations see **Extended Data Fig**. **4c**). TAD boundaries were defined using insulation square sizes of 500kb, as 150-kb genomic regions centered on lowest insulation score windows. Boundaries were considered different if separated by >50kb edge-to-edge. Cell-type specific TAD borders show low average insulation scores in the scell type where they are found, with highly significant differences at central border window with all other cell types (****p < 1.8e-07 for all comparisons; Mann-Whitney test).

c, Genes at cell-type specific TAD borders have enriched ontologies relevant for their cell type functions. GO terms represented are the top four enriched and one selected GO (over-representation as measured by Z-Score; permuted p-values <0.05, 2000 permutations; Z-Score >1.96).

d,e, Contact density and immunoGAM matrices for *Grik2* and *Dscam* which overlap with cell-type specific TAD borders in PGNs and DNs, respectively. A general loss of contact density, or 'melting', is seen in the cell type where these genes are most expressed.

f, The top expressing long genes have significantly lower insulation scores compared to the lowest expressed genes, at both the TSS and TES, in DNs and PGNs, while mESCs are lower at the TSS only, and OLGs show no detectable difference (*p < 0.05, ***p < 0.001, n.s. respectively; Mann-Whitney test at TSS or TES windows). Average insulation score profiles at the gene TSS or TES for genes >400kb in length, using insulation square size 500kb. The top and bottom 20% expressing genes were determined using the length-normalized number of reads covering the gene body (length-scaled RNA Reads per Million; lsRRPM).

g, Domain melting determined with the MELTRON pipeline at long genes (>400 kb, 479 genes). The distributions of insulation score values (100-1000kb squares) were compared between brain cell types and mESCs. The maximum distance between the distributions was determined using one-sided Kolmogorov-Smirnov testing and a domain melting score was calculated based on the distribution dissimilarity. Genes were considered to melt if the melting score computed across their coding region was > 5 ($p < 1x10^{-5}$).

h, Melting occurs at cell-type specific genes, which have significantly higher expression than other long-genes that do not melt in the corresponding cell type. Association with high expression is most pronounced in PGNs and DNs, compared with OLGs (**p<0.01, ****p<0.0001; Wilcoxon rank-sum test).

i, Melting genes tend to have higher density of open chromatin regions (length-scaled ATAC-seq RPM values; lsARPM), especially in PGNs and DNs, and to a minor extent in OLGs (*p<0.05, ****p<0.0001; Wilcoxon rank-sum test).

As TAD formation and border reorganisation have been previously shown to contain active genes or relate with gene activation[8,31], we asked whether the rewired TAD boundaries contained expressed and/or cell type-specific genes. We found that 52-55% of boundaries specific to the brain cell types contain expressed genes respective to each cell type, with lowest frequency of 38% for mESC-specific boundaries (**Extended Data Fig. 4f**). Genes found at unique TAD boundaries are important for several cell processes, and are enriched for GO terms relevant for the

290 specialized functions of each cell type (**Fig. 2c, Supplemental Table 7**). For
291 example, PGN-specific TAD borders contain genes with roles in 'membrane
292 depolarization' and 'cognition', including the modulation of the response to neuronal
293 activation[32], while DN-specific borders contain genes important for dopaminergic
294 differentiation, regulation of tyrosine hydroxylase activity and dopamine synthesis,
295 and G-protein coupled receptor mediation of dopaminergic cell signalling[33,34].
296
297 ***Domain reorganization is accompanied by loss of chromatin contacts at highly***
298 ***expressed genes in brain cells***
299 We noticed that domain reorganisation is especially apparent at long genes in
300 both PGNs and DNs (**Fig. 2d,e**). Many neuronal genes which are involved in
301 specialized cell processes, such as synaptic plasticity, are often longer than 400kb,
302 contain several alternative promoters, and produce many isoforms due to alternative
303 initiation sites and complex RNA processing[35]. For example, *Grik2* is a 700kb-long
304 gene, which encodes for the glutamate ionotropic receptor kainate type subunit 2, an
305 epilepsy risk gene[36] with gain-of-function mutations in intellectual disability[37]. In
306 PGNs, the genomic region containing the *Grik2* gene undergoes a general loss of
307 contact density compared to mESCs, especially around the TSS and TES (**Fig. 2d**).
308 Similar examples are found for DNs, such as at *Dscam*, a 600kb-long gene which
309 encodes the Down syndrome cell adhesion molecule, and is disrupted in Down
310 syndrome and intellectual disability[38]. *Dscam* is highly expressed in DNs, with high
311 chromatin accessibility across its entire gene body (**Fig. 2e**).
312
313 To assess whether changes in contact density are related to expression of
314 long genes (> 400kb), we compared the average insulation score at the TSS and
315 TES of long genes with the highest or lowest 20% of expression levels in each cell

316 type (**Fig. 2f**). To capture transcriptional activity over the whole gene body, gene
317 expression was measured from the length-normalized number of reads covering
318 exons and introns (length-scaled RNA Reads per Million; lsRRPM; see **Methods**).
319 Highly expressed, long neuronal genes had significantly lower insulation scores at
320 TSSs, throughout the gene bodies, and especially at TESs in both DNs and PGNs. In
321 contrast, the most expressed long genes in OLGs and mESCs showed a small
322 increase in decondensation, only at the TSS in mESCs, without detectable insulation
323 loss through gene bodies or TESs.
324
325 The general loss of contact density observed by immunoGAM at highly
326 expressed long neuronal genes is reminiscent of the decondensation, or 'melting', of
327 higher-order chromatin structures observed by imaging approaches upon
328 transcriptional activation at fly polytene chromosome puffs[39-41], or at tandem arrays of
329 the mammary tumour virus (MMTV) promoter driving a *ras* reporter[42]. Such large
330 decondensation events have not been reported using genome-wide ligation-based
331 approaches such as Hi-C, possibly because they capture proximal contacts between
332 genomic regions without directly measuring their physical distances[13]. Approaches to
333 detect 'domain melting' will help identify the genes that undergo large-scale
334 topological changes, and their association with the highest levels of expression, while
335 providing additional genomic features that can inform on how disease-associated
336 genomic variations result in neurodevelopmental or degenerative disorders[43,44].
337
338 ***Detection of domain melting in long (>400kb) genes***
339 To unbiasedly detect domain melting events at long genes genome-wide, we
340 devised the MELTRON pipeline to statistically detect differences in chromatin contact
341 density at genomic regions of interest. MELTRON calculates the differences between

342 cumulative probabilities of insulation scores across a range of genomic scales (100-
343 1000kb) in two cell types or treatments. We established a 'domain melting score', as
344 the significance value of the differences between the maximum distances of
345 cumulative probabilities between two cell types within the region of interest, here
346 considered the genomic regions containing long (> 400kb) genes (**Fig. 2g**). We
347 calculated domain melting scores for all long genes (479 genes) for all the brain cell
348 types in comparison to mESCs, and considered a threshold of melting score of > 5 to
349 define melted genes (Kolmogorov-Smirnov test between cell types, $p < 1x10^{-5}$; **Fig.**
350 **2h**, **Supplemental Table 8**). We found that *Grik2* has melting scores of 12 and 26
351 (replicates 1 and 2, respectively) between mESCs and PGNs (**Extended Data Fig.**
352 **5a**), *Dscam* has melting scores of 38 and 50 (replicates 1 and 2, respectively)
353 between mESCs and DNs. We also detected domain melting between OLGs and
354 mESCs, for example in the *Magi2* gene with a domain melting score of 73 (**Extended**
355 **Data Fig. 5a-b**). Domain melting scores were well correlated between replicates in
356 both PGNs and DNs (Pearson correlation, $p < 0.0001$; **Extended Data Fig. 5c**).
357

358 ***Melting of long (>400kb) genes is related with high transcription and open***
359 ***chromatin state***

360 To investigate whether domain melting is broadly related with transcriptional
361 activity, we compared gene melting scores to transcription over the whole gene body
362 (calculated by lsRRPM as above; **Fig. 2h**). Genes with melting scores > 5 have
363 significantly higher transcription levels in all three brain cell types, compared to long
364 genes that are not melted, most pronounced in PGNs and DNs (Wilcoxon rank-sum
365 test; **Fig. 2h**, see **Extended Data Fig. 5d** for PGN and DN biological replicates). We
366 also investigated whether chromatin accessibility was related with domain melting
367 (Wilcoxon rank-sum test; **Fig. 2i**, see **Extended Data Fig. 5e** for biological

368 replicates). Chromatin accessibility was also calculated across the whole gene body
369 using the length-normalized number of reads covering both exons and introns
370 (length-scaled ATAC Reads per Million; lsARPM; see **Methods**). Melted genes
371 contain significantly higher chromatin accessibility throughout their gene bodies in all
372 three brain cells, compared to long genes that are not melted, especially in PGNs
373 and DNs (Wilcoxon rank-sum test; **Fig. 2i**).
374
375 Transcription of many long neuronal genes has been found especially
376 sensitive to topoisomerase inhibition in topotecan treatments of *ex-vivo* neuronal
377 cultures, suggesting that the high expression of long genes is especially dependent
378 on the resolution of topological constraints[45]. We tested whether melted genes were
379 amongst those shown to be topotecan sensitive, and found that half (55%) of genes
380 with the highest melting scores in any cell type or replicate (top 3% melting scores,
381 24 of 44 melted genes) were previously found to be sensitive to topoisomerase
382 inhibition[45], compared to only 16% of genes with intermediate melting scores (42 of
383 261), and 16% of the other genes tested (27 of 174; χ^2 p-value = 5.0×10^{-9}; **Extended**
384 **Data Fig. 5f**).
385
386 To understand how melting events relate with other aspects of chromatin
387 folding, we investigated compartment A/B transitions, and found that melting genes
388 often belong to compartment A in both mESCs and the associated brain cell type
389 (43-58%), especially when the gene is more highly transcribed in both cell types
390 (cluster 4, **Extended Data Fig. 5g**). We also found that melting genes in OLGs and
391 DNs are significantly less likely to belong to a lamina-associated or nucleolus-
392 associated domain (LADs or NADs, respectively) in mESCs, according to published
393 LAD[46] and NAD[47] mapping, though there was no association in PGNs (**Extended**

394 **Data Fig. 5h-i**). Therefore, domain melting of long genes is not trivially associated
395 with a transition from a heterochromatic state of long genes in compartment B, at
396 LADs or NADs, to an open chromatin state in compartment A, although such events
397 can occur at some genes, such as *Magi2* in OLGs or *Dscam* in DNs (**Supplemental**
398 **Table 8**).
399
400 Overall, we discovered that long genes in brain cell types can become highly
401 decondensed, or melted, when they are most highly transcribed and/or with high
402 chromatin accessibility. Domain melting occurs independently of compartment
403 transitions, in fact it is more likely to occur in genes that are upregulated in neurons,
404 but already present within active chromatin environments in mESCs, and not directly
405 associated with the nuclear lamina and the nucleolar surface.
406
407 ***Melting is associated with large-scale 3D chromatin reorganisation in neurons***
408 We next examined domain melting in more detail in two neuronal genes,
409 Neurexin-3 (*Nrxn3*) and RNA Binding Fox-1 Homolog 1 (*Rbfox1*), both of which are
410 highly sensitive to topoisomerase I inhibition, which prevents resolution of
411 transcription-induced chromatin supercoiling[45]. *Nrxn3* is 1.54Mb long, encodes a
412 membrane protein involved in synaptic connections and plasticity, and for which
413 genetic variation is associated with alcohol dependence and autism spectrum
414 disorder[43,48]. In mESCs, *Nrxn3* spans two TADs with high contact density and has
415 barely detectable transcription and accessibility. *Nrxn3* localizes in a B compartment,
416 and associates with the nuclear lamina and/or the nucleolus, in mESCs (**Fig. 3a**). In
417 DNs, it extensively melts, throughout its gene body (melting scores of 48 and 49, in
418 replicates 1 and 2, respectively), is most highly transcribed and with highest
419 chromatin accessibility than in the other cell types analysed, and belongs to

420 compartment A (see **Extended Data Fig. 6a** for all cell types and replicates). The
421 *Rbfox1* gene is 1.1 Mb long, and encodes an RNA-binding protein that regulates
422 alternative splicing, with mutations linked to epilepsy and autism spectrum
423 disorder[49,50]. In mESCs, it is found within a large domain with high contact density,
424 which maps within compartment A while having very low expression, low chromatin
425 accessibility, membership in a NAD, and partially in a LAD (**Fig. 3b**). In PGNs,
426 *Rbfox1* undergoes extensive loss of contact density (melting scores of 65 and 39 in
427 replicates 1 and 2, respectively), especially within a large (~500kb) genomic region
428 centred on its first TSS, coinciding with its highest level of expression between cell
429 types, with high accessibility, and continued membership in an A compartment
430 (**Extended Data Fig. 6b**).
431
432 ***Modelling the domain melting of Nrxn3 in DNs shows extensive genome***
433 ***decompaction***
434 To better understand how the loss of domain structure could influence 3D
435 chromatin folding of long, highly expressed genes, we used a polymer-physics-based
436 approach (PRISMR)[51] to generate ensembles of 3D models of the *Nrxn3* region in
437 mESCs and DNs from the GAM matrices (**Fig. 3c**). 3D models were validated by
438 applying an "*in-silico* GAM" approach[13] by randomly sectioning each 3D model, and
439 reconstructing *in-silico* GAM chromatin contact matrices that closely resemble the
440 experimental data (Pearson r = 0.72 in mESCs, and 0.79 in DNs; **Extended Data**
441 **Fig. 6c**).
442

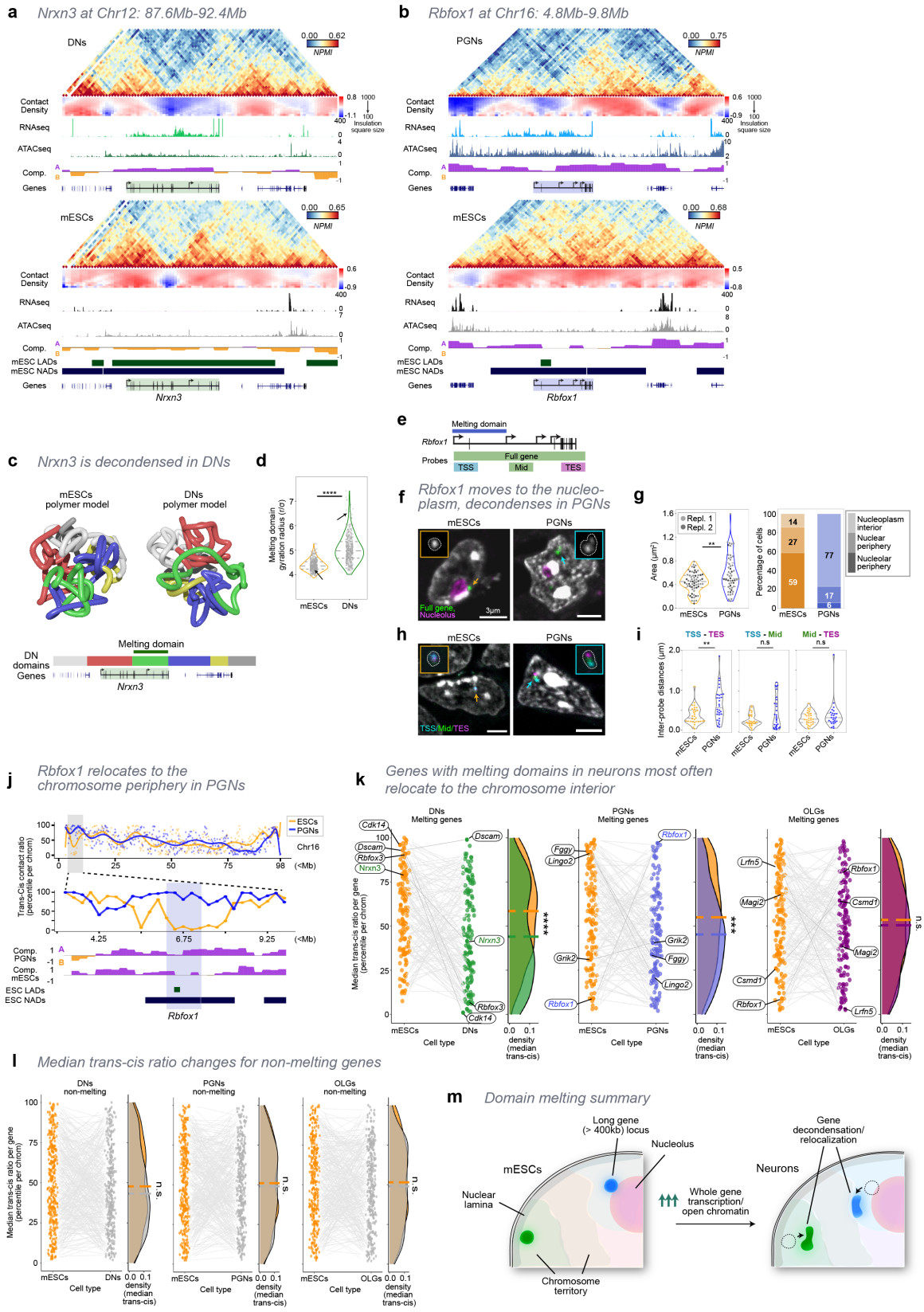

Figure 3. Extensive decondensation and re-localization of genes with melting domains.

a, GAM contact matrices show the *Nrxn3* locus, a long (1.5Mb) gene which is highly expressed and has highly accessible chromatin in DNs, but not in mESCs (50-kb resolution; chr12: 87,600,000-92,400,000). *Nrxn3* occupies two TADs with strong contacts in mESCs,

449 *(Fig. 3 legend, cont.)* which lose contact density in DNs. Compartment tracks are shown for
450 each cell type, and lamina- and nucleolus-associated domains (LADs and NADs) for mESCs.
451 Scale bars range between 0 and the 99th percentile of NPMI values for each cell type.

