Extensive long-range polycomb interactions and weak compartmentalization are hallmarks of human neuronal 3D genome

Ilya A. Pletenev[1,§], Maria Bazarevich[1,§], Diana R. Zagirova[1,§], Anna D. Kononkova[1,2], Alexander V. Cherkasov[1], Olga I. Efimova[1], Eugenia A. Tiukacheva[3-7], Kirill V. Morozov[1], Kirill A. Ulianov[1], Dmitriy Komkov[7], Anna V. Tvorogova[7], Vera E. Golimbet[8], Nikolay V. Kondratyev[8], Sergey V. Razin[4,7], Philipp Khaitovich[1], Sergey V. Ulianov[4,7,*], Ekaterina E. Khrameeva[1,*]

[1] Skolkovo Institute of Science and Technology, 121205 Moscow, Russia
[2] A.A. Kharkevich Institute for Information Transmission Problems, 127051 Moscow, Russia
[3] Moscow Institute of Physics and Technology, 141700 Moscow, Russia
[4] Department of Molecular Biology, Faculty of Biology, M.V. Lomonosov Moscow State University, 119991 Moscow, Russia
[5] CNRS UMR9018, Institut Gustave Roussy, 94805 Villejuif, France
[6] Koltzov Institute of Developmental Biology, Russian Academy of Sciences, 119334 Moscow, Russia
[7] Institute of Gene Biology, Russian Academy of Sciences, 119334 Moscow, Russia
[8] Mental Health Research Center, 115522 Moscow, Russia

[§] Contributed equally
[*] Correspondence to: sergey.v.ulyanov@gmail.com and e.khrameeva@skoltech.ru

Lead contact: Ekaterina E. Khrameeva, e.khrameeva@skoltech.ru

Abstract

Chromatin architecture regulates gene expression and shapes cellular identity, particularly in neuronal cells. Specifically, polycomb group (PcG) proteins enable establishment and maintenance of neuronal cell type by reorganizing chromatin into repressive domains that limit the expression of fate-determining genes and sustain distinct gene expression patterns in neurons. Here, we map the 3D genome architecture in neuronal and non-neuronal cells isolated from the Wernicke's area of four human brains and comprehensively analyze neuron-specific aspects of chromatin organization. We find that genome segregation into active and inactive compartments is greatly reduced in neurons compared to other brain cells. Furthermore, neuronal Hi-C maps reveal strong long-range interactions, forming a specific network of PcG-mediated contacts in neurons that is nearly absent in other brain cells. These interacting loci contain developmental transcription factors with repressed expression in neurons and other mature brain cells. But only in neurons, they are rich in bivalent promoters occupied by H3K4me3 histone modification together with H3K27me3, which points to a possible functional role of PcG contacts in neurons. Importantly, other layers of chromatin organization also exhibit distinct structure in neurons, which corroborates with increased cohesin levels in neurons and therefore might be attributed to elevated neuronal loop extrusion activity.

Introduction

Mammalian genomes possess a complex 3D architecture reflecting a composite interplay between structure and functionality. Structures of different scales appear to be formed by distinct mechanisms and presented hierarchically within each other[1]. Traditionally, four levels of chromatin organization are distinguished: chromosome territories, compartments, Topologically Associating Domains (TADs), and chromatin loops. Chromosome territories are distinct nuclear areas, preferentially occupied by different interphase chromosomes. At coarse-grain level, chromosomes are segregated into two compartments, A and B, comprising regions of similar epigenetic states - “active” and “inactive” chromatin, respectively[2]. The formation of A and B compartments is thought to be driven by a combination of factors including the distribution of active and repressive chromatin marks[3]. At mid-range distances, topological domain structure is formed by high density of contacts inside the same TAD and relatively high insulation between adjacent TADs[4]. TADs are suggested to be main functional regulatory domains by modulating contacts between enhancers and promoters. Elevated frequency of contacts within the TAD mediates a physical interaction between enhancer-promoter pairs, while high insulation at TAD borders

can restrict such interactions for pairs located in neighboring TADs[5]. At the finest scale, chromatin is organized into loops that bring promoters close to their regulatory elements such as enhancers. Chromatin loops are thought to be formed through the cohesin-mediated loop extrusion restricted by CTCF and other factors[6–9].

Distinct mechanisms facilitate chromatin feature formation at multiple levels. Besides conventional types described above, the current research[10–17] has revealed an additional level of interactions mediated by polycomb group (PcG) proteins. PcG proteins are a family of transcriptional repressors that have first been described as capable of gene silencing maintenance during development and cellular differentiation. PcG proteins are organized into two epigenetic complexes, the Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2). PRC2 catalyzes the trimethylation of lysine residue on histone H3 (H3K27me3), which contributes to chromatin compaction and provides a stable and heritable mark that can be passed through cell generations[18,19]. PRC1 is responsible for maintaining the silenced state of genes by recognizing and binding to the H3K27me3 modification produced by PRC2[14,20]. While the role of PcG proteins has been mostly investigated in the context of the development, their involvement in the maintenance of chromatin organization in mature cells is only starting to be elucidated. Specifically, PcG proteins mediate contacts between H3K27me3-enriched regions located up to hundreds of megabases apart[10,13–15,17]. The formation of extreme long-range *cis*- and even *trans*-interactions indicates that PcG-mediated contacts represent a distinct hierarchical level of the 3D genome. Whereas the general function of long-range PcG-mediated interactions remains unclear, their dynamic reorganization is found to be important during neural differentiation in mice[13].

Neurons represent a fundamental component of the nervous system. They are characterized by complex morphological and functional properties that require the establishment of specific gene expression programs[21]. However, the complexity of the brain tissue and limitation of sample availability constraint the range of conducted chromatin studies focused on main types of brain cells. Few studies compared genome-wide neuronal chromatin organization with non-neuronal cells but they unveiled the presence of neuronal-specific chromatin features at multiple levels of the genome organization[22,23].

Altogether, though some unique features of chromatin organization in neurons have been previously shown, a comprehensive comparison of the chromatin architecture in neurons and non-neuronal cell types is still lacking. Moreover, none of the previous works focus on long-range polycomb interactions as a distinct feature of the neuronal 3D genome. Here, we use fluorescence-activated nuclear sorting (FANS) to separate neurons and non-neuronal cells obtained from the Wernicke's area (BA22p) of four human postmortal brain samples, and apply an optimized Hi-C protocol[24] to construct chromatin interaction maps for neuronal and non-neuronal

cells. We observe a prominent decrease of chromatin compartmentalization in neurons accompanied by the presence of neuron-specific long-range cis and trans contacts between H3K27me3-marked loci. These long-range interactions could shed a new light on the involvement of PcG proteins in the 3D organization of mature genomes in the brain.

Results

Whole-genome maps of chromatin folding in human neuronal and non-neuronal cells

We applied fluorescence-activated nuclei sorting (FANS) to isolate neuronal and non-neuronal cells from the left posterior superior temporal gyrus (Wernicke's area - BA22p) based on staining with antibodies against neuron-specific marker NeuN[24] (Suppl. Fig. 1). 5-kb resolution Hi-C maps were constructed for the NeuN(+) and NeuN(-) nuclei obtained from four human individuals (Fig. 1A, Suppl. Table 1). NeuN(+) and NeuN(-) cells demonstrate a clear separation in the principal component analysis (PCA) of Hi-C maps (41% of the explained variance, Fig. 1B), suggesting global differences in chromatin architecture between neuronal and non-neuronal cells. Similar analysis combining our data with publicly available Hi-C maps obtained for a different brain region[22] confirms separation of NeuN(+) and NeuN(-) cells (32% of the explained variance, Suppl. Fig. 2A) and serves as an additional quality control of our data. Incorporation of publicly available Hi-C data from other human organs into the PCA analysis reveals that NeuN(+) cells possess the most distinctive chromatin organization (Suppl. Fig. 2B), while NeuN(-) cells are located closer to other organs and cell lines. We note that the NeuN(-) cell population is highly heterogeneous as it comprises a mix of endothelial cells, microglia, astrocytes, oligodendrocytes, and other glial cell types. Therefore, NeuN(-) Hi-C maps represent an average picture of non-neuronal cells, and do not reflect chromatin organization in particular glial cell types. Thus, we further focus on NeuN(+)-specific chromatin organization features and use NeuN(-) cell population as a baseline for comparisons.

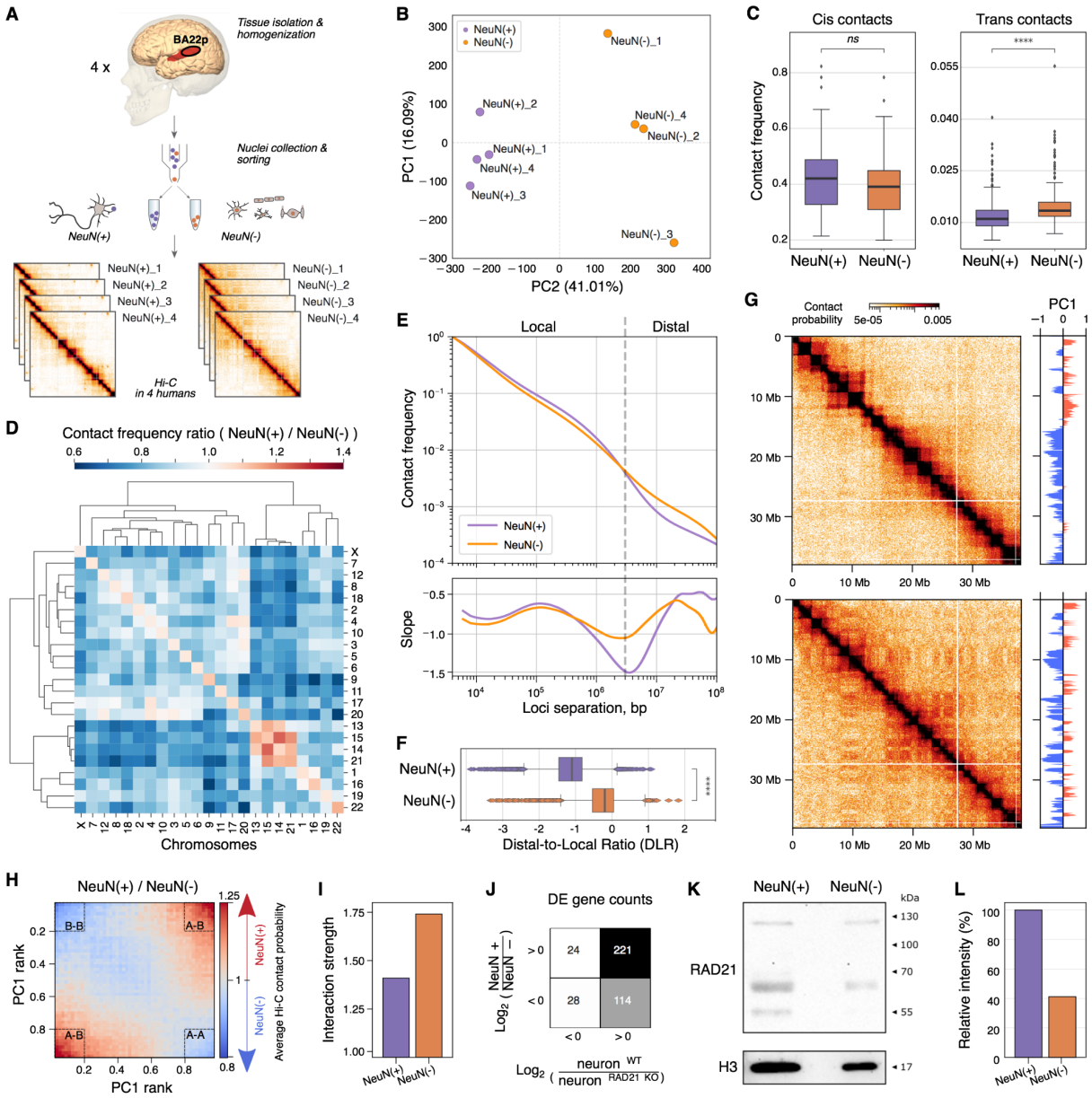

Fig. 1. Global chromatin organization in NeuN(+) and NeuN(-) cells. (A) Anatomical localization of the analyzed brain region, experimental procedure and the design of this study. Hi-C experiments were performed in NeuN(+) and NeuN(-) cells isolated using FACS from the Wernicke's speech area of four human individuals. (B) Principal component analysis plot based on the Insulation Score (IS) variation among all produced Hi-C maps. Colors represent NeuN(+) and NeuN(-) cells, here and in panels C-F. (C) Average interactions within all chromosomes (cis contacts, left panel) and between all pairs of chromosomes (trans contacts, right panel) calculated separately for NeuN(+) and NeuN(-) cells. Asterisks indicate Wilcoxon test p-values: **** - $p < 0.00001$, ns - $p > 0.05$. (D) Ratio of interactions within and between all chromosomes (NeuN(+)/NeuN(-)). (E) Polymer scaling plot showing average interaction frequencies at various genomic distances, as well as the first derivative of this plot demonstrating the slope. (F) Distal (>3 Mb) to local (<3 Mb) contact ratio (DLR) calculated for every 100-kb genomic region, demonstrating that chromosomes in neurons are more compact at a local scale and less compact at a large scale. Asterisks represent Wilcoxon test p-value < 0.00001. (G) A fragment of Hi-C map with compartment eigenvector (PC1) for NeuN(+) (left) and NeuN(-) (right) cells. Positive values of PC1 correspond to the A compartment, while negative values - to the B compartment. (H) NeuN(+) / NeuN(-) ratio of average Hi-C contact probability (observed over expected) of genomic regions arranged by the corresponding PC1 rank (saddle plot). (I) Compartment interaction strength calculated as the highest intra-compartment interactions divided by the lowest

inter-compartment interactions. (J) Contingency table showing the number of upregulated or downregulated genes in NeuN(+) compared to NeuN(-) (from Rizzardi et al.[25]) or NeuN(+) WT compared to NeuN(+) RAD21 KO (from Calderon et al.[26]). (K) Western blot analysis of RAD21 protein and histone H3 (the loading control) in NeuN(+) and NeuN(-) nuclei. The exposure time was the same for both proteins. (L) Quantification of RAD21 expression level. Each bar represents total RAD21 intensity values normalized to histone H3 values and presented as a percentage of the neuronal value.

Among global differences in chromatin organization, the analysis of *cis*- and *trans*-contact frequencies reveals a significant decrease in the neuronal *trans*-contacts compared to the non-neuronal ones (Wilcoxon test p-value $< 10^{-5}$, Fig. 1C), indicating that chromosome territories in NeuN(+) cells are more pronounced. This observation could be explained by the larger nuclear volume in neurons compared to glial cells[27] (Suppl. Fig. 1), potentially providing an additional free space in the nucleus for better spatial segregation of chromosomes. Increased nuclear size could also lead to a global decondensation of chromatin in neurons to occupy the free nuclear space, yet we observe an opposite tendency for cis-contacts (Fig. 1C). Furthermore, recent studies report that the global chromatin structure is robust to dramatic nuclear volume expansion and contraction, with a striking maintenance of loops, TADs, active and inactive compartments, and even chromosome territories[6,28]. Therefore, we can conclude that the nuclear size alone is unlikely to explain the global differences in genome organization between neuronal and non-neuronal cells. At the same time, in NeuN(+) cells we observe a remarkable increase in contact frequency between acrocentric chromosomes 13, 14, 15 and 21 containing nucleolus organizer regions (NORs) crucial for the nucleolus formation (Fig. 1D). This observation suggests that the nucleolus is more pronounced in neurons and is in line with microscopy studies reporting larger well-defined nucleolus as a cytological feature of neurons compared to glial cells[29]. The nucleolus is prominently present in both NPCs, which have highly active ribosomal biogenesis[30], and neurons that are still able to grow despite being non-dividing cells[31].

To further explore chromatin architecture at the chromosomal scale, we analyzed intra-chromosomal contact frequencies at different genomic distances and calculated the contact scaling as the average interaction frequency over distance separating two genomic loci, which can reveal general properties of the chromosome fiber[32]. Contact scaling plots show an increase in short-range and a decrease in long-range chromatin interactions in NeuN(+) cells compared to NeuN(-) (Fig. 1E), indicating that chromosomes in neurons are more compact at a local scale and less compact at a large scale. To estimate the significance of these differences, we employed a contact-scaling metric calculated separately at every 100 kb region across the genome, the distal (>3 Mbp) to local (<3 Mbp) contact ratio (DLR)[33]. A pronounced decrease in DLR (ΔDLR = -0.9) confirms the significance of chromatin

interactions shift towards shorter ranges in neurons (Wilcoxon test p-value < 10^{-5}, Fig. 1F).

Weak compartmentalization is a characteristic feature of chromatin in NeuN(+) cells

Remarkably different contact frequency at large genomic distances between NeuN(+) and NeuN(-) cell populations potentially implies different chromatin compartmentalization in these cells. Indeed, both visual comparison and eigendecomposition of Hi-C maps suggests different positioning and weaker chromatin compartmentalization in neurons compared to non-neuronal cells (Fig. 1G, Suppl. Fig. 3,4). Specifically, we observe decreased intra-compartment (A-A and B-B) interactions and increased inter-compartment interactions (A-B) in neurons compared to NeuN(-) cells (Fig. 1H,I). Consistently, correlations between NeuN(-) PC1 and NeuN(-) chromatin states are higher than between NeuN(+) PC1 and NeuN(+) states (Suppl. Fig. 5).

Stronger compartmentalization pattern that better correlates with chromatin states could be a consequence of a reduction in loop-extruding activity of structural maintenance of chromosomes (SMC) complex cohesin[34]. To test this, we compare the abundance of cohesin in NeuN(+) and NeuN(-) cells by reanalyzing previously published RNA-seq data and examining expression level of cohesin subunits and related proteins in NeuN(+) and NeuN(-) cells[25] and in neurons upon knock-out of cohesin subunit RAD21[26]. In general, differences in gene expression patterns in RAD21 KO neurons resemble the ones between NeuN(+) and NeuN(-) cells. Moreover, we observe that the direction of expression changes in cells with RAD21 KO is similar to the neuron-specific expression (Fig. 1J). Further, an increased abundance of cohesin subunits and related proteins in neurons as compared to non-neuronal cells is also supported by publicly available single-cell RNA-seq (Suppl. Fig. 6A, Suppl. Table 2) and proteome analysis[35]. Specifically, cohesin subunits RAD21, Stag1/2 and Smc1a/3 show an elevated level in neurons according to the mouse brain proteome data, while there is an opposite trend in glial cells (Suppl. Table 3). We thus conclude that relatively weak compartmentalization of compartment structure in NeuN(+) cells might be driven by higher levels of active cohesin complexes, in line with previously observed antagonism between loop extrusion and compartmentalization[34].

TAD borders in NeuN(+) cells are enriched with neuron-specific genes

As cohesin-dependent loop extrusion is the major mechanism of TAD formation in mammals, we assume that higher cohesin abundance in neurons might affect TAD

profile. Visual inspection of Hi-C maps reveals a number of locus-specific differences in TAD structure between neuronal and non-neuronal cells, specifically in a vicinity of genes involved in the neuron-specific processes such as the maintenance of synapse and transport systems, alternative splicing regulation and others (Fig. 2A). To analyze TAD profiles in NeuN(+) and NeuN(-) cells systematically, we performed TAD calling using the insulation score (IS) algorithm and identified 2194 NeuN(+)-specific and 2218 NeuN(-)-specific TADs borders, along with 2162 common TADs borders (Fig. 2B). We note that, though many TAD borders will be further referred to as cell-specific, they might in fact insulate chromatin in all cell types but with different strength. On average, TADs are more pronounced in neurons as compared to non-neuronal cells (Fig. 2C,D). Increased contact density at TAD level is in agreement with the enrichment of short-range chromatin interactions (<3 Mbp) in neuronal cells (Fig. 1D).

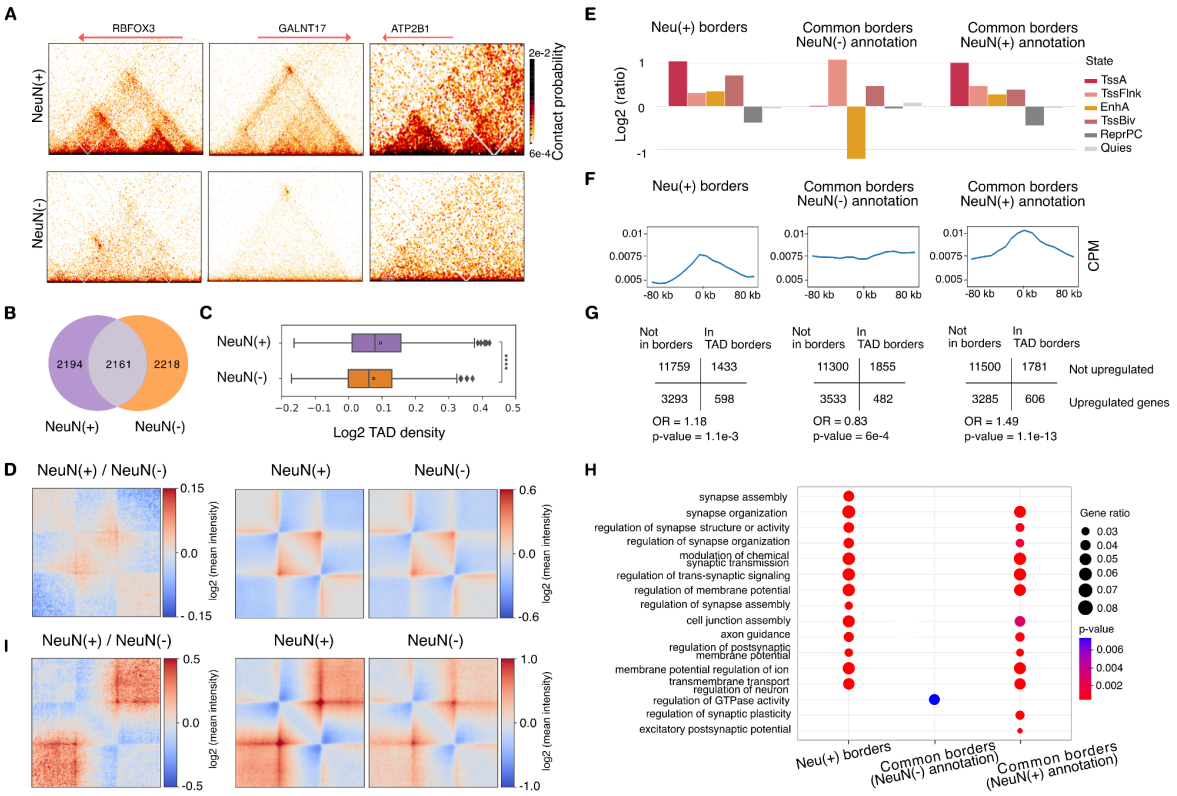

Fig. 2. Cell-type-specific differences of TADs in NeuN(+) and NeuN(-) cells. (A) Examples of differential TAD profiles at neuronal DE genes. *RBFOX3, GALNT17, ATP2B1* genes location are shown. *RBFOX3* produces the neuronal nuclei (NeuN) antigen that is used as a neuronal marker in the study. NeuN is a RNA binding protein involved in the regulation of alternative splicing of pre-mRNA. *GALNT17* encodes a glycosyltransferase enzyme involved in the post-translational modification of neuronal proteins. This gene has been shown to be essential for the proper development and function of the nervous system. *ATP2B1* is a gene that is involved in neuronal signal transmission encoding a calcium pump that is important for the regulation of calcium levels in neurons. (B) Venn diagram of the intersection of neuronal and non-neuronal TAD borders identified based on the insulation profiles. (C) Box plot of TAD densities for neurons and non-neuronal cells. Asterisks indicate Wilcoxon test p-values: **** - $p < 0.0001$ (D) Average TAD. (E) Grouped stacked bar plots of the chromatin states

coverage within TADs borders. The ratio of each state is normalized to the total coverage within the genome in the corresponding cell type. Chromatin states annotation includes six distinct chromatin states, specifically EnhA (active enhancer), TssA (active promoters), TssBiv (bivalent promoters), TssFlnk (promoter flanking region), ReprPC (polycomb repression region), and Quies (other regions). (F) Expression profiles around the cell-type-specific and common borders. (G) Confusion tables with Fisher test p-values for the enrichment of genes upregulated in NeuN(-) and NeuN(+) cells within corresponding TAD borders. DE genes were determined as genes upregulated in the selected cell type with FC > 1.5. (H) GO terms enrichment (shown with dots sizes) among upregulated genes placed in neuronal-specific or common TAD borders. (I) Average large TADs located within lamina-associated domains (LADs).