452 **b,** GAM contact matrices show the *Rbfox1* locus, a long (1.1Mb) gene, highly expressed and
453 containing highly accessible chromatin in PGNs, but not in mESCs (50-kb resolution; chr16:
454 4,800,000-9,800,000). *Rbfox1* has high contact density in mESCs, which is lost in PGNs at
455 the gene TSS, where melting occurs. Scale bars are adjusted to a range between 0 and the
456 99th percentile of NPMI values for each cell type.

457 **c,** Ensembles of polymer models were produced for the *Nrxn3* locus in mESCs and DNs
458 from experimental GAM data using PRISMR modelling[51]. Colour bar shows the position of
459 domains in DNs and are used to colour the polymers shown. In mESCs, the *Nrxn3* locus
460 polymer shows a globular structure with subdomains that interact with one another. In DNs,
461 the *Nrxn3* locus decondenses, especially at the green region covering the 5' end of the gene
462 and the most decondensed domain, which becomes extended and loops away from the
463 neighbouring regions.

464 **d**, The green domain has significantly higher gyration radii in DNs than mESCs (****$p <$
465 0.0001; Mann-Whitney test, arrows indicate where exemplar models are in the distribution;
466 (n= 450 polymers).

467 **e,** Cryo-FISH probes were designed to cover the whole *Rbfox1* gene, or to specifically target
468 the TSS, middle of the coding region (mid) or the end of the gene (TES; see **Supplementary**
469 **Table 9** for the list of probes).

470 **f,** Cryo-FISH using the full-gene probe shows that *Rbfox1* (pseudo-coloured green) occupies
471 small, rounded foci in mESCs, often located at the nucleolus periphery (visualised by
472 nucleophosmin 1 immunostaining, purple). In PGNs, *Rbfox1* occupies larger and more
473 extended foci, and relocates away from the nucleolar periphery. Scale bars, 3 μm.

474 **g,** *Rbfox1* occupies significantly larger areas in PGNs than mESCs (**$p < 0.01$; Mann-
475 Whitney test; n=52 and 63, respectively; two experimental replicates). Most *Rbfox1* foci
476 localize at the nucleolar periphery in mESCs, and away from the nucleolus in PGNs.

477 **h,** Decondensation of the *Rbfox1* gene is confirmed in cryo-FISH experiments that target
478 TSS, Mid and TES regions. Insets show that the TSS, Mid and TES regions overlap in
479 mESCs, whereas in PGNs they separate from one another. Scale bars, 3 μm.

480 **i,** The *Rbfox1* TSS region is significantly separated from its TES in PGNs, compared to
481 mESCs (**$p < 0.01$; Mann-Whitney test).

482 **j,** The *Rbfox1* locus gains contacts with other chromosomes in PGNs, compared to mESCs.
483 Trans-cis contact ratios were determined by the mean ratio between trans NPMI scores and
484 cis NPMI scores (250kb genomic bins), and normalizing each ratio as a percentile for each
485 chromosome. Inset (grey shaded region) shows a 7Mb region (Chr16: 3,000,000-
486 10,000,000) containing the *Rbfox1* gene (blue shaded region).

487 **k,** Melting genes are more likely to gain intra-chromosomal contacts in PGNs and DNs R1,
488 but not OLGs, compared to mESCs (***$p<0.001$, ****$p<0.0001$; Wilcoxon rank-sum test).
489 Median trans-cis contact ratios were calculated for each gene with domain melting in DNs,
490 PGNs, or OLGs, and compared to mESCs.

491 **l,** Non-melting genes show no preference for changes in trans-cis contact ratios between
492 brain cells and mESCs.

493 **m,** Summary scheme representing the melting of long (>400kb) genes in neurons when they
494 are highly transcribed and/or have open chromatin in neurons, resulting in large-scale gene
495 decondensation. Genes with domain melting are less likely to be associated with the
496 nucleolus or lamina in mESCs, though melting can be accompanied by re-localization away
497 from these repressive areas (blue locus; e.g. for *Rbfox1* in PGNs). Domain melting is most
498 likely to result in increased intra-chromosomal contacts (green locus; e.g. for *Nrxn3* in DNs),
499 though more rarely can also become more peripheral to their chromosome (blue locus; e.g.
500 *Rbfox1* in PGNs).

501

502 We inspected single polymer models, coloured according to the structure of
503 domains found in DNs, where the red and green domains correspond to the first and
504 second TADs that contain the *Nrxn3* gene. mESC models showed a structure
505 containing several globular domains which were highly intermingled, including the
506 green and red domains (**Supplemental movie 1**; see **Extended Data Fig. 6d** for
507 additional example models). In contrast, in DNs, the second green TAD, containing
508 the melted domain, becomes highly extended (**Supplemental movie 2**). Across the
509 ensemble of DN polymer models, the green melting domain shows significantly
510 higher gyration radii compared to the same genomic region in mESCs (450 models
511 were generated in each case; Mann-Whitney test, $p = 1.1 \times 10^{-92}$; **Fig**. **3c**). In contrast,
512 the regions upstream and downstream of the *Nrxn3* gene (grey and blue domains)
513 tend to condense, as seen by marked decreases in gyration radii in DNs (Mann-
514 Whitney test, $p < 0.0001$; **Extended Data Fig. 6e**), which can also be seen directly in
515 the contact matrices (**Fig. 3a**), and that may be related to transcription-induced
516 chromatin supercoiling induced by high *Nrxn3* transcriptional activity, and/or to
517 changes in gene expression in the neighbouring domains.

518

519 ***Single-cell imaging of Rbfox1 reveals extensive domain melting accompanied***
520 ***by large-scale gene re-localization away from the nucleolar periphery***

521 To further investigate the prevalence of domain melting across a population of
522 neuronal cells, we determined the folding and nuclear positioning of the *Rbfox1* gene
523 in mESCs and PGNs by cryo-FISH[3,52], an approach where fluorescence in-situ
524 hybridization is performed directly on thin cryosections obtained from cells or tissues.
525 For PGNs, cryo-FISH experiments were performed with the same samples as used
526 for immunoGAM (PGNs R1). We used FISH probes produced from 8 contiguous

527 bacterial artificial chromosomes (BACs) to cover the entire 1.1-Mb length of *Rbfox1*
528 (**Fig. 3e**; the list of probes is found in **Supplemental Table 9**). To identify the position
529 of nucleoli, we performed immunostaining prior to FISH, using antibodies against
530 nucleophosmin (NPM1), a nucleolar marker[53]. In mESCs, *Rbfox1* is seen as a small
531 spherical focus (**Fig. 3f**, **Extended Data Fig. 6f).** Quantitative image analysis shows
532 that *Rbfox1* occupies an average area of 0.44 ± 0.17 μm^2 in mESCs, and is most
533 often found at the nucleolar surface (59%) or at the nuclear periphery (27%; **Fig. 3g**).
534 In contrast, in PGNs, *Rbfox1* becomes highly decondensed and elongated, with a
535 significantly higher area compared to mESCs (0.59 ± 0.31 μm^2; Mann-Whitney test p
536 < 0.01, n=63 and 52 for mESCs and PGNS, respectively), and relocalizes to the
537 nucleoplasm interior (77%).
538
539 To visualize the decondensation of the *Rbfox1* gene in finer detail, we used
540 specific BAC probes that distinguish the TSS, middle, or the TES of the gene, and
541 measured the distance between the three regions. In mESCs, TSS, middle and TES
542 regions often overlap one another, while in PGNs each region becomes more clearly
543 separated and the whole gene visibly decondensed (**Fig. 3h**, **Extended Data Fig.**
544 **6g**). Remarkably, the distance between the center of mass of the TSS and TES
545 regions is significantly larger in PGNs compared to mESCs (**Fig. 3i**; 0.65 ± 0.41 μm
546 and 0.37 ± 0.22 μm, respectively; Mann-Whitney test $p < 0.01$; n=31 and 30 for
547 mESC and PGNs respectively). The increased distance between the TSS and TES
548 also recapitulates the marked decrease in chromatin contacts observed directly in the
549 GAM matrices (**Fig. 3b**).
550
551
552

553 ***Melting genes often show increased contacts with their own chromosome in***
554 ***DNs and PGNs***

555 The extensive changes in nuclear localization of *Rbfox1* in PGNs,
556 accompanied by large-scale decondensation, lead us to ask whether domain melting
557 is generally related with re-localization of genes to the interior of their chromosome
558 territory or towards contacts with other chromosomes, and how these parameters
559 related with a preferred LAD or NAD association in mESCs. First, we calculated a
560 *'trans-cis* contact ratio' for each 250kb window along all chromosomes (see
561 **Methods**). A higher *trans-cis* contact ratio indicates that the region is more peripheral
562 and establishes more contacts with other chromosomes relative to the rest of its own
563 chromosome. Indeed, we found that the *Rbfox1* gene in chromosome 16 has higher
564 median *trans-cis* values in PGNs than mESCs (97 and 9, respectively, averaged from
565 all 250kb windows covering *Rbfox1*), indicating that *Rbfox1* re-localizes towards its
566 chromosome periphery in PGNs (**Fig. 3j**). In contrast, *Nrxn3* re-localizes towards its
567 chromosome interior in DNs (**Extended Data Fig**. **6h**).
568

569 When we considered all melted genes, we found that they can undergo all
570 possible transitions in *trans-cis* ratios between mESCs and brain cells, though they
571 more often undergo a decrease in *trans-cis* values in DNs or PGNs (Mann-Whitney
572 test, $p < 0.0001$ and $p < 0.001$, respectively; **Fig. 3k**, see **Extended Data Fig. 6i-j** for
573 replicates), but not in OLGs. The tendency for lower trans-cis ratios in DNs and
574 PGNs was a feature of melting genes, and not observed at long genes that are
575 expressed but do not melt (**Fig. 3l**; see **Extended Data Fig. 6k** for replicates).
576 Interestingly, the tendency for decreased trans-cis ratios of melting genes in DNs or
577 PGNs was independent of NAD association in mESCs (**Extended Data Fig. 6l**), in

578 contrast with genes with lower trans-cis values that do not melt, which tend to be
579 NAD associated in mESCs (**Extended Data Fig. 6m**).
580
581 We find that the decondensation of long melting genes between mESCs and
582 brain cells often coincides with extensive reorganisation of their contacts with their
583 own and other chromosome territories, with a preference for domain melting to occur
584 alongside increased intra-chromosomal contacts, as seen for *Nrxn3* in DNs, but also
585 with increased trans-cis contacts, as observed with the *Rbfox1* gene in PGNs (**Fig.**
586 **3m**). Together, polymer modelling from GAM data and single-cell imaging highlight
587 that domain melting is a previously unappreciated topological feature of very long
588 genes, robustly captured by immunoGAM, when genes are highly expressed, or
589 highly accessible, in brain cell types.
590
591 ***Cell-type specific chromatin contacts contain binding motifs of differentially***
592 ***expressed transcription factors***
593 To explore whether the extensive rearrangements in chromatin topology found
594 in specialized brain cells are related with changes in the landscape of cis-regulatory
595 elements, we next focused on differences in specific chromatin loops between the
596 two neuronal cell types, PGNs and DNs. We were especially interested in finding
597 whether they involve differentially expressed genes and contain accessible chromatin
598 regions characterized by putative binding sites for neuron-specific transcription
599 factors (TFs). First, we determined strong differential chromatin contacts by
600 extracting the top 5% cell-type specific contacts[11], which occur within 5Mb and have
601 NPMI values above 0.15 (**Fig. 4a**). We found that each 50kb window contains only
602 1326 ± 881 or 1270 ± 1270 bp of accessible chromatin, in PGNs and DNs,
603 respectively (**Extended Data Fig. 7a**). Next, we searched accessible regions within

604 the top 5% differential contacts for binding motifs of TFs that are expressed in PGNs
605 and/or DNs (218 TFs; R-Log > 2.5 in pseudo-bulk scRNA-seq). To investigate cell-
606 type specific loops, we prioritized TFs that are differentially expressed between
607 PGNs and DNs (**Extended Data Fig. 7b, Supplemental Table 10**), and high
608 coverage in at least 5% of differential contacts, finding 16 DN-specific TFs and 32
609 PGN-specific TFs (**Extended Data Fig. 7c**). Reassuringly, many of the differentially
610 expressed TFs with highest coverage in differential contacts have known roles in
611 neuronal development, e.g. *Neurod1* and *Foxa1*[54,55] and activation, e.g. *Egr1* and
612 *Neurod2*[56,57]. We prioritized for further analyses the TF-feature pairs found in PGN or
613 DN differential contacts according to whether they were most enriched in each cell
614 type (five TF motif pairs per cell type), or with a high ability to distinguish contacts
615 between cell types based on Information Gain (ten TF-motif pairs; see **Extended**
616 **Data Fig. 7d** for full pipeline and criteria). This selection resulted in nineteen TF-motif
617 contact pairs (from 1275 possible homotypic and heterotypic combinations from
618 expressed TFs), which contain combinations of 14 TF motifs, with the *Neurod2*-
619 *Neurod2* motif pair being found by both PGN enrichment and Information Gain
620 criteria (**Extended Data Fig. 7e**, see **Supplemental Table 11** for all TF pairs).
621
622

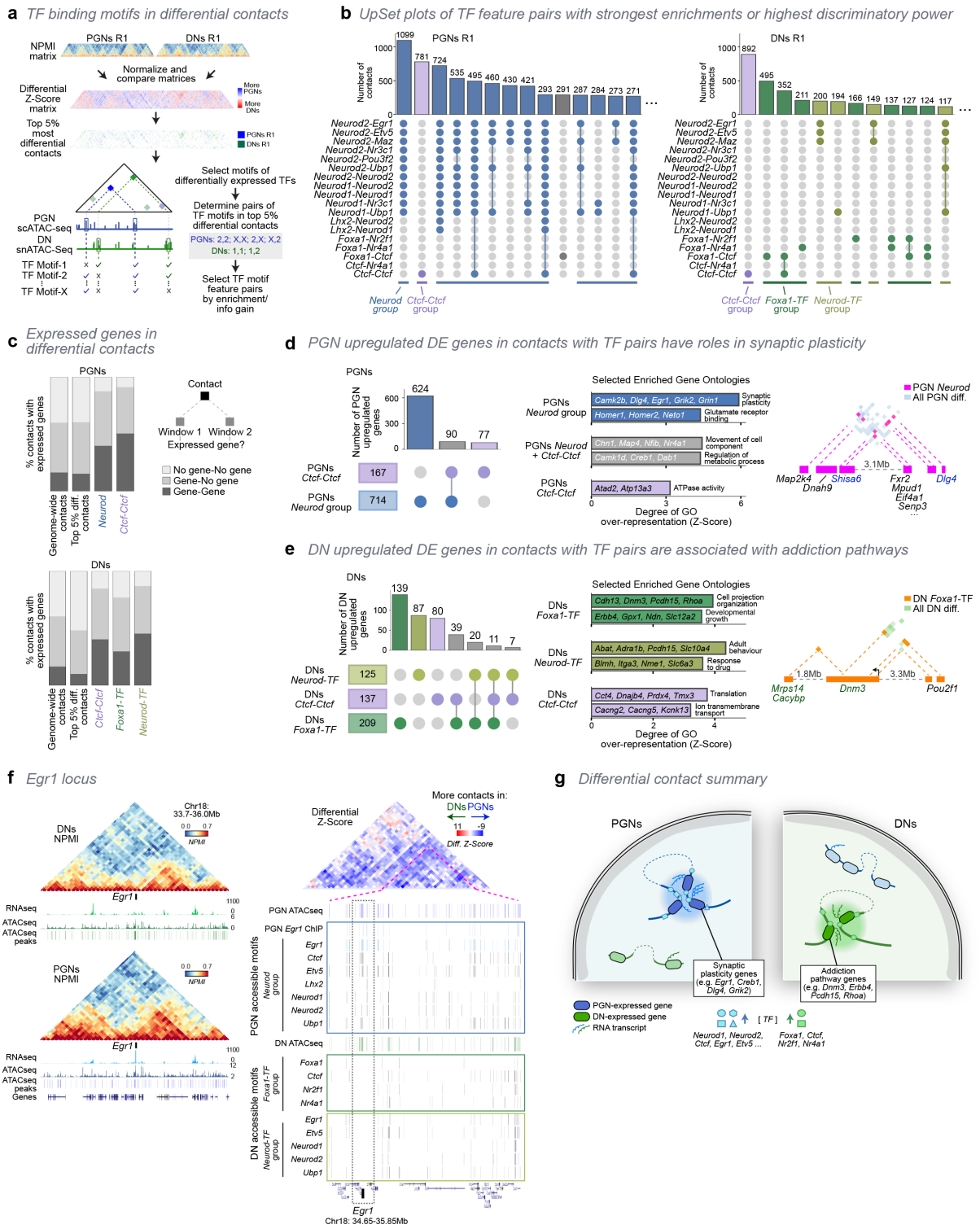

Figure 4. Neuron-specific genes establish PGN- and DN-specific contacts, which contain rich networks of putative TF binding sites within open chromatin regions.