We therefore proceed with the analysis of epigenetic and structural properties of common and cell type-specific TAD borders in order to find a connection between TADs border positions and cell-type-specific transcription. First, we examine chromatin states predicted by ChromHMM for NeuN(+) and NeuN(-) cells[36] and find that chromatin states are distributed not equally relative to TAD borders in neurons (Fig. 2E). In NeuN(+) cells, both common and neuron-specific TAD borders are significantly enriched with active chromatin states defined by the presence of H3K4me3 and/or H3K27ac histone modifications (TssA, EnhA, TssFlnk), and depleted with repressed chromatin state defined by H3K27me3 (ReprPC). In line with these findings, we observe the increase in mean expression profiles around common and cell-type-specific borders in neurons in contrast to non-neuronal common borders (Fig. 2F). Next, we examine the distribution of genes differentially expressed (DE) in NeuN(+) and NeuN(-) cells, according to previously published data[25]. In line with active chromatin state profiles, common as well as cell-specific TAD borders in neuronal cells are significantly enriched in DE genes upregulated in neurons (Fig. 2G) and strongly associated with neuron-specific GO terms (Fig. 2H). Remarkably, lists of GO categories are nearly identical for common and NeuN(+)-specific borders. Collectively, these findings suggest the association between TADs borders and transcription in neuronal cells.

Furthermore, we segregated TADs into two distinct groups based on their coverage with lamina-associated domains (LADs), that are chromatin regions attached to the nuclear envelope[37]. LADs positions were approximated based on the lamin B1 occupancy sites in the developing human forebrain (gestation week 20)[38]. Subsequently, two groups of TADs located outside or within LADs were further classified into three size categories. Remarkably, large neuronal TADs nestled within LADs display distinct characteristics when compared to other TAD groups in NeuN(+) and the corresponding group in NeuN(-) (Fig. 2I, Suppl. Fig. 7). Specifically, we observe increased interactions at their borders, pronounced extrusion tracks, and increased interactions with surrounding TADs. Notably, borders of large TADs residing in LADs are enriched with GO terms associated with neuronal development (Suppl. Fig. 8). This finding suggests that the positioning of genes related to neuronal

development within TADs covered by LADs may be crucial for repressing their expression in mature cells. However, the presence of these genes at the borders could also indicate their conditional activation, potentially occurring during specific developmental stages.

Loops in NeuN(+) cells facilitate the formation of gene-enhancer hubs

Prominent differences in TAD profile and gene expression patterns between NeuN(+) and NeuN(-) cells imply differential looping between regulatory elements of the genome. Visual inspection of Hi-C maps reveals a number of differential loops, in particular between neuronal enhancers and genes with neuron-specific expression, for example, *ENC1* and *GRIN3A* (Fig. 3A). Genome-wide annotation of loops in 5-kb resolution Hi-C maps identifies 1078 shared loops and 1545 NeuN(+)-specific loops (Fig. 3B). We observe that the average loop intensity across cell-specific and common loops is the same for NeuN(+) and NeuN(-) cells (Suppl. Fig. 9A,B). Notably, neuronal loops are significantly longer than loops in non-neuronal cells (Fig. 3C). This observation aligns with the higher cohesin levels found in neurons and positive association between loop length and the cohesin residence time[39].

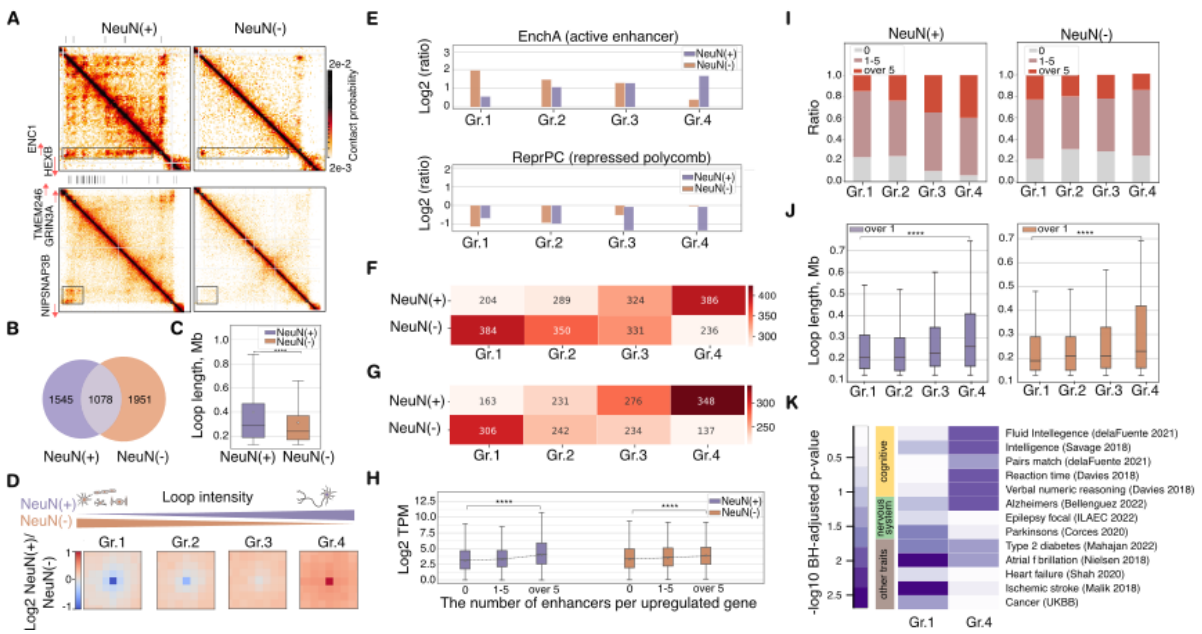

Fig. 3. Loops in NeuN(+) and NeuN(-) cells. (A) Fragments of 5-kb resolution Hi-C maps for regions of neuronal-specific gene interactions with enhancers. Enhancer tracks are shown at the top of Hi-C maps. (B) Venn diagram of the intersection of neuronal and non-neuronal loop positions. (C) Box plot of loop length distributions. Asterisks indicate Wilcoxon test p-values: **** - $p < 0.00001$. (D) Average loops within four groups defined based on the increase of NeuN(+)/NeuN(-) intensity ratio. (E) EnchA and ReprPC chromatin states in anchors of four loop groups calculated relative to the average across the genome as log2(ratio). (F) The number of upregulated genes placed within four loop groups. DE genes were determined as genes upregulated in the selected cell type with FC > 1.5. (G) The number of enhancer - promoter loops within four loop groups. (H) Box plot gene expression

grouped based on the number of interactions with enhancers ("0" – no interactions, "1-5" and "over 5" – 1-5 or more than 5 interactions). Asterisks indicate Wilcoxon test p-values: **** - $p < 0.00001$. (I) Ratio of genes with three types of interaction with enhancers within four loop groups. (J) Box plots of loop lengths mediating interactions between a gene and at least one enhancer. Asterisks indicate Wilcoxon test p-values: **** - $p < 0.00001$. (K) Linkage disequilibrium (LD) score regression p-values calculated based on GWAS studies[40–49] for NeuN(-) and NeuN(+) loops. UKBB - UK Biobank[50].

We then grouped shared loops based on their intensity ratio between neuronal and non-neuronal cells and determined groups 1 and 4 as strongest loops in NeuN(-) and NeuN(+) cells (hereinafter these loops are referred to as NeuN(-) and NeuN(+) loops, respectively, Fig. 3D). We first annotated chromatin states at loop anchors and observed that NeuN(+) loops are markedly enriched with NeuN(+) active enhancers and depleted with Polycomb-repressed chromatin state defined in NeuN(+) (Fig. 3E). Similarly, NeuN(-) loops are enriched with NeuN(-) active enhancers and depleted with NeuN(-) Polycomb-repressed chromatin state. Furthermore, both NeuN(-) and NeuN(+) groups are enriched with DE genes upregulated in NeuN(-) and NeuN(+) cells, respectively (Suppl. Fig. 9C, Fig. 3F), potentially suggesting the regulatory role of these contacts.

In addition, we show that both NeuN(-) and NeuN(+) loops are enriched with enhancer – promoter interactions (Fig. 3G). Moreover, genes interacting with multiple enhancers are on average expressed at a higher level as compared to other genes (Fig. 3H), and NeuN(+) loops are enriched with genes interacting with multiple enhancers (Fig. 3I). The presence of gene - multiple enhancers interactions especially pronounced in differential neuronal loops could be connected with a wider range of loop lengths (Fig. 3J), presumably contributing to establishing interaction on a larger spectrum of distances between genes and enhancers. A higher length of NeuN(+) loops is in line with our previous observation on the elevated cohesin level in neurons. We thus conclude that the NeuN(+)-specific transcription program might be characterized by the formation of "multivalent" active transcription hubs[51].

Furthermore, we applied Linkage Disequilibrium (LD) score regression to the NeuN(-) and NeuN(+) loops to understand how these loops are related to the human heritable traits (Fig. 3K; Suppl. Fig. 10). Intriguingly, we find a pronounced and significant enrichment of NeuN(+) loop anchors with cognitive traits exclusively. By contrast, NeuN(-) loop boundaries are enriched with non-brain-related traits, such as autoimmune, blood and cardiovascular disorders, in line with previous studies reporting general enrichment of disease-related genes at loop boundaries[52,53]. Collectively, these findings suggest that NeuN(+)-specific loops might regulate specialized transcription programs associated with human cognitive functions.

NeuN(+) chromatin is shaped around networks of long-range contacts between H3K27me3 loci

A remarkable feature of the NeuN(+) Hi-C map is the presence of bright dots located far away from the main diagonal and thus representing strong long-range contacts (we further refer to these features as "neuronal dots", Fig. 4A, Suppl. Fig. 11). We manually annotated 214 genomic loci that correspond to neuronal dots and then mapped significantly high Hi-C interactions of these loci using FitHiC2 software[54] to get neuronal dot annotation. We observe that Hi-C contact frequency of neuronal dots, on average, is much higher compared to contact frequency of the same pairs of regions in NeuN(-) (Fig. 4B). Hi-C contact frequency of neuronal dots decays with genomic distance remarkably slower than the genome-average (Fig. 4C) pointing at the presence of a molecular mechanism ensuring juxtaposition of dot anchors through large genomic distances. Neuronal dots form networks that include up to 11 loci (Fig. 4D). These networks can involve inter-chromosomal interactions, which are present almost exclusively in neurons (Fig. 4E).

Comparison with available epigenetic profiles[36] reveals that 92% (197/214) of neuronal dot regions overlap with H3K27me3 ChIP-seq peaks. H3K27me3-marked genomic regions are known to be enriched at lamina-associated domain (LAD) boundaries[37]. Therefore, we hypothesize that H3K27me3 positioning near lamina increases the probability of neuronal dot regions to get in contact. Analysis of laminB1 occupancy sites in the developing human forebrain (gestation week 20) as a proxy for LADs[38] reveals that the stronger interacting neuronal dot loci have the greater abundance of LADs within genomic regions encompassing these loci (Fig. 4F). By contrast, LADs occupancy is strongly depleted at the precise positions of interacting H3K27me3-marked loci.