a, GAM contacts from PGNs and DNs were normalized and subtracted to produce a matrix of cell-type differential contacts (differential Z-Score matrix). The top 5% differential contacts were considered within 0.05-5 Mb, and overlapped with accessible chromatin regions to identify candidate regulatory elements; these cover on average 1.3±0.9 or 1.3±1.3kb in PGNs and DNs, respectively (**Extended Data Fig. 7a**). Next, we searched for putative transcription factor (TF) binding sites exclusively within the accessible windows. Out of 218 TFs expressed in PGNs and DNs, we focused on discovering binding sites for TFs which are *(Fig. 4 legend, cont.)* differentially expressed in PGN and DN, and which are well

634 represented in the differential contact windows (50 TFs; **Extended Data Fig**. **7c**). Contacts
635 containing pairs of TF motifs which were either enriched in PGNs or DNs (5 highest
636 enrichment scores in each cell type) or with the most discriminatory power (10 highest
637 information gain scores) were further investigated (**Extended Data Fig**. **7e**). R1, mouse
638 replicate 1.
639 **b**, Multiple TF feature pairs coincide in the same PGN (*left*) or DN (*right)* differential contacts.
640 Most abundant groups of contacts are shown for each cell type (for all combinations see
641 **Extended Data Fig**. **7f**).
642 **c**, Differential contacts with the most enriched TF feature pairs in PGNs and DNs contain
643 expressed genes in both contacting windows.
644 **d**, Differential contacts involving the most abundant TF feature pairs in PGNs contain
645 differentially expressed genes with PGN-specific roles. (*Right*) Matrix shows differential
646 contacts established between PGN-upregulated genes. Gene names highlighted in blue are
647 upregulated in PGNs; other genes are expressed.
648 **e**, Differential contacts containing the most abundant TF feature pairs in DNs also contain
649 differentially expressed genes with roles in DN-specific functions. (*Right*) Matrix shows
650 contacts between DN-upregulated genes. Gene names highlighted in green are upregulated
651 in DNs; other genes are expressed.
652 **f**, In DNs, the *Egr1* gene establishes weak contacts with a large downstream domain
653 containing expressed genes. In PGNs, *Egr1* is upregulated and shows stronger interactions
654 within the downstream domain. The differential contact matrix shows increased PGN-specific
655 contacts in the whole region surrounding *Egr1*. The *Egr1*-containing TAD (chr18: 34,650,000-
656 35,850,000) has multiple putative TF-binding sites found within PGN accessible regions. DNs
657 also contain multiple ATAC-seq peaks but many fewer TF motifs, most notably surrounding
658 the Egr1 gene (*grey dashed box*). Scale bars range between 0 and the 99th percentile of
659 NPMI values for each cell type.
660 **g**, Summary scheme representing PGN- and DN-specific contacts which contain distal genes
661 related to synaptic plasticity in PGNs, and drug addiction in DNs. The specific contacts are
662 enriched for open chromatin regions containing binding sites for TFs that are differentially
663 expressed in the two cell types.
664

665 As many of the chosen TF pairs were redundant, we next searched if they
666 coincided in the same differential contacts (**Fig. 4b**, full list shown in **Extended Data**
667 **Fig. 7f**). In PGNs, the most abundant sets of contacts contain putative binding sites
668 for *Neurod1* and/or *Neurod2* homo- and heterodimers, plus combinations of eight
669 different TFs: *Egr1, Etv5, Lhx2, Maz, Nr3c1, Pou3f2,* and *Ubp1*, giving a combined
670 total of 5572 contacts and representing the most common combinations, which we
671 call the *Neurod* group of contacts. In DNs, *Neurod1* and *Neurod2* appear only as
672 heterodimers, in a combined total of only 660 contacts with *Egr1*, *Etv5*, *Maz* and
673 *Ubp1*. Instead, the most frequent TF-motif pair in DNs, and second most in PGNs, is
674 a homodimer of *Ctcf* (892 and 781 contacts, respectively). In addition to its roles in
675 chromatin regulation[58], *Ctcf* regulates adult neuronal pathways related to learning

676 and memory formation[59], and is disrupted in several neurodevelopmental disorders[60].
677 After *Ctcf-Ctcf* contacts, the most abundant group of DN-specific contacts contain
678 *Foxa1* heterotypic contacts with *Ctcf*, *Nr2f1* or *Nr4a1* (*Foxa1-TF* group; combined
679 total of 1612 contacts). All groups of TF-containing contacts span all genomic
680 distances considered (50kb-5Mb; **Extended Data Fig. 7g**), and capture strong
681 contacts compared to their surrounding (200kb) regions, or in distance-matched
682 random contacts (**see Methods**; **Extended Data Fig. 8a**). These strong contact
683 loops show similar local enrichment to those identified between *Neurod2* ChIP-seq
684 binding sites in neocortical PGNs[8]. We note that the selected differential contacts
685 rarely coincide with two TAD borders, even when considering the *Ctcf-Ctcf* contacts
686 (**Extended Data Fig. 8b**). The selected contacts often involve windows that belong to
687 A compartment in PGNs (71-73%) or DNs (56-62%), compared to all top differential
688 contacts (45% and 30%) or to the genome-wide distribution (44% and 45%, in PGNs
689 and DNs, respectively; **Extended Data Fig. 8c**).
690
691 The connectivity between multiple TF motifs in PGNs, and between *Foxa1* or
692 *Neurod* and specific TFs in DNs, could also be observed in the differential contacts
693 without pre-selecting specific TFs (**Extended Data Fig. 8d-e**). We built networks
694 based on the co-occurrence of motifs for all 50 differentially expressed TFs as nodes
695 in the top 5% differential contacts (as edges) of PGNs or DNs, followed by
696 community detection (see **Methods**, **Extended Data Fig. 8d**). In strong networks
697 that represent the top 20% of the total TF-containing differential contacts (i.e. with a
698 minimum of 15,833 and 5,400 contacts in PGNs and DNs, respectively), we find the
699 same strong connectivity in PGNs and DNs as in the TF motif pair analysis
700 (**Extended Data Fig. 8e, Supplemental Table 12**).

701 ***Cell-type specific contacts contain differentially expressed genes involved in***
702 ***synaptic plasticity and addiction pathways***

703 Next, we investigated whether the cell-type specific contacts in each of the TF-
704 motif groups contain genes with neuronal-specific functions. Remarkably, the TF-
705 motif groups contained many contacts with expressed genes in both contacting
706 windows (30-45% in DNs, 40-50% in PGNs), more commonly than genome-wide or
707 in all top 5% contacts (16% or 10-15%, respectively; **Fig. 4c**). The observed overlaps
708 between TF-based groups of contacts with expressed genes in both contacting
709 windows was highly significant (1000 random circular permutations of PGN and DN
710 expressed gene regions; empirical $p < 0.001$ in all cases; **Extended Data Fig. 8f**).
711 The observation that expressed genes are found in both contacting windows, for both
712 PGNs and DNs, suggests that exploring differential contacts containing differentially
713 expressed TF-motifs is a good strategy to discover cell-type specific hubs of genes
714 regulated by similar cohorts of TFs. Furthermore, many genes found in the TF-based
715 contacts are also differentially expressed (1490 and 975 genes in PGNs and DNs,
716 respectively, of 3537 differentially expressed genes; **Extended Data Fig. 8g**).
717

718 We next investigated whether the genes found at the specific contacts
719 containing open chromatin regions with binding motifs for differentially expressed TFs
720 have specific roles in each type of neuron. We focused on genes that are
721 upregulated in the cell type where they have stronger contacts. In PGN-specific
722 contacts, most PGN-upregulated genes are found in the *Neurod* contacts (624 of
723 1693 upregulated genes), with only a smaller proportion found in *Ctcf-Ctcf* contacts
724 (77), or sharing both groups of TFs (90; **Fig. 4d**). Each group of genes has GO terms
725 related to synaptic plasticity, the regulation of calcium transport, and energy
726 metabolism, respectively. For example, two of the PGN-upregulated genes, *Dlg4*,

727 important for long-term potentiation[61], and *Shisa6*, critical for preventing de-
728 sensitization of AMPA receptors during plasticity[62]**,** are separated by ~3.5 Mb and
729 part of a hub of *Neurod* contacts that involve several other activity-related genes,
730 including *Map2k4*, involved in the JNK signalling pathway[63], and *Dnah9,* part of
731 disrupted synaptic networks in major depressive disorder and Alzheimer's
732 disease[64,65] (**Extended Data Fig. 8h** shows the differential contact matrix). These
733 results are consistent with models that invoked the role of TFs in the formation of
734 transcription factories or hubs[66,67].
735
736 A similar picture is found in the selected TF-group of contacts in DNs. DN-
737 upregulated genes are exclusive to the *Foxa1*-TF (139 of 1844 upregulated genes),
738 *Neurod*-TF (87) or *Ctcf-Ctcf* (80) contacts, with 77 genes shared in different
739 combinations between groups (**Fig. 4e**). Their GO terms are related to behaviour,
740 synaptic organization and function, with many proteins in these groups involved in
741 addiction pathways (e.g. nicotine and alcohol dependence). For example, *Dnm3*
742 encodes Dynamin-3, a scaffolding protein involved in synapse formation, is
743 upregulated in VTA DNs following prenatal alcohol or nicotine exposure[68] and has
744 altered protein expression in an alcohol dependence paradigm[69]. *Dnm3*, which
745 contains a melting domain in DNs (score = 36 and 31 in replicates 1 and 2,
746 respectively), makes *Foxa1*-TF containing contacts (separated by ~1.8Mb) with the
747 DN-upregulated genes *Mrps14* (a mitochondrial protein downregulated after nicotine
748 exposure[70]) and to *Cacynp* (a Calcyclin-binding protein upregulated following alcohol
749 exposure[71]). *Dmn3* also makes specific contacts between its promoter region and
750 *Pou2f1* (separated by ~3.3Mb), which was identified as a transcription co-factor in a
751 model of alcohol dependence[72] (**Extended Data Fig. 8i** shows the differential contact
752 matrix).

753 ***Upregulation and regional condensation of the Egr1 locus in PGNs coincides***
754 ***with accessible chromatin containing TF motifs specific to PGN contacts***
755 Curiously, *Egr1*, an immediate early gene upregulated in activated neurons[56],
756 establishes PGN-specific contacts that contain accessible regions covered by *Egr1*
757 itself, together with *Neurod* motifs (**Fig. 4f**). In DNs, *Egr1* (a short 5.2kb gene) is
758 expressed (R-log = 6.8), and found at a TAD border with interactions with its right-
759 sided TAD. As expected from the frequent activation of PGNs, either stochastically or
760 during ongoing short-term memory formation[15,16], *Egr1* is highly upregulated (R-log =
761 9.5) in PGNs compared to DNs, and strikingly it gains increased contacts with the
762 TAD on its right (**Fig. 4f,** see **Extended Data Fig. 8j** for replicate data). Furthermore,
763 the general increase in PGN-specific contacts across the whole *Egr1*-containing TAD
764 correspond with an increased density of accessible chromatin peaks which contain
765 complex combinations of TF motifs belonging to the *Neurod* group of contacts,
766 especially at the *Egr1* gene itself (**Fig. 4f**), which also include *Ctcf* and *Lhx2*. In
767 contrast, in DNs, *Egr1* and the surrounding genes do not contain accessible regions
768 with motifs for the DN-specific contacts (e.g. *Foxa1*, *CTCF*, *Nr2f1*, *Nr4a1*). The
769 binding of EGR1 protein to its own promoter and surrounding region is confirmed in
770 published ChIP-seq data in cortex PGNs[73]. The increased regional condensation of
771 *Egr1* when most highly transcribed has been reported in other systems, such as the
772 α-globin locus during erythroid lineage commitment[74], and may be a feature of
773 shorter genes when they become very active, and/or depend on the transcriptional
774 and regulatory activity of the neighbouring genomic regions.
775
776 Together, our analyses of strong differential chromatin contacts specific for
777 different neuron types contain putative binding sites for TFs differentially expressed
778 between cell-types (**Fig. 4g**). TF-motif containing contacts form inter-connected hubs

779 which bring together distal genes, often differentially expressed between the two
780 types of neurons, and involved in specialized neuronal functions, such as synaptic
781 plasticity in PGNs, or drug addiction in DNs. Genomic regions containing multiple TF
782 motifs can also lead to cell-type specific local condensation of contacts (e.g. for
783 *Egr1*), which may be important for the regulation of transcriptional levels.
784

785 ***Extensive A/B compartment changes occur between mESCs and brain cells,***
786 ***and relate with the upregulation of neuronal genes***

787 As a last exploration of the specialization of chromatin contacts across whole
788 chromosomes, we investigated large-scale changes of compartment A/B domains in
789 mESCs and brain cells. Comparisons of normalized compartment eigenvectors show
790 broad changes in A/B compartmentalization between all cell types (**Fig. 5a**, see
791 **Extended Data Fig. 9a,b** for replicates). The compartment eigenvector values are
792 least correlated between brain cells and mESCs, and best between neuronal
793 biological replicates (Pearson's correlation, **Extended Data Fig. 9c**). A large fraction
794 of the genome remains in compartment A (40%, 2325 windows), in contrast with a
795 smaller proportion remaining in compartment B (15%, 898 windows) in all cell types
796 (**Fig. 5b**). Next, 12% of the genome changes from compartment B in mESC to A in all
797 the brain cells (transition B-mESC → A-brain cells) or between compartment A in
798 mESCs to B in brain cells (7%, B-mESC → A-brain cells; rarer transitions are shown
799 in **Extended Data Fig. 9d**). The extent of compartment transitions is similar across
800 chromosomes, with few exceptions, such as chromosome 7 which is highest for B →
801 A transitions from mESCs to brain cells (**Extended Data Fig. 9e**). Regardless of the
802 changes in compartment membership, the mean and total genomic lengths occupied
803 contiguously by A or B compartments are similar across cell types (**Extended Data**
804 **Fig. 9f**).

Figure 5. Clusters of sensory receptor genes more often belong to B compartments in brain cells and establish megabase-range interactions.

a, Open and closed chromatin compartments (A and B, respectively) display different genomic distributions in mESCs, OLGs, PGNs and DNs. Mouse replicate 1 (R1) is shown.

b, Most genomic windows share membership to compartments A, followed by B, in all cell types. The most frequent compartment changes occur from compartment B in mESCs to A in all brain cells (pink box), followed by changes from A in mESCs to B in all brain cells (blue box; for all combinations see **Extended Data Fig. 9d**).

c, Top enriched (over-representation as measured by Z-Score) gene ontologies (GOs) for genes that increase expression in all brain cells relative to mESCs and move from compartment B in mESCs to compartment A in brain cells (*pink box*); and for genes that

817 *(Fig. 5 legend, cont.)* decrease expression in brain cells and move to compartment B,
818 compared to mESCs (*blue box*). All enriched GOs had permuted p-values = 0.
819 **d**, Top enriched GOs for genes that remain silent in all cell types but gain membership to
820 compartment B in brain cells. Most genes are *Olfr* and *Vmn* sensory receptor cluster genes.
821 A higher proportion of *Olfr* and *Vmn* genes are found in B compartments in brain cells,
822 compared to mESCs. All enriched GOs had permuted p-values = 0.
823 **e**, GAM contact matrices containing *Vmn* and orphan receptor genes show large patches of
824 strong interactions separated by 10 Mb between B compartments in OLGs, PGNs and DNs,
825 but not mESCs. Shaded green regions indicate interacting regions. Scale bars range
826 between 0 and the 99th percentile of NPMI values for each cell type.
827 **f**, GAM contact matrices show interactions between an *Olfr/Vmn* gene cluster and a second
828 *Olfr* cluster (green shaded regions) separated by 25 Mb. The contacts between the two
829 receptor clusters are strongest in OLGs, where the B compartment is strongest. Scale bars
830 range between 0 and the 99th percentile of NPMI values for each cell type.
831 **g**, Summary diagram. The 3D genome is extensively reorganized in brain cells to reflect its
832 gene expression specialization. (i) Contacts are rearranged at multiple scales, where
833 formation of new TAD borders can coincide with genes important for cell specialization in all
834 cell types. (ii) Domain melting occurs at very long genes which are highly transcribed and
835 with high chromatin accessibility in brain cells. (iii) The most specific contacts in neurons
836 contain complex networks of binding sites of neuron-specific transcription factors. Contacts
837 bridge genes expressed in the neurons where the contacts are observed, with specialized
838 functions, such as in synaptic plasticity (PGNs) and addiction (DNs). (iv) Finally, B
839 compartments contain clusters of sensory receptor genes silent in all cell types which form
840 strong contacts across tens of megabases.
841

842 To investigate how compartment transitions between mESCs and the brain
843 cell types relate with gene expression changes, we clustered genes according to
844 compartment transition and expression in each cell type. Within the regions with
845 transitions from B-mESC → A-brain cells, we find 335 genes which share functions
846 and are more strongly expressed in all the brain cell types than in mESCs (**Extended**
847 **Data Fig. 10a**). These genes have enriched GO terms such as “behaviour”,
848 “regulation of synaptic plasticity” and “gated ion channel activity” (**Fig. 5c**). In regions
849 that undergo A-mESC → B-brain cells transitions, we remarkably find that most
850 genes are silent in all cell types (572/715 genes), and only 50 are more expressed in
851 mESCs (**Extended Data Fig. 10b**). Enriched GO terms for these 50 genes include
852 transcriptional regulation and transcription factor activity, and include many *Zfp*
853 genes (**Fig. 5c**), which encode zinc-finger binding proteins with roles in the
854 epigenetic regulation of pluripotency factors, and silencing of active genes during the
855 exit from pluripotency[75].

856 ***Clusters of sensory receptor genes are more strongly associated with B***
857 ***compartments in brain cells and establish megabase-range interactions***
858 We were intrigued by the large number of silent genes that undergo A-mESC
859 to B-brain-cell transitions, enriched for sensory receptor genes such as *Vmn*
860 (vomeronasal; 149 genes out of 572 silent genes in the group) and *Olfr* (olfactory;
861 179 genes), and which are most often found in gene clusters[76,77] (**Fig. 5d**). Although
862 silent, only 35% of *Vmn* genes (104/290) and 66% of *Olfr* genes (718/1095) are
863 associated with compartment B in mESCs, compared to 82-96% and 72-85% in brain
864 cells, respectively (**Fig. 5d**).
865
866 We explored how the stricter compartment B association of sensory receptors
867 in brain cells was related with chromatin contact formation in regions containing *Vmn*
868 and *Olfr* clusters. Matrices and compartments on chromosome 7 (which contains the
869 most *Vmn* and *Olfr* genes combined) showed that *Vmn* genes and/or *Olfr* genes are
870 often involved in strong clusters of contacts, which can span up to 50 Mb, and are
871 only observed in brain cells (**Fig**. **5e,f**, **Extended Data Fig. 10c**). For *Olfr* clusters in
872 **Fig. 5f**, the contacts between the two receptor clusters are stronger in OLGs where
873 their B compartment eigenvectors are also stronger. These clusters of strong
874 contacts are also weakly present in PGNs, but not in DNs where the second cluster
875 is in the A compartment. The long-range clusters of contacts on chromosome 17,
876 previously shown in **Fig. 1d,** also contain both *Vmn* and *Olfr* genes. We quantified
877 the strength of long-range contacts between *Vmn* or *Olfr* genes at distances >3 Mb in
878 non-adjacent compartments, by comparing the top 20% of distance-normalized Z-
879 Scores. The brain cell types show significantly stronger long-range contacts between
880 B-compartment regions containing *Vmn*, or to a lesser extent *Olfr* genes (**Extended**
881 **Data Fig**. **10d**). These results suggest that sensory genes are not only more locally

882 compartmentalized in heterochromatic B compartments in brain cells, but they also
883 form stronger contacts in 3D with other B compartments across tens of megabases,
884 a conformation that may be important to achieve stricter repression mechanisms in
885 the brain cells where these receptor genes are not expressed.
886
887 ***Discussion***
888 Here, we adapted the GAM method to capture genome-wide chromatin
889 conformation states of selected cell populations in the mouse brain that perform
890 specialized functions. Using immunoGAM, we selected nuclear slices from specific
891 brain cells by immunofluorescence prior to laser microdissection, and characterized
892 the relationship between 3D chromatin organization and specialized gene expression
893 programs from single mice using intact tissues. We discovered dramatic
894 reorganization of chromatin topology which reflects cell function and specialization at
895 multiple genomic scales, from specific contacts to compartment changes that span
896 tens of megabases (**Fig. 5g**). We found extensive domain restructuring with many
897 TAD borders being unique to each cell type and occurring at genes relevant for brain
898 cell specialization. Future work with higher resolution GAM datasets will enable
899 further analyses of the reorganisation of domains at finer genomic scales to
900 investigate changes in sub-TADs, shown previously to occur following cell
901 commitment to neuronal lineages[78], and whether the presence of neuronal genes at
902 TAD borders relates with their high expression in the absence of cell division, or are
903 important for interactions with cis-regulatory elements and transcriptional activation.
904
905 We find that the 3D structure of very long genes (>400kb to 2.2 Mb) with
906 highest expression levels and/or accessible chromatin in brain cells is largely
907 reorganized, by extensive decondensation, or ‘melting’, in structures reminiscent of

908 fly chromatin puffs induced by heatshock[79-81]. In neurons, melted genes often show
909 increased contacts within their own chromosome. Single-cell imaging of *Rbfox1*
910 showed that its overall decondensation in PGNs is most prominent between its
911 transcription start and end site, also reminiscent of increased spatial distance
912 reported for shorter genes between their promoters and enhancers, such as *Pax6*
913 with its enhancers when the gene is most expressed[82], or of PARylation-dependent
914 decreases in *Shh*-enhancer proximity during neuronal differentiation[83]. Many long
915 genes are regulated in a specialized manner in terminally differentiated cells, for
916 example by the activity of topoisomerases[45], by long stretches of broad H3K27ac and
917 H3K4me1 which act as enhancer-like domains[84], by large transcription loops[85],
918 and/or through repressive mechanisms such as DNA methylation[86]. The regulation of
919 long genes is further complicated by intricate splicing dynamics[35], which require
920 highly adaptive responses based on neuronal activation state. In fact, for many of the
921 genes we highlight, including *Nrxn3*, *Rbfox1, Grik2,* and *Dscam*, genetic variants are
922 associated with or directly result in neuronal diseases[36-38,43-44,48-50]. Thus,
923 understanding how genome folding relates with the response to environmental
924 challenges is increasingly important to further our understanding of the mechanisms
925 of neurological disease.
926
927 Our results suggest that chromatin architecture could be used to make
928 predictions of functional or disease states in a cell type. Rich cell-type specific
929 networks of contacts contain expressed genes with specialized functions in neurons
930 and are enriched for putative binding sites for TFs expressed differentially between
931 the cell types analysed. Motifs for several neurodevelopmental TFs were frequent
932 within the accessible chromatin (ATAC-seq) peaks in the most differential contacts in
933 PGNs and DNs, including *Neurod1, Neurod2* and *Foxa1*. These TF motifs can

934 connect hubs of differentially expressed genes with specialized functions in neurons,
935 consistent with known roles of these TFs in adult maintenance and in the regulation
936 of synaptic responses[54-57]. DN-specific loops contained heterotypic *Foxa1* or
937 *Neurod1/2* binding motifs, present in contact hubs containing genes related to the
938 response to drug exposure and addiction paradigms. DNs in the midbrain VTA are
939 the first cells in the brain to respond to a range of addictive substances, including
940 ethanol, amphetamines, nicotine and cocaine[87,88]. Therefore, future studies will be
941 important to understand the relationship between DN-specific chromatin landscapes,
942 disease-associated genetic variation and the regulation of these critical genes, with
943 potential implications for the onset of addiction.
944
945 In contrast with DNs, the PGN-specific contacts were enriched in complex
946 combinations of putative TF binding motifs, which connect hubs of differentially
947 expressed genes with roles in synaptic plasticity. PGNs frequently respond to
948 environmental cues during long-term potentiation (LTP) or depression (LTD), either
949 by external stimuli or by stochastic and intrinsic activation[15,16]. Strikingly, PGN-
950 specific contacts at the *Egr1* gene, involved in the early activation of LTP, were also
951 marked by accessible sites containing its own binding motifs, suggesting
952 mechanisms of self-activation by which expression of a TF increases contacts of the
953 TF with other regions containing its own binding sites. For example, positive
954 feedback loops of BDNF self-signalling at local synapses have been proposed as a
955 mechanism of LTP induction[89]. Finally, predictive polymer modelling has shown that
956 transcription factor binding can result in the self-assembly of specialized clusters
957 reminiscent of nucleoplasmic transcription factories[66,67]. Together with reports that *de*
958 *novo* chromatin looping can accompany transcriptional activation[7], our work suggests
959 that coordinated transcription-factor binding at apparent distant locations in the linear

960 genomic scale, but in close contact due to the 3D chromatin landscape, may be
961 critical for LTP induction.
962
963 We observed that highly repressed *Olfr* and *Vmn* genes are found in B
964 compartment regions in brain cells, which form strong cis-contacts tens of mega-
965 bases away. These strong B compartment contacts were especially noteworthy, as
966 *Olfr* genes can have non-sensory organ functions in neurons related to specificity of
967 neuronal activation[90]. *Olfr* genes are highly repressed in mature olfactory sensory
968 neurons, where they form a large inter-chromosomal hub to regulate specificity of
969 single *Olfr* gene activation[91,92]. Long-range interactions between *Olfr* genes were
970 independently found in CpG-poor regions of adult cortical neurons[93]. Our results
971 expand and complement these observations, highlighting that long-range
972 mechanisms of *Olfr* repression may also reflect their specialized roles in different cell
973 types of the brain. Tighter B compartmentalization of *Vmn* and *Olfr* genes may be
974 important for their repression in the brain cell types, as *Olfr* genes have been
975 reported to be stochastically activated and mis-expressed in cortical PGNs, in several
976 neurodegenerative diseases[90].
977
978 Finally, our work shows that immunoGAM is uniquely positioned to probe
979 questions related to cell state and specialized function within the brain or in other
980 complex tissues, as local tissue structure is maintained. ImmunoGAM has the
981 potential to be applied in multiple cell types within the same tissue while retaining the
982 geographic positions of each cell, a prospect that can further deepen our
983 understanding of coordinated interactions between cells in complex diseases. As
984 immunoGAM requires only small cell numbers to capture complex differences in
985 genomic topology (~1000 cells), from single individuals, it may also have prognostic

986 value in highly precious clinical patient samples, and to provide critical insights into
987 the etiology and progression of neurological disease. Collectively, our work revealed
988 that cell specialization in the brain and chromatin structure are intimately linked at
989 multiple genomic scales.
990

991 **References**

992 1. Jung, I., *et al.* A compendium of promoter-centered long-range chromatin
993 interactions in the human genome. *Nat. Genet.* **51**, 1442-1449 (2019).
994
995 2. Rao, S.S., *et al.* A 3D map of the human genome at kilobase resolution
996 reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
997
998 3. Beagrie, R.A., *et al.* Complex multi-enhancer contacts captured by genome
999 architecture mapping. *Nature* **543**, 519-524 (2017).
1000
1001 4. Quinodoz, S.A., *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D
1002 Genome Organization in the Nucleus. *Cell* **174**, 744-757 (2018).
1003
1004 5. Dixon, J.R., *et al.* Chromatin architecture reorganization during stem cell
1005 differentiation. *Nature* **518**, 331-336 (2015).
1006
1007 6. Fraser, J., *et al.* Hierarchical folding and reorganization of chromosomes are
1008 linked to transcriptional changes in cellular differentiation. *Mol. Syst. Biol.* **11**,
1009 852 (2015).
1010
1011 7. Beagan, J.A., *et al*. Three-dimensional genome restructuring across
1012 timescales of activity-induced neuronal gene expression. *Nat. Neurosci.* **23**,
1013 707-717 (2020).
1014
1015 8. Bonev, B., *et al.* Multiscale 3D Genome Rewiring during Mouse Neural
1016 Development. *Cell* **171**, 557-572 (2017).
1017
1018 9. Walczak, A., *et al.* Novel higher-order epigenetic regulation of the Bdnf gene
1019 upon seizures. *J. Neurosci.* **33**, 2507-2511 (2013).
1020
1021 10. Harabula, A., & Pombo, A. The dynamics of chromatin architecture in brain
1022 development and function. *Curr. Opin. Genet. Dev.* **67**, 84-93 (2021).
1023
1024 11. Beagrie, R.A., *et al.* Multiplex-GAM: genome-wide identification of chromatin
1025 contacts yields insights not captured by Hi-C. Preprint at *BioRxiv*
1026 https://doi.org/10.1101/2020.07.31.230284 (2021).
1027
1028 12. Van den Brink, S.C., *et al.* Single-cell sequencing reveals dissociation-induced
1029 gene expression in tissue subpopulations. *Nat. Methods* **14**, 935-936 (2017).
1030
1031 13. Fiorillo, L., *et al.* Comparison of the Hi-C, GAM and SPRITE methods by use
1032 of polymer models of chromatin. *Nature Methods* **18**, 482-490 (2021).
1033
1034 14. Huges, E.G., Orthmann-Murphy, J.L., Langseth, A.J., & Bergles, D.E. Myelin
1035 remodeling through experience-dependent oligodendrogenesis in the adult
1036 somatosensory cortex. *Nat. Neurosci.* **21**, 696-706 (2018).
1037
1038 15. Stackman Jr., R.W., Cohen, S.J., Lora, J.C., & Rios, L.M. Temporal
1039 inactivation reveals that the CA1 region of the mouse dorsal hippocampus

plays an equivalent role in the retrieval of long-term object memory and spatial memory. *Neurobiol. Learn. Mem.* **133**, 118-128 (2016).

16. Combe, C.L., Canavier, C.C., & Gasparini, S. Intrinsic Mechanisms of Frequency Selectivity in the Proximal Dendrites of CA1 Pyramidal Neurons. *J. Neurosci.* **38**, 8110-8127 (2018).

17. Keiflin, R., & Janak, P.H. Dopamine Prediction Errors in Reward Learning and Addiction: From Theory to Neural Circuitry. *Neuron* **88**, 247-263 (2015).

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

19. Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N., & Takeshi, Y. CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell Rep.* **2**, 345-357 (2012).

20. Monahan, K., *et al.* Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-α gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **23**, 9125-9130 (2012).

21. Yu, Y., Suo, L., & Wu, Q. Protocadherin α gene cluster is required for myelination and oligodendrocyte development. *Zoological Research.* **33**, 362-366 (2012).

22. Furusho, M., Dupree, J.L., Nave, K-A., & Bansal, R. Fibroblast Growth Factor Receptor Signaling in Oligodendrocytes Regulates Myelin Sheath Thickness. *J. Neurosci.* **32**, 6631-6641 (2012).

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

24. 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).

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

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

27. Cusanovich, D.A., *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324 (2018).

28. Sinnamon, J.R., *et al.* The accessible chromatin landscape of the murine hippocampus at single-cell resolution. *Genome Res.* **29**, 857-869 (2019).

29. Ben-Shalom, R., *et al.* Opposing Effects on Na_V 1.2 Function Underlie Differences Between SCN2A Variants Observed in Individuals With Autism Spectrum Disorder or Infantile Seizures. *Biol. Psychiatry* **82**, 224-232 (2017).

30. Lamar, T., *et al. SCN3A* deficiency associated with increased seizure susceptibility. *Neurobiol. Dis.* **102**, 38-48 (2017).

31. Wang, Q., Sun, Q., Czajkowsky, D.M., & Shao, Z. Sub-kb Hi-C in D. melanogaster reveals conserved characteristics of TADs between insect and mammalian cells. *Nat. Commun.* **9**, 188 (2018).

32. Petrovic, M.M., *et al.* Metabotropic action of postsynaptic kainate receptors triggers hippocampal long-term potentiation. *Nature Neurosci.* **20**, 529-529 (2017).

33. Zhang, D., *et al.* Protein kinase C delta negatively regulates tyrosine hydroxylase activity and dopamine synthesis by enhancing protein phosphatase-2A activity in dopaminergic neurons. *J. Neurosci.* **16**, 5349-5362 (2007).

34. Xu, Y., & Fisher, G.J. Receptor type protein tyrosine phosphatases (RPTPs) – roles in signal transduction and human disease. *J. Cell Commun. Signal.* **6**, 125-138 (2012).

35. Ray, T.A., *et al.* Comprehensive identification of mRNA isoforms reveals the diversity of neural cell-surface molecules with roles in retinal development and disease. *Nat. Commun.* **11**, 3328 (2020).

36. Xiong, S., Wang, Y., Li, H., & Zhang, X. Interaction among GRIK2 gene on epilepsy susceptibility in Chinese children. *Acta Neurol. Scand.* **139**, 540-545 (2019).

37. Guzmán, Y.F., *et al.* A gain-of-function mutation in the *GRIK2* gene causes neurodevelopmental deficits. *Neurol. Genet.* **3**, e129 (2017).

38. Guidi, S., *et al.* Neurogenesis impairment and increased cell death reduce total neuron number in the hippocampal region of fetuses with Down syndrome. *Brain Pathol.* **18**, 180-197 (2008).

39. Zuckerkandl, E. Gene control in eukaryotes and the c-value paradox "excess" DNA as an impediment to transcription of coding sequences. *J. Mol. Evol.* **9**, 73-104 (1976).

40. Nechaev, S., *et al.* Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in Drosophila. *Science* **327**, 335-338 (2012).

41. Boldyreva, L.V., *et al.* Protein and Genetic Composition of Four Chromatin Types in Drosophila melanogaster Cell Lines. *Curr. Genomics* **18**, 214-226 (2017).

42. Müller, W.G., Walker, D., Hafer, G.L., & McNally, J.G. Large-scale chromatin decondensation and recondensation regulated by transcription from a natural promoter. *J. Cell Biol.* **154**, 33-48 (2001).

43. Vaags, A.K., *et al.* Rare Deletions at the Neurexin 3 Locus in Autism Spectrum Disorder. *Am. J. Hum. Genet.* **90**, 133-141 (2012).

44. Martinez-Mir, A., *et al.* Genetic study of neurexin and neuroligin genes in Alzheimer's disease. *J. Alzheimers Dis.* **35**, 403-412 (2013).

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

46. Peric-Hupkes, D., *et al.* Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. *Mol. Cell* **38**, 603-613 (2010).

47. Bizhanova, A., Yan, A., Yu, J., Zhu, L.J., & Kaufman, P.D. Distinct features of nucleolus-associated domains in mouse embryonic stem cells. *Chromosoma* **129**, 121-139 (2020).

48. Hishimoto, A., *et al*. Neurexin 3 polymorphisms are associated with alcohol dependence and altered expression of specific isoforms. *Hum. Mol. Genet.* **16**, 2880-2891 (2007).