To further study the interplay between heterochromatin and neuronal dots, we analyzed ChIP-seq data for H3K9me3, a histone mark of a constitutive heterochromatin. We compared the fetal brain H3K9me3 data as a proxy for NeuN(+) with astrocyte H3K9me3 data as a proxy for NeuN(-)[55]. We observed that neuronal dot regions are enriched with H3K9me3 in astrocytes, but not in the fetal brain (Fig. 4G). This might indicate that neuronal H3K27me3-marked loci are less occupied by H3K9me3 in NeuN(+) than in NeuN(-), which could have affected the intensity of dot loci.

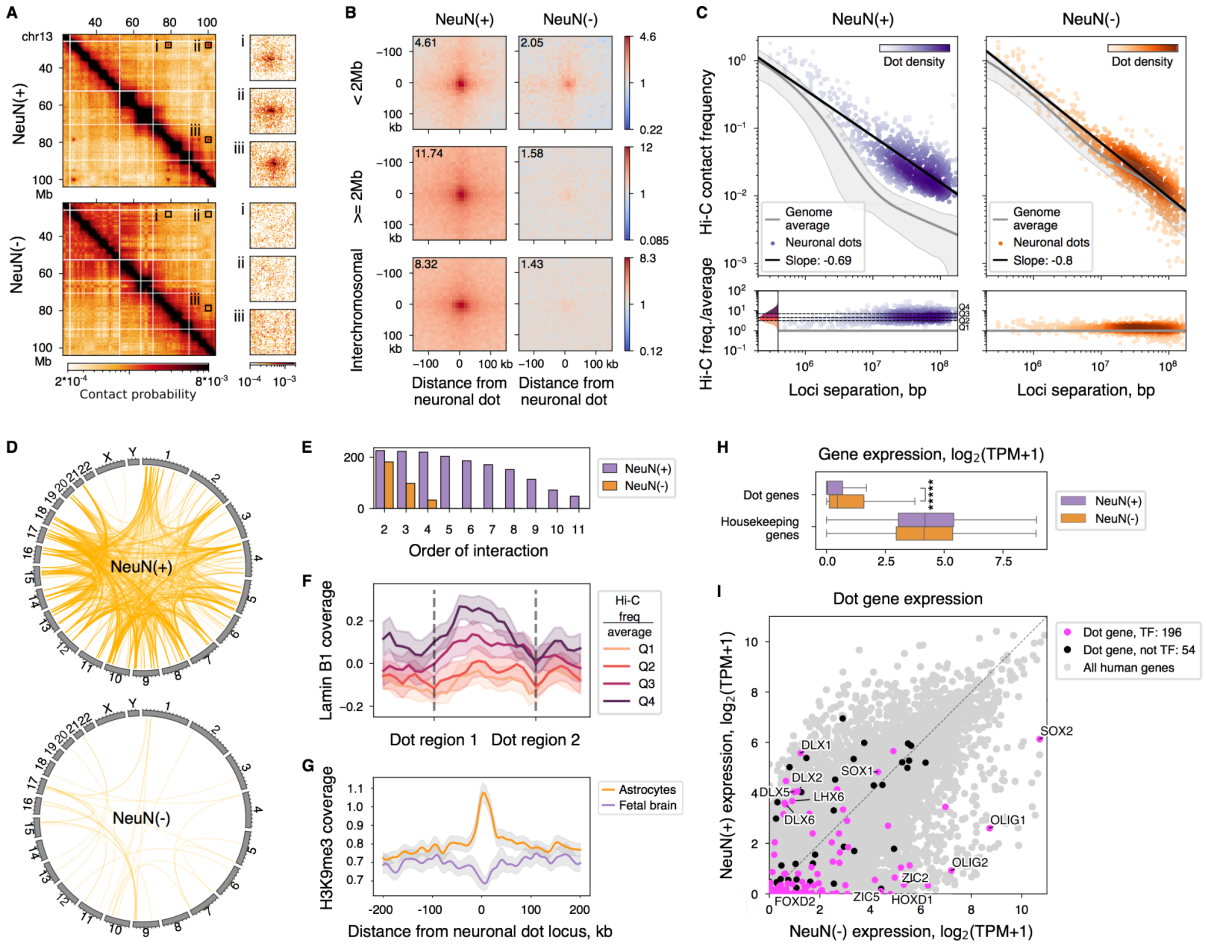

Fig. 4. Features of neuronal dots. (A) Fragment of a Hi-C map with neuronal dots present in NeuN(+) (left), but absent in NeuN- (right). (B) Average Hi-C signal (observed over expected) of neuronal dots (left) and corresponding pairs of neuronal dot loci in NeuN(-) (right). Value in the corner corresponds to the central pixel. (C) Left: contact scaling of neuronal dots (purple), approximated by a linear regression - black line. Right: contact scaling of pairs of neuronal dot loci in NeuN(-) (orange), approximated by a linear regression - black line. Whole-genome expected is the same as in Fig. 1E. (D) Significant interchromosomal interactions of neuronal dot loci in NeuN(+) and NeuN(-). Significance is defined by FitHiC2, $p < 0.05$. (E) Statistics for the size (order) of cliques (networks) made by neuronal dots (purple bars) or neuronal dot loci in NeuN(-) (orange bars). (F) Average Lamin B1 ChIP-seq read coverage profile between pairs of neuronal dot anchors. Horizontal axis is scaled so that separation between each pair of anchors is constant. Anchor locations are marked by gray dotted lines. Pairs of anchors are separated into four quantiles based on corresponding Hi-C interaction intensity (normalized by expected at given loci separation). Areas around the coloured lines - standard deviations. (G) Average H3K9me3 ChIP-seq read coverage profile at neuronal dot anchors. Area around lines - 95% confidence interval. (H) Expression of protein-coding genes located within anchors of neuronal dots compared with expression of housekeeping genes. (I) Expression of protein-coding genes located within anchors of neuronal dots. Genes are divided into TFs (magenta) and non-TFs (black). Grey dots - expression of all human protein-coding genes.

Most loci brought together by neuronal dots (171/214) reside near the transcription start sites (TSS) of protein-coding genes (Suppl. Fig. 12). Interestingly, the composition of chromatin states in these TSSs is different between NeuN(+) and

NeuN(-): in neurons the largest fraction of TSSs is “bivalent”, i.e. occupied by H3K27me3 together with H3K4me3, whereas in non-neurons most TSSs are occupied by H3K27me3 without H3K4me3 (Suppl. Fig. 13). This difference is specific to neuronal dots, since the genome-wide ratio of “bivalent” and “H3K27me3-only” TSSs is similar in NeuN(+) and NeuN(-) (Suppl. Fig. 14).

Out of all protein-coding genes within neuronal dot loci, 80% are transcription factors (hereby referred to as “dot TFs”). Most dot TFs are repressed in the adult brain in both NeuN(+) and NeuN(-), though in NeuN(+) the expression is significantly lower (Fig. 4H,I). Gene Ontology analysis showed that dot TFs are enriched with biological functions related to the organism development (Suppl. Fig. 15). Indeed, some dot TFs belong to gene families known to be involved in organism development, e.g., HOX[56], SOX[57] and DLX[58].

To further investigate the role of dot TFs, we analyzed the public data for human gene expression during development[59] and found that most dot TFs are expressed in at least one cell type in the developing organism (Suppl. Fig. 16). We conclude that dot TFs are expressed during development and must be repressed in the mature brain.

To study the localization of neuronal dots within the nucleus, we performed immunohistochemistry analysis of the cortex nuclei in the BA22p brain area of four human individuals. The results reveal H3K27me3-immunopositive patches adjacent to the nucleolus, which is defined by a prominent decrease of DAPI intensity (Fig. 5A). Typically, there are several H3K27me3-immunopositive patches within astrocyte nuclei (Fig. 5B), whereas neurons display a single, highly intense H3K27me3 patch next to the nucleolus (Fig. 5C). These patches are found at an average distance of 1.67±0.25 μm (14.3±3.7% of the nucleus diameter) from the nucleolus center in neurons and presumably represent nucleolus caps defined previously[60,61].

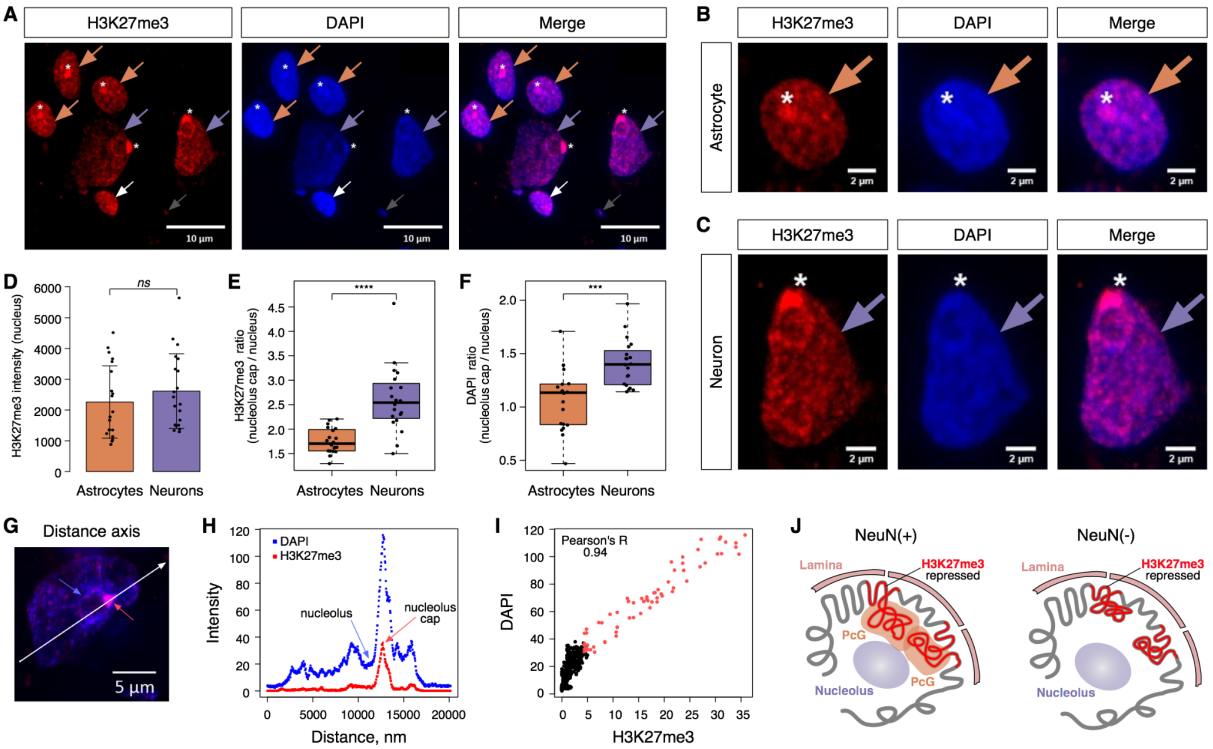

Fig. 5. Immunohistochemistry analysis of the human cortex nuclei. (A-C) H3K27me3 immunostaining in the cell nuclei of the human cortex: H3K27me3 is red, DNA is blue. Orange arrows mark astrocytes (B), violet arrows – neurons (C), white arrow – oligodendrocyte, gray arrow – microglial cell, asterisks – H3K27me3-immunopositive patches (presumably, nucleolar caps). Scale – 10 microns (A) and 2 microns (B-C). (D) H3K27me3 intensity in the whole nucleus. Bars represent means, whiskers - standard deviations, dots - individual data points. *ns* - P-value > 0.05, one-sided Wilcoxon test. (E-F) H3K27me3 (E) and DAPI (F) intensity ratio between the nucleolar cap and the whole nucleus. *** - P-value < 10^{-4}, **** - P-value < 10^{-5}, one-sided Wilcoxon test. (G) An example of a neuron with the distance axis represented by the white arrow. The red arrow marks the nucleolar cap, blue arrow - nucleolus, here and in panel H. (H) Colocalization of H3K27me3 (red dots) and DAPI (blue dots) intensity along the distance axis defined in panel G. (I) Correlation between H3K27me3 and DAPI signals (Pearson's R=0.94, p-value<0.05). Nucleolar cap is highlighted in red. (J) Schematic of neuronal dots overlapping with H3K27me3 loci (colored in red) enriched at LAD boundaries and bound by PcG proteins (colored in orange).