49. Lee, J-A., *et al.* Cytoplasmic Rbfox1 regulates the expression of synaptic and autism-related genes. *Neuron* **89**, 113-128 (2016).

50. Lal, D., *et al.* Extending the phenotypic spectrum of RBFOX1 deletions: Sporadic focal epilepsy. *Epilepsia* **56**, e129-133 (2015).

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

52. Branco, M.R., & Pombo, A. Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations. *PLoS Biol.* **4**, e138 (2006).

53. Valdez, B.C., *et al.* Identification of the Nuclear and Nucleolar Localization Signals of the Protein p120. *J. Biol. Chem.* **269**, 23776-23783 (1994).

54. Pristerà, A., *et al.* Transcription factors FOXA1 and FOXA2 maintain dopaminergic neuronal properties and control feeding behavior in adult mice. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4929-E4938 (2015).

55. Pataskar, A, *et al.* NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J.* **35**, 24-45 (2015).

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

57. Wilke, S.A., *et al.* NeuroD2 regulates the development of hippocampal mossy fiber synapses. *Neural Dev.* **7**, 9 (2012).

58. Barrington, C., *et al.* Enhancer accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type specific genome topology. *Nat. Commun.* **10**, 2908 (2019).

59. Sams, D.S., *et al.* Neuronal CTCF Is Necessary for Basal and Experience-Dependent Gene Regulation, Memory Formation, and Genomic Structure of BDNF and Arc. *Cell Rep.* **17**, 2418-2430 (2016).

60. Konrad, E.D.H., *et al.* CTCF variants in 39 individuals with a variable neurodevelopmental disorder broaden the mutational and clinical spectrum. *Genet. Med.* **21**, 2723-2733.

61. Bustos, F.J., *et al.* Epigenetic editing of the Dlg4/PSD95 gene improves cognition in aged and Alzheimer's disease mice. *Brain* **140**, 3252-3268 (2017).

62. Klaassen, R.V., *et al.* Shisa6 traps AMPA receptors at postsynaptic sites and prevents their desensitization during synaptic activity. *Nat. Commun.* **7**, 10682 (2016).

63. Coffey, E.T. Nuclear and cytosolic JNK signalling in neurons. *Nat. Rev. Neurosci.* **15**, 285-299 (2014).

64. Yoshino, Y., Roy, B., Kumar, N., Mukhtar, M.S., & Dwivedi, Y. Molecular pathology associated with altered synaptic transcriptome in the dorsolateral prefrontal cortex of depressed subjects. *Transl. Psychiatry* **11**. 73 (2021).

65. Li, P., *et al.* Epigenetic dysregulation of enhancers in neurons is associated with Alzheimer's disease pathology and cognitive symptoms. *Nat. Commun.* **10**, 2246 (2019).

66. Brackley, C.A., Johnson, J., Kelly, S., Cook, P.R., & Marenduzzo, D. Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucleic Acids Res.* **44**, 3503-3512 (2016).

67. Brackley, C.A., *et al.* Ephemeral Protein Binding to DNA Shapes Stable Nuclear Bodies and Chromatin Domains. *Biophys. J.* **112**, 1085-1093 (2017).

68. Kazemi, T., *et al.* Investigating the influence of perinatal nicotine and alcohol exposure on the genetic profiles of dopaminergic neurons in the VTA using miRNA–mRNA analysis. *Sci. Rep.* **10**, 15016 (2020).

69. Gorini, G., Roberts, A.J., & Mayfield, R.D. Neurobiological Signatures of Alcohol Dependence Revealed by Protein Profiling. *PLoS One* **8**, e82656 (2013).

70. Wang, J., *et al.* Genome-Wide Expression Analysis Reveals Diverse Effects of Acute Nicotine Exposure on Neuronal Function-Related Genes and Pathways. *Front. Psychiatry* **2**, 5 (2011).

71. Bell, R.L., *et al.* Gene expression changes in the nucleus accumbens of alcohol-preferring rats following chronic ethanol consumption. *Pharmacol. Biochem. Behav.* **94**, 131-147 (2010).

72. Repunte-Canonigo, V., *et al.* Identifying candidate drivers of alcohol dependence-induced excessive drinking by assembly and interrogation of brain-specific regulatory networks. *Genome Biol.* **16**, 68 (2015).

73. Sun, Z., *et al.* EGR1 recruits TET1 to shape the brain methylome during development and upon neuronal activity. *Nat. Commun.* **10**, 3892 (2019).

74. Oudelaar, A.M., *et al.* Dynamics of the 4D genome during in vivo lineage specification and differentiation. *Nat. Commun.* **11**, 2722 (2020).

75. Kwak, S., *et al.* Zinc finger proteins orchestrate active gene silencing during embryonic stem cell differentiation. *Nucleic Acids Res.* **46**, 6592-6607 (2018).

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

77. 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).

78. Phillips-Cremins, J.E., *et al.* Architectural protein subclasses shape 3-D organization of genomes during lineage commitment. *Cell* **153**, 1281-1295 (2013).

79. Ashburner, M., & Bonner, J.J. The induction of gene activity in drosophila by heat shock. *Cell* **17**, 241-254 (1979).

80. Lis, J.T., Simon, J.A., & Sutton, C.A. New heat shock puffs and β=galactosidase activity resulting from transformation of Drosophila with an *hsp70-lacZ* hybrid gene. *Cell* **35**, 403-410 (1983).

81. Berman, B.P., *et al.* Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol.* **5**, R61 (2004).

82. Buckle, A., Brackley, C.A., Boyle, S., Marenduzzo, D., & Gilbert, N. Polymer Simulations of Heteromorphic Chromatin Predict the 3D Folding of Complex Genomic Loci. *Mol. Cell* **72**, 786-797 (2018).

83. Benabdallah, N.S., *et al.* Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol. Cell* **76**, 473-484 (2019).

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

85. Leidescher, S., *et al.* Spatial organisation of transcribed eukaryotic genes. Preprint at *BioRxiv* https://doi.org/10.1101/2020.05.20.106591 (2021).

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

87. Lüscher, C., & Ungless M. The mechanistic classification of addictive drugs. *PLoS Med.* **3**, e437 (2006).

88. Lüscher, C., & Malenka R.C. Drug-evoked synaptic plasticity in addiction: from molecular changes to circuit remodeling. *Neuron* **69**, 650-663 (2011).

89. Hao, L., Yang, Z., & Lei, J. Underlying Mechanisms of Cooperativity, Input Specificity, and Associativity of Long-Term Potentiation Through a Positive Feedback of Local Protein Synthesis. *Front. Comput. Neurosci.* **12**, 25 (2018).

90. 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).

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

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

93. Tan, L., *et al.* Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell* **184**, 741-758 (2021).

94. 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).

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

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

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

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

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

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

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

102. Lex, A., *et al.* UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* **20**, 1983-1992 (2014).

103. 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).

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

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

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

107. 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).

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

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

110. 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.

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

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

113. 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).

114. Picelli, S., *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24**, 2033-2040 (2014).

115. Corces, M.R., *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959-962 (2017).

116. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032-2034 (2015).

117. Granja, J.M., *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403-411 (2021).

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

119. 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).

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

121. Fiorillo, L. *et al.* Inference of chromosome 3D structures from GAM data by a physics computational approach. *Methods* S1046-2023(18)30485-7. (2019).

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

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

124. Schindelin, J., *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676-682 (2012).

125. Kulakovskiy, I.V., *et al.* HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252-D259 (2018).

126. *Discovering Knowledge in Data*. D.R. Larose and C.D. Larose. Wiley, Haboken, New Jersey (2014).

127. Traag, V.A., Waltmann, L., & van Eck, N.J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

128. Kelly, S.T. leiden: R implementation of the Leiden algorithm. R package version 0.3.3, https://github.com/TomKellyGenetics/leiden (2019).

1443 **Acknowledgments**

1444 The authors thank Sheila Q. Xie and Azhaar Ashraf for help processing midbrain
1445 samples, Marek Bartosovic for help with manuscript revisions, Michael Gotthardt for
1446 providing animals and Rene Jüttner for help with dissections of midbrain samples for
1447 the snATAC-seq, Catherine Baugher for developing the TF motif analysis pipeline,
1448 and the Pombo lab members for helpful discussions. We thank Caroline Braeuning of
1449 the Scientific Genomics Platform for scientific and technical support, and Anja Schütz
1450 and her team of the Protein Production and Characterization Platform, both at the
1451 Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin. We
1452 thank Andrew Adey, who kindly provided a bam file for the published PGN scATAC-
1453 seq analysis.

1454 AP and AA acknowledge support from the Helmholtz Association (Germany). AP and
1455 MN acknowledge support from the National Institutes of Health Common Fund 4D
1456 Nucleome Program grants U54DK107977 and 1UM1HG011585, and the Berlin
1457 Institute of Health (BIH). AP and LZR acknowledge support by the Deutsche
1458 Forschungsgemeinschaft (DFG; German Research Foundation) International
1459 Research Training Group (IRTG2403). AP acknowledges support from the Deutsche
1460 Forschungsgemeinschaft (DFG; German Research Foundation) under Germany's
1461 Excellence Strategy – EXC-2049 – 390688087. GCB acknowledges European Union
1462 Horizon 2020/European Research Council Consolidator Grant (EPIScOPE no.
1463 681893), Swedish Research Council (no. 2015-03558; 2019-01360), Swedish Brain
1464 Foundation (no. FO2017-0075), Knut and Alice Wallenberg Foundation (grant 2019-
1465 0107), The Swedish Society for Medical Research (SSMF, grant JUB2019), Ming
1466 Wai Lau Centre for Reparative Medicine and Karolinska Institutet. GD and GA
1467 acknowledge support from the Austrian FWF through DK W1206 "Signal Processing
1468 in Neurons" and SFB F44 "Cell Signaling in Chronic CNS Disorders", P25014-B24.

1469 MAU acknowledges funding by the Medical Research Council (UK) (U120085816)
1470 and a Royal Society University Research Fellowship. MN thanks support from
1471 CINECA ISCRA Grant HP10CYFPS5 and HP10CRTY8P, by computer resources at
1472 INFN and Scope at the University of Naples (MN). LW acknowledges the support of
1473 Ohio University's GERB program. IH was supported by a Boehringer Ingelheim
1474 Fonds PhD fellowship, and ETT by an EMBO short-term fellowship (ASTF 336-
1475 2015). II-A was supported by a Long-Term Fellowship from the Federation of
1476 European Biochemical Societies (FEBS). SC is a GABBA PhD fellow supported by
1477 the FCT (Fundação para a Ciência e Tecnologia; PD/BD/135453/2017).
1478

1479 **Author Contributions**

1480 AP designed the concept for this work; WWN, MM, AA, EJP and AP collected animal
1481 tissues; IH, WWN, LS, AK, MM, and RK produced GAM datasets; IH, LS, AK, WWN,
1482 and MM optimised the experimental immunoGAM protocol; WWN, AK, CJT, II-A, EI
1483 and TMS developed computational pipelines for bioinformatics and QC analyses of
1484 GAM data; AK and II-A performed QC analyses of GAM data; TMS developed the
1485 NPMI normalization of GAM data and performed bias analysis; WWN, AK and DS
1486 performed bioinformatics analyses of GAM data; DS developed the domain melting
1487 analysis, with consultation from CJT; DS devised the MELTRON pipeline and
1488 performed the domain melting analysis of GAM data; DS developed the trans-cis
1489 ratio analysis; CJT and WWN developed the differential contact analysis; WWN
1490 performed differential contact analysis; YZ performed TF motif finding enrichment,
1491 network, and community analyses in differential contacts; WWN performed post-hoc
1492 analyses of TF motif enrichment in differential contacts; ETT performed mESC
1493 culture experiments; ETT and AAK produced single-cell RNAseq data; LZR and DS
1494 performed RNA-seq analysis; LZR performed differential gene expression analysis;

1495 LZR optimised the single-nucleus ATAC-seq protocol, produced and analyzed bulk
1496 and single-cell ATAC-seq data; SB optimized polymer modeling and performed
1497 PRISMR analysis; SB and AMC produced models for polymer modeling; AMC
1498 performed statistical analyses of polymer models; LF and FM performed in-silico
1499 GAM experiments; IH and SC performed and analyzed fluorescence *in-situ*
1500 hybridization (FISH) experiments; WWN, DS, IH, LZR, YZ, AK, TMS, SB, AMC, and
1501 MM produced the final data plots; SAT supervised scRNA-seq experiments; MM, GA,
1502 GD, MU and GCB supervised animal tissue collection and provided animal samples;
1503 AP supervised GAM experiments and bioinformatics analyses, cryo-FISH
1504 experiments and analysis, and ATAC-seq experiments and analysis; VF, AP, and AA
1505 supervised the scRNA-seq analysis; MN supervised the polymer modeling and in-
1506 silico GAM; LW supervised the TF motif and network analyses; WWN, AP, AK, IH,
1507 DS, LZR, LW, YZ, GCB, MM, SB, AMC, and MN contributed to the interpretation of
1508 the results; WWN wrote the first draft of the manuscript; WWN designed the figures
1509 and illustrations, with contributions from AP, DS and IH; WWN, AP, AK, IH, DS, LZR,
1510 YZ, LW, GCB, SC, SB, and AMC wrote the manuscript. All authors provided critical
1511 feedback and helped revise the manuscript. The authors consider WWN, AK and IH
1512 to have contributed equally to this work. The authors also consider LZR and DS to
1513 have contributed equally to this work.
1514

1515 **Competing interests**

1516 In the past 3 years, SAT has acted as a consultant for Genentech and Roche, and is
1517 a remunerated member of Scientific Advisory Boards of Biogen, GlaxoSmithKline
1518 and Foresite Labs.

1519 AP and MN hold a patent on ‘Genome Architecture Mapping’: Pombo, A., Edwards,

1520 P. A. W., Nicodemi, M., Beagrie, R. A. & Scialdone, A. Patent PCT/EP2015/079413

1521 (2015).

1522

1523 **Materials & Correspondence**

1524 Correspondence and material requests should be addressed to ana.pombo@mdc-

1525 berlin.de or warren.winick-ng@mdc-berlin.de

1526

Methods

Animal maintenance

Collection of GAM data from dopaminergic neurons was performed using one C57Bl/6NCrl (RRID: IMSR_CR:027; WT) mouse which were purchased from Charles River, and from one TH-GFP (B6.Cg-Tg(TH-GFP)21-31/C57B6) mouse, obtained as previously described[94,95]. 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 single-nucleus ATAC-seq (snATAC-seq) data from midbrain VTA was performed using male C57Bl/6Nl (RRID: IMSR_CR:027; WT) mice, ages 7 and 9 weeks, which were a kind gift from Michael Gotthardt. All procedures involving mice for snATAC-seq approved by the regulations of local animal care committee (LAGeSo, Berlin, Germany) and the animal welfare officers of the Max Delbrück Center for Molecular Medicine (MDC

Collection of GAM data from somatosensory oligodendrocyte cells was performed using Sox10::Cre-RCE::loxP-EGFP[96] animals which were obtained by crossing Sox10::Cre animals[97] on a C57BL/6j genetic background with RCE::loxP-EGFP animals[98] 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. One male mouse was 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 two 19 week-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[99]. 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

1571 WT, TH-GFP, and *Satb2*$^{flox/flox}$ mice were anaesthetised under isoflurane (4
1572 %), given a lethal IP injection of pentobarbital (0.08 µl; 100 mg/ml; Euthatal), and
1573 transcardially perfused with 50 ml of ice-cold phosphate buffered saline (PBS)
1574 followed by 50-100 ml of 4% depolymerised paraformaldehyde (PFA; Electron
1575 microscopy grade, methanol free) in 250 mM HEPES-NaOH (pH 7.4-7.6).
1576 Sox10::Cre-RCE::loxP-EGFP animals were sacrificed with a ketaminol/xylazine
1577 intraperitoneal injection followed by transcardial perfusion with 20 ml PBS and 20 ml
1578 4% PFA in 250 mM HEPES (pH 7.4-7.6). From C57Bl/6NCrl or TH-GFP mice, brains
1579 were removed and the tissue containing the VTA were dissected from each
1580 hemisphere at room temperature, and quickly transferred to fixative. For *Satb2*$^{flox/flox}$
1581 mice, the CA1 field hippocampus was dissected from each hemisphere at room
1582 temperature. For Sox10Cre/RCE mice, brain tissue containing the somatosensory
1583 cortex was dissected at room temperature. Following dissection, tissue blocks were
1584 placed in 4% paraformaldehyde (PFA) in 250 mM HEPES-NaOH (pH 7.4-7.6) for
1585 post-fixation at 4°C for 1 h. Brains were then placed in 8% PFA in 250mM HEPES
1586 and incubated at 4°C for 2-3 h. Tissue blocks were then placed in 1% PFA in 250
1587 mM HEPES and kept at 4°C until tissue was prepared for cryopreservation (up to 5
1588 days, with daily solution changes).
1589
1590 ***Cryoblock preparation and cryosectioning***
1591 Fixed tissue samples from different brain regions were further dissected to
1592 produce ~1.5x3 mm tissue samples suitable for Tokuyasu cryosectioning[3] (**Extended**
1593 **Data Fig. 1a**), at room temperature in 1% PFA in 250 mM HEPES. For the
1594 hippocampus, the dorsal CA1 region was further isolated. Approximately 1-3 mm x 1-
1595 3 mm blocks were dissected from all brain regions and were further incubated in 4%
1596 PFA in 250 mM HEPES at 4°C for 1 h. The fixed tissue was transferred to 2.1 M
1597 sucrose in PBS and embedded 16-24 h, at 4°C, positioned at the top of copper stub
1598 holders suitable for ultracryomicrotomy and frozen in liquid nitrogen. Cryopreserved
1599 tissue samples are kept indefinitely immersed under liquid nitrogen.
1600
1601 Frozen tissue blocks were cryosectioned with a Ultracryomicrotome (Leica
1602 Biosystems, EM UC7), with an approximate thickness of 220-230nm[3]. Cryosections
1603 were captured in drops of 2.1 M sucrose in PBS solution suspended in a copper wire
1604 loop and transferred to 10 mm glass coverslips for confocal imaging, or onto a 4.0
1605 µm polyethylene naphthalate (PEN; Leica Microsystems, 11600289) membrane on
1606 metal framed slides for laser microdissection.
1607
1608 ***Immunofluorescence detection of GAM samples for confocal microscopy***
1609 For confocal imaging, cryosections were incubated in sheep anti-tyrosine
1610 hydroxylase (TH, 1:500; Pel Freez Arkansas LLC, Cat#P60101-0), mouse anti-pan-
1611 histone H11-4 (1:500; Merck, Cat#MAB3422) or chicken anti-GFP (1:500; Abcam
1612 Cat#ab13970) followed by donkey anti-sheep or goat anti-chicken IgG conjugated
1613 with AlexaFluor-488 (for TH and GFP; Invitrogen) or donkey anti-mouse IgG
1614 conjugated with AlexaFluor-555 or AlexaFluor-488 (for pan-histone; Invitrogen).