While the total level of H3K27me3 signal is similar between the neuronal and astrocyte nuclei (Fig. 5D), H3K27me3 intensity is significantly higher in the neuronal nucleolar caps compared to astrocyte nucleolar caps (Fig. 5E). In addition, the neuronal nucleolus caps demonstrate significantly higher DAPI signal than astrocyte nucleolar caps, relative to the total DAPI signal within the corresponding cell type nuclei (Fig. 5F). This observation aligns with the strong neuronal enrichment of long-range contacts, represented by neuronal dots in Hi-C maps, which overlap with H3K27me3 ChIP-seq peaks.

Moreover, H3K27me3 and DAPI show remarkable colocalization in the neuronal nuclei. There are prominent peaks at the position of the nucleolus cap, which are juxtaposed in both the H3K27me3 and DAPI intensity profiles (Fig. 5G,H, Suppl.

Fig. 17). Consistently, the intensity profiles of H3K27me3 and DAPI are highly correlated in the neuronal nuclei (Pearson's R = 0.88±0.04, p-value<0.05, Fig. 5I).

Collectively, this immunohistochemistry analysis lends support to our Hi-C data indicating the juxtaposition of H3K27me3-marked dot anchors over large genomic distances specifically in neurons. These dots form networks that are predominantly present in neurons. The H3K27me3 immunostaining further suggests their preferential co-positioning near the nucleolus as a singular H3K27me3-immunopositive patch, presumably a nucleolus cap.

Genes from human neuronal dots are coupled with PcG proteins in mouse

Neuronal dot regions overlap with H3K27me3, a histone modification that is deposited by Polycomb repressive complex 2 (PRC2) and can be bound by PRC1[62]. Therefore, we hypothesized that H3K27me3 and polycomb-group (PcG) proteins are involved in the formation of neuronal dots. In mice, H3K27me3-bound loci establish long-range interactions in embryonic stem cells (ES), neural progenitor cells (NPC) and cortical neurons (CN)[13]. We thus explore if mouse orthologs of human dot genes are implicated in H3K27me3-coupled long-range interactions. Indeed, many dot genes overlap with the H3K27me3 mark in mice (219, 169 and 192 out of 245 genes for ES, NPC and CN correspondingly) and form long-range interactions.

Moreover, alongside H3K27me3, the mouse orthologs of dot genes are enriched with the ChIP-seq signal of RING1B (Suppl. Fig. 18), a component of the PRC1 complex, which is known to be involved in polycomb-mediated contacts[14]. To further investigate the role of RING1B in the repression of genes within neuronal dots, we analyzed a publicly available dataset on RING1A/B knockout in mouse motor neurons[63]. We found that RING1A/B knockout leads to the upregulation of 53% (125) of dot genes and downregulation of only 4% (10) of dot genes (Suppl. Fig. 19A). This observation suggests that RING1A or RING1B proteins likely contribute to the repression of dot genes in human neurons.

Since the H3K27me3 mark is deposited by the PRC2 complex, we hypothesized that PRC2 might play an important role in neuronal dot formation. Accordingly, PRC2 knockout in mouse striatal neurons leads to upregulation of TFs, many of which can be positively auto-regulated[64]. Furthermore, we find that a large fraction of TFs upregulated in PRC2 knockout overlaps with the neuronal dot TFs (Suppl. Fig. 19B), suggesting that PRC2 is indeed involved in the dot formation.

Consequently, there is compelling evidence supporting the involvement of PcG proteins, specifically RING1B, in the formation of neuronal dots..

Discussion

Our results suggest that chromatin folding in neurons of the human brain possess several distinct features. First, at the large scale, chromosomes in neurons are more segregated compared to non-neuronal cells as revealed by calculation of a fraction of inter-chromosomal contacts. Intuitively, this observation correlates with the expanded volume of neuronal nuclei[27] as the increase in the nucleus volume provides an additional space for chromatin. Following this logic, individual chromosomes are expected to expand to fill up the free space. Yet, we observe an opposite direction of differences in the total level of intra-chromosomal contacts between NeuN(+) and NeuN(-) cells. Moreover, it has been reported that artificial nuclei expansion[28] does not change global chromatin spatial organization, including chromosome territories. Therefore, factors other than the nuclear volume might contribute to global differences in genome organization between neuronal and non-neuronal cells that we observe.

At the local scale, short-range chromatin interactions (< 3 Mbp) are more abundant in neurons. Consistently, TADs are more prominent in neurons, and chromatin loops are, on average, longer. The formation of TADs and loops is facilitated by the loop extrusion machinery[65,66]. According to public transcriptome and proteome data, proteins involved in loop extrusion, such as the cohesin complex (RAD21, SMC1A, SMC3, STAG1, STAG2), CTCF, cohesin loading factors (NIPBL, MAU2) and unloading factors (WAPL, PDS5A), are more abundant in neurons, and genes encoding these proteins are more highly expressed (Suppl. Fig. 6A, Suppl. Tables 2A, 3). Therefore, higher cohesin activity may impact neuronal chromatin and increase the prominence of TADs in neurons.

The retention of cohesin at TAD borders allows genes and regulatory elements located within these regions to interact with distant targets[67]. We observe that neuronal-specific TAD borders and chromatin loops are enriched with genes upregulated in neurons, as well as with active histone modifications associated with active genes and enhancers. These findings indicate that increased cohesin activity plays an important role in the regulation of gene expression in neurons and complies with previous findings[23,26]. For instance, Calderon et al. demonstrated that the deletion of RAD21, a subunit of the cohesin complex, leads to incorrect expression of genes responsible for neuronal development in mouse immature cortical and hippocampal neurons. As we have shown, changes in gene expression upon cohesin knockout correlate with differences between NeuN(+) and NeuN(-) expression, highlighting the importance of cohesin in establishing and maintaining neuronal identity.

Besides neuronal-specific TAD borders, we study the common TAD borders that are enriched with neuron-specific genes and active chromatin marks in neurons. Consequently, we propose that the composition of TAD borders in different brain cell

types may align with neuronal transcriptional activity. Potentially, the presence of remnants of neuronal-specific chromatin organization patterns in non-neuronal cells might be attributed to the sequential order of cellular differentiation. Indeed, common precursors first differentiate into neurons, while gliogenesis occurs at a later stage[68]. Conforming to the proposed hypothesis, the majority of TAD boundaries exhibit a conservative positioning throughout the developmental transition from ESCs to neural cell types, with only a small fraction of TAD boundaries emerging or mitigating[69].

Distal intra-chromosomal contacts (> 3 Mbp) decay fast in neurons. The slope of contact decay with distance reaches -1.5 in comparison with -1 in non-neurons. This difference cannot be explained by the increased nucleus size in neurons as Sanders et al. demonstrate, on the contrary, slower contact decay with distance upon nuclear volume swelling[28]. But they show that increased nucleus size leads to compartment strength weakening, similar to what we and others[23] observe in neuronal Hi-C. However, nuclear swelling does not affect the genomic positioning of compartments[28], while compartments in neurons are largely different from non-neurons. Therefore, we propose that the observed compartmentalization differences in NeuN(+) cells, such as lower strength, altered positioning and lower correlation with histone marks, may be attributed to the more active loop extrusion in neurons counteracting compartmentalization, similar to what Schwarzer et al. observe upon NIPBL knockout[34].

A remarkable observation of our Hi-C analysis is the presence of bright long-range interactions between neuronal H3K27me3-enriched regions - neuronal dots - that are formed both in *cis* and in *trans*. Despite the rapid decay in the whole-genome expected contact frequency in neurons, neuronal dots are able to establish contacts even at large genomic distances up to 100 Mb. These regions form large interaction networks, comprising up to 11 distant loci. Our immunostaining experiments support these observations by demonstrating that a large portion of H3K27me3-marked chromatin in neurons is concentrated in a spatially constrained patch near the nucleolus. We hypothesize that this patch could be formed by loci involved into the neuronal dot network (Fig. 5J).

Within the neuronal dots, a remarkable proportion of genes (about 80%) encode transcription factors (TFs), the majority of which are expressed in neural or non-neural tissues during development but are downregulated in mature NeuN(+) and NeuN(-) cells. Since the H3K27me3 mark is associated with polycomb repression, we suggest that PcG proteins repress TF genes within the neuronal dots. This hypothesis is supported by two studies conducted in mice[63,64], and demonstrating that knockouts of PRC2 or RING1A/B proteins in mouse neurons lead to the upregulation of many mouse orthologs of dot genes. Furthermore, expression of PcG genes and abundance of PcG proteins are higher in neurons compared to NeuN(-) cells according to publicly available scRNA-seq and proteomic data (Suppl. Fig. 6B, Suppl. Table 2B,4).

RING1B protein, a member of Polycomb repressive complex 1 (PRC1), is of particular interest since it is required for long-range polycomb interactions[14]. Reanalysis of public data form mouse differentiating neurons[13] shows that: 1) many orthologs of dot genes are occupied by RING1B, 2) some of these RING1B-occupied orthologs form long-range interactions in embryonic stem cells, neural progenitor cells and differentiated neurons. This raises a possibility that RING1B binds to neuronal dot loci and is involved in neuronal dot formation.

Another peculiar feature of neuronal dot loci is that 60% of genes located at their anchors have bivalent promoters, i.e. their transcription start sites (TSSs) are simultaneously occupied by H3K27me3 and H3K4me3. In non-neurons only ~30% of these genes have bivalent promoters. Such "bivalent" genes in neurons were thoroughly investigated by Ferrai et al., who compared "bivalent" and "H3K27me3-only" genes and demonstrated that upon PRC2 enzymatic inhibition "bivalent" genes get upregulated, while "H3K27me3-only" genes do not[70]. Thus, it is possible that "bivalent" repressed genes have more potential for activation than "H3K27me3-only" genes. If this hypothesis is true, then an interesting question for further investigation is why some genes in mature neurons, but not in NeuN(-) cells have this potential to become activated.

Is there a link between neuronal dots and "bivalency" of dot genes? "Bivalent" promoters are known to be occupied by poised RNA polymerase II (RNAP)[70,71], which is dependent on RING1A and RING1B ubiquitination[71]. Thus, both long-range interactions and "bivalency" could be a consequence of RING1A and/or RING1B occupancy at promoters of neuronal dot genes.

Are neuronal dots functionally relevant? Do these interactions affect gene expression at corresponding loci? Kraft et al. demonstrated that long-range PcG-mediated interactions facilitate H3K27me3 spreading, such that the deletion of one interacting locus leads to gene activation in the other[17]. While we observe significant downregulation of gene expression at neuronal dot loci compared to NeuN(-), in general, genes within these regions are repressed in both groups. This makes sense, since considered genes are enriched with developmental functions and probably should be inhibited in all types of mature cells.

What biological mechanisms can facilitate such strong PcG-mediated interactions in neurons? One possible explanation is that neurons reside in a post-mitotic state, unlike non-neurons. In general, chromatin organization is largely altered during the cell cycle[72]. Ma and Buttitta demonstrated that, in *Drosophila*, H3K27me3 histone mark clustering increases for cells with slower proliferation and reaches maximum for post-mitotic cells[73]. They further show that delaying cell cycle exit disrupts H3K27me3 clustering. Thus, a neuronal post-mitotic state might facilitate interactions between H3K27me3-enriched genomic regions.

Interestingly, we observe that the higher abundance of LADs between, but not within, neuronal dot loci correlates with stronger Hi-C interactions. It is possible that chromatin anchoring to nuclear lamina facilitates interactions of the neighboring chromatin regions. An interesting open question is the interplay between PcG-mediated interactions and LADs/compartmentalization. Siegenfeld et al. developed a LIMe-Hi-C experiment that allows to simultaneously capture genome 3D structure and lamina association, and demonstrated that inhibition of EZH2 protein, a catalytic subunit of PRC2, leads to increased compactization of heterochromatin attached to the nuclear lamina[74]. They suggested that PcG-mediated interactions and LAD interactions are antagonistic processes. This hypothesis agrees with our NeuN(+) Hi-C data where we observe neuronal dots and reduced B compartment strength compared to NeuN(-).