1615
1616 For PGNs, cryosections were washed (3x, 30 min total) in PBS, permeabilized
1617 (5 min) in 0.3% Triton X-100 in PBS (v/v) and incubated (2 h, room temperature) in
1618 blocking solution (1% BSA (w/v), 5% FBS (w/v, GibcoTM Cat#10270), 0.05% Triton
1619 X-100 (v/v) in PBS). After incubation (overnight, 4ºC) with primary antibody in
1620 blocking solution, the cryosections were washed (3-5x; 30 min) in 0.025% Triton X-
1621 100 in PBS (v/v) and immunolabeled (1 h, room temperature) with secondary
1622 antibodies in blocking solution, followed by 3 (15 min) washes in PBS. Cryosections
1623 were then counterstained (5 min) with 0.5 µg/ml 4',6'-diamino-2-phenylindole (DAPI;
1624 Sigma-Aldrich® Cat#D9542) in PBS, and then rinsed in PBS and water. Coverslips
1625 were mounted in Mowiol® 4-88 solution in 5% glycerol, 0.1M Tris-HCl (pH 8.5).
1626
1627 The number of SATB2-positive cells present in the hippocampal CA1 area of
1628 the *Satb2*$^{flox/flox}$ control mice was determined by counting nuclei positive for the
1629 SATB2 immunostaining (1:100; Abcam, Cat#ab10563678). To avoid counting the
1630 same nuclei, only every 30th ultrathin section cut through the tissue was collected,
1631 and the remaining sections discarded. Twenty-five nuclei were identified in the
1632 pyramidal neurons layer per image in the DAPI channel only and SATB2-positive
1633 cells counted. We confirmed that within the CA1 layer most cells (96%) are pyramidal
1634 glutamatergic neurons (data not shown).
1635
1636 For DNs and OLGs, cryosections were washed (3x, 30min total) in PBS,
1637 quenched (20 min) in PBS containing 20mM glycine**,** then permeabilized (15 min) in
1638 0.1% Triton X-100 in PBS (v/v). Cryosections were then incubated (1h, room
1639 temperature) in blocking solution (1% BSA (w/v), 0.2% fish-skin gelatin (w/v), 0.05%
1640 casein (w/v) and 0.05% Tween-20 (v/v) in PBS). After incubation (overnight, 4ºC)
1641 with the antibody in blocking solution, the cryosections were washed (3-5x; 1 h) in
1642 blocking solution and immunolabeled (1 h, room temperature) with secondary
1643 antibodies in blocking solution, followed by 3 (15 min) washes in 0.5% Tween-20 in
1644 PBS (v/v). Cryosections were then counterstained with 0.5 µg/ml DAPI in PBS, then
1645 rinsed in PBS. Coverslips were mounted in Mowiol® 4-88.
1646
1647 Digital images were acquired with a Leica TCS SP8-STED confocal
1648 microscope (Leica Microsystems) using a 63x oil-immersion objective (NA = 1.4) or a
1649 20x oil-immersion objective, using pinhole equivalent to 1 Airy disk. Images were
1650 acquired using 405 nm excitation and 420-480 nm emission for DAPI; 488 nm
1651 excitation and 505-530 nm emission for TH or GFP; and 555 nm excitation and 560
1652 nm emission using a long-pass filter at 1024x1024 pixel resolution. Images were
1653 processed using Fiji (version 2.0.0-rc-69/1.52p), where adjustments included the
1654 optimization of the dynamic signal range with contrast stretching.
1655
1656 ***Immunofluorescence detection of GAM samples for laser microdissection***
1657 For laser microdissection, cryosections on PEN membranes were washed,
1658 permeabilized and blocked as for confocal microscopy, and incubated with primary
1659 and secondary antibodies as indicated above except for the use of higher

1660 concentrations of primary antibodies, as follows: anti-TH (1:50), anti-pan-histone
1661 (1:50) or anti-GFP (1:50). Secondary antibodies were used at the same
1662 concentration. Cell staining was visualized using a Leica laser microdissection
1663 microscope (Leica Microsystems, LMD7000) using a 63x dry objective. Following
1664 detection of cellular sections of the cell types of choice containing nuclear slices
1665 (nuclear profiles; NPs), individual NPs were laser microdissected from the PEN
1666 membrane, and collected into PCR adhesive caps (AdhesiveStrip 8C opaque; Carl
1667 Zeiss Microscopy #415190-9161-000). We used multiplexGAM[11] where three NPs
1668 are collected into each adhesive cap, and the presence of NPs in each lid was
1669 confirmed with a 5x objective using a 420-480 nm emission filter. Control lids not
1670 containing nuclear profiles (water controls) were included for each dataset collection
1671 to keep track of contamination and noise amplification of whole genome amplification
1672 and library reactions, and can be found in **Supplemental Table 2.**
1673
1674 ***Whole genome amplification of nuclear profiles***
1675 Whole genome amplification (WGA) was performed using an in-house
1676 protocol. Briefly, NPs were lysed directly in the PCR adhesive caps for 4 (or 24h, for
1677 160/585 GAM samples from DNs R1) at 60°C in 1.2x lysis buffer (30 mM Tris-HCl pH
1678 8.0, 2 mM EDTA pH 8.0, 800 mM Guanidinium-HCl, 5 % (v/v) Tween 20, 0.5 % (v/v)
1679 Triton X-100), containing 2.116 units/ml QIAGEN protease (Qiagen, Cat#19155**).**
1680 After protease inactivation at 75°C for 30 min, the extracted DNA was amplified using
1681 random hexamer primers with an adaptor sequence. The pre-amplification step was
1682 done using 2x DeepVent mix (2x Thermo polymerase buffer (10x), 400**µ**m dNTPs, 4
1683 mM $MgSO_4$ in ultrapure water), 0.5 µM GAT-7N primers (5′- GTG AGT GAT GGT
1684 TGA GGT AGT GTG GAG NNN NNN N) and 2 units/µl DeepVent® (exo-) DNA
1685 polymerase (New England Biolabs, M0259L) in the programmable thermal cycler for
1686 11 cycles. Primers that anneal to the general adaptor sequence were then used in a
1687 second exponential amplification reaction to increase the amount of product. The
1688 exponential amplification was done using 2x DeepVent mix, 10 mM dNTPs, 100 µM
1689 GAM-COM primers (5′-GTG AGT GAT GGT TGA GGT AGT GTG GAG) and 2
1690 units/µl DeepVent (exo-) DNA polymerase in the programmable thermal cycler for 26
1691 cycles. For a small number of NPs from DNs (see **Supplemental Table 1**), WGA
1692 was performed using a WGA4 kit (Sigma-Aldrich) using the manufacturer's
1693 instructions; the recent formulation of this kit is no longer suitable for GAM data
1694 production from sub-cellular nuclear slices.
1695
1696 ***GAM library preparation and high-throughput sequencing***
1697 Following WGA, the samples were purified using SPRI beads (0.725 or 1.7
1698 ratio of beads per sample volume). The DNA concentration of each purified sample
1699 was measured using the Quant-iT® Pico Green dsDNA assay kit (Invitrogen #P7589)
1700 according to manufacturer's instructions. GAM libraries were prepared using the
1701 Illumina Nextera XT library preparation kit (Illumina #FC-131-1096) following the
1702 manufacturer's instructions with an 80% reduced volume of reagents. Following
1703 library preparation, the DNA was purified using SPRI beads (1.7 ratio of beads per
1704 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 2.**

Tn5-based libraries are preferred for GAM data sequencing to increase fragment sequence variation, which helps avoid the need for dark cycles in the current Illumina machines. This choice greatly reduces the cost of sequencing and decreases the frequency of noise reads from absent windows seen with the previous protocol[3].

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[100]. All non-uniquely mapped reads, reads with mapping quality <20 and PCR duplicates were 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[101]. 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 2**). Most GAM libraries passed the quality control analyses (86-96% in each dataset; **Extended Data Fig. 1b-d**). 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. 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 determined per collection plate of <0.4 (Jaccard index). The number of samples in each cell type passing quality control is summarized in **Extended Data Fig. 1d**. Following quality control analysis, we noted that the 160 (of 585) DN R1 samples incubated with lysis buffer for 24h had decreases in orphan windows (median = 26% and 36% for 24h and 4h, respectively) and increases in total genome coverage (median = 9% and 6% for 24h and 4h, respectively). Though these differences were minor, we recommend 24h lysis for future work.

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

For mESCs, GAM datasets were downloaded from 4D Nucleome portal (https://data.4dnucleome.org/). 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 (Merck, Cat# SF001- B). The list of 4DN sample IDs is provided in **Supplemental Table 1**.

Visualization of pairwise chromatin contact matrices

To visualize GAM data, contact matrices were calculated using pointwise mutual information (PMI) for all pairs of windows genome-wide. PMI describes the difference between the probability of a pair of genomic windows being found in the same NP given both their joint distribution and their individual distributions 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:

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

For visualization of contact matrices, scale bars are adjusted in each genomic region displayed to a range between 0 and the 99th percentile of NPMI values for each cell type.

Insulation score and topological domain boundary calling

TAD calling was performed by calculating insulation scores in NPMI GAM contact matrices at 50 kb resolution, as previously described[3,11]. The insulation score was computed individually for each cell type and biological replicate, with insulation square sizes ranging from 100 to 1000 kb. TAD boundaries were called using a

1795 500kb-insulation square size and based on local minima of the insulation score.
1796 Within each dataset, boundaries that were touching or overlapping by at least 1
1797 nucleotide were merged. Boundaries were further refined to consider only the
1798 minimum insulation score within the boundary and 1 window on each side, to
1799 produce a 3 bin "minimum insulation score" boundary. In comparisons of boundaries
1800 between different datasets, boundaries were considered different when separated by
1801 at least one genomic bin, i.e. if they are separated by at least 50 kb. In **Fig. 2b,** we
1802 considered the boundary coordinate as the genomic window within a boundary with
1803 the lowest insulation value. TAD border coordinates for all cell types can be found in
1804 **Supplemental Table 3**, and the full range of insulation scores (100 - 1000kb) for all
1805 cell types can be found in **Supplemental Table 4**. UpSet plots for TAD border
1806 overlaps, compartments, and TF motif analyses were generated either with custom
1807 Python or R scripts, or using the *UpSetR* tool[102].
1808
1809 ***Identification of compartments A and B***
1810 For compartment analysis, matrices of co-segregation frequency were
1811 determined by the ratio of independent occurrence of a single positive window in
1812 each sample over the pairwise co-occurrence of pairs of positive windows in a given
1813 pair of genomic windows[3]. GAM co-segregation matrices at 250 kb resolution were
1814 assigned to either A- or B-compartments, as previously described[3]. Briefly, each
1815 chromosome was represented as a matrix of observed interactions O(i,j) between
1816 locus i and locus j (co-segregation), and separately for E(i,j) where each pair of
1817 genomic windows are the mean number of contacts with the same distance between
1818 i and j. A matrix of observed over expected values O/E(i,j) was produced by dividing
1819 O by E. A correlation matrix C(i,j) is produced between column i and column j of the
1820 O/E matrix. Principal component analysis was performed for the first 3 components
1821 on matrix C, before extracting the component with the best correlation to GC content.
1822 Loci with PCA eigenvector values with the same sign that correlate best with GC
1823 content were called A compartments, while regions with the opposite sign were B
1824 compartments. For visualizations and Pearson correlations between datasets,
1825 eigenvector values on the same chromosome in compartment A were normalized
1826 from 0 to 1, while values on the same chromosome in compartment B were
1827 normalized from -1 to 0. Compartments were considered common if they had the
1828 same compartment definition within the same genomic bin. Compartment changes
1829 between cell types were computed after considering compartments that were
1830 common between biological replicates, unless otherwise indicated.
1831
1832 To identify and visualize gene expression differences among genes in
1833 changing compartments, K-means clustering was performed on triplicate pseudo-
1834 replicates of each cell type using a custom Python script (**Extended Data Fig. 10a-**
1835 **b**). The number of clusters were determined using the elbow method, with k-means =
1836 6 for genes in B compartment in mESCs and A in brain cells, and k-means = 5 for A
1837 compartment in mESCs and B in brain cells.
1838
1839 ***mESC cell culture for single-cell RNA-seq and ATAC-seq***

1840 Mouse embryonic stem cells from the 46C clone, derived from E14tg2a and
1841 expressing GFP under *Sox1* promoter[103], were a kind gift from D. Henrique (Instituto
1842 de Medicina Molecular, Faculdade Medicina Lisboa, Lisbon, Portugal). mESCs were
1843 cultured as previously described[26], i.e. cells were routinely grown at 37°C, 5% (v/v)
1844 CO_2, on gelatine-coated (0.1% v/v) Nunc T25 flasks in Gibco® Glasgow's MEM
1845 (Invitrogen, Cat# 21710082), supplemented with 10% (v/v) Foetal Calf Serum (FCS;
1846 BioScience LifeSciences, Cat# 7.01, batch number 110006) for scRNA-seq or
1847 Gibco® Fetal Bovine Serum (FBS; Invitrogen, Cat# 10270-106, batch number
1848 41F8126K) for ATAC-seq, 2000 U/ml Leukaemia inhibitory factor (LIF, Millipore, Cat#
1849 ESG1107), 0.1 mM beta-mercaptoethanol (Invitrogen, Cat# 31350-010), 2 mM L-
1850 glutamine (Invitrogen, Cat# 25030-024), 1 mM sodium pyruvate (Invitrogen, Cat#
1851 11360070), 1% penicillin-streptomycin (Invitrogen, Cat# 15140122), 1% MEM Non-
1852 Essential Amino Acids (Invitrogen, Cat# 11140035). Medium was changed every day
1853 and cells were split every other day. mESC batches all tested negative for
1854 Mycoplasma infection, performed according to the manufacturer's instructions
1855 (AppliChem, Cat#A3744,0020). Before collecting material for scRNA-seq or ATAC-
1856 seq, cells were grown for 48h in serum-free ESGRO Complete Clonal Grade Medium
1857 (Merck, Cat# SF001- B) supplemented with 1000 U/ml LIF, on gelatine (Sigma, Cat#
1858 G1393-100ml)-coated (0.1% v/v) Nunc 10 cm dishes, with a medium change after
1859 24h.
1860
1861 ***Single-cell mRNA library preparation***
1862 Two batches (denoted batch A and B, respectively) of single-cell mRNA-seq
1863 libraries were prepared according to Fluidigm manual "Using the C1 Single-Cell Auto
1864 Prep System to Generate mRNA from Single Cells and Libraries for Sequencing".
1865 Cell suspension was loaded on 10-17 microns C1 Single-Cell Auto Prep IFC
1866 (Fluidigm, Cat# 100-5760, kit Cat# 100-6201). After loading, the chip was observed
1867 under the microscope to score cells as singlets, doublets, multiplets, debris or other.
1868 The chip was then loaded again on Fluidigm C1 and cDNA was synthesised and pre-
1869 amplified in the chip using Clontech SMARTer kit (Takara Clontech, Cat# 634833). In
1870 batch B, we included Spike-In Mix 1 (1:1000; Life Technologies, Cat# 4456740), as
1871 from the Fluidigm manual. Illumina sequencing libraries were prepared with Nextera
1872 XT kit (Illumina, Cat# FC- 131-1096) and Nextera Index Kit (Illumina, Cat# FC-131-
1873 1002), as previously described[104]. Libraries from each microfluidic chip (96 cells)
1874 were pooled and sequenced on 4 lanes on Illumina HiSeq 2000, 2x100bp paired-end
1875 (batch A) or 1 lane on Illumina HiSeq 2000, 2x125bp paired-end (batch B) in the
1876 Wellcome Trust Sanger Institute Sequencing Facility (**Supplemental Table 13**).
1877
1878 ***ScRNA-seq data processing: mapping and expression estimates***
1879 To calculate expression estimates, mRNA-seq reads were mapped with STAR
1880 (Spliced Transcripts Alignment to a Reference, v2.4.2a)[105] and processed with
1881 RSEM, using the 'single-cell-prior' option (RNA-Seq by Expectation- Maximization,
1882 v1.2.25)[106]. The references provided to STAR and RSEM were the GTF annotation
1883 from UCSC Known Genes (mm10, version 6) and the associated isoform-gene
1884 relationship information from the Known Isoforms table (UCSC), adding information

1885 for ERCCs sequences in samples from batch B. Tables were downloaded from the
1886 UCSC Table browser (http://genome.ucsc.edu/cgi- bin/hgTables) and, for ERCCs,
1887 from ThermoFisher website (www.thermofisher.com/order/catalog/product/4456739).
1888 Gene-level expression estimates in "Expected Counts" from RSEM were used for the
1889 analysis.
1890
1891 ***ScRNA-seq data processing: quality control***
1892 Cells scored as doublets, multiplets, or debris during visual inspection of the
1893 C1 chip were excluded from the analysis. Datasets were also excluded if any of the
1894 following conditions were met: <500,000 reads (calculated using sam-stats from ea-
1895 utils.1.1.2-537)[107]; <60% of reads mapped (calculated with sam-stats); <50% reads
1896 mapped to mRNA (picard-tools-2.5.0, broadinstitute.github.io/picard/); >15% of reads
1897 mapped to chrM (sam-stats); if present, >20% of reads mapped to ERCCs (sam-
1898 stats). Following processing, 98 single cells passed quality thresholds in the final
1899 dataset. Correlations between previously published mESCs (clone 46C) mRNA-seq
1900 bulk[26] and the single-cell RNAseq mESC transcriptomes were performed to assess
1901 quality of the single cell data. Correlations were performed as previously
1902 described[108]. Average single cell expression was highly correlated with bulk RNA-seq
1903 (**Extended Data Fig. 3c**).
1904
1905 ***ScRNA-seq integration and analysis***
1906 To utilize published single cell transcriptomes from brain cell types of interest,
1907 we selected p21-22 OLGs[23], p22-32 CA1 PGNs[24], and p21-26 VTA DNs[25], based on
1908 the cell type and subtype definitions provided in the respective publications. The
1909 matrices of counts provided in each publication, along with the single-cell mESC
1910 transcriptomes produced which passed QC, were normalized applying the
1911 LogNormalize method and scaled using Seurat[109]. The scaled data was used for a
1912 PCA analysis, followed by processing through dimensionality reduction using Uniform
1913 Manifold Approximation and Projection (UMAP)[110] for visualization purposes, using
1914 the Seurat R package[109], with default parameters. Single mESC transcriptomes from
1915 batch A and B clustered together, and were pooled for further analyses. Genes that
1916 could not be mapped to the chosen reference GTF were removed (UCSC; accessed
1917 from iGenomes July 17, 2015;
1918 https://support.illumina.com/sequencing/sequencing_software/igenome.html.
1919
1920 To generate bigwig tracks for visualization, raw fastq files from each cell type
1921 were pooled into one fastq file. Reads were mapped to the mouse genome (mm10)
1922 using STAR with default parameters but --outFilterMultimapNmax 10. Bam files were
1923 sorted and indexed using Samtools[111] and normalized (RPKM) bigwigs were
1924 generated using Deeptools[112] bamCoverage. To account for differences in the
1925 number of technical replicates in OLG samples, cells were divided into groups by
1926 number of runs (1, 2 and 6). The median of the reads for the group with the lowest
1927 sequencing depth was used as a threshold to normalize the other groups (i.e. the
1928 rest of the fastq files were randomly downsampled to that number of reads). The
1929 three groups of raw reads were pooled together and processed by applying the same

1930 method as for the other cell types. Pseudo-bulk expression was determined using the
1931 Regularized Log (R-log) value for each gene (**Extended Data Fig. 3f-g).** In each cell
1932 type, only the genes with R-log values ≥2.5 in all pseudo-bulk replicates were
1933 considered expressed.
1934
1935 ***Differential gene expression analysis***
1936 For differential expression analysis for all cell types, pseudo-bulk replicate
1937 samples were obtained by randomly partitioning the total number of single cells per
1938 dataset into three groups and pooling all unique molecular identifiers (UMIs) per gene
1939 of cells belonging to the same replicate. To determine differentially expressed (DE)
1940 genes, all 6 possible pairwise comparisons between samples were performed using
1941 DEseq2 with default parameters[113]. Genes classified as DE in at least one
1942 comparison were considered for further analysis (adjusted p-value < 0.05; Benjamini-
1943 Hochberg multiple testing correction method). A summary table for the DE
1944 expression analysis of all cell types can be found in **Supplemental Table 10**. For
1945 transcription factor motif analysis, only the DE genes obtained from the comparison
1946 between DNs and PGNs were considered for further analysis (**Extended Data Fig.**
1947 **7b-c**).
1948
1949 ***Tn5 purification***
1950 The pTXB1 plasmid carrying the Tn5-intein-CBD fusion construct with the
1951 hyperactive Tn5 protein containing the E54K and L372P mutations was obtained
1952 from Addgene (plasmid #60240). Tn5 expression and purification was performed
1953 similarly as previously described[114], except that the final storage buffer was 50 mM
1954 HEPES-KOH pH 7.2, 0.8 M NaCl, 0.1 mM EDTA, 1 mM DTT and 55 % glycerol.
1955
1956 ***Tn5 adapter mix preparation***
1957 To generate 100 μM adapter mix, 200 μM Tn5MErev (5'-
1958 [phos]CTGTCTCTTATACACATC) was mixed with of 200 μM Tn5ME-A (5'-
1959 TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG; Adapter_mixA, 1:1 ratio).
1960 Separately, 200 μM Tn5MErev was mixed with 1 vol of 200 μM Tn5ME-B (5'-
1961 GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG; Adapter_mixB, 1:1 ratio). The
1962 two mixtures were incubated for 5 min at 95°C and gradually cooled to 25°C at a
1963 ramp rate of 0.1°C/s. Finally, the Adapter_mixA was mixed with Adapter_mixB at a
1964 1:1 ratio for a final 100μM adapter mix.
1965
1966 ***mESC ATAC-seq library preparation***
1967 ATACseq libraries were generated from ~75,000 mESC nuclei following the
1968 Omni ATAC protocol[115] with a modified transposition reaction: TAPS-DMF buffer (50
1969 mM TAPS-NaOH, pH 8.5, 25 mM MgC_2, 50% DMF), 0.1 % Tween-20, 0.1%
1970 Digitonin, in PBS. 3 μl of the Tn5 mix (5.6 μg Tn5 and 0.143 volume of 100 μM
1971 adapter mix) were added to the transposition reaction mix. Libraries were prepared
1972 as described in the Omni ATAC protocol. The final library was sequenced with an

1973 Illumina NextSeq 500 machine, according to manufacturer's instructions, using
1974 paired-end 75 bp reads (150 cycles).
1975
1976 ***VTA isolation for single-nucleus ATAC-seq***
1977 For collection of single-nucleus ATAC-seq (snATAC-seq), male C57Bl/6Nl
1978 (RRID: IMSR_CR:027; WT) mice, ages 7 and 9 weeks, were sacrificed by cervical
1979 dislocation. Brains were removed and the tissue containing the midbrain VTA was
1980 dissected from each hemisphere at room temperature and rapidly frozen on dry ice.
1981 Frozen midbrain was kept at -80°C until further processing.
1982
1983 ***DNs snATAC-seq library preparation***
1984 Two 10X Genomics scATAC-seq libraries from midbrain VTA, VTA-1 and
1985 VTA-2 (ages 7 and 9 weeks, respectively), were generated from midbrain VTA
1986 according to the 10X Genomics manual "Nuclei Isolation from Mouse Brain Tissue for
1987 Single Cell ATAC Sequencing Rev B" for flash frozen tissue with minor adjustments.
1988 Briefly, 500 μl 0.1x lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$,
1989 1% BSA, 0.01% Tween-20, 0.01% Nonidet P40 Substitute, 0.001% Digitonin) was
1990 added to the frozen samples and immediately homogenized using a pellet pestle
1991 (15x), followed by 5 min incubation on ice. The lysate was pipette mixed 10x, then
1992 incubated 10 min on ice. Finally, 500 μl of chilled wash buffer (10 mM Tris-HCl, pH
1993 7.4, 10 mM NaCl, 2 mM $MgCl_2$, 1% BSA, 0.1% Tween-20) was added to the lysed
1994 cells, and the suspension was passed through 30 μm CellTrics strainers (Th Geyer,
1995 cat# 7648779). The final ~500 μl nuclei suspension was stained with DAPI (final
1996 concentration 0.03 μg/ml) for ~5 min.
1997
1998 Around 200,000 DAPI positive events were sorted using a BD FACSAria III
1999 Flow Cytometer with 70-µm nozzle configuration with sample and sort collection
2000 device cooling set to 4°C into 300 μl Diluted Nuclei buffer (commercial buffer from
2001 10X Genomics) in a 1.5 ml Eppendorf tube. A first gate excluded debris in a
2002 FSC/SSC plot. A consecutive, second gate in a DAPI-A/DAPI-H-plot was used to
2003 exclude doublets and nuclei with incomplete DNA content (BD FACSDiva Software,
2004 version 8.0.2). The collected nuclei were centrifuged at 500x*g* for 5 min at 4°C and
2005 resuspended in 20 µl Diluted Nuclei Buffer. The nucleus concentration was
2006 determined using a Countess II FL Automated Cell Counter in DAPI fluorescence
2007 mode. snATACseq libraries were prepared following the Chromium Next GEM Single
2008 Cell ATAC Reagent Kits v1.1 User Guide. Briefly, nuclei were loaded on a
2009 microfluidics chip together with RT reagents, beads with oligo-dT tags and oil to
2010 create an emulsion. Afterwards, the reverse transcription reaction takes place inside
2011 the droplets. The barcoded cDNA is recovered from the emulsion, amplified and
2012 cleaned using a bead purification process. The cDNA is then using for library
2013 construction including enzymatic fragmentation, adapter ligation and sample index
2014 PCR. Libraries were sequenced with either an Illumina NextSeq 500 machine using
2015 paired-end 75 bp reads (for VTA-1, 150 cycles), or by NovaSeq 6000 using paired-
2016 end 75 bp reads (for VTA-2, 100 cycles).

2017
2018 ***ATAC-seq data processing: mapping, processing and QC***
2019 For bulk mESC ATAC-seq, paired-end reads were mapped to the mouse
2020 genome (mm10) using Bowtie with the following parameters: --minins 25 --maxins
2021 2000 --no-discordant --dovetail --soft-clipped-unmapped-tlen. Low quality mapped
2022 reads (MQ < 30) and mitochondrial reads were removed. Duplicated reads were
2023 removed with Sambamba[116] (version 0.6.8). Reads passing quality checks were
2024 converted to BAM format for further analyses.
2025
2026 For VTA snATAC-seq, paired-end reads were demultiplexed and mapped to
2027 the mouse genome (mm10) using the 10X Genomics Cellranger software. The two
2028 VTA snATAC-seq libraries were analyzed using ArchR software[117]. Doublets were
2029 removed following default parameters in ArchR. Next, low-quality cells (identified as
2030 TSS enrichment score < 4 and < 2500 unique fragments per cell) were removed for
2031 further analyses.
2032
2033 Next, dimensionality reduction was performed using the Latent semantic
2034 indexing (LSI) dimensionality reduction method from ArchR, with default parameters
2035 (except: iterations = 10, resolution =0.2, varFeatures = 60000). The ArchR
2036 addHarmony function was ysed to run the Harmony algorithm for batch correction
2037 with default parameters, followed by clusters calling. Gene scores were determined
2038 as specified by ArchR[117]. DNs were identified as the cluster with higher gene scores
2039 for *Th*, a well-known DN marker, and confirmed by additional DN marker expression
2040 (e.g. *Lmx1b*, *Foxa2*, *Foxa1*, *Slc6a3*). The DN cluster is composed of 216 cells in total
2041 (113 from VTA-1 and 103 from the VTA-2). UMI duplicates were collapsed to one
2042 fragment. To visualize an approximation for gene expression, gene scores were
2043 calculated using the createArrowFiles (addGeneScoreMat = TRUE) function in
2044 ArchR.
2045
2046 ***Processing of published OLG and PGN scATAC-seq***
2047 scATAC-seq bam files for OLGs were downloaded from the sciATAC-seq *in*
2048 *vivo* atlas of the mouse brain[27]. Next, reads were extracted from the bam file which
2049 corresponded to cells from the cluster identified as oligodendrocytes from prefrontal
2050 cortex (458 cells), to produce a pseudo-bulk ATAC .bam file. The original data,
2051 mapped to the mm9 genome, was converted to mm10 using the liftOver tool from
2052 UCSC utilities (https://genome.ucsc.edu/cgi-bin/hgLiftOver).
2053
2054 scATAC-seq datasets were obtained from hippocampal PGNs[28]. A bam file
2055 containing all cell types was kindly supplied by Andrew Adey (Molecular and Medical
2056 Genetics, Oregon Health & Science University, Portland, OR, USA). Reads were
2057 extracted from the bam file which corresponded to the NR1 PGN population (270
2058 cells), to produce a pseudo-bulk ATAC .bam file.
2059
2060 ***Generation of normalized ATAC-seq bigwig tracks***

2061 A size factor normalization was applied to generate ATAC-seq bigwig tracks
2062 comparable between mESCs, OLGs, PGNs and DNs. First, a count matrix was
2063 generated for all TSS regions (± 250 bp), which contained reads from at least two of
2064 the four cell types. The TSS list was extracted from the genes.gtf file included in the
2065 cell ranger reference data (refdata-cellranger-atac-mm10-1.2.0l;
2066 https://support.10xgenomics.com/single-cell-
2067 atac/software/pipelines/latest/advanced/references#overview). To calculate size
2068 factors, the TSS count matrix was processed through DESeqDataSetFromMatrix and
2069 estimateSizeFactors from the DESeq2 package[113]. For all cell types, the scale factor
2070 (SF) = (cell type size factor)-1.
2071
2072 Each pseudobulk ATAC-seq .bam file from mESCs, PGNs and OLGs were
2073 converted to bedGraph format using the genomeCoverageBed function from
2074 bedtools[101] with the following parameters: -pc -bg -scale SF. For DNs, ATAC-seq
2075 fragment files were converted to bedGraph format using the genomeCoverageBed
2076 function from bedtools[101] with the following parameters: -g chrom.sizes -bg -scale SF.
2077 The mm10 chrom.sizes file was downloaded from UCSC using fetchChromSize from
2078 UCSC utilities (http://hgdownload.soe.ucsc.edu/admin/exe/). The bedGraph files
2079 were then converted to bigwig using the bedGraphToBigWig function from UCSC
2080 utilities.
2081
2082 ***DN and PGN ATAC-seq peak calling***
2083
2084 ATAC-seq peaks were called in DNs following the Iterative Overlap Peak
2085 Merging Procedure described in the ArchR package[117]. First, two pseudo-bulk
2086 replicates were generated by running the *addGroupCoverages* function and then
2087 reproducible peaks were called using the *addReproduciblePeakSet* function. For
2088 PGNs, peaks for the NR1 cluster were obtained from Sinnamon *et al.* 2019[28]. For
2089 further analyses, peaks were considered positive for if they were found in at least
2090 10% of single-nuclei (> 10 nuclei in DNs; > 13 cells in PGNs).
2091
2092
2093 ***RNA and ATAC-seq length-scaled ATAC reads per million***
2094 To calculate length-scaled RNA reads per million (lsRRPM) for 479 long
2095 genes (> 400kb), the mESC bam file (paired-end) was read using
2096 *readGAlignmentPairs* function from the GenomicAlignments function from the
2097 GenomicAlignments package in R
2098 (https://bioconductor.org/packages/release/bioc/html/GenomicAlignments.html). For
2099 published single-cell datasets (OLGs, PGNs, DNs; single-end libraries), bam files
2100 were loaded using the *readGAlignments* function from the *GenomicAlignment*
2101 package. Due to very long length of some reads, all BAM fragments were resized to
2102 the 5' end base pair to avoid overlapping with multiple features. Next, the following
2103 formula was used to compute lsRRPM values for each cell-type and per gene:
2104