In conclusion, neuronal genome organization is largely different from non-neuronal one. Many biological processes, including loop extrusion, compartmentalization, long-range interactions, are likely to work differently in neurons and probably facilitate correct gene expression in this cell type.

Methods

Tissue collection

Post-mortem human brain samples were obtained via National BioService Russian Biospecimen CRO (St. Petersburg, Russia). Informed consent for using human tissues for research was obtained in writing form from all donors or their next of kin by the tissue provider bank. Sampled brain tissue of all subjects was defined as healthy by medical pathologists. All subjects had sudden death with no prolonged agony state and had no history of neurologic and psychiatric disorders or alcohol and drug abuse. Post-mortem brain frontal blocks 1-1,5 cm thick were stored at -80°C warped in a foil, in zip-lock packages. For sample dissection, The Atlas of the Human Brain (AHB)[75] was used to locate the area of interest, a posterior part of the superior temporal gyrus of the left hemisphere (posterior Brodmann area 22 (BA22p), Wernicke's area). Frontal blocks with surfaces nearest to AHB level: 60 (MNI: -36.57; ICL: 34.82; MCP: 20.08) were placed into the cryostat Leica CM1950 chamber with -20°C before dissection for 15-20 min on custom-made aluminum support. A tissue piece of approximately 150 mg weight was cut out from a selected area using a metal scalpel. Dissected samples were then collected with tweezers, put into tubes, and placed in dry ice immediately. All materials used during dissection (scalpels, tweezers, tubes) were sterile and chilled using dry ice before use.

Nuclei isolation

Nuclei isolation from brain tissue and further NeuN antibody immunostaining were performed according to the previously published STAR protocol[24] with optimization for human brain tissue. Briefly, we dissected and fixed 200-300 mg of frozen brain tissue with a 1% formaldehyde-based buffer and homogenized it with a plastic pestle on ice. The brain homogenate was mixed for 10 minutes at room temperature, then the crosslinking reaction was quenched with 2M Glycine and centrifuged at 1100g +4 °C during 5 min. After that, the pellet was washed several times with NF1 buffer (Tris-HCl 10 mM, pH 8.0; EDTA 1 mM; MgCl2 5 mM; Sucrose 100 mM; Triton X100 0.5%), homogenized with Dounce homogenizer 10–30 times on ice and filtered to the new tube. Then sucrose mix (1.2M Sucrose) was layered on homogenized samples to isolate nuclei with ultracentrifugation in a sucrose gradient (1600g, 5 min, +4 °C). The nuclei pellet was resuspended in PBST with 5% BSA and 3% bovine serum, the isolated nuclei were stained with NeuN antibody (1:10000 anti-NeuN-525, FITC) for at least 1 hour, and then resuspended in PBST with 5% BSA. Before FANS (Fluorescence-Activated Sorting of fixed Nuclei) procedure nuclei were filtered with a 35-µm cell strainer into the Falcon™ Round-Bottom Polystyrene Test Tubes.

The sorted NeuN-positive and NeuN-negative nuclei populations were collected and used in amounts of 200,000-400,000 sorted nuclei per sample for Hi-C libraries preparation.

Hi-C library preparation

Preparation of bulk Hi-C libraries from 11 samples of human brain tissue was performed as previously described[2,76] with optimal changes to brain tissue[24]. Approximately 200,000-400,000 sorted nuclei for the sample were lysed in a solution of lysis buffer containing 150 mM Tris-HCI pH7.5-8, 150 mM NaCI, 0.5% (v/v) NP-40, 1% (v/v), Triton-X100, 1x Protease Inhibitor (CALBIOCHEM, #539137) on ice for 15 min, centrifuged at +4 °C 5000g for 7 min, the pellet was resuspended and washed in 1.12X restriction enzyme DPN II buffer. After the final wash, the pellet suspension in 1.12X DPN II buffer with added 0.3% SDS has incubated for 60 min 1400 rpm at +37°C. Further, the reaction was quenched with 1.8% Triton X-100 for 60 min at 1400 rpm at +37°C, and the 500U DPN II restriction enzyme was added, the suspension was left overnight at 1400 rpm at +37°C. After that, the additional amount of 200U DPN II restriction enzyme was added to the digested nuclei and incubated for the next 120 min at 1400 rpm at +37°C. After that, the restriction was unactivated at +65°C for 20 min. The digested DNA samples were washed and resuspended in 1.2XNEB2, and after that digested ends were marked with biotin during the incubation in the mix of 1M biotin-14dATP, 10x NEB2, 10mM dCTP, 10mM dGTP, 10mM dTTP, 5U/ul Klenow (NEB #M0210S) for 90 min 900 rpm at +37 °C. The

resulting blunt ends were washed with 1X T4 DNA ligase buffer and chromatin fragments were proximally ligated with 50 U T4 DNA ligase (Fermentas) for 6 h 1400 rpm at +37 °C. Then the 0.5% SDS with 20 µg/ml proteinase K was added to samples for overnight incubation at +65 °C. The following step was DNA purification with phenol–chloroform extraction and precipitation with 2.5 volumes of 96% ethanol, 0.1 volumes of 3M sodium acetate, 20 µg/ml glycogen, and 9.4 µg/ml tRNA for 60 min at −80 °C. After the centrifugation step at 20000g for 20 min, +4°C the pellet was dissolved in 10 mM Tris-HCl pH 8.0 with the addition of RNase A and incubated for 45 min at +37°C. The next step was DNA sample purification on 2.5 volumes of magnetic AMPure XP beads with the following DNA elution in 10 mM Tris-HCl for 15 min at +55°C. Eluted DNA was fragmented by sonication with Power 15 regime, 4 repeats for 30 s with 3 min break. Sonicated DNA samples were then concentrated on Amicon columns 14000g/5 min with double 10 10 mM Tris-HCl pH 8.0 washing. The resulting biotin-marked fragments were isolated using Dynabeads MyOne Streptavidin C1 (Invitrogen #65001) and diluted in TWB buffer (1M Tris-HCl pH 8.0, 50 mM EDTA, 5M NaCl, 1% Tween). All subsequent steps were performed on that bead-bound DNA fraction including DNA ends repair with repair mix (10X ligase NEB buffer, 10 mM dNTP mix, T4 DNA Pol, T4 PNK, 1 ul Klenow) for 30 min at room temperature, and PolyA tail adding reaction with a mix containing 10X NEB2, 10 mM dATP, Klenow (exo-) and incubation for 30 min at +37°C and then for 20 min at +65°C. Then beads were washed with 1X T4 ligase buffer for 3 min at room temperature 900 rpm, and the ligation reaction was performed with a ligation mix of 10X T4 ligase buffer, adapter, 5 U T4 ligase (Fermentas) for 2.5 h at room temperature. Finally, libraries were PCR amplified using Illumina forward and reverse primers and KAPA HiFi HotStart for 16 cycles to select the appropriate cycle for amplified DNA fragments collection (the 6, 9, 12, and 15 cycles were collected and checked for DNA quantity on chosen cycle). Last, up to the chosen number of cycles Hi-C libraries were PCR amplified, and resulting fragments were purified using 1.5 volumes AMPure XP magnetic beads.

Hi-C data processing

Raw reads of each Hi-C library were mapped to the human hg38 reference genome with the distiller-nf pipeline (https://github.com/open2c/distiller-nf, v.0.3.3)[77] which utilizes BWA for genome mapping procedure, ensuring valid reads filtering, 'binning', and generating list of valid pair contacts which transformed into binned matrices of Hi-C interactions.

Most of the Hi-C analysis was performed on Hi-C matrices merged by cell type. Merging was done at 1 kbp resolution using cooler.merge_coolers() function of a cooler library[78], v.0.8.11. The resulting merged Hi-C maps were sampled to the equal number of contact pairs using cooltools.sample() function of a cooltools library[79],

v.0.5.1. To avoid possible bias due to Hi-C ligation artifacts, the values at the main diagonal were removed prior to sampling using cooler and custom python code. Resulting .cool files with Hi-C matrices were ICE normalized[80] using "cooler balance" command with default parameters.

Hi-C compartments

Merged Hi-C maps in .cool format[78] at 50 kbp resolution were filtered to remove consecutive genomic regions (bins) that are empty in at least one of two matrices. Hi-C compartments were called using cooltools software (v0.5.1)[79], eigs_cis() function with parameters: n_eigs=5, sort_metric='spearmanr'. ChIP-seq H3K27ac histone abundance profile was used as a phasing track. Chromosomes Y and M and short chromosomal arms: chr13_p, chr14_p, chr15_p, chr21_p, chr22_p - were ignored. For neurons, the principal component (PC) that correlated best with H3K27ac histone abundance was used to analyze compartments. For non-neurons, we visually compared the first two PCs that correlated best with H3K27ac and for each chromosomal arm chose the PC that best described the Hi-C matrix. Saddle plots were calculated using cooltools.saddle() function with parameters: contact_type='cis', n_bins=50, qrange=(0.025, 0.975), min_diag=4. The code for saddle plots was adapted from cooltools tutorial: https://cooltools.readthedocs.io/en/latest/notebooks/compartments_and_saddles.html. To calculate compartment interaction strength at Fig. 1I, the mean intensity of the top 20% of corresponding interactions obtained from the saddle matrix were used.

To calculate correlations of Hi-C maps PC1 with chromatin states, the genome was split into 100 kbp bins. Chromatin states annotation for neuronal and non-neuronal cells was retrieved from Dong et al.[36]. The annotation includes six distinct chromatin states: 'Quies', 'ReprPC', 'TssBiv', 'TssFlnk', 'TssA', and 'EnhA'. For each genomic bin, the proportion of the bin occupied by each chromatin state was calculated. Obtained values of chromatin state proportions were used to calculate pairwise Spearman's rank correlation coefficients between states and PC1 values. The correlations between PC1 and chromatin states were calculated for neurons and non-neurons separately.

Insulation score profiling and TAD calling

Insulation score (IS) was calculated genome-wide for 15-kb resolution NeuN(+) and NeuN(-) Hi-C maps with the following parameters: min_frac_valid_pixels=0.82, min_dist_bad_bin=10, window = 120_000. Principal component analysis was based on the obtained IS profile variation among all produced Hi-C maps, as well as similarly re-analyzed publicly available NeuN(+), NeuN(-) Hi-C maps (Hu et al., 2021) and other tissue and cell line Hi-C maps[81,82].

The identification of TAD relied on the genome-wide insulation score identification. TAD calling on Hi-C data was implemented with cooltools software (v0.5.2)[79]. Module "insulation" available in the software utilizes one of the common algorithms for insulation profile calculation that is based on diamond-window score. Insulation score was defined genome-wide on NeuN(+) and NeuN(-) maps at 15 kbp resolution with the following parameters: min_frac_valid_pixels=0.82, min_dist_bad_bin=10, window = 120_000. TAD boundaries were defined based on the border strength with the application of the "Li" threshold from scikit-image[83].

NeuN(+) and NeuN(-) Insulation tables retrieved from cooltools were converted to bedGraph format by extraction columns representing chromosome, start or end of the interval and logarithm of the insulation score. Then, bedGraphToBigWig binaries available at UCSC Genome Browser web-page were applied to convert insulation score files into bigWig format[84].

Identification of cell-specific TAD borders

TAD is referred to in this work as the cell-type-specific in one of two cases: 1 - its border has no intersection with all other TADs identified in another type of cell, 2 - ratio of border strengths between NeuN(+) and NeuN(-) cells is not within Q1 to Q3. Identification of the cell-specific TAD borders relied on several steps implemented with the bedtools package[85]. First, each TAD was widened by 1.5 bins (22.5 kbp) at both sides from the original TAD border with "bedtools slop -b 22500". This step ensures that TAD borders error calculation will not significantly affect the intersection and precision of differential TADs identification. Then, cell-specific .bed files with increased TAD borders were intersected with "bedtools intersect -a glia1.bed -b glia2.bed > all_glia.bed" to account for the possible variations between samples. These steps resulted in the TAD borders presented in glial and neuronal samples. Finally, these files were subtracted from each other to identify differential TAD borders with "bedtools intersect -a all_glia.bed -b all_neuron.bed -v > glial_unique.bed". As the results of the procedure, part of the neuron-specific and glia-specific TAD borders were identified along with common TADs.

Common TADs were inspected further to identify borders whose strength is altered significantly between cell types. First, the ratio of border strength was calculated for each individual border and the following identification of the quartiles of ratio distribution. Borders with ratio absolute value more than third quartile were added to respective cell-specific TAD borders.

Analysis of TAD covered with LADs

LaminB1 ChIP-seq data from the developing human forebrain (gestation week 20) (Ahanger et al. 2021) was used as a proxy for LADs positions. The genomic coordinates of laminB1 peaks were intersected with the positions of TADs in NeuN(+)

and NeuN(-). TADs that intersected with laminB1 peaks, covering more than 99% of their size, were categorized as being located within LADs. Conversely, TADs with 0% coverage were classified as positioned outside of LADs.

Size groups of TADs were identified based on the distribution of TAD sizes across all samples. Specifically, TADs were categorized into three distinct groups based on their sizes in relation to the lower and upper one-third quantiles of the distribution.

Annotation of loops

Loops positions were identified with the "cooltools.dots" module of cooltools software (v0.5.2)[79]. Hi-C maps with the resolution of 5 kbp were utilized. The following parameters were set for loop calling: max_loci_separation = 10_000_000, max_nans_tolerated = 1, lambda_bin_fdr = 0.05, clustering_radius = 20_000, cluster_filtering = None.

Loops density identification

The loops union was obtained by gathering loop positions found in NeuN(+) or NeuN(-). Then, off-diagonal pileup was determined individually for each loop with "cooltools.pileup" module of cooltools software (v0.5.2). This module allows to extract chromatin density information with the window centered at the pixel with one anchor as a left loop coordinate and another anchor as a right loop coordinate. Then, the ratio between NeuN(+) and NeN- densities of the central pixels were defined for each loop individually. Distribution of ratios of central pixel densities were used to identify loop groups based on quartiles. Thus, four groups representing loops with central pixel density ratio laying lower Q1 (Group 1), between Q1 and Q2 (Group 2), between Q2 and Q3 (Group 3) or higher Q3 (Group 4) were further examined with different approaches.

LD score regression analysis in loop anchors

To study the link between these loops and human heritable traits, we applied LD score regression to the anchors of loops in Groups 1 and 4 defined above using the stratified LD score regression method, which allows partitioning of heritability from GWAS summary statistics. We obtained GWAS summary statistics files for a total of 234 traits from either the United Kingdom Biobank (UKBB) or from individual GWAS studies. We followed the LD score regression tutorial available at https://github.com/bulik/ldsc/wiki, using default parameters as suggested. The results are presented as LDSC enrichment p-values, which were adjusted for multiple comparisons with the Benjamini-Hochberg method.

Loops annotation with differentially expressed genes

Loops containing genes that are differentially expressed (DE) between NeuN(+) and NeuN(-) in their anchors became other types of loop groups. Cell-specific loop positions were intersected with the position of DE genes possessing the respective direction of log2FC. Specifically, loops with log2FC > 0.58 were selected for NeuN(+) loop study, and genes with the opposite expression change were utilized for another cell type. The function "pair_to_bed" of pybedtools[86] was utilized for retrieval of overlaps between a .bedpe file and .bed. Thus, two groups were identified for NeuN(+) and NeuN(-) based on the positioning of DE genes: loops containing DE genes in one or both anchors and loops with no intersection with DE genes.

Chromatin features annotation with chromatin states

Chromatin states annotation for neuronal and non-neuronal cells was retrieved from Dong et al.[36]. For annotation of TADs, each border was firstly intersected with the chromatin states data. Then, the percent of coverage with each of six states was calculated. The final results are presented as the log2-normalized ratio of the percent coverage of the selected state found in the border to the genome-wide percent coverage of this state. To study the association of the TAD density with enrichment of active or inactive states, we identified two groups of TADs with the increased ratio of ReprPC or EnhA states. Initially, the upper quartile of the distribution of ReprPC or EnhA states coverages within TAD borders of the corresponding cell type was identified. Then, TADs with exceeding coverage were separated into ReprPC or EnhA groups for further density estimation. Significance of change between groups was assessed with a two-sided Mann-Whitney-Wilcoxon test.

Loops annotation was performed in two modes. First, each loop anchor was annotated independently following the strategy described for TADs. Then, loops were characterized based on the chromatin states assigned to the left and right anchor.

Analysis of long-range H3K27me3 interactions

Annotation of neuronal dots was done in two steps. First, we performed manual annotation of bright dot regions on a Hi-C map using HiGlass visualization software [87]. Second, we filtered our manual selection to keep a subset of significantly enriched Hi-C interactions. For this, we used FitHiC2 software (v.2.0.8)[54,88] with parameters: "-r 100000", "--contactType All". To calculate the *bias* values, we used the inverse of "weight" values obtained after iterative correction procedure and normalized them to the average of 1:

$$bias = \frac{weight^{-1}}{mean(weight^{-1})},$$

where $mean(weight^{-1})$ is calculated over all genomic bins. The resulting FitHiC2 interactions were filtered for q-value < 0.05.

On Fig. 4B, FitHiC2 filtering was not used, and all pairs of manually annotated regions were considered. Average interaction was calculated for merged Hi-C maps at 10 kbp resolution using coolpup.py software (v.1.0.0)[89].

Publicly available ChIP-seq data for H3K27me3-enriched regions in ENCODE broadPeak format were used[36].

To find cliques (networks) of NeuN(+) and NeuN(-) interactions, networkx library (v.2.6.3) was used[90]. *N* loci form a clique of size *N* if each pair of these loci forms a significant contact according to FitHiC2. Diagrams from Fig. 4D were plotted using pyCircos python library (https://github.com/ponnhide/pyCircos).

The list of housekeeping genes was obtained from HRT Atlas v1.0 database[91]. The list of transcription factors was obtained from Lambert et al.[92].

RNA-seq data analysis

Publicly available RNA-seq data for post-mortem brain samples (region BA9) with isolated NeuN(+) and NeuN(-) nuclei was used[25]. Fastq files were processed using nf-core/rnaseq pipeline[93,94], v.3.8.1, with the following parameters: “--aligner hisat2”, “--genome hg38”, “-profile singularity”. The resulting BAM files were sorted by name using samtools sor [95], v.1.6. Number of read counts per gene was calculated with HTSeq software[96], v.2.0.2, using htseq-count command with parameters: “-t exon”, “--mode intersection-nonempty”, “-s yes” - and GENCODE gene annotation[97], v.41. Protein-coding genes and lncRNAs were further analyzed. Differential expression was calculated using DEseq2 package[98], v.1.36.0. Genes with at least five read counts in at least 5 samples were kept. Two variables: cell type and patient id - were added to DEseq2 design formula. Differentially expressed genes with adjusted p-value < 0.05 and $|\log2FoldChange| > 0.58$ were further used. Gene Ontology enrichment analysis was performed on protein-coding genes using clusterProfiler package[99,100], v.4.4.4, with enrichGO function and the following parameters: “OrgDb = org.Hs.eg.db”, “pAdjustMethod = "BH"”, “pvalueCutoff = 0.05”.

Protein extracts preparation and immunoblotting

FANS-sorted nuclei were incubated in RIPA buffer (150 mM NaCl, 1% Triton X-100, 0,5% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HC (pH=8.0), 0,1mM PMSF, and protease inhibitors) for 30 min on ice. The protein extracts then were sonicated at power 15 regime for 10 s with a VirSonic 100 ultrasonic cell disruptor (Virtis). The protein concentration was measured with the Bradford Assay (Sigma-Aldrich). Aliquots of the extracts were separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis and transferred onto polyvinylidene difluoride membranes (Amersham/GE Healthcare). The membranes were incubated with primary antibodies in a blocking solution (2% BSA in PBS containing 0.1% Tween 20 (PBS-T)) overnight at 4℃. After washing with PBS-T (10 min, three times), the membranes were incubated with secondary anti-rabbit antibodies (horseradish peroxidase-conjugated) for 1 hour at room temperature. The membranes were washed in PBS-T again (10min, five times). The binding was visualized using a Pierce ECL plus western blotting substrate. Band intensities were quantified using ImageJ software[101].

Immunohistochemistry analysis

For fluorescent immunohistochemistry, 20 μm frozen sections were cut from samples of the posterior portion of superior temporal gyrus (BA22p) from four mentally healthy persons (Supplementary Table 1). Sections were thaw-mounted onto Superfrost Plus microscope slides, dried at +36°C on the hot plate for 30 min and fixed with 10% neutral buffered formalin for 7 min followed by washing in phosphate-buffered saline (0.1M PBS pH 7.4) three times for 5 min. Then sections were permeabilized and preincubated in 1% Triton X-100 in PBS (PBST) with 5% normal donkey serum (NDS) for 1 h. Reaction with a mixture of primary antibodies (rabbit anti-H3me3K27 and mouse anti-NeuN) (Supplementary Table 5) was performed in the blocking buffer (2.5% NDS in PBST with 0, 01% sodium azide) for 24 h at +4°C in Shandon Coverplates. Following washing and incubation with a mixture of secondary antibodies (anti-mouse Alexa Fluor 488 and anti-rabbit Alexa Fluor 568 conjugate) for 2 h at room temperature (Supplemental Table 5). After a wash in PBS slides were mounted with Fluoromount aqueous mounting medium with added blue fluorescent nuclear counterstain DAPI, coverslipped and sealed with nail polish. No staining was seen in control sections processed without the primary antibody, except lipofuscin granules, located exclusively in the cytoplasms of the pyramidal neurons, visible in all three channels, which were differentiated by the characteristic combinatorial color in the images.

Images were obtained by Zeiss LSM 800 AiryScan with C Plan-Apochromat 40x/1,3 Oil DIC UV-VIS-IR objective with Zeiss Immersol 518 F oil at Core Centrum of IDB RAS. Super-resolution AiryScan imaging allowed us to get images with 200

nm lateral resolution. Image analysis was performed with Zeiss ZEN 2.3 at Skoltech BioImaging and Spectroscopy Core Facility. For the neuronal and non-neuronal nuclei, area, channel intensity, colocalization coefficients and profile calculation were calculated based on unbiased principles.

Data availability

Hi-C data generated in this study were deposited to the GEO database under the accession number … Reviewer token: … Processed ChIP-seq H3K27me3 and chromHMM chromatin states data were obtained from syn25716684. Processed ChIP-seq H3K9me3 data were obtained from ENCODE website (https://www.encodeproject.org/). Accession numbers are: ENCFF557YDE for astrocytes and ENCFF985OQZ for the fetal brain data. Lamin B1 ChIP-seq data were obtained from syn25931622. RNA-seq data on human NeuN(+) and NeuN(-) cells were obtained from GSE96615. scRNA-seq on human fetal gene expression was obtained from descartes.brotmanbaty.org, "matrix of normalized gene expression values". snRNA-seq data on post-mortem human brain cortex was provided by Allen Institute for Brain Science (2022) in Allen Cell Types Database. The dataset is available from https://portal.brain-map.org/atlases-and-data/rnaseq/human-multiple-cortical-areas-smart-seq. Publicly available dorsolateral prefrontal cortex (DLPFC), lung, spleen and bladder Hi-C data were obtained from GSE87112, hESCs - from GSE52457, mESCs and mouse NPC - from 4D Nucleome Data Portal, accession numbers 4DNESDXUWBD9 and 4DNESJ9SIAV5, differentiated mouse neurons - from GSE96107, human NeuN(+) and NeuN(-) cells from BA9 cortex area - from syn21760712.