2105 (1) lsRRPM = (number of overlaps between RNA fragments and long gene
2106 body)/(gene length/10^6 * total number of RNA fragments/10^6).
2107
2108 To calculate length-scaled ATAC reads per million (lsARPM) for 479 long
2109 genes (> 400kb), concordant paired-end fragments were extracted for all cell types
2110 using the readGAlignmentPairs function from the GenomicAlignments package in R
2111 with the following total number of fragments: 37261746 (mESC), 2121258 (OLG),
2112 4594229 (PGN) and 8939526 (DN). Next, the following formula was used to compute
2113 lsARPM values for each cell-type and per gene:
2114
2115 (2) lsARPM = (number of overlaps between ATAC fragments and long gene
2116 body)/(gene length/10^6 * total number of ATAC fragments/10^6).
2117
2118
2119 ***Gene Ontology***
2120 Gene ontology (GO) term enrichment analysis was performed using GOElite
2121 (version 1.2.4)[118]. In **Extended Data Fig. 3l**, DN snATAC-seq marker genes were
2122 extracted with the *getMarkerFeatures* function from ArchR with default parameters.
2123 Marker genes were selected as genes with log2FC > 1 and FDR <0.01 in the DN
2124 cluster (total: 973 genes). All unique genes were used as the background GO
2125 dataset. In **Fig. 2c**, all genes expressed in at least one cell type, annotated to mm10,
2126 were used as the background dataset. In **Figs. 4d, 4e**, all genes expressed in PGNs
2127 or DNs were used as the background dataset, and in **Figs. 5c**, **5d** all unique genes
2128 were used. Default parameters were used for the GO enrichment: GO terms that
2129 were enriched above the background (significant permuted p-values <0.05, 2000
2130 permutations) were pruned to select the terms with the largest Z-score (>1.96)
2131 relative to all corresponding child or parent paths in a network of related terms
2132 (genes changed >2). GO terms which had a permuted p-value ≥ 0.01, contained
2133 fewer than 6 genes per GO term, or from the 'cellular_component' ontology were not
2134 reported in the main figures. A full list of unfiltered GO terms can be found in
2135 **Supplemental Table 7**.
2136
2137 ***Melting domain discovery***
2138 To assess gene insulation differences, insulation square values at ten length
2139 scales (100 – 1000 kb) were calculated for genes > 400kb in length (n = 479).
2140 Cumulative probability distributions of insulation square values were calculated for
2141 each dataset, and the brain cells were compared to mESC probability distributions for
2142 each gene by computing the maximum distance between the distributions and
2143 applying a Kolmogorov-Smirnov test with a custom script in R. P-values were
2144 corrected for multiple testing using the Bonferroni method, and -log10 transformed to
2145 obtain a 'domain melting score'. Domain melting scores for each gene in each
2146 comparison can be found in **Supplemental Table 8**. For visualization, empirical
2147 cumulative probabilities and insulation score values were smoothed using a
2148 Guassian kernel density estimate (adjust = 0.3).
2149

2150 ***Calculation of trans-cis contact ratio***

2151 To determine the interaction strength of contacts to all (trans) somatic
2152 chromosomes relative to interaction strength to their own (cis) chromosome, cis- and
2153 trans- NPMI normalized matrices were calculated at 250kb resolution. Bins detected
2154 in less than 3%, or more than 75%, of 3NP samples were removed from the analysis.
2155 To be sensitive to outliers, NPMI values of both cis- ($NPMI_C$) and trans- ($NPMI_T$)
2156 contacts for every bin were summarized with the arithmetic mean. The trans-cis
2157 contact ratio was then obtained by the following formula:

2158 (1) Trans-cis contact ratio = ($\Sigma(NPMI_T)$/genomic bins(n_T))) /
2159 ($\Sigma(NPMI_C)$/genomic bins(n_C)))

2160 Trans-cis values of bins spanning long genes were summarized with the
2161 median.

2162

2163 ***Modeling and in-silico GAM***

2164 To reconstruct 3D conformations of the *Nrxn3* locus we employed the Strings
2165 & Binders Switch (SBS) polymer model of chromatin[119,120]. In the SBS, a chromatin
2166 region is modelled as a self-avoiding chain of beads, including different binding sites
2167 for diffusing, cognate, molecular binders. Binding sites of the same type can be
2168 bridged by their cognate binders, thus driving the polymer folding. The optimal SBS
2169 polymers for the *Nrxn3* locus in mESCs and DNs were inferred using PRISMR, a
2170 machine-learning-based procedure which finds the minimal arrangement of the
2171 polymer binding sites best describing input pairwise contact data, such as Hi-C[51] or
2172 GAM[121]. Here, PRISMR was applied to the GAM experimental data by considering
2173 the NPMI normalization on a 4.8Mb region around the *Nrxn3* gene (Chr12:
2174 87,600,000-92,400,000; mm10) at 50kb resolution in mESCs and DNs. The
2175 procedure returned optimal SBS polymer chains made of 1440 beads, including 7
2176 different types of binding sites, in both cell types.

2177

2178 Next, to generate thermodynamic ensembles of 3D conformations of the
2179 locus, molecular dynamics simulations were run of the optimal polymers, using the
2180 freely available LAMMPS software[122]. In those simulations, the system evolves
2181 according to Langevin equation, with dynamics parameters derived from classical
2182 polymer physics studies[123]. Polymers are first initialized in self-avoiding
2183 conformations and then left to evolve in order to reach their equilibrium globular
2184 phase[119]. Beads and binders have the same diameter σ =1, expressed in
2185 dimensionless units, and experience a hard-core repulsion by use of a truncated
2186 Lennard-Jones (LJ) potential. Analogously, attractive interactions are modelled with
2187 short-ranged LJ potentials[119]. A range of affinities between beads and cognate
2188 binders were sampled in the weak biochemical range, from $3.0K_BT$ to $8.0K_BT$ (where
2189 K_B is the Boltzmann constant and T the system temperature). In addition, binders
2190 interact unspecifically with the polymer with a lower affinity, sampled from $0K_BT$ to
2191 $2.7K_BT$. For sake of simplicity, the same affinity strengths were used for all different
2192 binding site types. Total binder concentration was taken above the polymer coil-

2193 globule transition threshold[119]. For each of the considered cases, ensembles of up to
2194 ~450 distinct equilibrium configurations were derived. Full details about the model
2195 and simulations are discussed in Barbieri *et al.*[119] and Chiariello *et al.*[120].
2196
2197 *In-silico* GAM NPMI matrices were obtained from the ensemble of 3D
2198 structures by applying the *in-silico* GAM algorithm[13], here generalized to simulate the
2199 GAM protocol with 3NPs per GAM sample and to perform NPMI normalization.
2200 Specifically, the same number of slices were used as in the GAM experiments,
2201 249x3NPs for mESCs and 585x3NPs for DNs. Pearson correlation coefficients were
2202 used to compare the *in-silico* and experimental NPMI GAM matrices.
2203
2204 Example of single 3D conformations were rendered by a third-order spline of
2205 the polymer beads positions, with regions of interest highlighted in different colors. To
2206 quantify the size and variability of the 3D structures in mESCs and DNs, the average
2207 gyration radius (R_g) was measured from the selected domains encompassing and
2208 surrounding the *Nrxn3* gene, expressed in dimensionless units σ in **Fig. 3d**,
2209 **Extended Data Fig. 6d**.
2210
2211 ***Cryosections for fluorescence in-situ hybridization experiments***
2212 Fixed and cryopreserved hippocampal CA1 tissue and mESCs were
2213 cryosectioned as previously described (see Methods *Cryoblock preparation and*
2214 *cryosectioning*) with an approximate thickness of 400 nm and transferred to glass
2215 coverslips (thickness Nr1.5, diameter 10mm) coated with Laminin (Sigma-Aldrich®,
2216 Cat# P8920) according to the manufacturer's instructions, for the 3-colour
2217 fluorescence in-situ hybridization (FISH) experiment (TSS-Mid-TES), or washed in
2218 100% ethanol and autoclaved for the immunofluorescence-whole gene FISH
2219 experiment (nucleolus, *Rbfox1*).
2220
2221 ***BAC probes labelling and precipitation***
2222 BACs targeting the *Rbfox1* locus (see **Supplementary Table 9**) were
2223 obtained from BACPAC Resources Center (https://bacpacresources.org/) and
2224 amplified from glycerol stocks using a MIDIprep kit (NucleoBond® Xtra BAC
2225 purification kit, Machery-Nagel, Cat# 740436). Purified BACs were labelled using a
2226 nick translation kit (Abbott Molecular, Cat# 7J0001) according to manufacturer's
2227 instructions and the following fluorophores (all Invitrogen, Thermo Fisher Scientific):
2228 ChromaTide™ Alexa Fluor™ 488-5-dUTP (Cat# C11397), ChromaTide™ Alexa
2229 Fluor™ 568-5-dUTP (Cat#C11399) and Alexa Fluor™ 647-aha-dUTP (Cat# A32763).
2230 Labelled BAC probes were coprecipitated with yeast tRNA (20 µg/µl final
2231 concentration; Invitrogen Cat# AM7119) and mouse *Cot-1* DNA (3 µg/µl final
2232 concentration; Invitrogen Cat# 18440-016) overnight at -20 °C. After clean up in 70 %
2233 ethanol, the probes were dissolved in 100 % deionized formamide (for 1 h; Sigma
2234 Cat# F9037), before adding (1:1) a 2xhybridization mix (20 % dextran sulfate, 0.1 M
2235 Phosphate buffer in 4x SSC; mixing for 1 h), denatured (10 min, 80 °C), and
2236 reannealed (30 min, 37 °C) prior to hybridization.
2237

2238 ***Immunolabeling prior to FISH***
2239 Immunofluorescence labelling of the nucleolus was performed as described
2240 above (see Methods *Immunofluorescence detection for confocal microscopy*) by
2241 incubating the cryosections overnight (at 4°C) with a mouse monoclonal antibody
2242 anti-nucleophosmin B23 (a kind gift from Harris Busch[53]), followed by incubation (1h)
2243 with donkey antibodies raised against mouse IgG conjugated with AlexaFluor-555
2244 (Invitrogen). Prior to cryo-FISH, the bound antibodies were fixed (1 h, 4 °C) in 8%
2245 depolymerized PFA (EM-grade) in 250 mM HEPES-NaOH (pH 7.6), and rinsed in
2246 PBS.
2247
2248 ***Cryo-FISH***
2249 Cryo-FISH was performed as previously described[3,52] with few modifications.
2250 Briefly, cryosections were washed (30 min) in 1xPBS, rinsed with 2xSSC (Sigma
2251 Cat#S6639), and incubated (2h, 37 °C) in 250 µg/ml RNase A (Sigma Cat#R4642) in
2252 2xSSC. After washing in 2xSSC, cryosections were treated (10 min) with 0.1 M HCl,
2253 dehydrated in ethanol (30 %, 50 %, 70 %, 90 %, 100 % series, 3 min each on ice)
2254 and denatured (10 min) at 80 °C in 70 % formamide, 2xSSC, 0.05 M phosphate
2255 buffer (pH 7.4). Cryosections were dehydrated as described above, and overlaid on
2256 hybridization mixture on HybriSlip™ (Invitrogen Cat#H18202). After sealing with
2257 rubber cement, and incubation (48h, 37 °C) in a moist chamber, cryosections were
2258 washed (25 min, 42 °C) in 50 % formamide in 2xSSC, (30 min, 60 °C) in 0.1x SSC
2259 and (10 min, 42 °C) in 0.1 % Triton X-100 in 4xSSC. After rinsing with 1xPBS,
2260 coverslips were mounted in Vectashield® Mounting Medium (Anti Fading) w/ DAPI
2261 (Vector laboratories, Cat# H-1200).
2262
2263 ***Cryo-FISH Microscopy***
2264 Cryo-FISH images were collected sequentially with a Leica TCS SP8-STED
2265 confocal microscope (Leica Microsystems DMI6000B-CS) using the Leica Application
2266 Suite X Version 3.5.5.19976 and the HC PL APO CS2 63x/1.40 OIL objective (NA =
2267 1.4 Plan Apochromat) (see Methods *Immunofluorescence detection for confocal*
2268 *microscopy*) using the following settings: 405 nm excitation and 420-500 nm
2269 emission (for DAPI), 488 nm excitation and 510-535 nm emission (for probes labeled
2270 with ChromaTide™ AlexaFluor-488 and for nucleophosmin), 568 nm excitation and
2271 586-620 emission (for probes labeled with ChromaTide™ AlexaFluor-568), 647 nm
2272 excitation and 657-700 emission (for probes labeled with AlexaFluor-647), and 555
2273 nm excitation and 586-640 emission (for immunofluorescence labelling of
2274 nucleophosmin with AlexaFluor-555). All images were collected with a 4x zoom at
2275 1024x1024 pixels resolution (pixel size 0.0451 µm, resolution 22.1760 pixels/µm).
2276
2277 ***Cryo-FISH Image analysis***
2278 Images were analyzed using the Fiji software (version 2.0.0-rc-69/1.52p)[124].
2279 All images were pre-processed as previously described[52]. Genomic foci were
2280 identified visually and areas of the manually defined objects were measured using
2281 the Fiji-*Area* tool. For the cryo-FISH experiment combined with immunofluorescence,
2282 the location of genomic loci in relation to the nuclear lamina or nucleolus was

assessed based on the overlap of foci with the nucleolus (identified by nucleophosmin immunolabeling) or the nuclear lamina (as defined by the periphery of the DAPI staining) by at least three pixels. To determine the distance between the TSS, Mid and TES genomic foci, we took the centre of mass defined of the selected objects, as defined by Fiji-*Center of mass* function (the brightness-weighted average of the x and y coordinates of all pixels within the selected areas). Distances between the objects were measured using the Fiji-*Line* tool between the centres of mass defined for each object. Images for visualization in figure panels were processed using Fiji or Adobe Photoshop CS6, where adjustments included the optimization of the dynamic signal range with contrast stretching.

Determining differential contacts between GAM datasets

Significant differences in pairwise contacts between a pair of GAM datasets were determined as previously described with modifications[11]. Briefly, 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. Contacts were filtered to be within 0.5-5 Mb distance and above 0.15 NPMI, and NPMI contact frequencies at each genomic distance of each chromosome were normalized by computing a Z-score transformation, and a differential matrix (D) was derived by subtracting the two Z-score normalized matrices[11].

Transcription factor binding site analysis

To find transcription factor (TF) binding motifs present within specific contacts, significant differential contacts were determined for DNs and PGNs. Accessible regions within the differential contacts were determined using scATAC-seq for PGNs[28] and DNs, respectively. To account for methodological differences, including lower sequencing depth in PGN scATAC-seq data (**Extended Data Fig. 3j**), we considered only the peaks that occurred in >10% of cells (> 10 cells in DNs; > 13 in PGNs). Motif finding within accessible regions in significant contacts was performed using the Regulatory Genomics Toolbox (https://www.regulatory-genomics.org/motif-analysis/introduction/) with TF motifs (from the HOCOMOCO database, v11)[125] obtained for TFs expressed in either DNs or PGNs (R-log $\geq$ 2.5) to determine the percentage of windows containing each TF motif. Next, TF motifs were filtered based on (1) percentage of windows containing the motif (>5%) and (2) differential expression in either PGNs or DNs (-log10(adj. p value) > 3, see *Differential gene expression analysis* **Methods** section) resulting in 50 TF motifs for feature pair analysis (33 TF motifs from PGNs and 17 from DNs; see **Extended Data Fig**. **7b-c**).

Feature pairs associated with specific contacts were determined as previously described[11], and testing the 1275 combinations of motif pairs (1225 heterotypic motif pairs and 50 homotypic motif pairs). The number of contacts containing each pair of selected TF motifs (PGN_{TF} and DN_{TF}), together with the percentage of total significant differential contacts in PGNs and DNs (PGN and DN), were used to determine the enrichment score for all TF feature pair interactions (i.e., the ratio

between frequencies of contacts in PGNs or DNs, (PGN_{TF}/PGN) / (DN_{TF}/DN)). The effectiveness of a TF pair for discriminating between contacts from PGNs and DNs was assessed by using the Information gain measure[126]. Enrichment and Information gain for all TF feature pair interactions, as well as differential expression values for each TF motif (DNs compared to PGNs), can be found in **Supplemental Table 11.** The top feature pairs were extracted based on the highest Information gain (10 feature pairs), PGN enrichment (5 feature pairs), and DN enrichment (5 feature pairs) scores. Contact overlaps for top feature pairs were visualized using UpSet plots.

Network and community detection analysis of transcription factor binding sites in significant differential contacts

To determine the interconnectivity between different TF motifs found in accessible regions of significant differential contacts, the number of contacts for each pair of TF motifs (1275 pairs) was determined. After filtering pairs of TF motifs involved in fewer then 20% of the total contacts (15,833 and 5,400 contacts minimum in PGNs and DNs, respectively), a network was built for each cell type with TF motifs as nodes and number of contacts as weighted edges. The Leiden algorithm was used to detect communities of strongly interconnected nodes, using the *leiden* package in R[127,128], with a resolution of 1.01 for both PGNs and DNs (**Extended Data Fig. 8d**, **Supplemental Table 12**).

GAM aggregated contact plots

To visualize the average contact intensity for a set of genomic contacts, NPMI contact frequencies at each genomic distance of each chromosome were first normalized by computing a Z-score transformation. The resulting Z-Score values were determined for each contact, and for each contact in a 4-bin radius (50kb bins). For each chromosome, Z-score values for each set of contacts and for the surrounding bins were summarized by the arithmetic mean. Mean values computed for each chromosome were added together and divided by the number of chromosomes.

Data availability

Raw fastq sequencing files for all samples from DN, PGN and OLG GAM datasets, together with non-normalized co-segregation matrices, normalized pair-wised chromatin contacts maps and raw GAM segregation tables are available from the GEO repository under accession number GSE94364. Raw fastq sequencing files for mESC GAM datasets are available from 4DN data portal (https://data.4dnucleome.org/). The 4DN sample IDs for all samples used in the study are available in **Supplemental Table 1**.

Raw single cell mESC transcriptome data are available from ENA data portal (https://www.ebi.ac.uk/ena/browser/home). The ENA sample IDs for all samples used in the study are available in **Supplemental Table 13**. Raw single cell and bulk ATAC-seq fastq files for DNs and mESCs, respectively, are available from the GEO repository under accession number GSE174024, together with processed bigwig

files. A public UCSC session with all data produced, as well as all published data utilized in this study is available at: http://genome-euro.ucsc.edu/s/Kjmorris/Winick_Ng_2021_GAMbrainpublicsession.

Code availability

Processing and plotting scripts for MELTRON and insulation scores are available at: https://github.com/pombo-lab/Meltron/. Processing and plotting scripts for the trans-cis contact ratios are available at: https://github.com/pombo-lab/Meltron/.