Acknowledgements

We are grateful to the Core Centrum of IDB RAS and Skoltech BioImaging and Spectroscopy Core Facility for the excellent technical assistance. We thank the Center for Precision Genome Editing and Genetic Technologies for Biomedicine, IGB RAS for cell sorting. We thank Prof. Mikhail S. Gelfand (Skolkovo Institute of Science and Technology) and Dr. Aleksandra A. Galitsyna for helpful advice, Skoltech Genomics Core Facility and Prof. Maria Logacheva (Skolkovo Institute of Science and Technology) for sequencing and quality control of Hi-C libraries, Arkuda HPC cluster provided by the Skolkovo Institute of Science and Technology for computational resources, and National BioService Russian Biospecimen CRO (St. Petersburg, Russia) for providing brain samples. Lamin B1 ChIP-seq, human Hi-C of BA9 cortex area, H3K27me3 ChIP-seq and chromHMM chromatin states data for this publication

were obtained from NIMH Repository & Genomics Resource, a centralized national biorepository for genetic studies of psychiatric disorders. This study was supported by the Russian Science Foundation (grant number 21-74-10102 to E.K.).

References

1. Hansen, A. S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20–32 (2018).
2. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
3. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
4. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
5. Cavalheiro, G. R., Pollex, T. & Furlong, E. E. To loop or not to loop: what is the role of TADs in enhancer function and gene regulation? *Curr. Opin. Genet. Dev.* **67**, 119–129 (2021).
6. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6456–E6465 (2015).
7. Zhang, S., Übelmesser, N., Barbieri, M. & Papantonis, A. Enhancer–promoter contact formation requires RNAPII and antagonizes loop extrusion. *Nat. Genet.* **55**, 832–840 (2023).
8. Dequeker, B. J. H. *et al.* MCM complexes are barriers that restrict cohesin-mediated loop extrusion. *Nature* **606**, 197–203 (2022).
9. Showpnil, I. A. *et al.* EWS/FLI mediated reprogramming of 3D chromatin

promotes an altered transcriptional state in Ewing sarcoma. *Nucleic Acids Res.* **50**, 9814–9837 (2022).

10. Denholtz, M. *et al.* Long-Range Chromatin Contacts in Embryonic Stem Cells Reveal a Role for Pluripotency Factors and Polycomb Proteins in Genome Organization. *Cell Stem Cell* **13**, 602–616 (2013).
11. Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* **47**, 1179–1186 (2015).
12. Kundu, S. *et al.* Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol. Cell* **65**, 432-446.e5 (2017).
13. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e24 (2017).
14. Boyle, S. *et al.* A central role for canonical PRC1 in shaping the 3D nuclear landscape. *Genes Dev.* **34**, 931–949 (2020).
15. Du, Z. *et al.* Polycomb Group Proteins Regulate Chromatin Architecture in Mouse Oocytes and Early Embryos. *Mol. Cell* **77**, 825-839.e7 (2020).
16. Zhang, Y. *et al.* The Polycomb protein RING1B enables estrogen-mediated gene expression by promoting enhancer–promoter interaction and R-loop formation. *Nucleic Acids Res.* **49**, 9768–9782 (2021).
17. Kraft, K. *et al.* Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation. *Proc. Natl. Acad. Sci. U. S. A.* **119**, 1–10 (2022).
18. Eskeland, R. *et al.* Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. *Mol. Cell* **38**, 452–464 (2010).
19. Hugues, A., Jacobs, C. S. & Roudier, F. Mitotic Inheritance of PRC2-Mediated

Silencing: Mechanistic Insights and Developmental Perspectives. *Front. Plant Sci.* **11**, 262 (2020).

20. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
21. McKenzie, A. T. *et al.* Brain Cell Type Specific Gene Expression and Co-expression Network Architectures. *Sci. Rep.* **8**, 8868 (2018).
22. Hu, B. *et al.* Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders. *Nat. Commun.* **12**, (2021).
23. Rahman, S. *et al.* From compartments to gene loops: Functions of the 3D genome in the human brain. *bioRxiv* 1–45 (2021).
24. Eremenko, E. *et al.* FACS-based isolation of fixed mouse neuronal nuclei for ATAC-seq and Hi-C. *STAR Protoc.* **2**, 100643 (2021).
25. Rizzardi, L. F. *et al.* Neuronal brain-region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric trait heritability. *Nat. Neurosci.* **22**, 307–316 (2019).
26. Calderon, L. *et al.* Cohesin-dependence of neuronal gene expression relates to chromatin loop length. *eLife* **11**, 1–31 (2022).
27. Hodge, R. D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
28. Sanders, J. T. *et al.* Loops, topologically associating domains, compartments, and territories are elastic and robust to dramatic nuclear volume swelling. *Sci. Rep.* **12**, 4721 (2022).
29. García-Cabezas, M. Á., John, Y. J., Barbas, H. & Zikopoulos, B. Distinction of

Neurons, Glia and Endothelial Cells in the Cerebral Cortex: An Algorithm Based on Cytological Features. *Front. Neuroanat.* **10**, (2016).

30. Hetman, M. & Slomnicki, L. P. Ribosomal biogenesis as an emerging target of neurodevelopmental pathologies. *J. Neurochem.* **148**, 325–347 (2019).
31. Hetman, M. & Pietrzak, M. Emerging roles of the neuronal nucleolus. *Trends Neurosci.* **35**, 305–314 (2012).
32. Fudenberg, G. & Mirny, L. A. Higher-order chromatin structure: bridging physics and biology. *Curr. Opin. Genet. Dev.* **22**, 115–124 (2012).
33. Heinz, S. *et al.* Transcription Elongation Can Affect Genome 3D Structure. *Cell* **174**, 1522-1536.e22 (2018).
34. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
35. Sharma, K. *et al.* Cell type– and brain region–resolved mouse brain proteome. *Nat. Neurosci.* **18**, 1819–1831 (2015).
36. Dong, P. *et al.* Population-level variation in enhancer expression Identifies disease mechanisms in the human brain. *Nat. Genet.* **54**, 1493–1503 (2022).
37. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
38. Ahanger, S. H. *et al.* Distinct nuclear compartment-associated genome architecture in the developing mammalian brain. *Nat. Neurosci.* **24**, 1235–1242 (2021).
39. Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).

40. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).

41. Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).

42. Nielsen, J. B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234–1239 (2018).

43. Malik, R. *et al.* Genome‐wide meta‐analysis identifies 3 novel loci associated with stroke. *Ann. Neurol.* **84**, 934–939 (2018).

44. Mahajan, A. *et al.* Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat. Genet.* **54**, 560–572 (2022).

45. International League Against Epilepsy Consortium on Complex Epilepsies, Berkovic, S. F., Cavalleri, G. L. & Koeleman, B. P. *Genome-wide meta-analysis of over 29,000 people with epilepsy reveals 26 loci and subtype-specific genetic architecture*. http://medrxiv.org/lookup/doi/10.1101/2022.06.08.22276120 (2022) doi:10.1101/2022.06.08.22276120.

46. De La Fuente, J., Davies, G., Grotzinger, A. D., Tucker-Drob, E. M. & Deary, I. J. A general dimension of genetic sharing across diverse cognitive traits inferred from molecular data. *Nat. Hum. Behav.* **5**, 49–58 (2020).

47. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* **9**, 2098 (2018).

48. Corces, M. R. *et al.* Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020).

49. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).

50. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

51. Rieder, D., Trajanoski, Z. & McNally, J. G. Transcription factories. *Front. Genet.* **3**, (2012).

52. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369-1384.e19 (2016).

53. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).

54. Kaul, A., Bhattacharyya, S. & Ay, F. Identifying statistically significant chromatin contacts from Hi-C data with FitHiC2. *Nat. Protoc.* **15**, 991–1012 (2020).

55. Zhang, J. *et al.* An integrative ENCODE resource for cancer genomics. *Nat. Commun.* **11**, 3696 (2020).

56. Hubert, K. A. & Wellik, D. M. Hox genes in development and beyond. *Dev. Camb. Engl.* **150**, dev192476 (2023).

57. Stevanovic, M. *et al.* SOX Transcription Factors as Important Regulators of Neuronal and Glial Differentiation During Nervous System Development and

Adult Neurogenesis. *Front. Mol. Neurosci.* **14**, 654031 (2021).

58. Tan, Y. & Testa, J. R. DLX Genes: Roles in Development and Cancer. *Cancers* **13**, 3005 (2021).
59. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science* **370**, eaba7721 (2020).
60. Kitamura, A., Yuno, S., Muto, H. & Kinjo, M. Different aggregation states of a nuclear localization signal-tagged 25-kDa C-terminal fragment of TAR RNA/DNA-binding protein 43 kDa. *Genes Cells* **22**, 521–534 (2017).
61. Korsholm, L. M. *et al.* Recent advances in the nucleolar responses to DNA double-strand breaks. *Nucleic Acids Res.* **48**, 9449–9461 (2020).
62. Blackledge, N. P. & Klose, R. J. The molecular principles of gene regulation by Polycomb repressive complexes. *Nat. Rev. Mol. Cell Biol.* **22**, 815–833 (2021).
63. Sawai, A. *et al.* PRC1 sustains the integrity of neural fate in the absence of PRC2 function. *eLife* **11**, e72769 (2022).
64. Von Schimmelmann, M. *et al.* Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration. *Nat. Neurosci.* **19**, 1321–1330 (2016).
65. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
66. Mirny, L. & Dekker, J. Mechanisms of Chromosome Folding and Nuclear Organization: Their Interplay and Open Questions. *Cold Spring Harb. Perspect. Biol.* a040147 (2021) doi:10.1101/cshperspect.a040147.
67. Canzio, D. *et al.* Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin α Promoter Choice. *Cell* **177**, 639-653.e15

(2019).

68. Okano, H. & Temple, S. Cell types to order: temporal specification of CNS stem cells. *Curr. Opin. Neurobiol.* **19**, 112–119 (2009).

69. Kishi, Y. & Gotoh, Y. Regulation of Chromatin Structure During Neural Development. *Front. Neurosci.* **12**, 874 (2018).

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

71. Stock, J. K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat. Cell Biol.* **9**, 1428–1435 (2007).

72. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).

73. Ma, Y. & Buttitta, L. Chromatin organization changes during the establishment and maintenance of the postmitotic state. *Epigenetics Chromatin* **10**, 53 (2017).

74. Siegenfeld, A. P. *et al.* Polycomb-lamina antagonism partitions heterochromatin at the nuclear periphery. *Nat. Commun.* **13**, 4199 (2022).

75. Mai, J. *Atlas of the Human Brain. 4th Ed.* (2016).

76. Ulianov, S. V. *et al.* Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* **26**, 70–84 (2016).

77. Open2C *et al.* Pairtools: from sequencing data to chromosome contacts. *bioRxiv* 2023.02.13.528389 (2023) doi:10.1101/2023.02.13.528389.

78. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).

79. Open2C *et al. Cooltools: enabling high-resolution Hi-C analysis in Python*. http://biorxiv.org/lookup/doi/10.1101/2022.10.31.514564 (2022) doi:10.1101/2022.10.31.514564.

80. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).

81. Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* **17**, 2042–2059 (2016).

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

83. van der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).

84. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).

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

86. Dale, R. K., Pedersen, B. S. & Quinlan, A. R. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinforma. Oxf. Engl.* **27**, 3423–3424 (2011).

87. Kerpedjiev, P. *et al.* HiGlass: Web-based Visual Exploration and Analysis of Genome Interaction Maps. *Genome Biol.* **19**, 1–12 (2018).

88. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24**, 999–1011 (2014).

89. Flyamer, I. M., Illingworth, R. S. & Bickmore, W. A. *Coolpup.py:* versatile pile-up analysis of Hi-C data. *Bioinformatics* **36**, 2980–2985 (2020).

90. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring Network Structure, Dynamics, and Function using NetworkX. in *Proceedings of the 7th Python in Science Conference* (eds. Varoquaux, G., Vaught, T. & Millman, J.) 11–15 (2008).

91. Hounkpe, B. W., Chenou, F., de Lima, F. & De Paula, E. V. HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. *Nucleic Acids Res.* **49**, D947–D955 (2021).

92. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **175**, 598–599 (2018).

93. Patel, H. *et al.* nf-core/rnaseq: nf-core/rnaseq v3.10.1 - Plastered Rhodium Rudolph. (2023) doi:10.5281/zenodo.7505987.

94. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).

95. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).

96. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

97. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Res.* **49**, D916–D923 (2021).

98. 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).

99. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).

100. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).

101. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).